UAB
Universitat Autònoma de Barcelona

Accessing genetic variability in Spanish barleys through high-throughput sequencing

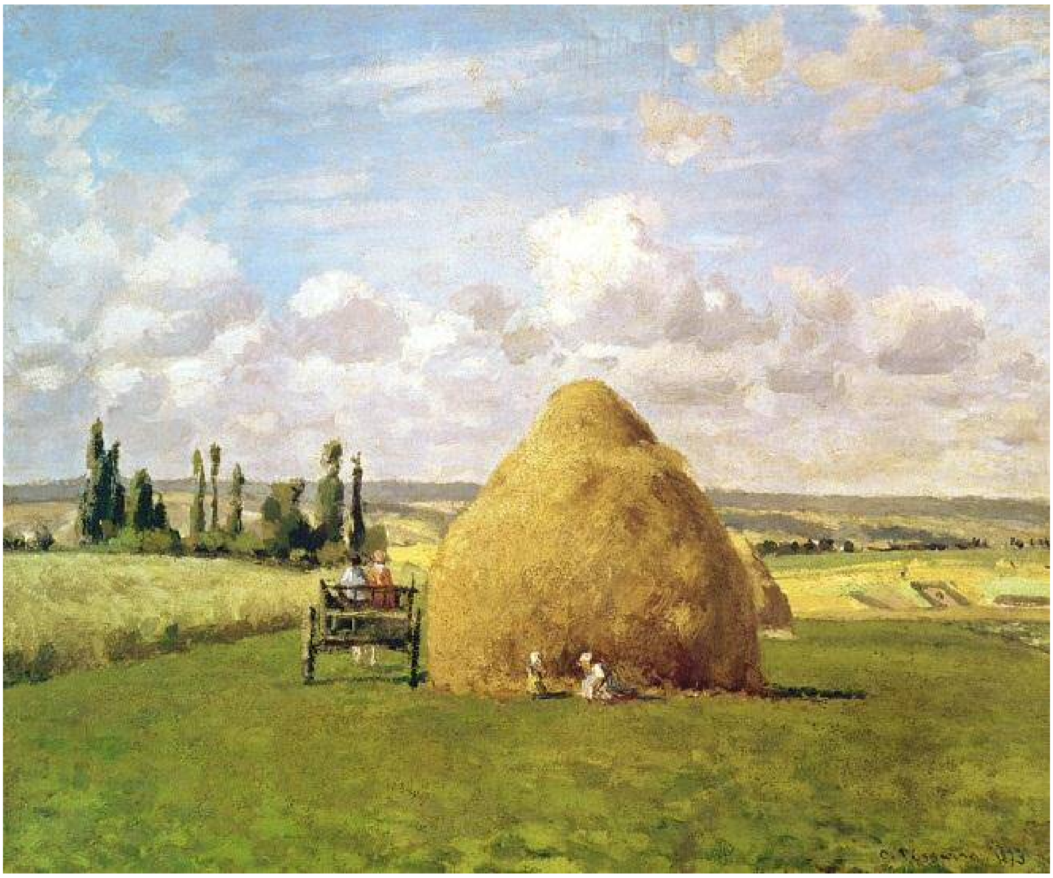
Carlos Pérez Cantalapiedra

CSIC
Consejo Superior de Investigaciones Científicas

UAB
Universitat Autònoma de Barcelona

Accessing genetic variability in Spanish barleys through high-throughput sequencing

Carlos Pérez Cantalapiedra



Zaragoza 2016

Universidad Autónoma de Barcelona
Facultad de Biociencias
Dpto. Biología Animal, Biología Vegetal y Ecología
Estudios de Doctorado en Biología y Biotecnología Vegetal

PhD Thesis

Accessing genetic variability in Spanish barleys through high-throughput sequencing

Research memory presented by Carlos Pérez Cantalapiedra to obtain the title of Doctor in Plant Biology and Biotechnology from Universidad Autónoma de Barcelona (UAB)

This work has been done at Estación Experimental de Aula Dei (EEAD), belonging to Consejo Superior de Investigaciones Científicas (CSIC), in Zaragoza

Co-director	Co-director	Tutor	PhD student
Dra. Ana María Casas Cendoya	Dr. Bruno Contreras Moreira	Dr. Josep Alluè	Carlos Pérez Cantalapiedra

Zaragoza, September 2016

Cover image: *The haystack pontoise*. Camille Pissarro 1873

To my mother, Elena,
to Javier,
to Álvaro,
and to I.

Agradecimientos (acknowledgements)

En primer lugar, quiero agradecerle todo al Dr. Bruno Contreras Moreira: la bienvenida (allá por un verano de 2010), la paciencia y la docencia, y la amistad, en las risas y el trabajo que compartimos. Sobre todo, quiero agradecerte tu profesionalidad, la confianza que depositaste en mí y, lo más importante, tu inmensa calidad como persona. No solo yo te estoy agradecido por mi parte.

Una gran confianza también han puesto en mí el Dr. Ernesto Igartua Arregui y la Dra. Ana María Casas Cendoya. A ambos quiero agradeceros también vuestra profesionalidad, y vuestra calidad como personas. He tenido dos padres y una madre en esto de la ciencia. Sería mucho peor profesional, y persona, de no haber compartido los últimos cinco años con vosotros. Habéis hecho todo lo posible por ayudarme a avanzar. Además, habéis demostrado una paciencia infinita para los chistes malos durante la comida. Doy fé.

También quiero agradecer al personal de los centros donde he realizado estancias FPI. Al Dr. Takeshi Itoh y todo el personal de la Bioinformatics Unit del NIAS de Tsukuba (ありがとうございます). Al Dr. Nils Stein y al Dr. Uwe Scholz, del IPK de Gatersleben. También de Alemania, al Dr. Martin Mascher y al Dr. Thomas Schmutzer (vielen Dank IPK). También agradecer al Dr. Paolo Ribeca y el Dr. Santiago Marco, por unas lecciones aceleradas en el CNAG, así como al Prof. Mariano Hernández, por la paciencia y el spray en la ULL. También al Dr. Pablo Vinuesa (UNAM), al Dr. Dragan Perovic (JKI), y Valerie y Emilio, por abrirme las puertas de sus casas. Muchas gracias también a Alba Farré Martínez (IAMZ) por la ayuda lingüística.

No puedo dejar de agradecer a todos los profesionales de la Estación Experimental de Aula Dei. A María Pilar Gracia, a Pili Asín, y a las chicas de la cebada. También a Chus y a Bibiana, por su esforzada colaboración. A mi compañera Arantxa Monteagudo, a la que espero no haber asustado demasiado. Muy especialmente me gustaría agradecer a Javi su amistad y su paciencia (sobre todo con los comederos y los zorzales). Este último año hubiera sido distinto; para mí, para la fauna aragonesa, y para la tienda de caramelos gajo.

Por último, agradecer su paciencia y ayuda incondicional a mi madre, Elena, y a Javier. A Álvaro PC, por muchas cosas desde pequeños. A Elena Paracuellos Planas y todas mis florecillas. A R. Longás, M. Frago; y a mi Loliña, a la que debo unos Kms y unas risas. A Carlos G. Seral, por tu apoyo, tu cariño, por ser como eres.

A CS, JMdT, y a I

Indexes

TABLE OF CONTENTS

Summaries 1

1. General introduction 7

1.1. Barley (Hordeum vulgare L.) 9

1.1.1. Importance and uses 9

1.1.2. Taxonomy and description of the species 10

1.1.3. Origin, domestication, and gene pools 14

1.1.4. Barley genomics 19

1.2. Breeding challenges and approaches 22

1.2.1. Disease resistance 23

1.2.2. Drought tolerance 24

1.2.3. Looking for stress tolerance genes 27

1.2.4. Breeding methods 29

1.3. High-throughput sequencing 31

1.3.1. Technologies 31

1.3.2. Applications and breeding 33

1.3.3. Data analysis: bioinformatics 35

1.4. References 40

2. Objectives 61

3. BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley 65

3.1. Introduction 67

3.2. Materials and Methods 68

3.2.1. Pipeline outline 68

3.2.2. Barley data configuration and application distribution 69

3.2.3. Genetic map construction 70

3.3. Results 71

3.3.1. Alignment of barley transcripts 71

3.3.2. Alignment of barley markers ... 71
3.3.3. Mapping of aligned markers to barley genetic/physical maps ... 72
3.3.4. Matching of genetic markers to barley genes ... 73
3.3.5. Validating genetic maps of barley populations ... 73
3.4. Discussion ... 76
3.5. References ... 79
4. A cluster of NBS-LRR genes resides in a barley powdery mildew resistance QTL on 7HL ... 83
4.1. Introduction ... 85
4.2. Materials and methods ... 86
4.2.1. Plant material and mapping population ... 86
4.2.2. Pathogen isolates and disease assessment ... 87
4.2.3. Exome sequencing ... 87
4.2.4. Identification and annotation of the BACs located within the QTL region ... 88
4.2.5. Finding and assembling heterozygous mapping regions ... 88
4.2.6. Validation of the genotypes found with the exome capture by PCR ... 88
4.2.7. Characterization of the new assembled sequence contigs ... 89
4.2.8. Real-Time PCR of the assembled sequence contigs ... 89
4.3. Results ... 89
4.3.1. Fine mapping of the resistance locus ... 89
4.3.2. Analysis of exome sequencing polymorphisms ... 90
4.3.3. Physical localization of the resistance locus ... 91
4.3.4. Searching for candidate genes in the reference cultivar Morex ... 91
4.3.5. Analysis of "heterozygous mappings" in Morex ... 95
4.3.6. De-novo assembly of exome sequence reads spanning the resistance locus ... 96
4.3.7. Validation and characterization of the new assembled sequence contigs ... 97
4.4. Discussion ... 99
4.5. References ... 103
5. Large differences in gene expression between elite barley cultivar Scarlett and a Spanish landrace under drought and

heat stress 109

5.1. Introduction 111

5.2. Materials and methods 112

5.2.1. Plant material and drought experiments 112

5.2.2. Measurement of phenotypic traits 112

5.2.3. RNA extraction and transcriptome sequencing 113

5.2.4. RNAseq data preprocessing and transcriptome assembly 113

5.2.5. Correction, validation and annotation of de novo transcriptomes 114

5.2.6. Analysis of gene expression 114

5.2.7. RT-qPCR validation 115

5.2.8. Functional annotation of differentially expressed isoforms 115

5.2.9. Comparison with related studies 116

5.2.10. Clustering and identification of cis-regulatory elements of co-expressed genes 116

5.3. Results 117

5.3.1. Growth of Scarlett and SBCC073 plants subjected to drought 117

5.3.2. Assembly and validation of Scarlett and SBCC073 transcriptomes 119

5.3.3. Analysis of gene expression 120

5.3.4. Differentially expressed isoforms in leaves under drought 121

5.3.5. Differentially expressed isoforms in leaves under mild drought and heat 122

5.3.6. Differentially expressed isoforms in young inflorescences in SBCC073 126

5.3.7. Differentially expressed isoforms in young inflorescences in Scarlett 126

5.3.8. Comparison with related studies 129

5.3.9. Analysis of co-expressed genes 131

5.4. Discussion 134

5.5. References 140

6. General discussion 149

7. Conclusions 155

8. Annexes 159

LIST OF FIGURES

Figure 1.1. Diagram of a typical grass plant 10

Figure 1.2. Barley inflorescences 12

Figure 1.3. Summary of barley development 13

Figure 1.4. Spikes and caryopses from several Spanish barley landraces 18

Figure 1.5. Comparative analysis between barley and *B. distachyon* 21

Figure 1.6. Bridge amplification of DNA 32

Figure 1.7. *De novo* assembly using de Bruijn graphs 36

Figure 1.8. Paired-end reads mapped to a reference genome 37

Figure 3.1. The BARLEYMAP pipeline 69

Figure 3.2. Percentage of sequences found by either Blastn or GMAP 71

Figure 3.3. Comparison of BARLEYMAP positions and genetic map 75

Figure 4.1. Fine mapping of the 7HL QTL 90

Figure 4.2. Analysis of BACs in MTP of FPC 591 93

Figure 4.3. NBS and LRR motifs found in the region of FPC 591 94

Figure 4.4. Heterozygous mappings (HMs) 96

Figure 4.5. Presence-absence genotypes for ELOC1 and ELOC2 98

Figure 5.1. Design of stress treatments, and leaf water potential patterns 118

Figure 5.2. De novo assembled genes confirmed in existing barley references 120

Figure 5.3. Comparison of RT-qPCR and RNAseq gene expression results 122

Figure 5.4. Number of differentially expressed isoforms and genes 123

Figure 5.5. Metabolic pathways and cellular processes with differentially expressed isoforms from leaves under mild drought and heat 124

Figure 5.6. Metabolic pathways and cellular processes with differentially expressed isoforms from Scarlett young inflorescences 125

Figure 5.7. Percentage of differentially expressed tags from other studies which were identified in the present work 131

Figure 5.8. Enriched DNA motifs in promoters of differentially co-expressed isoforms 132

LIST OF TABLES

Table 3.1. Genetic markers aligned with BARLEYMAP 72

Table 3.2. Comparison of different mapping approaches 73

Table 5.1. Physiological measurements of plants in the drought experiments 119

Table 5.2. Statistics of *de novo* and reference-guided assemblies 121

Table 5.3. Gene Ontology terms enriched in Scarlett young inflorescences 128

Table 5.4. Studies from the literature assessing protein or transcript expression changes in response to drought in barley 130

Table 5.5. Differentially expressed isoforms found in three or more previous studies.......... 133

Table 5.6. Predicted DNA motifs for differentially expressed transcription factors 134

Summaries

Summary

Barley is an important crop in the Mediterranean region, characterized by scarce and irregular rainfalls. In the Iberian Peninsula, it has been cultivated for thousands of years, leading to specific adaptations to prevalent biotic and abiotic stresses. These features, present in Spanish barley landraces, remain to be exploited in breeding.

High-throughput sequencing (HTS) has revolutionized plant research. It has made it possible to sequence the genomes of multiple organisms. The sequence-enriched physical map of barley was published in late 2012. A first step to exploit barley genomics, for practical purposes, was facilitating geneticists and breeders access to the barley physical map. This was the aim which led us to the development of Barleymap, a software tool which allows locating genetic markers in the barley physical-genetic map. This application effectively integrates and maps markers from different widely used barley genotyping platforms, and, in general, any marker with sequence information.

Another advantage of HTS is that diverse experimental setups can be used with different research objectives. Here, we used exome sequencing to fine-map a powdery mildew resistance QTL from a Spanish barley landrace. Exploiting a large mapping population, we were able to narrow down the position of the QTL to a single physical contig. Moreover, we could identify, and partially assemble, an expressed candidate gene. To achieve this, an array of bioinformatics approaches was applied to differentiate presence-absence variation, within a cluster of closely related genes of the NBS-LRR family.

Another powerful application of HTS is RNAseq, which allows sequencing whole transcriptomes, and gene expression assays can be performed with unprecedented power. We *de novo* assembled the transcriptomes of a drought susceptible elite barley cultivar and a drought resistant Spanish barley landrace. Then, we compared the expression changes, in leaves and developing inflorescences from both genotypes, under drought treatments. This revealed large differences in their responses to stress. A comparison with other drought gene expression studies on barley, and an analysis of transcription factors and *cis*-regulatory elements involved, provided new insights into the complex barley gene expression network under stress.

In summary, HTS has brought many new possibilities to plant research. To take full advantage of it, crosstalk between bioinformatics and genetics must be fostered to adapt the new genomic resources to specific needs.

Resumen

La cebada es un cultivo importante en la región mediterránea, caracterizada por escasas e irregulares precipitaciones. En la Península Ibérica, ha sido cultivada durante miles de años, surgiendo adaptaciones específicas a estrés. Estas características, presentes en las variedades locales españolas, permanecen sin ser explotadas en mejora.

La secuenciación de alto rendimiento (HTS, por sus siglas en inglés) ha revolucionado la investigación. Ha hecho posible secuenciar los genomas de múltiples organismos. El mapa físico de cebada, con secuencias asociadas, fue publicado a finales de 2012. Para sacar partido de estos recursos, había que facilitar el acceso a dicho recurso a genetistas y mejoradores. Este fue el objetivo que nos llevó a desarrollar Barleymap, una herramienta informática que permite localizar marcadores genéticos en el genoma de cebada. La aplicación integra y localiza marcadores de distintas plataformas de genotipado de cebada ampliamente utilizadas.

Otra ventaja de la HTS es que se pueden llevar a cabo distintos tipos de experimentos con distintos objetivos de investigación. Nosotros utilizamos la secuenciación del exoma para mapeo fino de un QTL de resistencia a oidio de una variedad local española. A partir de una gran población de mapeo, fuimos capaces de acotar la posición del QTL a un solo contig físico. Además, pudimos identificar, y ensamblar parcialmente, un gene candidato que se expresa. Para conseguir esto, una serie de enfoques bioinformáticos fueron aplicados para diferenciar variación de presencia-ausencia, en un grupo de genes relacionados de la familia NBS-LRR.

Otra aplicación poderosa de la HTS es RNAseq, que permite secuenciar transcriptomas completos, y llevar a cabo ensayos de expresión con una resolución sin precedente. Ensamblamos *de novo* los transcriptomas de un cultivar de cebada susceptible a sequía y de una variedad local española resistente. Comparamos los cambios de expresión, en hojas e inflorescencias en desarrollo de ambos genotipos, bajo tratamientos de sequía. Se revelaron grandes diferencias en sus respuestas a estrés. La comparación con otros trabajos de sequía en cebada, y el análisis de los factores de transcripción y elementos reguladores implicados proporcionó nuevos datos sobre la compleja red de expresión génica de cebada bajo estrés.

En resumen, la HTS trae muchas nuevas posibilidades. Para aprovecharla totalmente, se debe fomentar colaboración de bioinformáticos y genetistas, para adaptar los nuevos recursos genómicos a necesidades específicas.

Resum

L'ordi és un cultiu important a la regió mediterrània, caracteritzada per precipitacions escasses i irregulars. A la Península Ibèrica, ha estat conreat durant milers d'anys, permeten l’aparició d’adaptacions específiques a l’estrès. Aquestes característiques, presents en les varietats locals espanyoles, romanen sense ser explotades en la millora de cereals.

La seqüenciació d'alt rendiment (HTS, per les sigles en anglès) ha revolucionat la investigació fent possible la seqüenciació dels genomes de múltiples organismes. El mapa físic de l'ordi, amb seqüències associades, va ser publicat a finals de 2012. Per treure partit d'aquests recursos, calia facilitar-ne l'accés a genetistes i milloradors. Aquest va ser l'objectiu que ens va portar a desenvolupar Barleymap, una eina informàtica que permet localitzar marcadors genètics en el genoma de l’ordi. Aquesta aplicació integra i localitza marcadors de diferents plataformes de genotipat d'ordi àmpliament utilitzades.

Un altre avantatge de la HTS és que es poden dur a terme diferents tipus d'experiments amb diferents objectius d'investigació. Nosaltres fem servir la seqüenciació de l’exoma pel mapeig fi d'un QTL de resistència a l’oïdi d'una varietat local espanyola. A partir d'una gran població de mapeig, vam ser capaços de delimitar la posició del QTL a un contig físic. A més, vam poder identificar i ensamblar parcialment un gen candidat que s'expressa. Per aconseguir això, una sèrie aproximacions bioinformàtiques van ser aplicades per diferenciar la variació de presència-absència en un grup de gens de la família NBS-LRR.

Una altra aplicació poderosa de la HTS és RNAseq, que permet seqüenciar transcriptomes complets, i dur a terme assajos d'expressió amb una resolució sense precedent. Ensamblem de novo els transcriptomes d'un cultivar d'ordi susceptible a sequera i d'una varietat local espanyola resistent. Comparem els canvis d'expressió, en fulles i inflorescències en desenvolupament d'ambdós genotips, sota tractaments de sequera. Es van revelar grans diferències en les seves respostes a estrès. La comparació amb altres treballs de sequera en ordi, i l'anàlisi dels factors de transcripció i elements reguladors implicats va proporcionar noves dades sobre la complexa xarxa d'expressió gènica d'ordi sota estrès.

En resum, la HTS aporta moltes noves possibilitats. Per aprofitar-la totalment, s'ha de fomentar la col•laboració de bioinformàtics i genetistes, per adaptar els nous recursos genòmics a les necessitats específiques.

1. General introduction

1.1. Barley (Hordeum vulgare L.)

1.1.1. Importance and uses

Most people rely on grasses (rice, wheat, maize, barley, sorghum, oats) as sources for a major part of their diet, for feeding livestock and domestic animals, and as an important part of the urban and suburban landscape (Kellogg, 2001). Grasses grown to harvest their grains are known as cereals. Cereal crops, barley among them, have accompanied mankind throughout their history playing a relevant role in the development of agriculture, civilizations, and cultures (Ullrich, 2011).

Nowadays, the majority of barley production is used for **animal feeding**, mainly cattle and pigs, mostly as grain but also as forage. For this purpose, barley grain is a favorable source of starch and has a higher content of crude fiber and protein than other cereals (Verstegen et al., 2014).

A significant percentage of barley grain is used for **malting**, a process which dries germinated cereal grains, and which goes back to at least 8000 years ago in the Middle East and Egypt (Ullrich, 2011). Malt is used to produce alcoholic beverages, through brewing and distilling (beer, whiskey). Malting barley contains traditionally less protein than feed barley (Verstegen et al., 2014).

Food consumption represents only a small proportion of barley production. Grains can be cooked or milled for bread making. Although this use is relatively minor today, it has been important in past times and has remained a major food source for some cultures, mostly in Asia and North Africa (Newman and Newman, 2008). Renewed interest of barley as food in the developed world is due to an increasing emphasis to take advantage of the health benefits associated to whole grain consumption (Ames and Rhymer, 2008). Barley is the richer cereal source of β-glucans, and it has low glycemic index and high fiber content (Baik and Ullrich, 2008).

Another aspect of barley use is related with its **adaptability**, which allows growing barley in a wide range of environments, reaching high altitudes and latitudes (Graner et al., 2003). Barley is less limited by requirements of good soil fertility and suitable climatic conditions than the other major crops grown in the temperate zone (like wheat and maize), and it is economically viable at low levels of fertilization, including semi-arid areas. Therefore, it is a very important crop in Mediterranean regions, northern Europe, the Middle East, North Africa, and the Andean region of South America (Ullrich, 2011).

Furthermore, barley has been used as **experimental model** for the temperate cereals of the *Triticeae* tribe (wheat, rye, triticale) (Kumlehn and Stein, 2014). It has a long history as a prominent tool in genetics, and considerable research has been done on the origins of barley and crop domestication, on phylogeny and systematics, and as a model in physiological and anatomical topics, especially of the grain (Ullrich, 2011).

Regarding **production**, barley ranks fourth among cereal crops, after maize, rice and wheat, with almost 144 million tons obtained in 2013 (FAOSTAT, 2016). Around 62% of barley production comes from Europe, including Russia, whereas America and Asia produce close to 15% each. World average **yield** is 2.9 t·ha^{-1}, and ranges from around 8 t·ha^{-1}, under optimal conditions, to average yields of 1-2 t·ha^{-1} in African countries bordering the Sahara desert. This yield gap can be attributed mainly to water availability and nutritional inputs.

In **Spain**, barley is one of the major options in non-irrigated agriculture, and the first in production and area harvested (Ministerio de Agricultura Alimentación y Medio Ambiente, 2015). This was 2.8 million ha in 2014, which is close to 10% of the country's agricultural area. Average annual production was 6.9 million tons in the period from 1961 to 2014, with average annual yield ranging from 1.2 t·ha^{-1} to 3.7 t·ha^{-1}. Spain is one of the major producers of barley grain world-wide, after the Russian Federation, and close to France, Germany, Canada or Ukraine.

1.1.2. Taxonomy and description of the species

Barley belongs to the **family *Poaceae* (*Gramineae*)**, a group of monocotyledonous plants, commonly known as grasses, which evolved 70-55 million years ago (Kellogg, 2001). Economically, *Poaceae* is the most important plant family, since it encompasses species such as maize, rice, wheat, barley, sorghum, oats and millet. Natural grasses and bamboos are also included in this family, which in total comprises around 780 genera and 12,000 species (Christenhusz and Byng, 2016). Grasses may be annual or perennial herbs, rarely tree-like as the bamboos, and show an outstanding ecological success, covering more than one fifth of earth's land surface (Shantz, 1954; Watson, 1990). Morphologically (Figure 1.1), *Gramineae* plants develop cylindrical stems with hollow internodes, which are also referred to as culms. Leaves of grasses grow from the base of the plant, in alternate positions, and enclose the stems with their lower part, the sheath, which attaches to stem nodes. The upper part of the leaves, the blade, separates from the stem. It is a narrow, distichous, lanceolate-linear sheet, with parallel veins and entire margins. The epidermis of grasses contains long and short cells, silica bodies, stomata

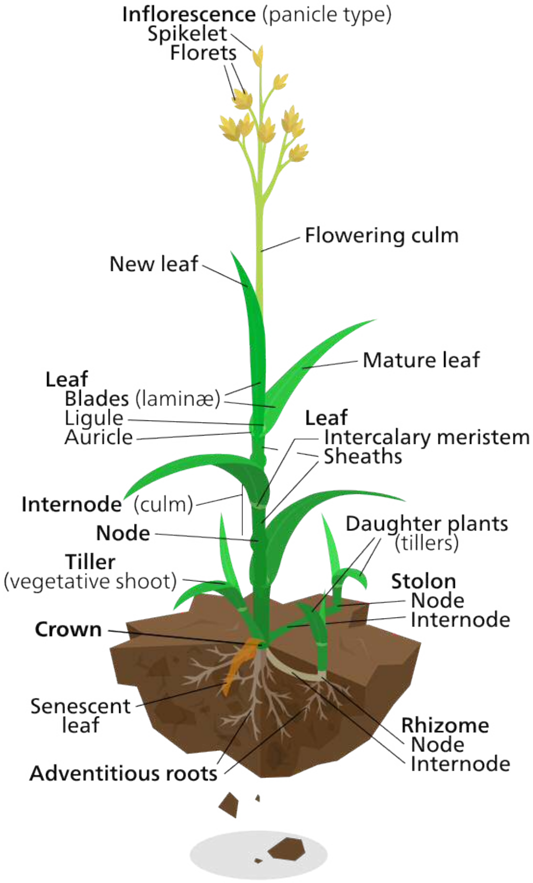


Figure 1.1. Diagram of a typical grass plant. The features of the different parts vary in the different genera and species. For example, *Hordeum* species have spike type inflorescences, instead of panicles. Image from Wikimedia Commons, by Kelvinsong (under CCA3.0 license).

with subsidiary cells, and dumb-bell shaped guard-cells (Watson, 1990). Another common feature of grasses is a membranous appendage which lies at the junction between sheath and blade, called the ligule. *Poaceae* inflorescences emerge from elongated stems, in the form of either panicles or spikes. These are groups of spikelets, which consist of two or fewer bracts at the base, the glumes, and one or more florets. Each floret holds a flower, with the perianth reduced to two scale-like lodicules, surrounded by two additional bracts: one external, the lemma, and one internal, the palea (Clayton, 1990). The flowers of grasses are usually hermaphroditic, and in most species the gynoecium has two stigmas and the androecium has three stamens (Kellogg, 2001). The fruit is a caryopsis, in which the seed coat covers the fruit wall, with abundant, starchy endosperm and a peculiar, laterally placed, embryo (Watson, 1990), configuring a structure which is unique among the flowering plants (Kellogg, 2001).

Within *Poaceae*, the genus *Hordeum* is part of the **tribe *Triticeae***, which belongs to the *Pooideae* subfamily of C3 grasses (Soreng et al., 2015). The *Triticeae* also includes a number of other important cereal crops, such as wheat (*Triticum* spp.), rye (*Secale cereale*), artificially developed triticale, and also many important forage and soil stabilization grass species (Sato et al., 2014). Morphologically, the *Triticeae* show open leaf sheaths, membranous ligules, sessile to almost sessile spikelets, and ovaries with a hairy top (Barkworth and Bothmer, 2009). The inflorescence form comprises a spike (also referred as ear), in contrast with the panicle formed in members of related tribes (Clayton, 1990). An elongation of the stem, the rachis, supports the inflorescence. The spike generally produces a single spikelet per rachis node, which can increase to three spikelets in a few species (Komatsuda, 2014) like, for example, species from the *Hordeum* genus. Each spikelet forms one to a few florets, each with the lemma, the palea, three anthers, and a multi-branched pistil. The mature spike can disarticulate in various forms, either by breakage above the lowest node of the spike, below or above each rachis node, or by breakage of the rachilla above the glumes (Sakuma et al., 2011). All the species in the tribe share the same basic chromosome number of x=7, with different levels of ploidy, with some species having complex genetic histories involving genome duplications and deletions or composite genomes, as that of bread wheat, which carries genomes of three species (Petersen et al., 2006), or that of *Hordeum* polyploid species (Brassac and Blattner, 2015).

The **genus *Hordeum*** consists of 33 species which originated in western Eurasia, and are endemic of the Northern Hemisphere, southern Africa and the southern cone of South America (Blattner, 2009). Ploidy ranges from diploid to hexaploid, with combinations of four basic genomes (Blattner, 2009, and references therein). In contrast with most *Triticeae* species, all *Hordeum* members develop three single-flowered spikelets per rachis node, with one central and two lateral florets, the latter being often sterile (Bothmer et al., 2003). Most of the *Hordeum* species are capable of inbreeding (Blattner, 2009) and, in mature spikes, disarticulation of the spikelets occurs above the rachis node (except for *H. bogdanii*) (Sakuma et al., 2011). Some species are annuals and some are perennials (Bothmer et al., 2003). Seed dispersal depends either on wind (small caryopses), or on animal carriers (large caryopses), varying with the species (for example, *H. vulgare* seeds are transported by animals) (Komatsuda, 2014).

Barley (*Hordeum vulgare* L.) is an annual, self-pollinating, diploid species, which stands 60-120 cm tall and is supported by two types of root systems: seminal and adventitious (Briggs, 1978; Reid, 1985). The base of the plant, the crown, is where adventitious roots, leaves and stems develop. A mature barley plant consists of a central stem and 2-5 branch stems (in spring genotypes), called tillers, each with 5 to 7 internodes (Reid, 1985). Barley leaves, typically ranging from 5 to 10 per stem, are 5-15 mm wide, with glabrous ligule and auricles, which envelop the stem and can be pigmented with anthocyanins (Gomez-Macpherson, 2000). Barley spikelets are attached directly to the rachis of the spike (Australian Government, 2008). Barley inflorescence (Figure 1.2) is classified as indeterminate because the rachis does not terminate in a spikelet (Reid, 1985). Depending on the variety, each lemma is extended as an awn, or more rarely a hood (Gomez-Macpherson, 2000). The sterile glumes in some varieties can develop in an awn, and awnless varieties are also known (Briggs, 1978). In hulled or husked varieties, the palea and lemma adhere to the caryopsis at maturity, whereas in hull-less or naked varieties, the palea and lemma are not attached and the caryopsis threshes free (Briggs, 1978; Reid, 1985). In wild barley, which carries two-rowed spikes, the lateral florets are sterile, yet visible, whereas in cultivated barley both two-rowed and six-rowed can be found, the latter with fertile lateral florets (Komatsuda et al., 2007). Each two-rowed spike may carry 15-30 kernels, whereas six-rowed varieties show 25-60 kernels per spike, in average (Briggs, 1978; Gomez-Macpherson, 2000). The caryopsis is oval, ridged, with rounded ends, and can be of different colors (Blattner, 2009). It is enclosed by the lemma and the palea, with the rachilla attached (Reid, 1985).

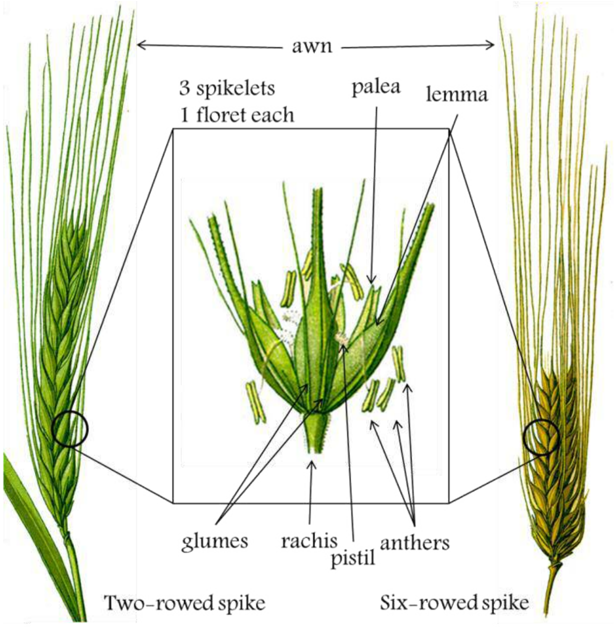


Figure 1.2. Barley inflorescences. Complete two-rowed (left) and six-rowed spikes (right) are shown. In the center, three spikelets are shown, one central, and two laterals. The latter are only developed, and fertile, in six-rowed spikes. Adapted from public image at https://commons.wikimedia.org/wiki/File:Illustration_Hordeum_vulgare1.jpg

Barley **development** (Figure 1.3) will be introduced in this work divided in two main stages, pre-anthesis development and anthesis (or flowering), as in Drosse et al. (2014), as this division is crucial for the agronomic features of the crop. Pre-anthesis development in temperate cereals has been divided into three phases based on morphological changes of the shoot apical meristem: the vegetative phase, the early reproductive phase and the late reproductive phase (González et al., 2002). During the vegetative phase, the seeds germinate, seedling roots emerge, and the coleoptile starts growing. Primary or seminal roots grow from the coleorhiza, branching and producing root hairs, whereas adventitious roots grow out of the crown (Reid, 1985). Once the coleoptile reaches the soil surface, the initiation of leaves and tillers is produced, and the vegetative phase continues until floral initiation, when the first reproductive *primordium* is formed (González et al., 2002). Whereas the main stem comes from the coleoptile, tillers arise from the lateral buds of that first culm, from the axils of lower leaves. This so called tillering stage is critical for the potential number of ears and grains, and fertilization during this period is decisive to set a maximum yield (Gomez-Macpherson, 2000). In the early reproductive phase, all the spikelets differentiate, until the formation of the terminal spikelet, when a few florets have differentiated. Floral initiation occurs first in the main culm and subsequently in the tillers. During the late reproductive phase, when the stem internodes elongate, the floret primordia reach their maximum number and then reach maturation (González et al., 2002; Drosse et al., 2014). As the spike grows in size within the flag leaf sheath, this last leaf of the stem undergoes swelling, a process which is known as booting (Gomez-Macpherson, 2000). During this process, some florets degenerate, while others reach the fertile stage at anthesis (Drosse et al., 2014). Afterwards, the ear emerges after the awns, an event recorded as an important agronomic trait called heading date.

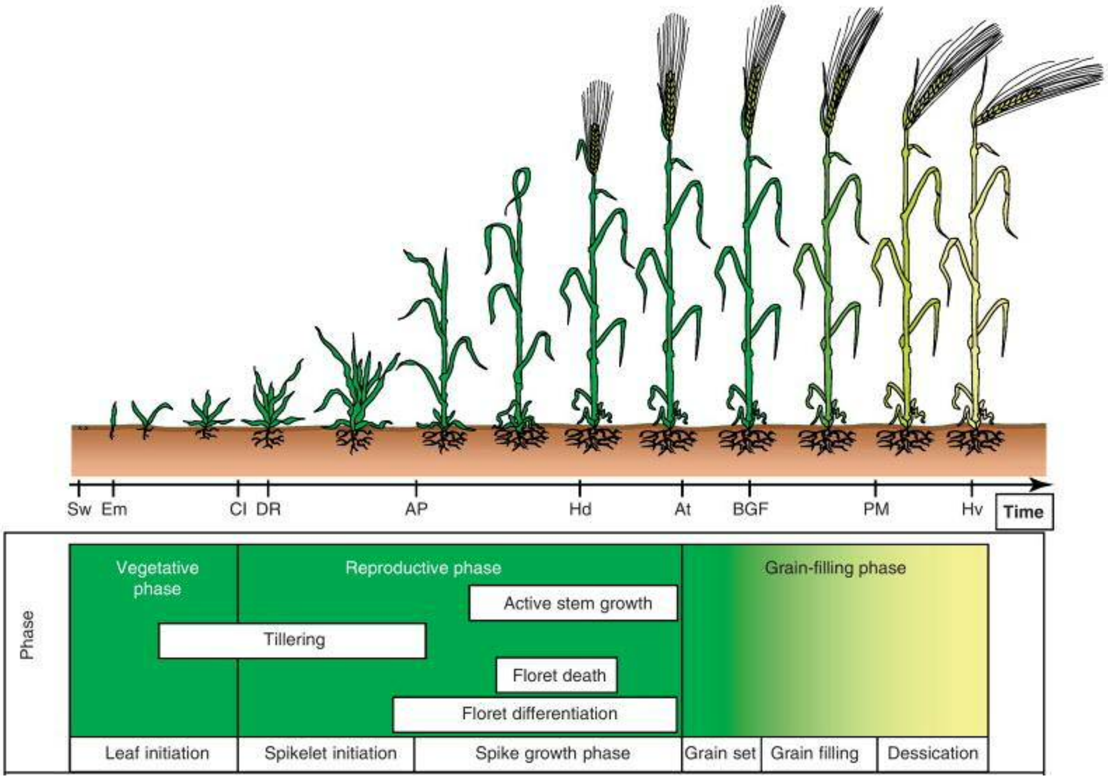


Figure 1.3. Summary of barley development. The typical aspect of the plant is shown on top, throughout development. The bottom diagram shows different processes, which take place during the main three development phases. AP, awn primordium; At, anthesis; BGF, begin grain filling; CI, collar initiation; DR, double ridge; Em, seedling emergence; Hd, heading time; Hv, harvest; PM, physiological maturity; Sw, sowing. Adapted from Sreenivasulu and Schnurbusch (2012), with permission (Elsevier license 3944180388212).

The duration of the vegetative phase, stem elongation and flowering time are affected by **environmental cues**. Barley is sensible to daylength, with number of leaves on the main shoot increasing under short days and reducing with long days (Wych et al., 1985). Daylength combines with temperature, and both interact with genotype, to determine the duration of the vegetative phase. The effect of temperature in flowering time is related with the accumulation of time exposed to low temperatures, a feature of temperate cereals which is known as vernalization (Griffiths et al., 1985). In contrast, the stem elongation phase is most sensitive to changes in photoperiod (Slafer et al., 2001). Genetic variation in both vernalization and photoperiod pathways was crucial for the successful expansion of barley cultivation from the Fertile Crescent to temperate climates (Drosse et al., 2014).

After the events from the late reproductive phase, the first stamen appears, and a new process, **flowering or anthesis**, commences. It takes about two days until all flowers are open. Barley florets open when the lodicules swell and force the lemma and palea apart. Then, the filaments of the three anthers elongate rapidly between the lemma and palea. There are barleys in which the lodicules cannot separate the lemma and palea. On these, cleistogamy, self-pollination within each single flower, takes place (Reid, 1985). After fertilization, the ovary continues to grow and differentiate, to become the barley kernel. Then, grain formation occurs, a phase known as grain filling, which is important for yield and for industrial quality. Such phase ends when the grain dries up, reaching maturity (Gomez-Macpherson, 2000). During this process barley plants senesce, drying and acquiring the yellow appearance typical of fields about to be harvested. A last process takes place, in wild barleys only, in which the brittle rachis disarticulates, and spikelets are excised from the plant, ready to be transported and germinate, when the right conditions show up.

1.1.3. Origin, domestication, and gene pools

Hordeum vulgare subsp. *spontaneum* (C. Koch) Thell., **wild barley**, is the ancestral form of cultivated barley (Bothmer and Komatsuda, 2011). The evolution of this wild plant in the Near East resulted in a complex biological specialization across the species range, which is associated with a large genetic diversity (Sato et al., 2014). This diversity facilitated morphological, physiological and functional adaptability to colonize primary and secondary habitats throughout the Fertile Crescent and in a range of most diverse environments (Graner et al., 2003). This subspecies is distributed in the eastern Mediterranean area, including parts of Greece, Turkey, Libya and Egypt, extending to the east up to West Pakistan (Bothmer et al., 2003).

Barley was one of the first domesticated cereals (Zohary et al., 2013). **Domestication** happened in the Fertile Crescent area of the Near East, and started about 10,000 years Before Present (BP), when mankind started to switch from hunter-gathering to cultivation as main food supply activity (Badr et al., 2000). As genetic discontinuity was observed between the Fertile Crescent and central Asia, the latter was proposed as a second origin of barley domestication (Bothmer and Komatsuda, 2011). The domestication process narrowed the diversity introducing a bottleneck, being wild barley a source of diversity for its cultivated form (Sato et al., 2014). This process fixed a series of agronomically valuable haplotypes. Some of the most relevant were early selected, and include the non-brittle rachis, the number of fertile florets in the spike, the flowering time or the hull type of caryopsis (Bothmer and Komatsuda, 2011; Sato et al., 2014).

Wild barley has **brittle rachis**, which promotes seed dispersal, whereas cultivated barleys have tough non-shattering rachis, preventing grain falling before harvesting of the spikes

(Bothmer and Komatsuda, 2011). That difference allows differentiating subsp. *spontaneum* from subsp. *vulgare* seeds in archeological grain specimens, by inspection of the disarticulation scars. The earliest remains of the *vulgare* subspecies, dated to 9500-8400 BP, were found in admixtures with subsp. *spontaneum* grain (Komatsuda, 2014). Two main genes are involved in the brittle rachis trait, *Btr1* and *Btr2* (Pourkheirandish et al., 2015), related with thickness of cell wall in the "constriction groove" were disarticulation occurs, with mutation in any of them causing the tough rachis which avoids grain falling (Komatsuda, 2014).

Regarding the **number of fertile florets** in the spike, we can differentiate barley with sterile lateral florets (also known as two-rowed barleys), and those in which the lateral florets are fertile and produce grain (called six-rowed barleys). The first is the exclusive phenotype in wild barley, and therefore it could be the ancestral form (Bothmer and Komatsuda, 2011). The latter, with fertile lateral spikelets, arose around 8,800-8,000 BP, as an important part of the domestication process (Komatsuda et al., 2007). The advantage of six-rowed type would not reside in grain yield, since although they produce three times as many grains as the two-rowed spike, they tend to tiller less freely and their grains are lighter on average (Komatsuda, 2014). Two-rowed barleys have better kernel performance, with high thousand kernel weight, lower protein content, and higher starch content. Preference of cultivation of six and two-rowed barleys is mostly due to historical reasons in the different countries (Verstegen et al., 2014). Six-rowed spikes are consequence of the loss of function of a transcriptional repressor gene *vrs1* (Komatsuda et al., 2007), which is expressed only in the lateral spikelets, while immature, and not in the central ones. Analysis of DNA sequences of the *vrs1* gene revealed different origins for six-rowed barley (Bothmer and Komatsuda, 2011). The six-row trait has appeared several times during barley cultivation, and can be used to trace barley spread throughout the world.

The spread of barley into different agricultural environments required adaptation of timing of flowering, which responds predominantly to day length and temperature (Cockram et al., 2011). Modulation of **flowering time** enables plants to optimize the use of the available resources in the place they grow (Laurie, 1997). Wild barleys have winter-habit, which means that they need vernalization, that is, the induction of the reproductive stage by exposure to a prolonged period of cold. The mutations required for the loss of the winter habit are thought to have occurred post-domestication (Saisho et al., 2011). As a result of selection, vernalization requirement in cultivated barleys ranges from winter to spring habit barleys. In the latter, flowering begins even without a period of cold. This range includes facultative barleys, with frost tolerance and a minimum vernalization requirement which can be sown either in autumn or spring, and intermediate barleys, with requirement of not-so-prolonged periods of cold, adapted to areas of mild winters (Casao et al., 2011a). Winter and intermediate barleys are sown in the autumn, and can withstand temperatures as low as -20 ^{o}C, whereas spring barleys do not require vernalization, and show a broad adaptation to different environments (Verstegen et al., 2014). These are sown when the cold period has ended, the end of winter in the Mediterranean or spring in the UK and northern Europe, allowing barley cultivation in higher latitudes, where winter cold would be harmful for seedlings. Winter varieties have a yield advantage, due to their longer growing season, but they cannot be cultivated in areas with very long periods of below zero temperatures (Verstegen et al., 2014), whereas spring varieties cannot be grown where the summer is too hot and dry to allow proper grain filling (Bothmer et al., 2003). Therefore, cultivation of winter or spring barleys is chosen depending on climatic conditions. The genetic control of vernalization relies on the genes *VrnH1* and *VrnH2*. Mutation of any one of these genes is sufficient to abolish the vernalization requirement (Komatsuda, 2014). In addition to temperature, flowering time also depends on photoperiod (Laurie, 1997). The expansion of barley into higher latitudes required lowering photoperiod sensitivity, since wild barleys require a 12 hours photoperiod to trigger the switch to reproductive stage (Komatsuda, 2014). Photoperiod sensitivity is affected under long day conditions by the *PpdH1* gene (Turner et al., 2005), and by *PpdH2*, under short days (Laurie et al., 1995; Szucs et al., 2006; Casao et al., 2011b). A mutation of the wild (sensitive) *PpdH1* allele was needed to allow spring cultivation and expansion of barley to central and northern Europe (Jones et al., 2008). Integration of photoperiod response and vernalization pathways is modulated by the gene *VrnH3*. The timing and strength, of the signals reaching this gene, produce an interaction which determines flowering time (Trevaskis et al., 2006).

Hull-less, naked or free-threshing barleys, those where the **hull** does not adhere to the caryopsis at maturity, are cultivated in many parts of the world, in particular in East Asia, Tibet, Nepal, India and Pakistan (Bothmer et al., 2003). Hull adherence depends on the formation of a lipid layer between the pericarp epidermis and the hull, and naked types date to around 8,000 BP (Komatsuda, 2014). This trait is controlled by the recessive gene *nud*, having naked barleys a large DNA deletion which includes an ethylene response factor (Taketa et al., 2008).

Other domestication traits affected the seed, which is the main product obtained from barley cultivation, including a reduced degree of dormancy, and a major increase in seed size and number (Komatsuda, 2014). **Dormancy**, which facilitates delaying germination in wild barley until favorable conditions are ensured, is a problem for cultivated barley, both in crop establishment and for the malting process, and therefore has been greatly reduced in domesticated barley. However, stringent selection against dormancy could increase pre-harvest sprouting, that is, germination of the seed while still on the mother plant (Prada et al., 2004). The main genes which have been associated with seed dormancy are SD1, which encodes an alanine aminotransferase (Sato et al., 2016b), and SD2, encoding mitogen-activated protein kinase kinase 3 (Nakamura et al., 2016), both on chromosome 5H. Hormones, like ABA and GA, are also involved (Komatsuda, 2014).

Due to the bottleneck produced during domestication, many polymorphisms are absent from elite varieties of most crops. This general statement is also true for barley, in which only a few landraces were the ancestors of modern European barley breeding germplasm (Melchinger et al., 1994). Therefore, **genepools** have yet to be fully exploited, either through classical breeding or aided by genetic engineering techniques, as a source of useful genes for barley improvement. In summary, barley gene pools can be classified as primary, secondary and tertiary (Bothmer et al., 2003). The tertiary gene pool includes all species of *Hordeum*, to which crossing is difficult and backcrossing almost impossible (Bothmer et al., 1983). The secondary gene pool includes *Hordeum* species whose gene transfer is possible but difficult in practice. This pool includes a single species, *H. bulbosum*, which shares the H genome with barley (Blattner, 2009) and crosses with some difficulty (Sato et al., 2014). The primary gene pool comprises cultivated barley (including elite cultivars or varieties, breeding lines and landraces), and wild barleys, in which gene transfer by crossing is easy.

Landraces are part of the primary genepool, and are the result of continuous multiplication of a population of a crop, once reached the equilibrium under a specific set of environmental conditions (Fischbeck, 2003). They have very rich and complex ancestry representing variation in response to many diverse stresses, and are vast resources for the development of future crops deriving many sustainable traits from their heritage (Newton et al., 2010). Barley landraces are still cultivated in Asia and North Africa, and have been used until recently in other areas , from coastal to mountainous regions (Fischbeck, 2003). In most places, landraces were replaced in a short time during the early decades of the twentieth century, and their diversity have been largely lost (Fischbeck, 2003). Nonetheless, others were collected, and some of the diversity was preserved, with an advantage in those regions were their replacement was delayed, like Spain or Italy (Sato et al., 2014; Casas et al., 2016).

Wild barley is adapted to a broad range of environments, including stable populations in deserts as well as in cold regions in Tibet, and represents a potential source of adaptive genetic diversity against abiotic and biotic stresses (Nevo and Chen, 2010). For example, populations in the Fertile Crescent have considerable genetic variation between populations, which is reflected in differences in physiological characteristics (Ellis et al., 2000).

Summarizing, wild barleys and landraces are thought to carry many polymorphisms which are absent from current barley cultivars, and the challenge is to make this variation available for crop improvement (Ellis et al., 2000).

The plant material studied in this work corresponds to **Spanish barley landraces**, which are now dedicated a more detailed explanation. In Spain, barley has been cultivated for at least 7,000 years, according to archaeological evidence (Zapata et al., 2004). Therefore, barley could have developed specific adaptations to the local environmental conditions. In the past century, more than two thousands of these landraces were collected, prior to the extensive introduction of modern varieties in the country, and maintained, along with a lower percentage of modern varieties, in the Spanish National Germplasm Bank (BNG). Many of those landraces were six-rowed, as this was the predominant type of barley traditionally grown in Spain, but there were also two-rowed barleys. From these set of genotypes, a Spanish barley core collection (SBCC) of 175 entries, 159 of them from local landraces, was developed (Igartua et al., 1998) to facilitate the exploration and utilization of their genetic diversity in breeding programs. In Figure 1.4, six- and two-rowed spikes, and grains, from a few Spanish accessions, are illustrated.

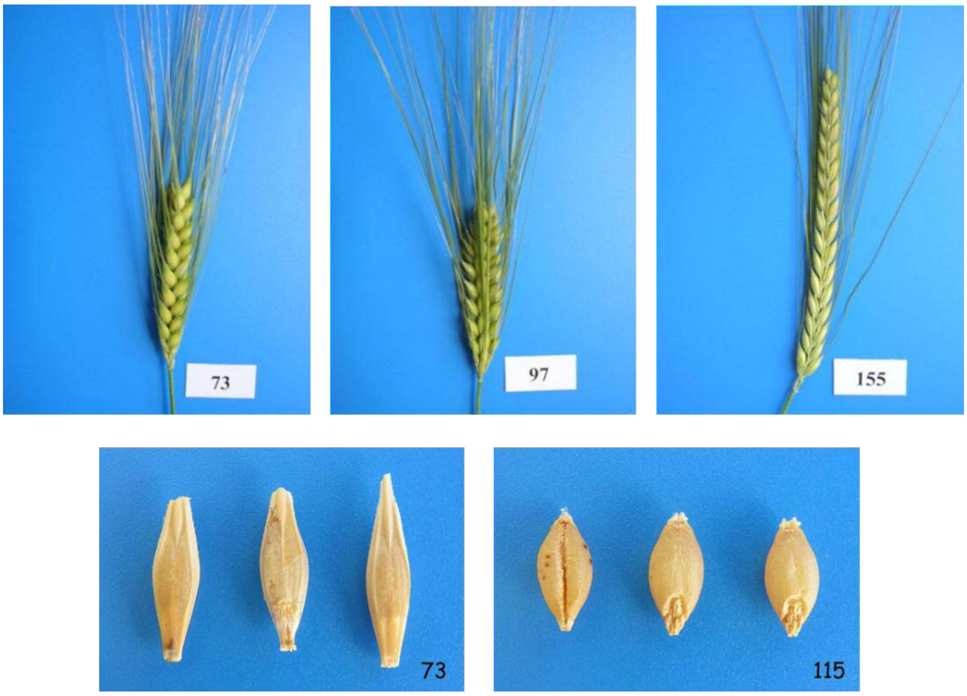


Figure 1.4. Spikes and caryopses from several Spanish barley landraces. Six-rowed spikes from landraces, SBCC073 (top left) and SBCC097 (top center), both studied in this work, are shown. A two-rowed spike is also shown (top right). Grains from SBCC073, a hulled barley, are shown (bottom left), along with hull-less, or naked, grains from SBCC115 (bottom right).

The availability of this compilation led to the first systematic genetic and morphological evaluations of the Spanish barley germplasm. The genetic singularity of Western Mediterranean barleys, including the Spanish ones, had been already highlighted (Tolbert et al., 1979). However, the **origin** of both two-rowed and six-rowed barleys in Spain remains to be revealed, and their **diversity** has started to be uncovered in the last decade. It was first proposed that Spanish barleys came from Moroccan wild barleys (Molina-Cano et al., 1987; Moralejo et al., 1994). Those Moroccan genotypes were stated to be weedy forms and segregation products, and not true wild forms (Badr et al., 2000; Bothmer and Komatsuda, 2011). Morphological and agronomical evaluation of the Spanish entries of the SBCC revealed a clear distinction between two- and six-rowed cultivars, and also between landraces and commercial varieties (Lasa et al., 2001). Genetic analyses led to suggest that six-rowed Spanish landraces derive from two different ancestral sources (Casas et al., 2005), and were more distant to the mainstream breeding genepool than Spanish two-rowed barleys (Yahiaoui et al., 2008). These populations were distributed according to geographic and climatic factors in the Iberian peninsula, with Spanish spring two-row barleys present in inland Northern Spain, a large group of Spanish six-row barleys from the warm areas of the South and the Mediterranean coast, and another large group of Spanish six-row barleys from the cooler highlands in the center of the peninsula (Yahiaoui et al., 2008). These Spanish barleys showed a significant grade of diversity, which could be related with genetic drift and with selection for adaptation to local constraints (Yahiaoui et al., 2008). This wealth of genetic diversity has been reflected in evaluations aimed to identify novel traits and donors for disease resistance (Silvar et al., 2010) and for abiotic stress tolerance (Yahiaoui et al., 2014).

1.1.4. Barley genomics

Barley is a true diploid, self-fertile, with a low number (2n=2x=14) of relatively large chromosomes (Taketa et al., 2003). The seven chromosomes are more or less metacentric, with five chromosomes without satellites (1H, 2H, 3H, 4H and 7H), very similar in length and arm ratios (Graner et al., 2011). The short arms are designated by the letter "S" and long arms by the letter "L" (for example, 7HL and 7HS for both arms of chromosome 7H) (Taketa et al., 2003).

The first barley **genetic maps** were based on morphological and disease resistance-based loci (Graner et al., 2011). The molecular age brought the publication of whole-genome maps using combination of restriction fragment length polymorphism (RFLP) (Kleinhofs et al., 1988) and polymerase chain reaction (PCR) (Shin et al., 1990) methods. RFLP-based markers were followed by faster and cheaper, not always so reliable, technologies, like random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-specific amplified polymorphisms (S-SAP), and simple sequence repeats (SSRs) (Graner et al., 2011). SSRs became the favorite of plant breeders for marker-assisted selection (MAS), due to its ease of use, co-dominant and multi-allelic nature, abundance in barley genomes, and transferability among diverse crosses (Kota et al., 2001). All those genotyping platforms were accompanied by their corresponding consensus genetic linkage maps, derived from different mapping populations (Graner et al., 2011). The next significant step was achieved by the use of new technologies, which increased the magnitude of markers from hundreds to thousands, including diversity array (DArT) markers and single nucleotide polymorphism (SNP)-based genotyping platforms (Graner et al., 2011). The first were mostly derived from actively expressed sequences, thanks to the use of PstI as restriction enzyme to get reduced genomic representation (Jaccoud et al., 2001). DArT markers were based on DNA hybridization, achieving around 2,000 polymorphic markers (Wenzl et al., 2006). SNPs in barley were mostly derived from expressed sequence tags (EST) sequences, obtained by traditional Sanger sequencing, and therefore they were also associated mostly with complementary DNA (cDNA), derived from coding sequences (CDS).

The motivation to develop low- or single-copy genetic markers, coupled with technical advances, derived in the first whole-genome scale **sequencing** efforts in barley. Over 500,000 barley ESTs, from cDNA 5′ and 3′ RNA ends, were obtained, and assembled in consensus sequences (also known as Unigenes). This led to the development of the first software platforms, to provide access to those resources, like the widely used HarvEST (Close et al., 2007). Genotyping platforms to exploit the availability of those SNPs were also made available, including the Illumina GoldenGate SNP assay, with several pilot assays, namely BOPA1 and BOPA2 (Close et al., 2009). Moreover, these sequence resources were used to design the Affymetrix 22K Barley1 GeneChip (Close et al., 2004) microarray to assess gene expression, which has been broadly exploited by the barley community (Stein, 2014). Later, full-length cDNA sequences (flcDNA) were obtained for barley cultivar Haruna Nijo (Sato et al., 2009; Matsumoto et al., 2011). This effort provided access to most exons of over 25,000 genes, further facilitating marker development (Thiel et al., 2003; Varshney et al., 2007). It served also the establishment of a new SNP-based genotyping platform, the Illumina

Infinium iSelect microarray (Comadran et al., 2012), which achieved almost 8,000 SNP markers. These flcDNA sequences have also been exploited for annotation of genome sequences (Mayer et al., 2012; Sato et al., 2016a).

The availability of full sequenced genomes, in the first decade of this century, propelled breakthrough advances in *Arabidopsis* (*Arabidopsis thaliana* L.) and rice (*Oryza sativa* L.) research (Bolger et al., 2014). Gaining access to genomics tools for many other plants, especially for important crops, became a major goal of their respective research communities. This goal was delayed for **barley genome**, by both its size and its redundancy (Feuillet et al., 2011). The seven barley chromosomes are estimated to contain 5.1 billion base pairs (Mayer et al., 2012). Around 80% of them correspond to repetitive DNA (Wicker et al., 2009). Nonetheless, the development of sequence resources has progressively provided key insights into the barley genome, while delivering new opportunities and perspectives for their application in the context of barley crop improvement (Stein, 2014). Moreover, the barley genome exhibits good marker order conservation, or synteny, with the other members of the *Triticeae* tribe as well as with rice, maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) (Graner et al., 2011). Therefore, while the complete reference genome of barley was not available, other related species were used as genomic models for the *Triticeae*, including rice and ***Brachypodium distachyon***. The latter is especially relevant as genomic model for temperate cereals, since it is closer to the *Triticeae* than to rice and maize (Vogel et al., 2006). It has a small genome of ca. 350 million base pairs (Mbp) (Huo et al., 2008), completely sequenced (International Brachypodium Initiative, 2010), with outstanding co-linearity with the *Triticeae* species (Bossolini et al., 2007). Moreover, it is a small-stature temperate grass, with self-fertility, rapid generation time, and simple growth requirements (Draper et al., 2001); and it is readily transformable (Garvin et al., 2008).

Despite largely hindering whole-genome sequencing, the large size of chromosomes of the *Triticeae* has an advantage. It allows using flow-cytometric sorting, a technique to isolate large chromosomes or chromosome arms (Doležel et al., 2012). This can be used to develop chromosome-specific resources (Doležel et al., 2007), and has multiple applications, including chromosome sequencing using high-throughput sequencing technologies (Doležel et al., 2012). Sequenced flow-sorted chromosome arms, coupled with synteny information (Figure 1.5), allowed obtaining 21,000 linearly ordered barley genes (Mayer et al., 2011). This resource, termed **genome-zippers**, serves as a genomic tool for molecular marker development and fine mapping efforts through synteny, in studies with organisms lacking a sequenced genome.

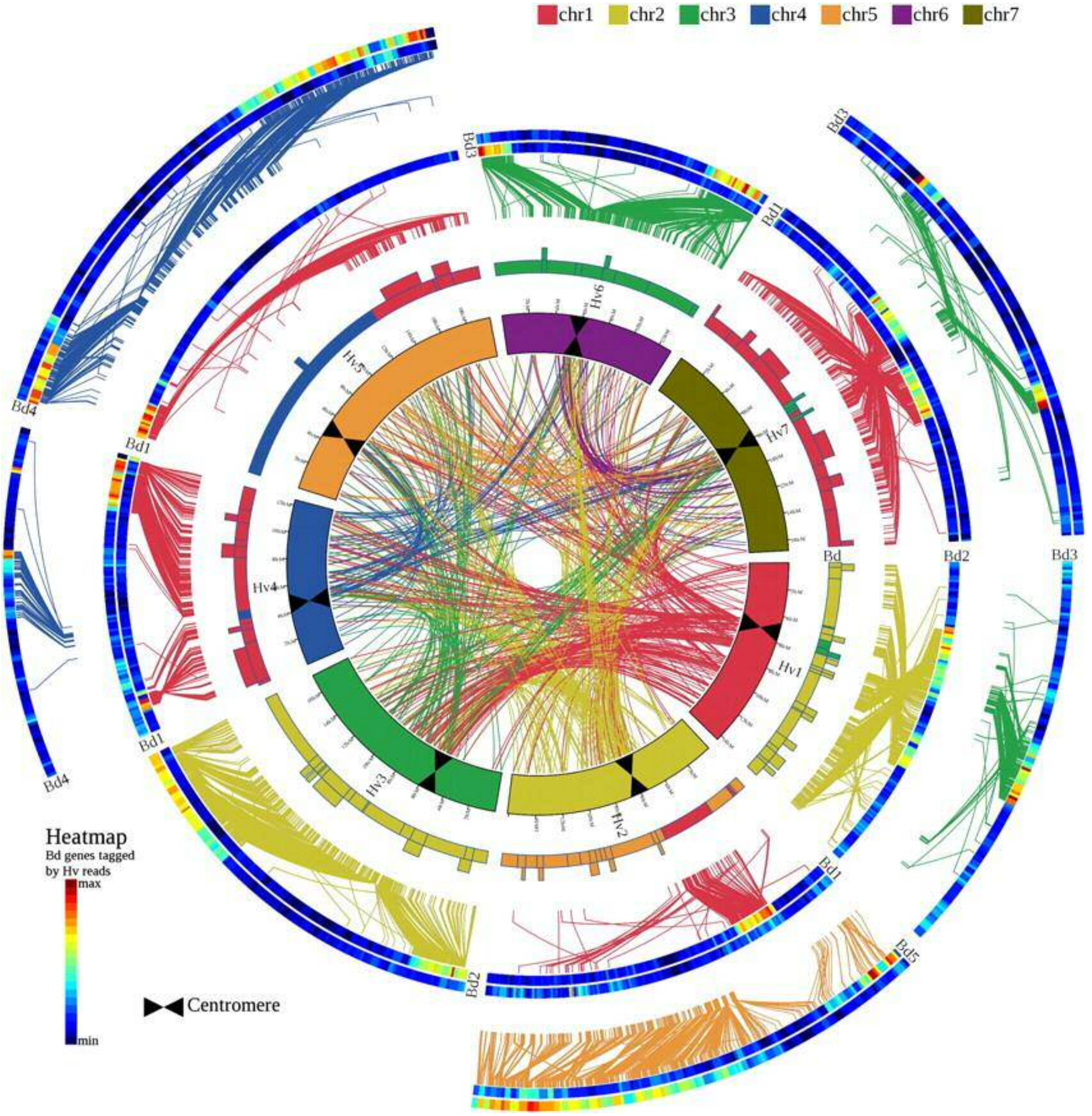


Figure 1.5. Comparative analysis between barley and *B. distachyon*. Synteny between both species is shown by the lines out of the inner circle, which link position of orthologous genes, on barley (Hv, colored) and *Brachypodium* (Bd, blue-to-red heatmap) chromosomes. The lines in the inner circle link positions, on barley chromosomes, of putative paralogous genes. Adapted from Mayer et al. (2011), with permission of American Society of Plant Physiologists (license 3944300161256).

In the way towards obtaining a **reference genome**, several genetic, molecular, and sequencing techniques were combined to develop the first barley physical map (Mayer et al., 2012). A dense genetic map (Comadran et al., 2012) was anchored to it, facilitating the association of sequence resources, in the form of sequenced bacterial artificial chromosomes (BAC) contigs, BAC-End sequences, whole-genome shotgun (WGS) contigs, obtained with high-throughput sequencing (HTS), and flcDNAs (Mayer et al., 2012). Population-based sequencing (POPSEQ) allowed anchoring further WGS contigs (Mascher et al., 2013a), and sequenced BAC contigs (Ariyadasa et al., 2014), to the physical map. This sequence-enriched physical map was also accompanied by draft sequence assemblies of WGS contigs from three barley cultivars (Barke, Bowman and Morex). The Morex WGS assembly was also enriched with annotation of gene models, by mapping transcript sequences obtained through high-throughput transcriptome sequencing (RNAseq). The development and release of these resources in 2012 boosted barley genomics, but they were not trouble-free. Lacking clone-by-clone sequencing of the minimum tiling path (MTP) of BACs, these assemblies were highly fragmented, presented functional and structural gene annotation of variable quality, and had abundant chimeric contigs. However, even access to partial genome sequence information is highly enabling for the development of new tools in applied crop research (Stein, 2014), as demonstrated by its use to develop new genomic tools for barley (Mascher et al., 2013b), as reference for other studies (Mascher et al., 2014; Pankin et al., 2014; Digel et al., 2015; Hübner et al., 2015; Cantalapiedra et al., 2016), or by the development of software and web services to facilitate accessing those resources (IBSC, 2013; Plant Genome and Systems Biology MIPS, 2013; The James Hutton Institute, 2014; Cantalapiedra et al., 2015; Colmsee et al., 2015; Kersey et al., 2016).

Further improvements of genomic resources in barley include an alternative sequencing of BACs from cultivar Morex (Muñoz-Amatriaín et al., 2015), and the WGS assembly of cultivar Haruna Nijo (Sato et al., 2016a). The latter represents a fourth sequenced genotype, and an improvement of the annotation of gene models. Finally, a new clone-by-clone sequenced MTP of cultivar Morex genome was recently made available to the barley community (M. Mascher, personal communication), and its description in peer-reviewed journals is imminent. This reference is expected to represent a first version of a barley finished genome, including larger contigs, longer and better assembled regulatory sequences and intergenic regions, accurate physical position of genes, and an improved annotation of gene models and isoforms. This new step will facilitate even further whole genome analyses, like genome-wide association studies, fine mapping efforts, and barley functional genetics and genomics. The availability of such genome will confirm barley as a genomic model plant for *Triticeae* research, and will enable breeders to develop new selection strategies, like genomic selection, which will accelerate barley improvement (Stein, 2014).

1.2. Breeding challenges and approaches

In recent decades, the productivity of barley has risen, due in part to genetic breeding progress. Yield increases have been accompanied with better yield stability, due to resistances against lodging, diseases, and insects (Friedt, 2011). Current breeding targets depend on the final use of the crop. For example, the quality aspect is the most important trait in malting barley, whereas starch, protein, and fiber content are important when directed towards livestock feeding. Yet, the main breeding target is grain yield, and barley breeders are challenged to develop new cultivars, allowing an economically viable production under increasingly unfavorable conditions (Verstegen et al., 2014). The main breeding targets for improving grain yield of barley are disease resistance and drought tolerance.

1.2.1. *Disease resistance*

Worldwide average yield losses, due to fungal and viral diseases, and insect pests, ranges between 20% and 30% (Weibull et al., 2003; Friedt, 2011). **Pathogen diseases** are battled against by cultivation of resistant varieties, combined with the use of appropriate agronomical practices. In many cases, resistant cultivars are the most cost-effective and environmentally benign means of controlling diseases (Paulitz and Steffenson, 2011). Moreover, improving barley resistances could reduce applications of chemicals, a general tendency which farmers must address (Friedt, 2011). In barley, the most important diseases, differing regionally and with season climate, are powdery mildew (caused by *Blumeria graminis* f.sp. *hordei*), speckled leaf blotch (caused by *Septoria passerinii*), scald (caused by *Rhynchosporium commune*), net and spot leaf blotch (caused by *Pyrenophora teres* f. *teres* and f. *maculata*, respectively), head blight (caused by *Fusarium graminearum* and *F. culmorum*), and stem rust and leaf rust (caused by *Puccinia graminis* f.sp. *tritici* and *Puccinia hordei, respectively*), all of them fungus; and barley yellow and mild mosaic viruses (BaYMV and BaMMV) (Schweizer, 2014).

The major challenge for breeders is obtaining **durable resistances**, ideally those which cope with a broad-spectrum of races from a given pathogen. Most pathogen species are composed of many races, and possess populations with swiftly changing dynamics (Brown, 1994; Wolfe and McDermott, 1994), capable of generating new virulence types at a rapid pace (Lee and Neate, 2007). Furthermore, under the dynamics of climate change, those pathogen populations may shift, and affect crops in regions in which their impact was traditionally limited.

Host resistance, in which only some genotypes of a plant species are resistant to a given pathogen, has been the primary means of controlling most barley diseases (Paulitz and Steffenson, 2011). For example, the effector-triggered immunity (ETI), traditionally associated to nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins, is a host resistance system for recognition of pathogen effectors, and effector-target complexes, which provides complete protection, but is usually race-specific and non-durable (Schweizer and Stein, 2011). These race-specific resistance genes, which operate through a gene-for-gene interaction against a particular pathotype (Flor, 1971), are often overcome by new pathogen races within a short period of time, through modification of effector proteins (Schweizer, 2014). Therefore, the use of one, or a few, resistant genes can lead to epidemics, due to "boom and bust" cycles (McDonald and Linde, 2002). Combining multiple resistance genes in a single cultivar (pyramiding or stacking of genes) is a sound approach for achieving a more stable resistance (Brown et al., 2001), since it might avoid strong selection of the pathogen (Brown et al., 1996). This strategy was only developed after molecular markers allowed genotypic differentiation of alleles, impossible to assess phenotypically. It requires the discovery of more genes recognizing conserved pathogen effectors, to be ultimately combined (Schweizer, 2014). Another means to exploit this kind of gene-for-gene resistance, in the near future, could be the generation of *in vitro* chimeric resistance genes, producing an artificial diversity which could be used to confer a broad spectrum of durable resistance (Paulitz and Steffenson, 2011).

However, as a means to avoid strong selection pressures on the pathogen, genes with partial resistance are better than those conferring complete resistance. Therefore, an alternative strategy to achieve durable resistance is the combination of **partial race-nonspecific resistances**. Partial resistance, was defined previously as "the resistance to epidemic built up" (Parlevliet and Ommeren, 1975). This incomplete protection depends on the allelic status of host genes, and operates against many races of a given pathogen species (Schweizer, 2014). Its durability and broad-spectrum allow increasing yield stability and sustainability, under field conditions. The difficulty of exploiting this kind of resistance lies in its polygenic nature, being inherited as several QTL, which depend on genotype-by-environment interaction (Schweizer and Stein, 2011).

A third type of resistance of plants against pathogens is called **nonhost resistance**. It corresponds to the resistance of entire plant species against the major part of existing pathogens (Heath, 2000). Indeed, most plant species are susceptible only to a few pathogens, considering the large list of potentially harmful diseases. It is unclear why a pathogen virulent on one species is nonpathogenic on others (Mysore and Ryu, 2004). Several mechanisms, and plant and cellular components, have been described to be involved in nonhost resistance (Gill et al., 2015). However, the molecular mechanisms involved, and the mode of inheritance of nonhost resistance, are under debate (Niks and Marcel, 2009; Schulze-Lefert and Panstruga, 2011; Niks et al., 2015). The few exceptional examples of single genes conferring long-lasting, broad resistance resistances, as *Rpg1,* against stem rust (Steffenson, 1992; Brueggeman et al., 2002), *mlo,* against powdery mildew (Jorgensen, 1992), and the line NDB112 to spot blotch (Steffenson et al., 1996), are examples of nonhost resistance (Humphry et al., 2006; Gill et al., 2015). However, other durable, broad resistances must be identified. Studying nonhost resistance is essential to understand plant defense mechanisms, and it was envisaged as a means for plant breeders to increase durability of disease resistance within host species (Heath, 2000). Moreover, the outcomes from nonhost resistance range from immunity to partial resistance, with varying degrees of efficacy (Bettgenhaeuser et al., 2014), which could be exploited in different breeding strategies, like those relying on partial resistance.

1.2.2. Drought tolerance

One of the major challenges for the present century is to provide food to an increasing worldwide population. To achieve this, enhancing crop yield, and yield stability, is essential. Breeding for yield requires conferring on crops tolerance to **abiotic stresses**. These stresses are already harmful in different regions worldwide, and include drought, heat, soil with excess of salt, cold, flooding, toxic substances, and shortage of mineral nutrition (Ceccarelli et al., 2004). Occurrence, severity, timing, and duration of stresses are different between regions, and vary from season to season. They seldom occur alone (Cattivelli et al., 2002), and are especially harmful under semiarid and drought-prone areas (Kishor et al., 2014). As with diseases, abiotic stresses faced by agronomists will change, or its impact may be aggravated, due to global change. Therefore, coping with abiotic stresses will require adaptation of agronomy in each region, including sowing different crops, or adapting the current ones to the new conditions (Cattivelli et al., 2011).

Drought is the most important abiotic stress (Boyer and Westgate, 2004), causing the greatest yield losses, both in developed and developing countries (Cattivelli et al., 2011). In the past century, genetic gain of yield, in absolute terms, and genetic progress have been less in regions suffering from drought stress (Slafer et al., 1994). A key challenge is to improve drought tolerance without limiting yield potential, and thus QTLs for stress-related traits coincident with QTLs for yield potential should be considered as priority targets for breeding (Cattivelli et al., 2011). In the Mediterranean areas, terminal drought, which takes place during the reproductive development of the plants, is especially relevant, due to irregular rainfalls, and hot and dry springs and summers (Ceccarelli et al., 1991; Kishor et al., 2014). In such regions, barley is one of the main crops. Therefore, improving its drought tolerance is a sensible breeding target. Fortunately, barley germplasm holds a high degree of genetic variability for stress tolerance (Stanca et al., 2003).

Among such diversity, it is important to recognize those features which could actually contribute to improve the performance of crops in the field. Not all the **strategies** which are effective from an adaptive point of view, for survival and successful reproduction of the individual, are suitable for breeding. The strategies shown by plants to cope with stress can be summarized in escape, avoidance (or resistance), and tolerance (Levitt, 1972; Mitra, 2001).

Escape is mainly related with adjusting (generally shortening) the life cycle of the plant, to avoid the most harmful hot and dry periods. In winter cereals, the plant anticipates flowering, which is reflected in lower measurements of flowering time (phenological measure) and heading date (agronomical measure). Variation among genotypes exists, and escape is an important strategy of genotypes adapted to Mediterranean conditions. However, earlier anthesis usually leads to lower potential yields. Therefore, too much earliness can be detrimental in the long run. Appropriate phenology for a region must take into account frequency and severity of terminal drought stress (Levitt, 1972; Mitra, 2001).

In turn, **avoidance** involves changes which the plant undergoes to maintain high tissue water potential. For example, closure of stomata, to reduce gas exchange, and avoid water loss through evapotranspiration, is a response often seen in leaves of plants under stress. However, a lower stomatal conductance implies lower respiration rates and reduced assimilation of carbon dioxide. It can lead to uncoupling of photosynthesis and carbon fixation rates, and over-heating of the photosynthetic apparatus, especially when drought turns up along with heat, which is very common in the field (Ceccarelli and Grando, 1996). Ideally, photosynthesis could be engineered to adjust it to environmental conditions, but this is as yet not possible (Blum, 2009; Ming et al., 2015). The reduced leaf photosynthesis could be compensated by remobilization of reserves for grain filling, which has been proposed as a criterion to select drought resistance genotypes (Blum, 1988; Slafer et al., 2005). Also, protection against active oxygen species (Reddy et al., 2004), which are a byproduct of altered metabolic processes, as excessive excitation energy in photosynthesis, is important in this kind of responses.

In contrast with resistance, drought **tolerance** is the ability to withstand water-deficit with low tissue water potential. Osmotic adjustment (Moinuddin et al., 2005) and effective use of water (Blum, 2009) are often associated to drought tolerance. Osmotic adjustment is achieved through accumulation of solutes (Serraj and Sinclair, 2002; Nguyen et al., 2004). It enables plants to maintain water absorption and cell turgor pressure, leading to sustained photosynthetic rate, and expansion growth (Ali et al., 1999). Effective use of water implies enhanced moisture conservation and acquisition, to be used for transpiration. It is favored improved water uptake provided by both osmotic adjustment and deep root systems (Blum, 2009). Deep roots are especially useful with terminal drought (Mitchell et al., 1996; Kirkegaard et al., 2007), and thus it is an interesting mechanism for winter cereals in the Mediterranean region. Valuable variation between genotypes could be found in development and architecture of roots, (Johnson et al., 2000; Nguyen et al., 2004), composition of the cuticle, permeability of the epidermis, regulation of expression and function of transporters of molecules (e.g. aquaporins), composition and restructuring of cell walls, and modulation of lipid content of cellular membranes (Xiong et al., 2002). Also, preserving proper folding of proteins is important, to maintain their optimal performance under abiotic stress (Wang et al., 2004). Those genotypes which mostly show tolerance strategies could be the most desirable target for breeders aiming to obtain better yield stability.

Drought tolerance has often been described as **a complex trait**, and, indeed, the molecular mechanisms of the response of plants to abiotic stress are still unknown. This, together with the gap between laboratory and field research, could be an explanation for the delayed development of drought-tolerant varieties compared to other traits (Yang et al., 2010). Nonetheless, single genes, as those controlling flowering time, plant height, ear type, and osmotic adjustment, may have important roles in the adaptation to drought-prone environments (Cattivelli et al., 2011). There are examples of successful improvement of abiotic tolerance of crops (Blum, 2011), by classical breeding (Rebetzke et al., 2002), by QTL introgression and marker-assisted selection (Courtois et al., 2003; Ribaut and Ragot, 2007), or by alteration of expression or transformation of single genes, in a few occasions with beneficial effects in the field (McKersie et al., 1996; Bahieldin et al., 2005; Hu et al., 2006; Xiao et al., 2007). In barley, QTLs related with drought stress have been identified working with mapping populations under different environments (Teulat et al., 2001; Baum et al., 2003; Diab et al., 2004; Talamé et al., 2004; Korff et al., 2008; Boudiar et al., 2016), and through association mapping (Comadran et al., 2011; Wehner et al., 2015). However, the meaningful advantage of these loci in the field has not been demonstrated. Moreover, the outcomes from studies based on gene expression, proteomics or metabolomics, show different results depending on the plant material used, the tissue and development stage assessed, and the mode of application and magnitude of the stress (Shaar-Moshe et al., 2015). Nonetheless, there are some key processes which appear to be often involved in responses to drought stress (e.g. heat-shock proteins, and abscisic-acid metabolism and signaling), and several fundamental signaling mechanisms are quite conserved among plant species (Nakashima et al., 2014; Gürel et al., 2016). This could facilitate transferring the knowledge gained in the model plants to crop species (Kishor et al., 2014).

1.2.3. Looking for stress tolerance genes

Discrimination of **susceptible and resistant genotypes** is required to perform more detailed analyses, with the goal of improving stress tolerance. Genotypes resistant to diseases are usually identified from screening impact on plants cultivated in the field, or by direct inoculation of isolated spores in the greenhouse, either to test for resistance to specific pathogens or as validation of field results (Silvar et al., 2010; Vasudevan et al., 2014). For abiotic stresses, tolerant genotypes have been traditionally identified empirically, and later by more sophisticated phenotype-based statistical analyses under stress. More recent statistical approaches, integrating environmental and genetic information into models, aim to identify key variables to estimate sensitivity and heritability of abiotic stress tolerance (Cattivelli et al., 2011).

After identifying stress-tolerant genotypes, the work must continue with efforts to locate **genomic loci or candidate genes** responsible of the tolerance. Approaches for this include the development of mapping populations, from crosses between contrasting genotypes, or association-mapping studies with collections of unrelated breeding lines. With the access to markers covering whole genomes, genome wide association studies (GWAS**)** were adopted in different plant species, to overcome some of the limitations of bi-parental linkage mapping, such as the limited genetic diversity assessed (Rafalski, 2002; Huang and Han, 2014). However, GWAS also has difficulties, including heritability of the trait under study, linkage disequilibrium levels, population structure, quality of phenotypic data, and sample size; all of which can affect resolution or validity of the detected associations (Korte and Farlow, 2013). Specifically, barley is a selfing species, with higher levels of LD than other species (Pasam et al., 2012), like maize, sorghum, and even other self-crossed species like rice. Moreover, some of the collections of genotypes used for GWAS in barley showed considerable population structure (Comadran et al., 2009). Several new crossing schemes of populations have been suggested to overcome the limitations of GWAS, including multi-parent advanced generation intercross (MAGIC) populations (Sannemann et al., 2015), and nested association mapping (NAM) populations (Barabaschi et al., 2016).

The previous methods provide clues about the genomic region in which the genetic features responsible for the trait of interest are located. Subsequent **fine mapping** is important, to clone the actual gene or genes which cause the different responses seen in the assessed genotypes. Cloning of genes has traditionally consisted on laborious cycles of population development, and identification of molecular markers, to narrow down the genomic segment containing the candidate gene. Numerous offspring lines had to be genotyped, looking for recombinant lines, and evaluated for the trait of interest, to locate recombinants with contrasting phenotypes. The availability of increasingly dense markers, from RFLPs to SSRs to SNPs, also has facilitated saturating genomic loci for fine mapping purposes (Stein et al., 2007). In the recent years, powerful techniques, like sequencing-based bulk segregant analysis or mapping-by-sequencing, are allowing to shorten the fine mapping process, providing a fast and powerful screening of recombinant lines (Varshney et al., 2014).

Further **evaluation of function** of a cloned gene is important to gain insight into the involvement of the gene in the phenotype changes, and allows establishing links between functional and molecular genetics.

Genetic transformation is a means to test function of genes, by introducing new genes or alleles into different genotypes. It represents an optimal approach for detailed elucidation of gene function (Friedt, 2011). It is regularly used in barley, and provides the potential to exploit the variability held in cultivated barley and wild barley germplasm, and even other species (Verstegen et al., 2014). Unfortunately, up to date only a few genotypes can be transformed with efficiency (Kumlehn et al., 2006; Hensel et al., 2008; Kumlehn et al., 2014).

High-throughput **transient induced gene silencing** (TIGS) is an alternative to test the involvement of a candidate gene in a function, and is being widely exploited in barley for evaluation of disease resistance genes, and related processes, like formation of callose or plant-fungi interaction during infection (Douchkov et al., 2005; Nowara et al., 2010; Pliego et al., 2013; Chowdhury et al., 2016). The only caveat is that is restricted to genes acting in the outermost cell layer of plant epidermis.

Also new **genome editing** techniques are stirring up functional genetics research, due to the ease of obtaining mutants for target genes with high specificity and accuracy, and even allowing to generate transgene-free mutants in hard-to-transform crop species (Zhang et al., 2016). Therefore, it represents an alternative to mutagenesis approaches, to standard breeding processes based on recombination, since allows generating new allelic variants, and to some aims of genetic transformation, as producing knock-out variants (Lawrenson et al., 2015). The initial approaches, such as zinc finger nucleases (ZFN) or transcription activator like effector nucleases (TALEN), are being shifted by the CRISPR (clustered regularly interspaced short palindromic repeats)/associated nuclease Cas9 system. The specificity of this system relies in CRISPR RNAs (crRNAs), and depends on hybridization of their sequence to the target. Therefore, these crRNAs can be designed to produce double-strand breaks at specific genomic sites, which subsequently lead to the introduction of a mutation at the DNA break site (Bortesi and Fischer, 2015). It allows generating insertions and deletions, but also gene stacking and allele substitutions, and even large deletions, are possible (Tsai et al., 2014). Genome quality of the target organism is very important, since genome editing relies on very accurate genome sequence information, particularly when the target gene is a member of a multigene family, or when there are homeologus copies in polyploid genomes (Barabaschi et al., 2016).

Genetic transformation and genome editing can be used to test function of genes, and also to generate new diversity. Indeed, genetic diversity could be one of the major limiting factors for further breeding progress (Friedt, 2011). Producing and cataloging mutant collections is indispensable to generate new diversity and make it available for research. In barley, **mutant collections** and mutant-based breeding programs exist from decades ago, and have been used to clone numerous genes reviewed in (reviewed in Druka et al., 2011). The systematic development of mutagenesis was limited by the lack of effective approaches of mutation screening, and by the basic knowledge of genes underlying the designated traits (Micke et al., 1990). Targeting local lesions in genomes (TILLING) combines chemical mutagenesis with genome-wide screening for point mutations in genes of interest (McCallum et al., 2000), and represents a powerful tool for reverse genetics. Researchers are now able to test the function of a gene of interest without relying on gene transformation. Collection of mutants are stored, holding pools of individuals which can readily be screened through PCR and sequencing (Slade and Knauf, 2005). In barley, TILLING populations are already available for several cultivars, including Optic (Caldwell et al., 2004), Barke (Gottwald et al., 2009), Morex (Talamè et al., 2008), and Lux (Lababidi et al., 2009). These TILLING populations are already being exploited to test function of genes and traits of interest (Bovina et al., 2011; Mascher et al., 2014; Sparla et al., 2014).

1.2.4. Breeding methods

Once the alleles conferring resistance have been identified, its **incorporation to elite varieties** is crucial to obtain improved cultivars. In the past, practical breeding approaches involved techniques like careful observation, precise testing, and conscious selection (Friedt, 2011). Being barley a natural self-pollinating crop, the overwhelming majority of current barley varieties are based on pure lines, that is, on crossing promising parental lines (elite material) to combine their favorable characteristics in the progeny (Lehmensiek et al., 2009). Classical breeding methods used in barley for over a century are pedigree breeding, mass selection, backcrossing, and (more recently) single seed descent, and combinations of them. Production of doubled haploids is a more recent technology in barley that has sped up breeding processes. By this technique, plants can reach homozygosis in one step, with the advantage of selection being applied on homozygous pure lines (Werner et al., 2007). In either breeding method, the progeny and the parents are tested in multi-location, multi-year, replicated trials to test yield and yield stability, and the process from cross to registration of a variety for its commercialization is a years-lasting process (Verstegen et al., 2014). Even so, a drawback of these approaches is the low efficiency attained when the estimated genetic effects are transferred to other genetic backgrounds (Lehmensiek et al., 2009). Hybrid breeding is also available in barley, and has resulted in the release of several barley varieties, and in a growing interest on the potential of hybrid varieties (Longin et al., 2012). The level of heterosis of barley (and wheat) is low in comparison with maize and rye, and control of pollination levels is difficult. However, barley hybrids have shown a commercial yield advantage of 7.6%, and higher barley productivity could be expected from further improvements in seed production and development of suitable parent lines (Verstegen et al., 2014).

Novel technologies have been brought to barley breeding in the last decades. The use of molecular markers, for identification and selection of promising lines and alleles, has seen an increase in the number of markers and resolution of the barley genetic maps, which in turn accelerates breeding processes. Once a locus, with significant effects, is identified, **marker-assisted selection** (MAS) can be used to accurately transfer the designated allele to an elite cultivar, based on the closest flanking markers (Xu and Crouch, 2007; Korell et al., 2008). MAS can also be used to combine several desirable alleles (Werner et al., 2005). In the case of large-effect QTLs (major genes), small QTL intervals are required for high efficiency of the introgression procedures, and, therefore, the resolution obtained from molecular markers should be as higher as possible. For complex traits, with small effect QTLs distributed throughout the genome, as is the case of many drought tolerance and nonhost resistance genes, both the identification and the transference of genes are difficult with MAS (Cattivelli et al., 2011).

Previously identified (cloned) genes are necessary to perform **gene transformation** (Tuberosa and Salvi, 2006), which can be used to develop cultivars with specifically modified traits (Friedt, 2011). However, in Europe, strong reservations against genetically modified (GM) crops are hindering its use. Moreover, transgenics have their own difficulties, like designing the necessary genetic features to introduce in the cultivar to improve, which requires good knowledge of the candidate gene and its regulation. Also, it involves leading with frequency and side effects of random mutations. In addition, gene transformation is not well suited for introducing many small effect QTLs.

A new approach, possible thanks to the availability of high-throughput technologies, allows adopting a totally different strategy. **Genomic selection** (GS) does not require mapping QTLs or genes, nor MAS, to lead to improved crops. In contrast with MAS, in which selection is applied over markers near the loci with specific desirable phenotypic effects, GS is based on the use of all available markers (requires a great number, covering the whole genome) as predictors of breeding value of a training, extensively phenotyped, population. The predictions made for the training population can be later extrapolated to larger populations (Heffner et al., 2009), without the need to perform further phenotyping, since it allows calculation of genomic estimated breeding values (GEBVs) of breeding materials using only genotypic data (Meuwissen et al., 2001). It allows selecting genotypes based on sets of small effect genes, which together lead to a high predicted breeding gain. Indeed, this has been frequently the base of the success of new crop varieties (Barabaschi et al., 2016). Similarly to GWAS, accuracy of genomic estimated breeding values (GEBVs) depends on the relationship between the training and the validation sets, the heritability of the trait, the marker density, and the rates of LD decay across the genome. GS was initially successful in animal breeding, and further evaluation needs to be done in plants. However, up to date GS has provided a higher accuracy in the estimation of GEBVs in plants than in animals, likely due to a narrower genetic base of breeding materials (Nakaya and Isobe, 2012). In barley, GS is currently being assessed thoroughly (Sallam et al., 2015).

Finally, the outstanding development of high-throughput genotyping methods highlight **phenotyping** as one of the current major bottlenecks for breeding progress (Fiorani and Schurr, 2013). Therefore, much research effort is being directed towards developing new high-throughput phenotyping methods, by using state-of-the-art technologies, including robotized greenhouse and data acquisition systems, integrated platforms of non-destructive sensors in controlled environments and to monitor field trials, and the latest algorithms and computer infrastructures for image recording, storage, and analysis (Barabaschi et al., 2016). This opens a new field for detailed and huge scale phenotyping, baptized phenomics (Houle et al., 2010).

High-throughput genotyping and phenotyping methods, and its integration with molecular biology knowledge from metabolomics, proteomics, and other 'omics', into systems biology, shape present and near future breeding. Hopefully, this will boost improvement of crops to achieve the goals of yield and yield stability which would be desirable. High-throughput sequencing and bioinformatics play an essential role in this **next generation breeding** (Tsai, 2013), contributing to empower polymorphism detection and genotyping, identification and fine mapping of candidate genes, and breeding through MAS and GS.

1.3. High-throughput sequencing

1.3.1. Technologies

High-throughput sequencing (HTS) has brought outstanding advances in the last decade. The success of HTS technologies relies on their capability to sequence an enormous amount of DNA strands. This is achieved by processing them in parallel. Their high-throughput and cost effectiveness have opened many opportunities to explore the relationships between genetic and phenotypic diversity with an unknown resolution (Mardis, 2008; Varshney et al., 2014). Moreover, HTS have introduced data analysis challenges, which resulted in a renaissance of bioinformatics-based sequence data analysis (Mardis, 2011).

In recent years different **HTS technologies** have emerged which share those features: parallel sequencing, cost effectiveness, and high-throughput. They differ in number, length, and quality of the sequences obtained. Also, errors or biases produced by HTS are different for each technology. The main competitors from what is now called second generation sequencing or next-generation sequencing were Roche 454, Illumina Genome Analyzer, and ABI SOLiD. The aim of the first NGS technologies were re-sequencing of a large number of samples and aligning them to an existing reference. Therefore, length of reads obtained from them, Roche/454 GS20 and Illumina GA, were initially very short, 100 and 24-35 nt, respectively, in comparison with traditional Sanger sequencing (Stein, 2014). Further improvements in 454 sequencers (Roche 454 GS-FLX+) yielded hundreds of thousands of reads, with length close to that obtained with regular Sanger assays. The main bias of the 454 technology, which proved to be very difficult to overcome, was the length of homopolymers, with error probability and magnitude increasing with their length (Balzer et al., 2010). Instead, Illumina sequencers have improved over the years. The initial Genome Analyzer provided short reads, below one hundred bases, and it was able to sequence up to millions of reads. The last version of Genome Analyzer (GAIIx), was able to provide paired-end reads up to 2x101 bp. The main observed biases in Illumina data were single nucleotide mismatches (Minoche et al., 2011). However, the frequency of errors was not far from that obtained with 454 (Luo et al., 2012a). ABI SOLiD had features similar to those of Genome Analyzer. However, instead of providing data with nucleotides, output from SOLiD was coded in so called color space. Color space supposedly allowed reducing errors by a technique of double checking each added base to the sequence. However, as most mismatches come from other procedures than sequencing (Schirmer et al., 2015), the advantage in comparison with Illumina technology was not largely significant (Shen et al., 2008). Moreover, color space made more difficult the development of standard tools, and many software developed for nucleotide space data was never available, or its support was dropped, for color space coding, limiting the availability of analysis software for their users (Pabinger et al., 2013). Eventually, the difficulty to deal with 454 homopolymer bias, the higher cost of ownership and maintenance of both 454 and SOLiD sequencers, in comparison with Illumina GA, and the constant improvement of Illumina sequencers (up to 2x300 bp in paired-end reads from MiSeq, up to 900 Gbp per run in the HiSeq sequencer series), which also provided easy and versatile protocols for library preparation, left Illumina sequencers as the market dominators.

Currently, other sequencing technologies are competing with Illumina, including Ion Torrent sequencers, which do without cameras and imaging analysis, since they use pH sensors directly coupled to digital microprocessors; and PacBio sequencers (Flusberg et al., 2010), which provide very long sequences (Berlin et al., 2015), and single-molecule real-time (SMRT) sequencing, with a more limited throughput than Illumina (Bashir et al., 2012). However, Illumina HiSeq sequencers, and their benchtop counterparts, remain the standard sequencing technologies for much sequencing studies. Comprehensive tables with features and current market state of HTS sequencers are annually updated at http://www.molecularecologist.com/next-gen-table-2-2016/.

A brief description of **Illumina sequencing-by-synthesis** is included here, as an example to understand how HTS reaches high-throughput, and given that it is the sequencing technology used in this work. Once that sample DNA is available, pre-sequencing procedures differ depending on the final application (RNAseq, exome capture, whole-genome sequencing, bisulfite-sequencing, ChIP-seq, …). One common step is adding short adaptors to the ends of the DNA strands. These adaptors allow each DNA strand to couple, by hybridization, to complementary adaptors attached to the surface of the sequencing plate (called flow cell in Illumina sequencers). Once that each DNA strand is linked to the flow cell, a series of PCR steps are carried out using each DNA strand as template. Each new produced strand will contain an adaptor sequence in their upper end, and will bend to bind the flow cell surface. This is called bridge-PCR (Figure 1.6). After a few PCR cycles, many clones of the same sequence will lay adjacent to each other, setting up a cluster of identical DNA sequences. Afterwards, a sequencing cycle commences by adding to the flow cell the four nucleotides, fluorescent-labeled, and blocked by a terminator, so that a single nucleotide, and no more, is added to each DNA strand. In the next step, the fluorescent

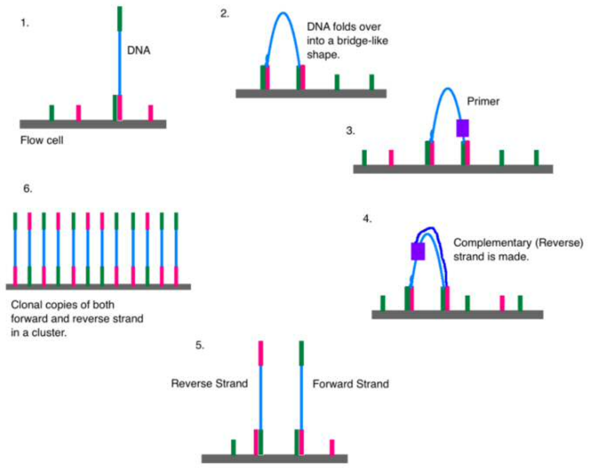

1.
DNA
Flow cell
2.
DNA folds over into a bridge-like shape.
Primer
3.
6.
4.
Complementary (Reverse) strand is made.
Clonal copies of both forward and reverse strand in a cluster.
5.
Reverse Strand
Forward Strand

Figure 1.6. Bridge amplification of DNA. 1: A DNA molecule (blue) binds to the adaptor (green), which is attached to the flow cell. 2: DNA bends to bind to another adaptor in the flow cell (red). 3-4: the complementary strand is synthetized with a primer and a DNA polymerase (violet square). 5: Both strands separate and further cycles of bending and synthesis can take place. 6. After several PCR cycles, a cluster of identical DNA strands is produced. Image from Wikimedia Commons, by DMLapato, under CCA-SA 4.0 International license.

signal, of the added nucleotides, is recorded by cameras. Since all the clones from a given cluster are expected to add the same nucleotide to their sequences, all of them will emit the same signal, and therefore the camera will be able to record such amplified fluorescence, from each of the clusters of the whole flow cell. Therefore, each image holds which single nucleotide was added in this cycle, to all the DNA strand clusters that are being sequenced in parallel. Remaining free nucleotides are washed out, and ends of DNA are unblocked by removing the terminators. Then, a new cycle starts by adding the four nucleotides again, which will be added to the DNA strands, and a new image is recorded. After the last cycle, the sequencer holds one image for each sequencing cycle. These images are preprocessed, and translated into nucleotide strings accompanied by base quality values (Mardis, 2008), which will be provided to the end user.

1.3.2. Applications and breeding

The main breakthrough obtained through HTS technologies is the ability to perform **whole genome sequencing** with much more ease and reduced cost than using traditional Sanger, either in a clone-by-clone or a whole-genome shotgun approach. In fact, an important step towards taking full advantage of genomics tools is the development of a reference genome for the species (Edwards et al., 2013). Fast and cheap whole genome sequencing using HTS has provided many finished and draft genomes, including those of several crops. As already mentioned, even incomplete sequencing of the largest and more repetitive-sequence containing genomes is providing valid tools for their respective genetic and breeding research communities. The knowledge of genome sequence facilitates traditional molecular essays, including primer design for PCR and RT-qPCR studies, looking for enzyme restriction cut sites, designing transformation clones, or defining accurate targets for TIGS and gene transformation. Moreover, the benefit of knowing the actual position of genes and molecular markers is invaluable for genetics and breeding.

There are organisms for which obtaining a complete genome is not feasible. This is often due to the size and repetitive content of the genome, but also for organisms which are not so important economically, or as research models, and do not count with much economical support. In those cases, HTS is still possible, and can provide great benefits. **Reduced representation sequencing** comprises different sequencing approaches which provide such access to HTS without a reference genome. For example, GBS is possible without such reference, and can be used to obtain numerous polymorphisms, mainly SNPs, for any organism, which can be used to produce dense genetic maps. Also, transcriptome sequencing (RNAseq), can be used to obtain a transcriptome reference, to be used for further re-sequencing efforts, providing useful information about the expressed fraction of the genome, including polymorphism detection, but also gene expression studies. Whenever a reference from a related species is available, targeted sequencing is another possibility. Although targeted sequencing is usually used as a re-sequencing method, the flexible specificity of hybridization probes, which are used to capture the DNA to be sequenced, could allow sequencing sequences between species, as exons from homologous genes (Mascher et al., 2013b) or genes from a given family (Jupe et al., 2013).

Whenever reference sequences are available, they provide a framework which opens the possibility to perform multiple sequencing experiments. **Re-sequencing** encompasses a series of experiments which take advantage of the availability of a sequence reference for a given organism. It is being facilitated by decreasing costs of HTS, but also by increasingly powerful computer infrastructures and software algorithms. Re-sequencing is revealing valuable genes and alleles hidden, until recently, in cultivars, landraces, mutagenized populations, and wild species (Säll, 1990).

One of the main applications of re-sequencing is **polymorphism detection**. HTS allows obtaining an immense number of markers, including SNPs, InDels, CNV, and PAVs. Such availability of markers can be exploited in linkage- and association-mapping studies, besides providing insights into diversity, evolution, and domestication of crops. Moreover, many of the discovered polymorphisms are being used to develop high-density marker platforms, especially those based on SNPs. Those variants can be obtained by genome sequencing, and also by reduced representation approaches, including RNAseq (Mortazavi et al., 2008), exome sequencing, capture and sequencing of custom targets (e.g. RenSeq (Jupe et al., 2013)), GBS (Mercer et al., 2014), RAD-seq (Elshire et al., 2011), DArTseq (Kilian et al., 2012), among others (Miller et al., 2007; Henry et al., 2014). In addition, HTS allows polymorphism detection and genotyping in a single step, and is replacing microarray platforms. In fact, most re-sequencing approaches, GBS as example, avoid some of the disadvantages of microarrays, like ascertainment bias (Mamanova et al., 2010; Moragues et al., 2010). The main limitation, in comparison with microarray based platforms, is that the latter are accompanied with ready-to-use results, whereas HTS data require specialized bioinformatics support to collect and interrogate the genotypic data (Waugh et al., 2014).

Re-sequencing is also helping **gene discovery**. It provides fast genome-wide screening of TILLING populations (TILLING-by-Sequencing) (Yang et al., 2016). Indeed, barley research based on mutagenesis is already being benefited by the recent advances in genomics (Salvi et al., 2014). Gene and QTL mapping is also being accelerated by HTS. For example, new mapping approaches combining HTS with bulk segregant analysis are particularly powerful (Schneeberger and Weigel, 2011; Abe et al., 2012). Also population sequencing (POPSEQ) provides fast and dense genotyping of mapping populations, leading to accurate QTL detection. Gene cloning through fine mapping, which is often hindered by the lack of polymorphic markers in the interval of the target QTL, benefits from the potential access to all the markers in the region. This allows direct identification of the differences between recombinant lines with divergent phenotypes. Moreover, fine mapping through sequencing, or mapping-by-sequencing, is not limited to standard molecular markers, and all the information from sequencing data can be exploited to identify the candidate genes (Mascher et al., 2014; Pankin et al., 2014). In summary, HTS benefits research both from a forward genetics and from a reverse genetics perspective (Salvi et al., 2014).

We will briefly cover here two of the most used re-sequencing approaches: exome sequencing and RNAseq.

Exome sequencing (Mascher et al., 2013b) is a targeted re-sequencing approach to sequence only the gene coding fraction of the genome. Therefore, it avoids investing resources in sequencing and analyzing data from most of the repetitive elements in the genome. The key step to perform exome sequencing is the capture of DNA strings from exons, isolating them from the rest of the genome. This is achieved through hybridization of probes designed over the exons annotated in a pre-existing sequence reference. Once the DNA from coding sequences is isolated, the next steps follow standard sequencing procedures. The reads obtained from the sequencer are analyzed through pipelines which usually involve mapping reads against the reference and a variant calling step to obtain the polymorphic markers and the genotypes of the samples included in the assay. Exome sequencing can be used in a straightforward way for fine mapping procedures, but also it could be used for QTL detection, or assembly of the coding fraction of genotypes and comparison of PAV between genotypes. The dependence of the capture probes on a preexisting reference makes exome sequencing susceptible to be affected by ascertainment bias.

RNAseq (Mortazavi et al., 2008) consists on sequencing the expressed fraction of the genome. Methodologically, the main difference with exome capture is that instead of capturing DNA, RNA is translated into cDNA, which will be sequenced afterwards. RNAseq is often used in gene expression assays. In this sense, it can be considered as a high-throughput version of RT-qPCR. Indeed, RT-qPCR of a few genes is usually applied as a validation step of RNAseq gene expression results. However, transcriptome sequencing data can be also used for polymorphism detection and even for genotyping, and thus it has a much broader application than RT-qPCR. As it happens with genotyping, microarrays are also being increasingly displaced by HTS for gene expression essays, in this case due to the large expression range and lack of ascertainment bias of RNAseq. One of the most relevant applications of RNAseq for breeding purposes could be the identification of expression QTLs (eQTLs), to unlock genetic variation due to changes in transcript abundance (Jackson et al., 2011).

1.3.3. Data analysis: bioinformatics

As in traditional Sanger sequencing, each HTS output sequence is called a read. In HTS solid plates, like Illumina flow cells, DNA is distributed, and therefore sampled, randomly. Therefore, **HTS output** reads are reported without sorting order or known relationship with other output sequences. However, many reads come from originally overlapping or adjacent DNA fragments, and became separated in the fragmentation step, during library preparation, previous to sequencing. Therefore, genomic location and relationship between different reads, as obtained from the sequencer, are unknown.

Analysis of HTS data often requires resolving original location of reads, its relationship, or both. Concatenating the reads which were adjacent in the original DNA is essential to obtain whole genome or transcriptome sequence references. The procedure to obtain concatenated reads is called assembling. Once an assembled reference genome or transcriptome, an assembly, is available, analysis of re-sequencing experiments usually requires locating the output reads in the reference. This procedure is called mapping, and it is essential for polymorphism detection and variant calling, analysis of gene expression, and other approaches which will be not covered here (e.g. ChIP-seq and bisulfite sequencing).

Basically, **assembling reads** consists in finding common fragments between two sequences, usually overlapping edges, and produce the consensus sequence, which is obtained from merging them into a single longer sequence called a contig (Staden, 1980). It is a procedure that was already performed with sequences obtained through traditional Sanger sequencing (Huang and Madan, 1999). However, the algorithms used to assemble Sanger sequencing reads faced a few, long and high quality, sequences. Thus, new algorithms were needed to cope with the large amount of short, or moderately long, reads obtained through HTS. HTS assembling algorithms can be divided in two main classes: *de novo* assembly and reference-guided assemblers. *De novo* assembly uses as input just the sequencing reads, without taking advantage of previously existing sequence references. In contrast, reference-guided assembly relies on a genome or a transcriptome, usually involving a mapping step previous to assembling. The advantage of *de novo* assembling is that it can be used without having a sequence reference available. Moreover, it lacks the errors induced by the natural differences between a sequence reference and the sample under study. However, finding the correct way to concatenate the reads can be a daunting, and even impossible, task, especially for large and repetitive genomes, since it is often unaffordable to resolve ambiguities. In turn, reference-guided assembly can be used to assemble reads resembling the linear DNA layout found in the reference. It has the drawback of not considering those sequences which are present in the sample but absent from the reference. A common hybrid approach is performing *de novo* assembly only with those reads which do not map properly to the reference (Digel et al., 2015; Yao et al., 2015).

Regarding underlying algorithms, the current most common approach is the exploration of **de Bruijn graph** (Figure 1.7). In this class of algorithms, each read is fragmented in k-mers, segments of length ‘k’ nucleotides, covering the whole read. Each of those k-mers is represented as a node in a graph, and adjacent k-mers in the original read are connected through graph edges. Whenever a k-mer from a read already exists in a previous analyzed read, they will share a node in the graph. After analyzing all the reads, a graph contains the nodes representing the k-mers from all of them, linked through edges. The resolution of each isolated subgraph, recovering the sequence of k-mers from each node, following the edges one-by-one, leads to the production of contigs. Since many reads will share

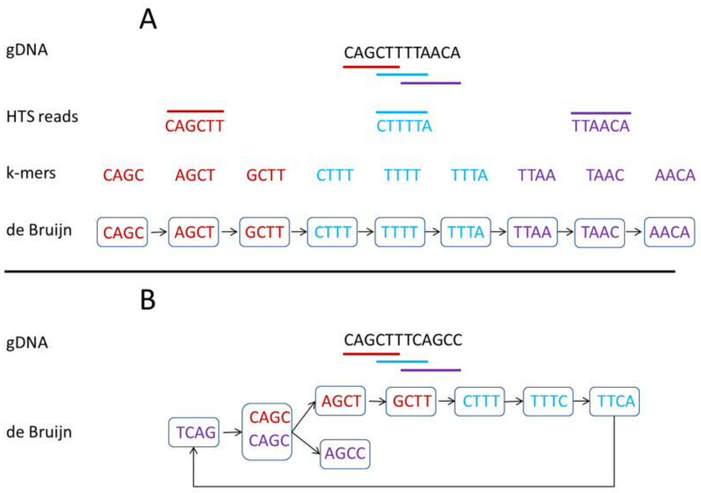


Figure 1.7. *De novo* assembly using de Bruijn graphs. A) A hypothetical genomic DNA (gDNA), with 12 bases, is sequenced producing three hypothetical HTS reads (red, blue, violet), of length 6 bases. Each read is fragmented in *k*-mers of length 4 (4-mers). Adjacent *k*-mers are linked in the de Bruijn graph. Going through the graph allows recovering the original DNA sequence. B) Repeats in the gDNA (CAGC in the example) produce more complex de Bruijn graphs, leading to ambiguous resolution.

common nodes in the graph, these algorithms require much less computer memory than previous algorithms, in which each read was treated separately. Moreover, these algorithms are fast, although in practice many include intermediate steps to resolve difficulties produced by repetitive sequences, and sequencing errors, being yet one of the most time consuming procedures of HTS data analysis. Examples of implementation of these algorithms are Velvet (Zerbino and Birney, 2008) or SOAPdenovo2 (Luo et al., 2012b), for genomic data; or Trinity (Haas et al., 2013), for transcriptomic data. Both Trinity and Velvet, include experimental modules to perform reference-guided assemblies. In the last years, the advent of long-read HTS technologies, as PacBio or long pseudo-read sequencing from Illumina, is bringing back the attention towards overlap-layout-consensus algorithms which cope with long reads (Koren et al., 2012; Li et al., 2012), as those which were originally developed for 454 data (e.g. SSAHA and Newbler).

Regarding **mapping** (Figure 1.8), it consists of matching HTS reads in a sequence reference. It is rather similar to sequence alignment, but in the latter the aim was usually locating the exact position of each nucleotide of the query sequences in the reference. The term mapping was originally proposed as different from alignment, since mapping did not report the exact position of each nucleotide base, but only the expected, approximate position of a sequence within another (for example, a locus within a chromosome, or a given gene in a transcriptome). This differentiation was especially important in HTS, since reads are much more numerous, and therefore an exact alignment was not necessary for all downstream applications. However, in practice the term mapping is used in HTS, even when most HTS mappers report sequence alignments as output. Anyhow, HTS mapping approaches try to distribute the task among several computer processors, and optimize their algorithms both in computing time and memory consumption, even when accuracy of the alignment is compromised, for a number of sequences below an acceptable statistical threshold. This need for optimization has led to the adoption of different data structures to represent either the reads or the sequence reference, generally the latter, as FM-index, suffix arrays or its compressed form, the Burrows Wheeler Transform (Lam et al., 2008). The latter is the most widely implemented nowadays in general purpose HTS mappers, as BWA (Li and Durbin, 2009), Bowtie2 (Langmead and Salzberg, 2012), or GEM (Marco-Sola et al., 2012).

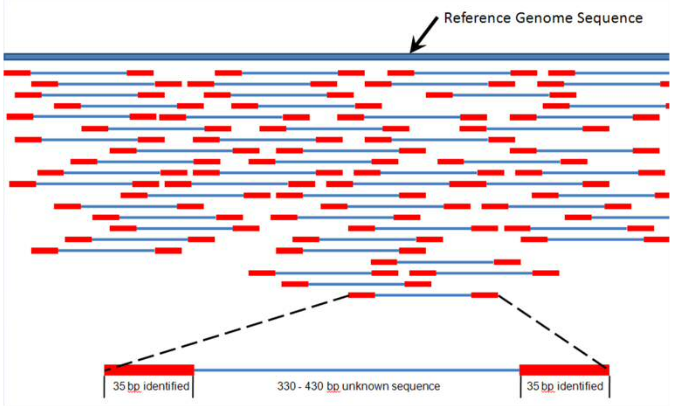


Figure 1.8. Paired-end reads mapped to a reference genome. Each read has sequenced ends (red) of length 35 base pairs (bp), with a non-sequenced insert of approximately known size (blue thin lines). Both sequenced ends and known insert size help aligning correctly the reads to the sequence reference (blue thick line). Image from Wikimedia Commons, under CC CC0 1.0 Universal Public Domain Dedication.

More specific mappers usually make use of the previous, but add different layers to manage heuristics of a specific problem to resolve. In the case of **mappings reads from RNAseq**, which must manage introns and splicing, there are different algorithms designed to map reads to a genome reference, like TopHat (Trapnell et al., 2009) or STAR (Dobin et al., 2013). When the reference is a transcriptome, there is no need, a priori, to cope with introns, and therefore general purpose mappers tend to be used (Haas et al., 2013). However, HTS algorithms, as HTS technologies, keep evolving at a fast pace. Here we will briefly explain the new approaches which have arose in the last few years to cope counting reads from RNAseq data. Note that the exact position of each read is not a requirement, for counting expression of a gene. It suffices to know that the read comes from that gene. Pseudoalignment strategies exploit this idea (Bray et al., 2016), removing the need to match each base in the query with their respective bases in the reference. Pseudoalignments can be obtained much faster than standard read alignments. Instead, reads are fragmented in k-mers and assigned to reference transcripts according to compatibility of their constituting k-mers. This idea is not so different to that of OLC mappers for long-reads (Ning et al., 2001). However, these approaches bring fast, and memory efficient algorithms, which moreover are showing improved accuracy to resolve splicing and paralogous isoforms, by putting together state-of-the-art algorithms and data structures, like k-mer compatibility classes and de Bruijn graphs (Bray et al., 2016), and suffix arrays combined with hash tables (Srivastava et al., 2016).

Finally, there are other data analysis steps, usually downstream of assembling and mapping, which depend on the final application of the HTS experiment. An important application of HTS in this work is **analysis of gene expression**. It comprises two main steps, after mapping reads to a reference: counting expression of each gene or transcript isoform, and testing for differential expression between samples. In this case, the main challenges which face these programs are adequately accounting for the reads which map to a given locus, and modeling expression data so that false positives and negatives are reduced when testing differential expression. Counting expression can be rather straightforward from mapping results. The major difficulty is how to consider multiple mapping reads. Some approaches do not count those reads, whereas others count them once for each target, and others reduce the counts in proportion to the number of targets hit. However, raw counts are not used in all gene expression analyses. When different samples have to be compared, raw counts can produce bias, mostly due to differences in sampling depth. Overcoming this requires normalizing counts, considering differences between samples, and also between the different loci (e.g. sampling reads from long genes is more frequent than sampling reads from short genes). Moreover, which measure should be used to normalize data has been also debated. The former normalized abundances, like FPKM (Trapnell et al., 2010), were put into question, since different samples were not directly comparable. More recent normalization values, as TPMs (Li and Dewey, 2011), try to make each sample equivalent in magnitude, so that they can be compared to each other in downstream analyses. After either raw counts or normalized values are generated, proper models need to be used to test differential expression. Many concepts from microarrays were translated into RNAseq initially. However, HTS data is intrinsically count data, and statistic models are different. Therefore, the statistical models have switched from Poisson models to the current most accepted Negative Binomial distributions (Robinson and Smyth, 2007; Anders and Huber, 2010), which aim to model the mean-variance interdependence. Difficulties for differential expression tests also come from the low number of samples analyzed in RNAseq experiments, being difficult to account for biological and technical variability (Robinson et al., 2010). Despite being cost efficient, absolute price of HTS is high, and the requirement of biological replicates to declare differentially expressed increases the number of samples that need to be sequenced. Therefore, the number of samples sequenced tends to be lower than what would be optimal, and this reduces statistical power to discriminate between false and true positives (Schurch et al., 2016).

1.4. References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30**,** 174-178.

Ali, M., Jensen, C.R., Mogensen, V.O., Andersen, M.N., and Henson, I.E. (1999). Root signalling and osmotic adjustment during intermittent soil drying sustain grain yield of field grown wheat. *Field Crops Res.* 62**,** 35-42.

Ames, N.P., and Rhymer, C.R. (2008). Issues surrounding health claims for barley. *J Nutrition* 138**,** 12375-12435.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11**,** R106.

Ariyadasa, R., Mascher, M., Nussbaumer, T., Schulte, D., Frenkel, Z., Poursarebani, N., et al. (2014). A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. *Plant Physiol.* 164**,** 412-423. doi: 10.1104/pp.113.228213

Australian Government (Australian Government), (2008). "The biology of *Hordeum vulgare* L. (barley)". Available online at: http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/content/barley-3/$FILE/biologybarley08.pdf (Accessed August 23, 2016)

Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H.H., et al. (2000). On the origin and domestication history of barley (*Hordeum vulgare*). *Mol Biol Evol* 17**,** 499-510.

Bahieldin, A., Hesham, H.T., Eissa, H.F., Saleh, O.M., Ramadan, A.M., Ahmed, I.A., et al. (2005). Field evaluation of transgenic wheat plants stably expressing the HVA1 gene for drought tolerance. *Physiol. Plant* 123**,** 421-427.

Baik, B., and Ullrich, S.E. (2008). Barley for food: characteristics, improvement, and renewed interest. *J. Cereal Sci.* 48**,** 233-242. doi: 10.1016/j.jcs.2008.02.002

Balzer, S., Malde, K., Lanzén, A., Sharma, A., and Jonassen, I. (2010). Characteristics of 454 pyrosequencing data - enabling realistic simulation with flowsim. *Bioinformatics* 26**,** i420-i425.

Barabaschi, D., Tondelli, A., Desiderio, F., Volante, A., Vaccino, P., Valè, G., et al. (2016). Next generation breeding. *Plant Sci.* 242**,** 3-13.

Barkworth, M.E., and Bothmer, R.v. (2009). "Scientific names in the *Triticeae*", in *Genetics and genomics of the Triticeae,* eds. C. Feuillet & G.J. Muehlbauer, (LLC: Springer Science+Business Media). doi: 10.1007/978-0-387-77489-3_1

Bashir, A., Klammer, A.A., Robins, W.P., Chin, C.S., Webster, D., Paxinos, E., et al. (2012). A hybrid approach for the automated finishing of bacterial genomes. *Nat Biotechnol* 30, 701-7. doi: 10.1038/nbt.2288

Baum, M., Grando, S., Backes, G., Jahoor, A., Sabbagh, A., and Ceccarelli, S. (2003). QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' x *H. Spontaneum* 41-1. *Theor. Appl. Genet.* 107**,** 1215-1225.

Berlin, K., Koren, S., Chin, C.S., Drake, J.P., Landolin, J.M., and Phillippy, A.M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* 33, 623-30. doi: 10.1038/nbt.3238

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., and Moscou, M.J. (2014). Nonhost resistance to rust pathogens - a continuation of continua. *Front. Plant Sci.* 5, 1-15. doi: 10.3389/fpls.2014.00664

Blattner, F.R. (2009). Progress in phylogenetic analysis and a new infrageneric classification of the barley genus Hordeum (Poaceae: Triticeae). *Breed Sci* 59, 471-480.

Blum, A. (1988). Improving wheat grain filling under stress by stem reserve mobilization. *Euphytica* 100, 77-83.

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res.* 112, 119-123. doi: 10.1016/j.fcr.2009.03.009

Blum, A. (2011). Drought Resistance and Its Improvement. 53-152. doi: 10.1007/978-1-4419-7491-4_3

Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B., and Mayer, K.F.X. (2014). Plant genome sequencing - applications for crop improvement. *Curr. Opin. Biotechnol.* 26, 31-37.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* 33, 41-52. doi: 10.1016/j.biotechadv.2014.12.006

Bossolini, E., Wicker, T., Knobel, P.A., and Keller, B. (2007). Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *Plant J.* 49, 704-717.

Bothmer, R.v., Flink, J., Jacobsen, N., Kotimäki, M., and Landström, T. (1983). Interspecific hybridization with cultivated barley (*Hordeum vulgare* L.). *Hereditas* 99, 219-244.

Bothmer, R.v., and Komatsuda, T. (2011). "Barley origin and related species", in *Barley: production, improvement, and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 14-61.

Bothmer, R.v., Sato, K., Komatsuda, T., Yasuda, S., and Fischbeck, G. (2003). "The domestication of cultivated barley", in *Diversity in barley (Hordeum vulgare),* eds. R.V. Bothmer, T.V. Hintum, H. Knüpffer & K. Sato, (Amsterdam, The Netherlands: Elsevier Science B.V.), 9-27.

Boudiar, R., Casas, A.M., Cantalapiedra, C.P., Gracia, M.P., and Igartua, E. (2016). Identification of quantitative trait loci for agronomic traits contributed by a barley (Hordeum vulgare) Mediterranean landrace. *Crop and Pasture Science* 67, 37. doi: 10.1071/cp15149

Bovina, R., Talamè, V., Salvi, S., Sanguineti, M.C., Trost, P., Sparla, F., et al. (2011). Starch metabolism mutants in barley: a TILLING approach. *Plant Genet. Resour. Characterization Utilization* 9, 170-173.

Boyer, J.S., and Westgate, M.E. (2004). Grain yields with limited water. *J. Exp. Bot.* 55**,** 2385-2394.

Brassac, J., and Blattner, F.R. (2015). Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferrer by Next-Generation Sequencing and *in silico* cloning of multiple nuclear loci. *Syst Biol* 64**,** 792-808. doi: 10.1093/sysbio/syv035

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnol.* 34**,** 525-527. doi: 10.1038/nbt.3519

Briggs, D.E. (1978). *Barley.* London: Chapman and Hall Ltd.

Brown, A.H.D., Garvin, D.F., Burdon, J.J., Abbott, D.C., and Read, B.J. (1996). The effect of combining scald resistance genes on disease levels, yield and quality traits in barley. *Theor. Appl. Genet.* 93.

Brown, J.K. (1994). Chance and selection in the evolution of barley mildew. *Trends Microbiol.* 2**,** 470-475.

Brown, W.M.J., Hill, J.P., and Velasco, V.R. (2001). Barley yellow rust in North America. *Annu. Rev. Phytopathol.* 39**,** 367-384.

Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., et al. (2002). The barley stem rust-resistance gene Rpg1 is a novel disease-disease gene with homology to receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* 99**,** 9328-9333.

Caldwell, D.G., McCallum, N., Shaw, P., Muehlbauer, G., Marshall, D.F., and Waugh, R. (2004). A structured mutant population for forward and reverse genetics in Barley (Hordeum vulgare L.). *Plant J.* 40**,** 143-150.

Cantalapiedra, C.P., Boudiar, R., Casas, A.M., Igartua, E., and Contreras-Moreira, B. (2015). BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol. Breeding* 15. doi: 10.1007/s11032-015-0253-1

Cantalapiedra, C.P., Contreras-Moreira, B., Silvar, C., Perovic, D., Ordon, F., Gracia, M.P., et al. (2016). A Cluster of Nucleotide-Binding Site–Leucine-Rich Repeat Genes Resides in a Barley Powdery Mildew Resistance Quantitative Trait Loci on 7HL. *The Plant Genome*. doi: 10.3835/plantgenome2015.10.0101

Casao, M.C., Igartua, E., Karsai, I., Lasa, J.M., Gracia, M.P., and Casas, A.M. (2011a). Expression analysis of vernalization and day-length response genes in barley (*Hordeum vulgare* L.) indicates that VRNH2 is a repressor of PPDH2 (HvFT3) under long days. *J. Exp. Bot.* 62**,** 1939-1949.

Casao, M.C., Karsai, I., Igartua, E., Gracia, M.P., Veisz, O., and Casas, A.M. (2011b). Adaptation of barley to mild winters: a role for PPDH2. *BMC Plant Biol.* 11**,** 164. doi: 10.1186/1471-2229-11-164

Casas, A.M., Gracia, M.P., and Igartua, E. (2016). "Cebada", in *Las variedades locales en la mejora genética de plantas,* eds. J.I. Ruiz De Galarreta, J. Prohens & R. Tierno: Neiker-Tecnalia), 119-131.

Casas, A.M., Yahiaoui, S., Ciudad, F.J., and Igartua, E. (2005). Distribution of MWG699 polymorphism in Spanish European barleys. *Genome* 48, 41-45. doi: 10.1139/g04-091

Cattivelli, L., Baldi, P., Crosatti, C., Grossi, M., Valè, G., and Stanca, A.M. (2002). "Genetic bases of barley physiological response to stressful conditions", in *Barley sciencei: recent advantages from molecular biology to agronomy of yield and qualità,* eds. G.A. Slafer, J.L. Molina-Cano, R. Savin, J.L. Araus & I. Romagosa, (New York: Food Product Press).

Cattivelli, L., Ceccarelli, S., Romagosa, I., and Stanca, M. (2011). "Abiotic stresses in barley: problems and solutions", in *Barley: production, improvement and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 282-306.

Ceccarelli, S., Acevedo, E., and Grando, S. (1991). Breeding for yield stability in unpredictable environments: single traits, interaction between traits, and architecture of genotypes. *Euphytica* 56, 169-185.

Ceccarelli, S., and Grando, S. (1996). Drought as a challenge for the plant breeder. *Plant Growth Regulation* 20, 149-155.

Ceccarelli, S., Grando, S., Baum, M., and Udupa, S.M. (2004). "Breeing for drought resistance in a changing climate", in *Challenges and strategies for dryland agriculture,* eds. S.C. Rao & J. Ryan, (Madison, WI: ASA and CSSA), 167-190.

Clayton, W. (1990). "The spikelet", in *Reproductive versatility in the grasses,* ed. G. Chapman, (Cambridge: Cambridge University Press), 32-51.

Close, T.J., Bhat, P.R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., et al. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10, 582. doi: 10.1186/1471-2164-10-582

Close, T.J., Wanamaker, S., Roose, M.L., and Lyon, M. (2007). "HarvEST: an EST database and viewing software", in *Plant bioinformatics: methods and protocols,* ed. D. Edwards, (Totowa, New Jersey: Humana Press), 161-77. doi: 10.1007/978-1-59745-535-0_7

Close, T.J., Wanamaker, S.I., Caldo, R.A., Turner, S.M., Ashlock, D.A., Dickerson, J.A., et al. (2004). A new resource for cereal genomics: 22K barley GeneChip comes of age. *Plant Physiol* 134, 960-8. doi: 10.1104/pp.103.034462

Cockram, J., Hones, H., and O´Sullivan, D.M. (2011). Genetic variation at flowering time loci in wild and cultivated barley. *Plant Genet Resour Characterization Utilization* 9, 264-267. doi: 10.1017/S1479262111000505

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U., et al. (2015). BARLEX - the barley draft genome explorer. *Mol. Plant* 8, 964-966. doi: 10.1016/j.molp.2015.03.009

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., et al. (2012). Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44, 1388-92. doi: 10.1038/ng.2447

Comadran, J., Russell, J.R., Booth, A., Pswarayi, A., Ceccarelli, S., Grando, S., et al. (2011). Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. *Theor Appl Genet* 122**,** 1363-73. doi: 10.1007/s00122-011-1537-4

Comadran, J., Thomas, W.T.B., Eeuwijk, F.A.v., Ceccarelli, S., Grando, S., Stanca, A.M., et al. (2009). Patters of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. *Theor. Appl. Genet.* 119**,** 175-187.

Courtois, B., Shen, L., Petalcorin, W., Carandang, S., Mauleon, R., and Li, Z. (2003). Locating QTLs controlling constitutive root traits in the rice population IAC 165 x Co39. *Euphytica* 134**,** 335-345.

Chowdhury, J., Schober, M.S., Shirley, N.J., Singh, R.R., Jacobs, A.K., Douchkov, D., et al. (2016). Down-regulation of the glucan synthase-like 6 gene (HvGsl6) in barley leads to decreased callose accumulation and increased cell wall penetration by Blumeria graminis f. sp. hordei. *New Phytol*. doi: 10.1111/nph.14086

Christenhusz, M.J.M., and Byng, J.W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* 261**,** 201-217. doi: 10.11646/phytotaxa.261.3.1

Diab, A.A., Teulat-Merah, B., This, D., Ozturk, N.Z., Benscher, D., and Sorrells, M.E. (2004). Identification of drought-inducible genes and differentially expressed sequence tags in barley. *Theor Appl Genet* 109**,** 1417-25. doi: 10.1007/s00122-004-1755-0

Digel, B., Pankin, A., and von Korff, M. (2015). Global Transcriptome Profiling of Developing Leaf and Shoot Apices Reveals Distinct Genetic and Environmental Control of Floral Transition and Inflorescence Development in Barley. *Plant Cell* 27**,** 2318-34. doi: 10.1105/tpc.15.00203

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29**,** 15-21.

Doležel, J., Kubaláková, M., Paux, E., Bartos, J., and Feuillet, C. (2007). Chromosome-based genomics in the cereals. *Chromosome Res.* 15**,** 51-66.

Doležel, J., Vrána, J., Safár, J., Bartos, J., Kubaláková, M., and Simková, H. (2012). Chromosomes in the flow to simplify genome analysis. *Funct. Integr. Genomics* 12**,** 397-416. doi: 10.1007/s10142-012-0293-0

Douchkov, D., Nowara, D., Zierold, U., and Schweizer, P. (2005). A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. *MPMI* 18**,** 755-761.

Draper, J., Mur, L.A.J., Jenkins, G., Ghosh-Biswas, G.C., Bablak, P., Hasterok, R., et al. (2001). A new model system for functional genomics in grasses. *Plant Physiol* 127**,** 1539-1555.

Drosse, B., Campoli, C., Mulki, A., and Korff, M.v. (2014). "Genetic control of reproductive development", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 81-99. doi: 10.1007/978-3-662-44406-1_5

Druka, A., Franckowiak, J., Lundqvist, U., Bonar, N., Alexander, J., Houston, K., et al. (2011). Genetic disecction of barley morphology and development. *Plant Physiol.* 155**,** 617-627.

Edwards, D., Batley, J., and Snowdon, R.J. (2013). Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet* 126**,** 1-11. doi: 10.1007/s00122-012-1964-x

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6**,** e19379.

Ellis, R.P., Forster, B.P., Robinson, D., Handley, L.L., Gordon, D.C., Russell, J.R., et al. (2000). Wild barley: a source of genes for crop improvement in the 21st century? *J. Exp. Bot.* 51**,** 9-17. doi: 10.1093/jexbot/51.342.9

FAOSTAT (2016). Available: http://faostat3.fao.org (Accessed July 7, 2016).

Feuillet, C., Leach, J.E., Rogers, J., Schnable, J.C., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 16**,** 77-88.

Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64**,** 267-291.

Fischbeck, G. (2003). "Diversification through breeding", in *Diversity in barley (Hordeum vulgare),* eds. R. Von Bothmer, T. Van Hintum, H. Knüpffer & K. Sato, (Amsterdam: Elsevier Science B.V.), 29-52.

Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9**,** 275-296.

Flusberg, B.A., Webster, D., Lee, J., Travers, K., Olivares, E., Clark, T.A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7**,** 461-465.

Friedt, W. (2011). "Barley breeding history, progress, objectives, and technology", in *Barley: production, improvement and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 160-220.

Garvin, D.F., Gu, Y.Q., Hasterok, R., Hazen, S.P., Jenkins, G., Mockler, T.C., et al. (2008). Development of genetic and genomic research for *Brachypodium distachyon*, a new model system for grass crop research. *Plant Genome* 1**,** S69-S84.

Gill, U.S., Lee, S., and Mysore, K.S. (2015). Host versus nonhost resistance: distinct wars with similar arsenals. *Phytopathology* 105**,** 580-587. doi: 10.1094/PHYTO-11-14-0298-RVW

Gomez-Macpherson, H. (2000). *Hordeum vulgare* [Online]. Available: http://ecoport.org/ep?Plant=1232&entityType=PL****&entityDisplayCategory=PL****0500 (Accessed August 23, 2016).

González, F.G., Slafer, G.A., and Miralles, D.J. (2002). Vernalization and photoperiod responses in wheat pre-flowering reproductive phases. *Field Crops Res.* 74**,** 183-195.

Gottwald, S., Bauer, P., Komatsuda, T., Lundqvist, U., and Stein, N. (2009). TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHoxI. *BMC Res. Notes* 2, 258.

Graner, A., Bjornstad, A., Konishi, T., and Ordon, F. (2003). "Molecular diversity of the barley genome", in *Diversity in barley (Hordeum vulgare),* eds. R.V. Bothmer, T.V. Hintum, H. Knüpffer & K. Sato, (Amsterdam, The Netherlands: Elsevier Science B.V.), 121-141.

Graner, A., Kilian, A., and Kleinhofs, A. (2011). "Barley genome organization, mapping and synteny", in *Barley: production, improvement and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 63-84.

Griffiths, F.E.W., Lyndon, R.F., and Bennett, M.D. (1985). The effects of vernalization on the growth of the wheat shoot apex. *Ann Bot* 56, 501-511.

Gürel, F., Özturk, Z.N., Uçarli, C., and Rosellini, D. (2016). Barley genes as tools to confer abiotic stress tolerance in crops. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01137

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494-512. doi: 10.1038/nprot.2013.084

Heath, M.C. (2000). Nonhost resistance and nonspecific plant defenses. *Curr. Opin. Plant Biol.* 3, 315-319.

Heffner, E.L., Sorrells, M.E., and Jannink, J. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1-12.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., et al. (2014). Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *Plant Cell* 26, 1382-1397. doi: 10.1105/tpc.113.121590

Hensel, G., Valkov, V., Middlefell-Williams, J., and Kumlehn, J. (2008). Efficient generation of transgenic barley: the way forward to modulate plant-microbe interactions. *J Plant Physiol* 165, 71-82. doi: 10.1016/j.jplph.2007.06.015

Houle, D., Govindaraju, D.R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Reviews Genet.* 11, 855-866.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., et al. (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12987-12992.

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531-551.

Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868-877.

Hübner, S., Korol, A.B., and Schmid, K.J. (2015). RNA-Seq analysis identifies genes associated with differential reproductive success under drought-stress in accessions of wild barley *Hordeum spontaneum*. *BMC Plant Biol.* 15**,** 134. doi: 10.1186/s12870-015-0528-z

Humphry, M., Consonni, C., and Panstruga, R. (2006). mlo-based powdery mildew immunity: silver bullet or simply non-host resistance? *Mol. Plant Pathol.* 7**,** 605-610.

Huo, N., Lazo, G.R., Vogel, J.P., You, F.M., Ma, Y., Hayden, D.M., et al. (2008). The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences. *Funct. Integr. Genomics* 8**,** 135-147.

IBSC (2013). *IPK Barley BLAST Server* [Online]. Available: http://webblast.ipk-gatersleben.de/barley/ (Accessed August 26, 2016).

Igartua, E., Gracia, M.P., Lasa, J.M., Medina, B., Molina-Cano, J.L., Montoya, J.L., et al. (1998). The Spanish barley core collection. *Genet Res Crop Evol* 45**,** 475-481. doi: 10.1023/A:1008662515059

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463**,** 763-8. doi: 10.1038/nature08747

Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29**,** E25.

Jackson, S.A., Iwata, A., Lee, S., Schmutz, J., and Shoemaker, R. (2011). Sequencing crop genomes: approaches and applications. *New Phytol.* 191**,** 915-925.

Johnson, W.C., Jackson, L.E., Ochoa, O., Wijk, R.v., Peleman, J., St. Clair, D.A., et al. (2000). A shallow-rooted crop and its wild progenitor differ at loci determining root architecture and deep soil water extraction. *Theor. Appl. Genet.* 101**,** 1066-1073.

Jones, H., Leigh, F.J., Mackay, I., Bower, M.A., Smith, L.M.J., Charles, M.P., et al. (2008). Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Mol. Biol. Evol.* 25**,** 2211-2219. doi: 10.1093/molbev/msn167

Jorgensen, J.H. (1992). Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica* 63**,** 141-152.

Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G.J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J* 76**,** 530-44. doi: 10.1111/tpj.12307

Kellogg, E.A. (2001). Evolutionary history of the grasses. *Plant Physiol* 125**,** 1198-1205. doi: 10. 1104/pp.125.3.1198

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., et al. (2016). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res* 44**,** D574-80. doi: 10.1093/nar/gkv1209

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 888**,** 67-89. doi: 10.1007/978-1-61779-870-2_5

Kirkegaard, J.A., Lilley, J.M., Howe, G.N., and Graham, J.M. (2007). Impact of subsoil water use on wheat yield. *Aust. J. Agric. Res.* 58**,** 303-315.

Kishor, P.B.K., Rajesh, K., Reddy, P.S., Seiler, C., and Sreenivasulu, N. (2014). "Drought stress tolerance mechanisms in barley and its relevance to cereals", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 161-179.

Kleinhofs, A., Chao, S., and Sharp, P.J. (Year). "Mapping of nitrate reductase genes in barley and wheat", in: *Proc. 7th Int. Wheat Genet. Symp.*, eds. T.E. Miller & R.M.D. Koebner (Cambridge, UK: Bath Press).

Komatsuda, T. (2014). "Domestication", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 37-54.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci U S A* 104**,** 1424-1429. doi: 10.1073/pnas.0608580104

Korell, M., Eschholz, T.W., Eckey, C., Biedenkopf, D., Kogel, K.H., Friedt, W., et al. (2008). Development of a cDNA-AFLP derived CAPS marker co-segregating with the powdery mildew resistance gene *Mlg* in barley. *Plant Breed.* 127**,** 102-104.

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* 30**,** 693-700. doi: 10.1038/nbt.2280

Korff, M.v., Grando, S., Del Greco, A., This, D., Baum, M., and Ceccarelli, S. (2008). Quantitative trait loci associated with adaptation to Mediterranean dryland conditions in barley. *Theor. Appl. Genet.* 117**,** 653-669.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9**,** 29.

Kota, R., Kumar, R., Varshney, R.K., Thiel, T., Dehmer, K.J., and Graner, A. (2001). Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* 135**,** 145-151.

Kumlehn, J., Gurushidze, M., and Hensel, G. (2014). "Genetic engineering", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 393-407.

Kumlehn, J., Serazetdinova, L., Hensel, G., Becker, D., and Loerz, H. (2006). Genetic transformation of barley (Hordeum vulgare L.) via infection of androgenetic pollen cultures with Agrobacterium tumefaciens. *Plant Biotechnol J* 4**,** 251-61. doi: 10.1111/j.1467-7652.2005.00178.x

Kumlehn, J., and Stein, N. (2014). "Preface", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), v-vi.

Lababidi, S., Mejlhede, N., Rasmussen, S.K., Backes, G., Al-Said, W., Baum, M., et al. (2009). Identification of barley mutants in the cultivar 'Lux' at the *Dhn* loci through TILLING. *Plant Breed.* 128**,** 332-336.

Lam, T.W., Sung, W.K., Tam, S.L., Wong, C.K., and Yiu, S.M. (2008). Compressed indexing and local alignment of DNA. *Bioinformatics* 24**,** 791-797.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9**,** 357-359. doi: 10.1038/nmeth.1923

Lasa, J.M., Igartua, E., Ciudad, F.J., Codesal, P., García, E.V., Gracia, M.P., et al. (2001). Morphological and agronomical diversity patters in the Spanish Barley Core Collection. *Hereditas* 135**,** 217-225. doi: 10.1111/j.1601-5223.2001.00217.x

Laurie, D., Parthchett, N., Bezant, J., and Snape, J. (1995). RFLP mapping of five major genes and eight quantitative trait loci controlling time in a winter x spring barley (Hordeum vulgare L.) cross. *Genome* 38**,** 575-585.

Laurie, D.A. (1997). Comparative genetics of flowering time. *Plant Mol Biol* 35**,** 167-177. doi: 10.1023/A:1005726329248

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Ostegaard, L., Patron, N., et al. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome Biol.* 16.

Lee, S.H., and Neate, S.M. (2007). Population genetic structure of *Septoria passerinii* in northern Great Plains barley. *Phytopathology* 97**,** 938-944.

Lehmensiek, A., Bovill, W., Wenzl, P., Langridge, P., and Appels, R. (2009). "Genetic mapping in the Triticeae", in *Genetics and genomics of the Triticeae,* eds. C. Feuillet & G. Muehlbauer, (Heidelberg: Springer), 201-236.

Levitt, J. (1972). *Responses of plants to environmental stresses.* New York: Academic Press.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12**,** 323. doi: 10.1186/1471-2105-12-323

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25**,** 1754-1760.

Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., et al. (2012). Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Brief Funct Genomics* 11**,** 25-37. doi: 10.1093/bfgp/elr035

Longin, C.F.H., Muehleisen, J., Maurer, H.P., Zhang, H., Gowda, M., and Reif, J.C. (2012). Hybrid breeding in autogamous cereals. *Theor Appl Genet* 125**,** 1087-1096.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. (2012a). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* 7, e30087.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012b). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1, 18.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., et al. (2010). Target-enrichment strategies for next-generation sequencing. *Nat. Methods* 7, 111-118. doi: 10.1038/nmeth.1419

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* 9, 1185-1188.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet* 24, 133-41. doi: 10.1016/j.tig.2007.12.007

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature* 470, 198-203. doi: 10.1038/nature09796

Mascher, M., Jost, M., Kuon, J., Himmelbach, A., Abfalg, A., Beier, S., et al. (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15, R78. doi: 10.1186/gb-2014-15-6-r78

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., et al. (2013a). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* 76, 718-27. doi: 10.1111/tpj.12319

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013b). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J.* 76, 494-505. doi: 10.1111/tpj.12294.

Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., et al. (2011). Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.* 156, 20-8. doi: 10.1104/pp.110.171579

Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 23, 1249-63. doi: 10.1105/tpc.110.082537

Mayer, K.F.X., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711-6. doi: 10.1038/nature11543

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. *Nat. Biotechnol.* 18, 455-457.

McDonald, B.A., and Linde, C. (2002). The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* 124, 163-180.

McKersie, B.D., Bowley, S.R., Harjanto, E., and Leprice, O. (1996). Water-deficit tolerance and field performance of transgenic alfalfa overexpressing superoxide dismutase. *Plant Physiol.* 111**,** 1177-1181.

Melchinger, A., Graner, A., Singh, M., and Messmer, M.M. (1994). Relationships among European barley germplasm: I. Genetic diversity among winter and spring cultivars revealed by RFLPs. *Crop Sci.* 34**,** 1191-1199.

Mercer, T.R., Clark, M.B., Crawford, J., Brunck, M.E., Gerhardt, D.J., Taft, R.J., et al. (2014). Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat Protoc* 9**,** 989-1009. doi: 10.1038/nprot.2014.058

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157**,** 1819-1829.

Micke, A., Donini, B., and Maluszynski, M. (1990). Induced mutations for crop improvement. *Mutat. Breed. Rev.* 7**,** 1-41.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007). Rapid and cost effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17**,** 240-248.

Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* 47**,** 1435-42. doi: 10.1038/ng.3435

Secretaría General Técnica (Ministerio de Agricultura Alimentación y Medio Ambiente), (2015). "Encuesta sobre superficies y rendimientos de cultivos: resultados nacionales y autonómicos". Available online at: http://www.magrama.gob.es/es/estadistica/temas/estadisticas-agrarias/espana2015web_tcm7-401244.pdf (Accessed August 25, 2016)

Minoche, A.E., Dohm, J.C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol.* 12**,** R112.

Mitchell, J.H., Fukai, S., and Cooper, M. (1996). Influence of phenology on grain yield variation among barley cultivars grown under terminal drought. *Aust. J. Agric. Res.* 47**,** 757-774.

Mitra, J. (2001). Genetics and genetic improvement of drought resistance in crop plants. *Curr. Sci.* 80**,** 758-763.

Moinuddin, A., Fischer, R.A., Sayre, K.D., and Reynolds, M.P. (2005). Osmotic adjustment in wheat in relation to grain yield under water deficit environments. *Agron. J.* 97**,** 1062-1071.

Molina-Cano, J.L., Fra-Mon, P., Salcedo, G., Aragoncillo, C., Roca de Togores, F., and García-Olmedo, F. (1987). Morocco as a possible domestication center for barley: biochemical and agromorphological evidence. *Theor Appl Genet* 73**,** 531-536.

Moragues, M., Comadran, J., Waugh, R., Milne, I., Flavell, A., and Russell, J.R. (2010). Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theor. Appl. Genet.* 120**,** 1525-1534.

Moralejo, M., Romagosa, I., Salcedo, G., Sánchez-Monge, R., and Molina-Cano, J.L. (1994). On the origin of Spanish two-rowed barleys. *Theor Appl Genet* 87**,** 829-836.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5**,** 621-628. doi: 10.1038/nmeth.1226

Muñoz-Amatriaín, M., Lonardi, S., Luo, M., Madishetty, K., Svensson, J.T., Moscou, M.J., et al. (2015). Sequencing of 15622 gene-bearing BACs clarifies the gene-dense regions of the barley genome. *Plant J.* 84**,** 216-227. doi: 10.1111/tpj.12959

Mysore, K.S., and Ryu, C. (2004). Nonhost resistance: how much do we know? *Trends Plant Sci.* 9**,** 97-104. doi: 10.1016/j.tplants.2003.12.005

Nakamura, S., Pourkheirandish, M., Morishige, H., Kubo, Y., Nakamura, M., Ichimura, K., et al. (2016). Mitogen-activated protein kinase kinase 3 regulates seed dormancy in barley. *Curr. Biol.* 26**,** 775-781.

Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00170

Nakaya, A., and Isobe, S.N. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110**,** 1303-1316.

Nevo, E., and Chen, G. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant Cell Environ* 33**,** 670-685. doi: 10.1111/j.1365-3040.2009.02107.x

Newman, R.K., and Newman, C.W. (2008). *Barley for food and health.* Hoboken, New Jersey: John Wiley & Sons.

Newton, A.C., Akar, T., Baresel, J.P., Bebeli, P.J., Bettencourt, E., Bladenopoulos, K.V., et al. (2010). Cereal landraces for sustainable agriculture. A review. *Agronomy Sustainable Development* 30**,** 237-269. doi: 10.1051/agro/2009032

Nguyen, T.T., Klueva, N., Chamareck, V., Aarti, A., Magpantay, G., Millena, A.C., et al. (2004). Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Mol Genet Genomics* 272**,** 35-46. doi: 10.1007/s00438-004-1025-5

Niks, R.E., Alemu, S.K., Marcel, T.C., and van Heyzen, S. (2015). Mapping genes in barley for resistance to Puccinia coronata from couch grass and to P. striiformis from brome, wheat and barley. *Euphytica* 206**,** 487-499. doi: 10.1007/s10681-015-1516-y

Niks, R.E., and Marcel, T.C. (2009). Nonhost and basal resistance: how to explain specificity? *New Phytol* 182**,** 817-28. doi: 10.1111/j.1469-8137.2009.02849.x

Ning, Z., Cox, A.J., and Mullikin, J.C. (2001). SSAHA: a fast search method for large DNA databases. *Genome Res.* 11**,** 1725-1729.

Nowara, D., Gay, A., Lacomme, C., Shaw, J., Ridout, C., Douchkov, D., et al. (2010). HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen Blumeria graminis. *Plant Cell* 22**,** 3130-41. doi: 10.1105/tpc.110.077040

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., et al. (2013). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* 15**,** 256-278.

Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., et al. (2014). Mapping-by-sequencing identifies HvPhytochrome C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics* 198**,** 383-396. doi: 10.1534/genetics.114.165613

Parlevliet, J.E., and Ommeren, A.v. (1975). Partial resistance of barley to leaf rust, *Puccinia hordei*. II. Relationship between field traisl, micro plot tests and latent period. *Euphytica* 24**,** 293-303.

Pasam, R.K., Sharma, R., Malosetti, M., Eeuwijk, F.A.v., Haseneyer, G., Kilian, B., et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 12**,** 16.

Paulitz, T.C., and Steffenson, B.J. (2011). "Biotic stress in barley: disease problems and solutions", in *Barley: production, improvement and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 307-354.

Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol Phylogenetics Evol* 39**,** 70-82. doi: 10.1016/j.ympev.2006.01.023

Plant Genome and Systems Biology MIPS (2013). *Genome View* [Online]. Available: http://pgsb.helmholtz-muenchen.de/plant/barley/fpc/index.jsp (Accessed August 26, 2016).

Pliego, C., Nowara, D., Bonciani, G., Gheorghe, D.M., Xu, R., Surana, P., et al. (2013). Host-induced gene silencing in barley powdery mildew reveals a class of ribonuclease-like effectors. *Mol Plant Microbe Interact* 26**,** 633-42. doi: 10.1094/MPMI-01-13-0005-R

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., et al. (2015). Evolution of the Grain Dispersal System in Barley. *Cell* 162**,** 527-39. doi: 10.1016/j.cell.2015.07.002

Prada, D., Ullrich, S.E., Molina-Cano, J.L., Cistué, L., Clancy, J.A., and Romagosa, I. (2004). Genetic control of dormancy in a Triumph/Morex cross in barley. *Theor Appl Genet* 109**,** 62-70. doi: 10.1007/s00122-004-1608-x

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5**,** 94-100.

Rebetzke, G.J., Condon, A.G., Richards, R.A., and Farquhar, G.D. (2002). Selection for reduced carbon isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. *Crop Sci.* 42**,** 739-745.

Reddy, A.R., Chaitanya, K.V., and Vivekanandan, M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* 161**,** 1189-1202.

Reid, D.A. (1985). "Morphology and anatomy of the barley plant", in *Barley,* ed. D.C. Rasmusson, (Madison, Wisconsin: American Society of Agronomy-Crop Science Society of America-Soil Science Society of America), 73-101.

Ribaut, J.M., and Ragot, M. (2007). Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *J. Exp. Bot.* 58**,** 351-360.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26**,** 139-40. doi: 10.1093/bioinformatics/btp616

Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23**,** 2881-2887.

Saisho, D., Ishii, M., Hori, K., and Sato, K. (2011). Natural variation of barley vernalization requirements: implication of quantitative variation of winter growth habit as an adaptive trait in East Asia. *Plant Cell Physiol.* 52**,** 775-784.

Sakuma, S., Salomon, B., and Komatsuda, T. (2011). The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant Cell Physiol* 52**,** 738-749. doi: 10.1093/pcp/pcr025

Salvi, S., Druka, A., Milner, S.G., and Gruszka, D. (2014). "Induced genetic variation, TILLING and NGS-based cloning", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 287-310.

Säll, T. (1990). Genetic control of recombination in barley. *Hereditas* 112**,** 171-178.

Sallam, A.H., Endelman, J.B., Jannink, J.L., and Smith, K.P. (2015). Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *The Plant Genome* 8**,** 0. doi: 10.3835/plantgenome2014.05.0020

Sannemann, W., Huang, B.E., Mathew, B., and Léon, J. (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol. Breed.* 35**,** 1-16.

Sato, K., Flavell, A., Russell, J., Börner, A., and Valkoun, J. (2014). "Genetic diversity and germplasm management: wild barley, landraces, breeding materials", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag).

Sato, K., Shin, I.T., Seki, M., Shinozaki, K., Yoshida, H., Takeda, K., et al. (2009). Development of 5006 full-length CDNAs in barley: a tool for accessing cereal genomics resources. *DNA Res* 16**,** 81-9. doi: 10.1093/dnares/dsn034

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J., and Itoh, T. (2016a). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Res.* 23**,** 21-28. doi: 10.1093/dnares/dsv033

Sato, K., Yamane, M., Yamaji, N., Kanamori, H., Tagiri, A., Schwerdt, J.G., et al. (2016b). Alanine aminotransferase controls seed dormancy in barley. *Nat. Communications* 7**,** 11625.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 43**,** e37. doi: 10.1093/nar/gku1341

Schneeberger, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* 16**,** 282-288.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci* 16**,** 117-25. doi: 10.1016/j.tplants.2011.01.001

Schurch, N.J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22**,** 839-51. doi: 10.1261/rna.053959.115

Schweizer, P. (2014). "Host and nonhost response to attack by fungal pathogens", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 197-235.

Schweizer, P., and Stein, N. (2011). Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. *MPMI* 24**,** 1492-1501. doi: 10.1094 / MPMI -05-11-0107

Serraj, R., and Sinclair, T.R. (2002). Osmolyte accumulation: can it really increase crop yield under drought conditions? *Plant Cell Environ.* 25**,** 333-341.

Shaar-Moshe, L., Hubner, S., and Peleg, Z. (2015). Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biol.* 15**,** 111. doi: 10.1186/s12870-015-0493-6

Shantz, H.L. (1954). The place of grasslands in the Earth's cover. *Ecology* 35**,** 143-145. doi: 10.2307/1931110

Shen, Y., Sarin, S., Liu, Y., Hobert, O., and Pe'er, I. (2008). Coparing platforms for *C. elegans* mutant identification using high-throughput whole-genome sequencing. *PLoS One* 3**,** e4012. doi: 10.1371/journal.pone.0004012.t001

Shin, J.S., Corpuz, L., Chao, S., and Blake, T.K. (1990). A partial map of the barley genome. *Genome* 33**,** 803-808.

Silvar, C., Casas, A.M., Kopahnke, D., Habekub, A., Schweizer, G., Gracia, M.P., et al. (2010). Screening the Spanish Barley Core Collection for disease resistance. *Plant Breeding* 129**,** 45-52. doi: 10.1111/j.1439-0523.2009.01700.x

Slade, A.J., and Knauf, V.C. (2005). TILLING moves beyond functional genomics into crop improvement. *Transgenic Res.* 14**,** 109-115.

Slafer, G.A., Abeledo, L.G., Miralles, D.J., Gonzalez, F.G., and Whitechurch, E.M. (2001). Photoperiod sensitivity during stem elongation as an avenue to raise potential yield in wheat. *Euphytica* 119**,** 191-197.

Slafer, G.A., Araus, J.L., Royo, C., and Moral, L.F.G. (2005). Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. *Annals of Applied Biology* 146**,** 61-70.

Slafer, G.A., Satorre, E.H., and Andrade, H. (1994). "Increases in grain yield in bread wheat from breeding and associated physiological changes", in *Genetic improvement of field crops,* ed. G.A. Slafer, (New York: Marcel Dekker), 1-67.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., et al. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *J Systematics and Evolution* 53**,** 117-137. doi: 10.1111/jse.12150

Sparla, F., Falini, G., Botticella, E., Pirone, C., Talamè, V., Bovina, R., et al. (2014). New starch phenotypes produced by TILLING in barley. *PLoS One* 9**,** e107779.

Srivastava, A., Sarkar, H., Gupta, N., and Patro, R. (2016). RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics* 32**,** i192-i200.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* 8**,** 3673-3694.

Stanca, A.M., Romagosa, I., Takeda, K., Lundborg, T., Terzi, V., and Cattivelli, L. (2003). "Diversity un abiotic stresses", in *Diversity in Barley (Hordeum vulgare L.),* eds. R.V. Bothmer, H. Knüpffer, T.V. Hintum & K. Sato, (Amsterdam: Elsevier).

Steffenson, B., Hayes, P.M., and Kleinhofs, A. (1996). Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teres* f. *teres*) and spot blotch (*Cochliobolus sativus*) in barley. *Theor. Appl. Genet.* 92**,** 552-558.

Steffenson, B.J. (1992). Analysis of durable resistance to stem rust in barley. *Euphytica* 63**,** 153-167.

Stein, N. (2014). "Development of sequence resources", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 271-285.

Stein, N., Prasad, M., Scholz, U., Thiel, T., Zhang, H., Wolf, M., et al. (2007). A 1,000-loci transcript map of the barley genome: new anchoring points for intergrative grass genomics. *Theor. Appl. Genet.* 114**,** 823-839.

Szucs, P., Karsai, I., Zitzewitz, J.v., Meszaros, K., Cooper, L.L., Gu, Y.Q., et al. (2006). Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley *Theor Appl Genet* 112**,** 1277-1285.

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., et al. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proc Natl Acad Sci U S A* 105**,** 4062-4067. doi: 10.1073/pnas.0711034105

Taketa, S., Linde-Laursen, I., and Künzel, G. (2003). "Cytogenetic diversity", in *Diversity in barley (Hordeum vulgare),* eds. R.V. Bothmer, T.V. Hintum, H. Knüpffer & K. Sato, (Amsterdam, The Netherlands: Elsevier Science B.V.), 97-119.

Talamè, V., Bovina, R., Sanguineti, M.C., Tuberosa, R., Lundqvist, U., and Salvi, S. (2008). TILLMore, a resource for the discovery of chemically induced mutants in barley. *Plant Biotechnol. J.* 6**,** 477-485.

Talamé, V., Sanguineti, M.C., Chiapparino, E., Bahri, H., Ben Salem, M., Forster, B.P., et al. (2004). Identification of *Hordeum spontaneum* QTL alleles improving field performance of barley grown under rainfed conditions. *Ann. Appl. Biol.* 144**,** 309-319.

Teulat, B., Merah, O., Souyris, I., and This, D. (2001). QTLs for agronomic traits from Mediterranean barley progeny grown in several environments. *Theor. Appl. Genet.* 103**,** 774-787.

The James Hutton Institute (2014). *morexGenes - barley RNA-seq database* [Online]. Available: https://ics.hutton.ac.uk/morexGenes/ (Accessed August 26, 2016).

Thiel, T., Michalek, W., Varshney, R.K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* 106**,** 411-422.

Tolbert, D.M., Qualset, C.O., Jain, S.K., and Craddock, J.C. (1979). A diversity analysis of a world collection of barley. *Crop Sci.* 19**,** 789-794.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25**,** 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28**,** 511-5. doi: 10.1038/nbt.1621

Trevaskis, B., Hemming, M.N., Peacock, W.J., and Dennis, E.S. (2006). HvVRN2 responds to daylength, whereas HvVRN1 is regulated by vernalization and developmental status. *Plant Physiology* 140**,** 1397-1405. doi: 10.1104/pp.105.073486

Tsai, C.-J. (2013). Next-generation sequencing for next-generation breeding, and more. *New Phytol.* 198**,** 635-637.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., et al. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* 32**,** 569-76. doi: 10.1038/nbt.2908

Tuberosa, R., and Salvi, S. (2006). Genomics-based approaches to improve drought tolerance of crops. *Trends Plant Sci* 11**,** 405-12. doi: 10.1016/j.tplants.2006.06.003

Turner, A., Beales, J., Faure, S., Dunford, R.P., and Laurie, D.A. (2005). The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310**,** 1031-1034.

Ullrich, S.E. (2011). "Significance, adaptation, production, and trade of barley", in *Barley: production, improvement and uses,* ed. S.E. Ullrich, (Chichester, West Sussex, UK: Wiley-Blackwell), 3-13.

Varshney, R.K., Marcel, T.C., Ramsay, L., Russell, J., Röder, M.S., Stein, N., et al. (2007). A high density barley microsatellite consensus map with 775 SSR loci. *Theor. Appl. Genet.* 114**,** 1091-1116.

Varshney, R.K., Terauchi, R., and McCough, S.R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12**,** e1001883. doi: 10.1371/journal.pbio.1001883

Vasudevan, K., Cruz, C.M.V., Gruissem, W., and Bhullar, N.K. (2014). Large scale germplasm screening for identification of novel rice blast resistance sources. *Front. Plant Sci.* 5**,** 505.

Verstegen, H., Köneke, O., Korzun, V., and Broock, R.v. (2014). "The world importance of barley and challenges to further improvements", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 3-19.

Vogel, J.P., Gu, Y.Q., Twigg, P., Lazo, G.R., Laudencia-Chingcuanco, D., Hayden, D.M., et al. (2006). EST sequencing and phylogenetic analysis of the model grass *Brachypodium distanchyon*. *Theor. Appl. Genet.* 113**,** 186-195.

Wang, W., Vinocur, B., Shoseyov, O., and Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci* 9**,** 244-52. doi: 10.1016/j.tplants.2004.03.006

Watson, L. (1990). "The grass family, *Poaceae*", in *Reproductive versatility in the grasses,* ed. G. Chapman, (Cambridge: Cambridge University Press), 1-31.

Waugh, R., Thomas, B., Flavell, A., Ramsay, L., Comadran, J., and Russell, J. (2014). "Genome-wide association scans (GWAS)", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 345-365.

Wehner, G., Balko, C., Enders, M., Humbeck, K., and Ordon, F. (2015). Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. *BMC Plant Biol.* 15**,** 125. doi: 10.1186/s12870-015-0524-3

Weibull, J., Walther, U., Sato, K., Habekub, A., Kopahnke, D., and Proeseler, G. (2003). "Diversity in resistance to biotic stresses", in *Diversity in barley (Hordeum vulgare),* eds. R.V. Bothmer, T.V. Hintum, H. Knüpffer & K. Sato, (Amsterdam, The Netherlands: Elsevier Science B.V.), 143-178.

Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E., et al. (2006). A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* 7**,** 206. doi: 10.1186/1471-2164-7-206

Werner, K., Friedt, W., and Ordon, F. (2005). Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). *Mol. Breed.* 16**,** 45-55.

Werner, K., Friedt, W., and Ordon, F. (2007). Localisation and combination of resistance genes against soil-borne viruses of barley (BaMMV, BaYMV) using doubled haploids and molecular markers. *Euphytica* 158**,** 323-329.

Wicker, T., Taudien, S., Houben, A., Keller, B., Graner, A., Platzer, M., et al. (2009). A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J.* 59**,** 712-722. doi: 10.1111/j.1365-313X.2009.03911.x

Wolfe, M.S., and McDermott, J.M. (1994). Population genetics of plant pathogen interactions: the example of the *Erysiphe graminis-Hordeum vulgare* pathosystem. *Annu. Rev. Phytopathol.* 32**,** 89-112.

Wych, R.D., Simmons, S.R., Warner, R.L., and Kirby, E.J.M. (1985). "Physiology and development", in *Barley,* ed. D.C. Rasmusson, (Madison, Wisconsin: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America), 103-125.

Xiao, B., Huang, Y., Tang, N., and Xiong, L. (2007). Over-expression of a *LEA* gene in rice improves drought resistance under the field conditions. *Theor. Appl. Genet.* 115**,** 5-46.

Xiong, L., Schumaker, K.S., and Zhu, J. (2002). Cell signaling during cold, drought, and salt stress. *Plant Cell* 14**,** S165-S183.

Xu, Y., and Crouch, J.H. (2007). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48**,** 391-407.

Yahiaoui, S., Cuesta-Marcos, A., Gracia, M.P., Medina, B., Lasa, J.M., Casas, A.M., et al. (2014). Spanish barley landraces outperform modern cultivars at low-productivity sites. *Plant Breeding* 133**,** 218–226. doi: 10.1111/pbr.12148

Yahiaoui, S., Igartua, E., Moralejo, M., Ramsay, L., Molina-Cano, J.L., Ciudad, F.J., et al. (2008). Patters of genetic and eco-geographical diversity in Spanish barleys. *Theor Appl Genet* 116**,** 271-282. doi: 10.1007/s00122-007-0665-3

Yang, S., Fresnedo-Ramírez, J., Wang, M., Cote, L., Schweitzer, P., Barba, P., et al. (2016). A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine. *Horticulture Res.* 3**,** 16002.

Yang, S., Vanderbeld, B., Wan, J., and Huang, Y. (2010). Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. *Mol Plant* 3**,** 469-90. doi: 10.1093/mp/ssq016

Yao, W., Li, G., Zhao, H., Wang, G., Lian, X., and Xie, W. (2015). Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biol* 16**,** 187. doi: 10.1186/s13059-015-0757-3

Zapata, L., Peña-Chocarro, L., Pérez-Jordá, G., and Stika, H.P. (2004). Early neolithic agriculture in the Iberian Peninsula. *J. World Prehistory* 18**,** 283-325.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly de Bruijn graphs. *Genome Res.* 18**,** 821-829.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Communications* 7**,** 12617. doi: 10.1038/ncomms12617

Zohary, D., Hopf, M., and Weiss, E. (2013). *Domestication of plants in the old world, fourth edition.* Oxford University Press.

2. *Objectives*

The **main objective** of this thesis is the adoption of new research avenues made available by sequencing and genomics to study the genetic variability of Spanish barleys, and to deliver new tools and genes to geneticists and breeders. The course of this work was coincident in time with the publication of the barley sequenced-enriched physical and genetic map, in 2012, and the availability of new barley genomic tools, like exome capture platforms, which were incorporated into the work plan. These are the specific objectives:

1. To integrate the genomic sequence resources available for barley in a software tool made to locate genetic markers within physical and genetic maps, emphasizing sensitivity and accuracy of the reported positions, and providing information about the genes in the surrounding loci.

2. To use high-throughput sequencing tools to accelerate gene cloning. As a case study, a powdery mildew resistance QTL, present in a Spanish landrace, was subjected to fine mapping and candidate gene identification, by exome capture and sequencing, of informative recombinant inbred lines from a large mapping population.

3. To gain new insights about the genetic features conferring yield advantage under drought to a Spanish barley landrace through transcriptome sequencing of plants subjected to long term drought and heat stresses.

3. BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley

3.1. Introduction

The main challenge of users of genomic data for applied purposes is the efficient use of the enormous amount of data generated by sequencing (Boller, 2013). To aid geneticists and breeders of the *Triticeae* crops, some of the most important species for food security, different tools and data repositories have been developed recently, like HarvEST (Close et al., 2007), the T3 toolbox (http://triticeaetoolbox.org) or the Genome Zippers (Mayer et al., 2011).

The public release of the sequence-enriched genetic and physical map of barley (*Hordeum vulgare* L.) is being exploited for different purposes and already benefits breeding programs and companies worldwide, which previously had to rely solely on genetic maps and synteny-driven predictions. However, the current genomic assemblies are highly fragmented, as barley contains a major fraction of repeated sequences which hinder the assembly process (IBSC, Mayer et al., 2012). Moreover, the anchored sequences come from different cultivars and sequencing methods, increasing the richness as well as the complexity of the reference map. In addition, another sequence-enriched map, based on one of the previous assemblies, has been published recently (POPSEQ, Mascher et al., 2013).

Due to that complexity, it can be a daunting task for plant breeders to place arbitrary nucleotide sequences within the barley genome and to identify nearby genes and genetic markers, useful for tasks such as genetic map assessment or map-based cloning. Furthermore, it is expected that some sequences will have multiple matches due to the presence of putative duplicated chromosome segments, paralogs and pseudogenes, as well as possible inconsistencies in the assembly (Muñoz-Amatriain et al., 2013; Poursarebani et al., 2013).

The described genomic patchwork is not exclusive of barley, as genomes from other species have been and are currently being assembled with the aid of sequence-enriched maps, especially since the advent of Next Generation Sequencing methods and when dealing with highly repetitive genomes. Examples of the last are some species related to barley: *Brachypodium distachyon* (International Brachypodium Initiative, 2010), *Aegilops tauschii* (Jia et al., 2013) and hexaploid wheat (*Triticum aestivum* L., Paux et al., 2008; Paux et al., 2012). Among dicots, examples include grapevine (*Vitis vinifera* L., Jaillon et al., 2007), potato (*Solanum tuberosum* L., Sharma et al., 2013) or allotetraploid cotton (*Gossypium hirsutum* L., Yu et al., 2014).

Here we present a generic software platform designed to exploit genetic and physical information from sequence-enriched maps. As such, it can be configured to work with different sequence databases and maps, and thus it may take advantage of re-sequencing data. The application can be used with two types of input:

1) DNA sequences, which are aligned to genome assemblies to estimate their likely genomic positions. Two strategies are supported, allowing users to map either: i) arbitrary genomic sequences and/or ii) transcripts or Expressed Sequence Tags (ESTs), allowing for possible introns in the alignment.

2) Standard marker identifiers, so that users can have immediate access to pre-computed positions of markers. For example, those widely used in high-throughput genotyping experiments for a given species.

The BARLEYMAP pipeline, available athttp://floresta.eead.csic.es/barleymap, provides researchers a simple mapping report with details on genetic and physical position of markers, as well as additional results with surrounding genes and known markers from other datasets. Here it is benchmarked and implemented as a web tool with barley data, although its use can be extended, with the standalone version, to any other species with similar genomic resources available.

3.2. *Materials and Methods*

3.2.1. *Pipeline outline*

The BARLEYMAP pipeline Figure 3.1a was mainly implemented in Python 2.6 and includes SplitBlast, a Perl script for distributing BLAST jobs (Contreras-Moreira and Vinuesa, 2013). It has two main commands: [Align sequences] and [Find markers]. The first one uses a batch of FASTA-formatted DNA sequences as input, which are aligned by means of Blastn:Megablast from the BLAST package (Altschul et al., 1997), GMAP (Wu and Watanabe, 2005) or both. The "auto" mode calls both programs sequentially: input sequences are first aligned by Blastn, and those which do not yield alignments over customizable sequence identity and query coverage thresholds are then passed to GMAP. Results from both programs are filtered. In the case of Blastn, only the alignments with the best bit score are kept. GMAP results with poor identity and coverage are also discarded, as well as those marked as chimera. The alignment step is performed against one or more sequence databases (DBs in Figure 3.1a). These can be queried independently, merging the results afterwards, or by using a hierarchical strategy, in which only those queries not found in one DB are searched in the next ones (Figure 3.1b). The [Find markers] command instead takes a list of query identifiers as input and retrieves their alignment targets from pre-computed datasets. For the mapping step, the positions of targets in one or more genetic/physical maps are looked up and transferred to the initial queries. Results that provide the same location for a given query are merged into a single record. Once map positions have been compiled, the output report is augmented with genes or genetic markers anchored to those genome regions. Finally, the user has toggle controls to append to the results the functional annotation of those genes, as well as the genes to which the additional markers hit.

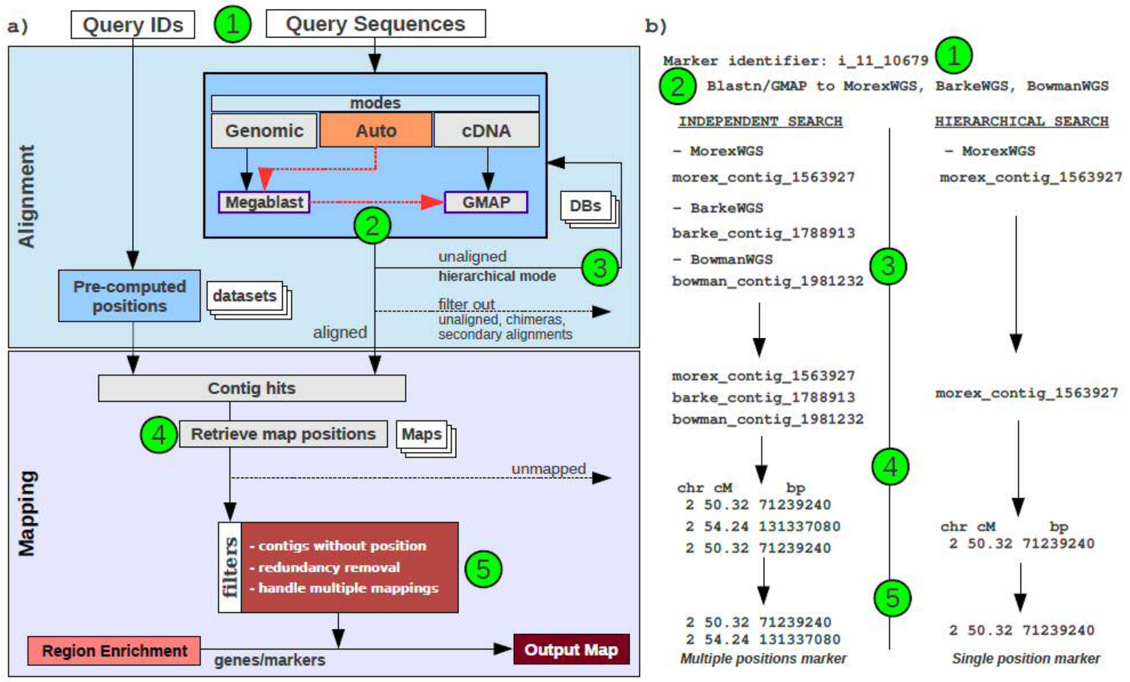


Figure 3.1. The BARLEYMAP pipeline. **a)** Two types of input can be queried: identifiers (query IDs) or FASTA sequences. The alignment modes allow to query for genomic and/or transcript sequences. The "auto" mode uses both Blastn:Megablast and GMAP (dotted arrows inside "modes" box). This will be repeated for each sequence reference (DB), independently, unless the hierarchical search is specified, in which case only unaligned queries will be searched in the remaining DBs. If those do not align against any DB, they will be discarded, along with secondary alignments, alignments without position (unmapped) and GMAP chimeras (dotted arrows). Alternatively, alignment targets can be recovered from pre-computed data. Map positions of the targets will be associated to the queries, and after several filtering steps, enrichment with surrounding genes and markers will be performed. Finally, annotation of genes maybe appended to the results. **b)** An example with marker i_11_10679, from the Infinium dataset. First, it is searched by means of sequence alignments against the barley shotgun assemblies. With the hierarchical search (right track), the marker is found in the Morex assembly, so no other DBs are queried. The position (chr: chromosome; cM: genetic position in centimorgan; bp: physical position in base pairs) of the Morex contig, which is the target of the alignment, is retrieved from the IBSC map and finally reported. If DBs are queried independently (left track), all the results are kept, and the position of such contigs retrieved. Finally, as the redundancy filter cannot distinguish between actual different positions and erroneous differences, it reports a marker with multiple positions. Circled numbers are used to relate the different steps from a) and b) flowcharts.

3.2.2. *Barley data configuration and application distribution*

BARLEYMAP was originally configured to work with barley data. Whole Genome Shotgun (WGS) assemblies of cultivars Morex, Barke and Bowman, as well as Morex Bacterial Artificial Chromosome (BAC) contigs and BAC-End sequences (BES) from Mayer et al. (2012), are employed as DBs. Genetic positions are retrieved separately from two recently published maps: the genetic/physical framework from the IBSC and the POPSEQ map of Morex contigs (Mascher et al., 2013). For the first one, mapping positions were obtained from the AC datasets and assigned to the DBs depending on the original source of the anchored sequence. As pre-calculated datasets, several collections of genetic markers were compiled: i) Infinium® iSelect 9K (Comadran et al., 2012), ii) DArTs™ (Wenzl et al., 2006), iii) DArTseq™ (Diversity Arrays Technology, Australia; Kilian et al., 2012) and iv) a set of SNPs generated via genotyping-by-sequencing (GBS) for the Oregon Wolfe Barley (OWB) population (Poland et al., 2012). All of them were aligned to the DBs by means of BARLEYMAP [Align sequences]. Cultivar Haruna Nijo full-length cDNAs (flcDNAs, Matsumoto et al., 2011) and HarvEST assembly 36 cDNA sequences (Close et al., 2007) were aligned to the DBs as well. 98% identity and 95% coverage were used as thresholds for the alignments in all cases, performing both Blastn and GMAP steps for aligning against every DB independently. For comparison purposes, the pre-previous datasets were also located using the hierarchical search with BARLEYMAP [Find markers] over the WGS assemblies (Morex, Barke and Bowman), BACs and BES references, in that order.

Finally, barley genes, including introns and up to 5,000 bp upstream of each transcript, were extracted from the Morex assembly, by means of custom scripts using the GTF data for High Confidence (HC) and Low Confidence (LC) genes from the MIPS FTP site (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data). Those two gene sets were used as targets for matching of all the markers from the pre-computed datasets. The same thresholds described above to align markers to the reference DBs were applied, using the hierarchical search to prioritize hits on the HC dataset. Functional annotations were also downloaded from the MIPS FTP site.

The standalone version of BARLEYMAP is distributed with the pre-computed barley datasets to support the [Find markers] mode without further requirements (the total package is ~15 MB). The attached documentation explains the configuration required to run the [Align sequences] mode and to add custom DBs, maps or datasets, including those from any other organism for which similar sequence-based mapping resources are available. The BARLEYMAP web application relies on a CherryPy web server to handle client requests, and enables the user to query all the barley resources described above. When several DBs are chosen by the user, the web application runs the hierarchical search by querying the WGS assemblies of cultivars Morex, Bowman and Barke; Morex BAC contigs and BES, in that order.

3.2.3. *Genetic map construction*

The performance of BARLEYMAP was benchmarked against a newly developed genetic map for the barley population SBCC073 x Orria. SBCC073 is a Spanish landrace-derived inbred line (from Archidona, Málaga, Spain), with high yield under drought (Yahiaoui et al., 2014). Orria [(((Api x Kristina) x M66.85) x Sigfrido's) x 79W40762] is a semi-dwarf cultivar selected in Spain from a CIMMYT nursery, which is highly productive across most Spanish regions. This cross was carried out within the Spanish National Breeding Program. This is a population of 101 BC_1F_5 lines, originally developed to carry out quantitative trait locus (QTL) studies, which was genotyped with a DArTseq™ GBS assay. One BC_1F_5 line was discarded on the basis of high percentages of heterozygous data. Therefore, the final mapping population comprised 100 lines. A genetic map was constructed in a two-step process, using first Joinmap 4 (Ooijen, 2006) and then MSTMap (Wu et al., 2008). Resulting linkage groups were assigned to barley chromosomes based on the genomic positions assigned by BARLEYMAP.

The same polymorphic SNP markers were also queried by means of BARLEYMAP [Find markers] to both IBSC and POPSEQ maps, in hierarchical mode, to obtain *in-silico* maps. Spearman rank correlations were calculated between positions in the resulting genetic map and positions in the genetic/physical maps of IBSC and POPSEQ, using GenStat 16 (Payne, 2009).

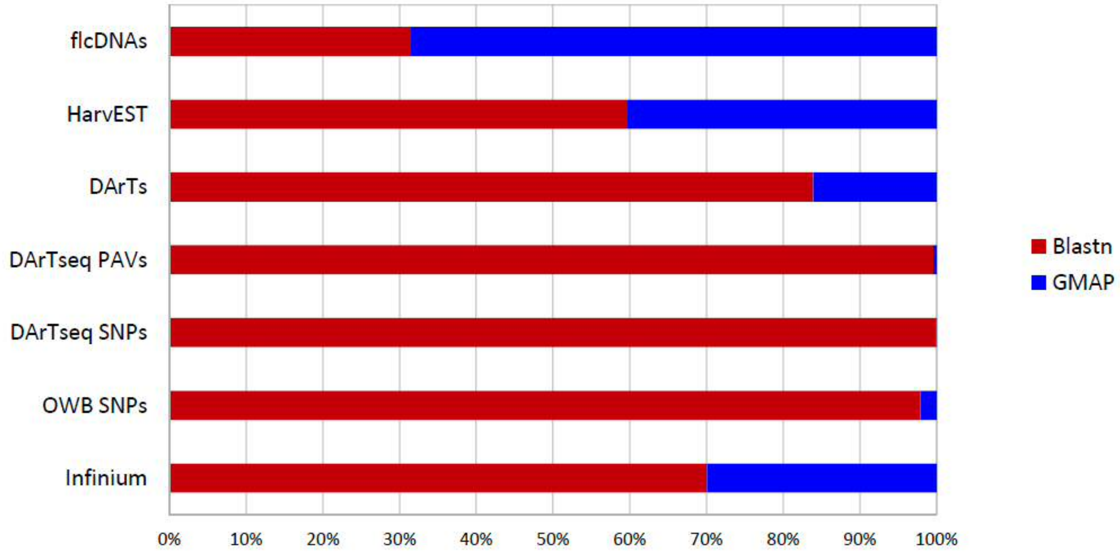


Figure 3.2. Percentage of sequences found by either Blastn or GMAP. The hierarchical method was used to align every dataset to barley sequence references.

3.3. *Results*

3.3.1. *Alignment of barley transcripts*

To test the alignment step of BARLEYMAP (Figure 3.1a), the "auto" mode was selected to match long transcripts against the WGS assemblies of cultivars Morex, Barke and Bowman, as well as against the BAC contigs and BES from the IBSC, in that order by means of the hierarchical search. Of 28,620 flcDNAs from cultivar Haruna Nijo (Matsumoto et al., 2011), 60% were successfully aligned, with 68.5% of the alignments obtained by GMAP (Figure 3.2). Applying the same method, at least one hit was found for 59% out of 70,148 HarvEST Unigenes, with almost 60% of them aligned by Blastn. 79% and 86% of the previous hits were matched against the first queried database, the WGS assembly of cultivar Morex. The rest, 3,578 and 5,725 queries respectively, could only be matched in the remaining references.

3.3.2. *Alignment of barley markers*

A second benchmark consisted on mapping diverse collections of genetic markers, described in Materials and Methods, which are widely used by geneticists and breeders:

1) 7,864 Infinium® iSelect SNPs.

2) 2,000 Diversity Array Technology presence-absence (PAV) markers (DArTs™).

3) 24,061 GBS markers, including both SNP and PAV markers (DArTseq™)

4) 34,396 GBS SNP markers from the OWB population.

As observed for transcripts, a significant number of Infinium (30%) and DArT (16%) markers could only be confidently aligned with GMAP (Figure 3.2). However, this proportion was tiny for GBS markers, especially for DArTseq SNPs, which were mostly aligned by Blastn. Nonetheless, around 1,400 OWB GBS markers were aligned by GMAP.

Although these markers are short DNA sequences, their alignments produced mostly single hits (over 98%) when searched independently in the WGS assemblies of cultivars Morex, Barke and Bowman. However, such percentage was smaller for BAC contigs and BES references (64% and 88%, respectively). Using the hierarchical method, this percentage was near 99% for every marker dataset (Table 3.1).

The databases yielding the highest number of aligned markers were the WGS assemblies, with those from cultivars Morex and Bowman being slightly more informative than the one from cultivar Barke. The number of markers aligned to BAC contigs and BES references was smaller in comparison. In all cases, the use of the hierarchical search method resulted in a larger number of markers available for position retrieval.

Table 3.1. Genetic markers aligned with BARLEYMAP. The hierarchical search method was used. The proportion of matched queries with a single alignment hit is shown as well.

Marker sets	Markers	Aligned (%)	Single target (%)
DArTs	2,000	1,340 (67.0)	1,334 (99.6)
DArTseq PAVs	15,526	7,498 (48.3)	7,456 (99.4)
DArTseq SNPs	8,535	6,876 (80.6)	6,832 (99.4)
OWB SNPs	34,396	22,992 (66.8)	22,731 (98.9)
Infinium	7,864	7,304 (92.9)	7,291 (99.8)
Total	68,321	46,010 (67.3)	45,644 (99.2)

3.3.3. *Mapping of aligned markers to barley genetic/physical maps*

Markers aligned to sequence DBs (Table 3.1) were then assigned genetic positions retrieved from the IBSC and POPSEQ sequence-enriched maps. While POPSEQ comprises only contigs from the Morex assembly, IBSC map positions can be retrieved for contigs from up to five different DBs. Thus, in the latter case, marker positions were obtained either i) by merging the results from their alignment to each DB independently or ii) from the hits obtained with the hierarchical method (see Materials and Methods). As summarized in Table 3.2, the highest number of markers was mapped to the IBSC map, with 59% of them having a single map position. In contrast, the POPSEQ results had the least number of mapped markers, but 99% of them had a single map position. Regarding the hierarchical search, it misses ~4,300 marker positions with respect to IBSC, but a large majority of the sequences mapped (99%) had a single map position, just as observed for POPSEQ.

A significant fraction of all the mapped markers lie on identical genetic positions and do not contribute to effectively resolve genomic intervals. Thus, considering only unique genetic locations, the hierarchical search method yields the maximum number of landmarks, with 6,908. This advantage of the hierarchical method when compared to the IBSC results comes at the cost of masking markers with multiple positions in different DBs. However, the information lost is mostly redundant, as revealed by the analysis of the positions of markers: for markers with multiple locations in the same DB reported by both search methods, 102 out of 140 (73%) lay in different chromosomes; for those removed by the hierarchical method (15,493) only 8% are in different chromosomes and most of the remaining are less than 5 cM apart.

Table 3.2. Comparison of different mapping approaches. Result of mapping all the 68,321 markers from Table 3.1 to the IBSC and POPSEQ maps. For IBSC, results obtained by the independent and hierarchical search strategies are shown.

Map / Search type	markers with map position	markers with single position	unique genetic positions
IBSC / Independent	38,528	22,891	5,675
POPSEQ / Morex assembly	30,330	30,232	2,721
IBSC / Hierarchical	34,203	34,063	6,908

3.3.4. *Matching of genetic markers to barley genes*

By taking the IBSC gene annotations, the sequences of genes, including introns and up to 5,000 bp upstream of each transcript, were obtained from the WGS assembly of cultivar Morex, yielding 62,426 HC and 69,299 LC sequences. A total of 68,321 markers from the datasets in Table 3.1 were matched to these gene sequences with the [Align sequences] command, hierarchical search and default parameters, as explained in Materials and Methods. Of these, 39.23% matched currently annotated genes, with 68% being HC genes.

3.3.5. *Validating genetic maps of barley populations*

The population SBCC073 x Orria yielded 2,483 polymorphic SNPs. These were filtered attending to presence of missing data (<10%), heterozygotes (<10%), or allelic frequency of the donor parent (SBCC073) over 75%. After filtering, 1,227 SNPs were used to construct a genetic map. In a first step, linkage groups were created with software Joinmap using the maximum likelihood algorithm. Then, in a second step, the distances between markers were recalculated based on the Kosambi's mapping function using MSTMap, which works more efficiently when the number of markers is large. A total of 11 linkage groups were thus identified, representing 4 whole chromosomes (1H, 3H, 4H and 5H) and 3 fragmented ones (chromosome 2H in 3 groups, chromosomes 6H and 7H in 2 groups each). Linkage groups were assigned to chromosomes, and the resulting genetic positions of the 1,227 SNP markers compared to the positions assigned to them by BARLEYMAP by hierarchically searching against either POPSEQ or IBSC references. Correlation analyses, summarized in Figure 3.3, reveal that loci order in the genetic map derived from the population is largely similar to the implicit ordering of positions automatically assigned by the [Find markers] command. The weighted averages obtained across linkage groups for POPSEQ and IBSC were 0.92 and 0.96, respectively. There were nonetheless three exceptions: i) a small linkage group made of 10 markers for which the genetic map is necessarily less consistent than for larger groups; ii) linkage group 4H and; iii) linkage group 6H.2. For these last two groups there was good agreement with only one of the two physical maps used, pointing out to local discrepancies between the data from IBSC and POPSEQ (see Figure 3.3).

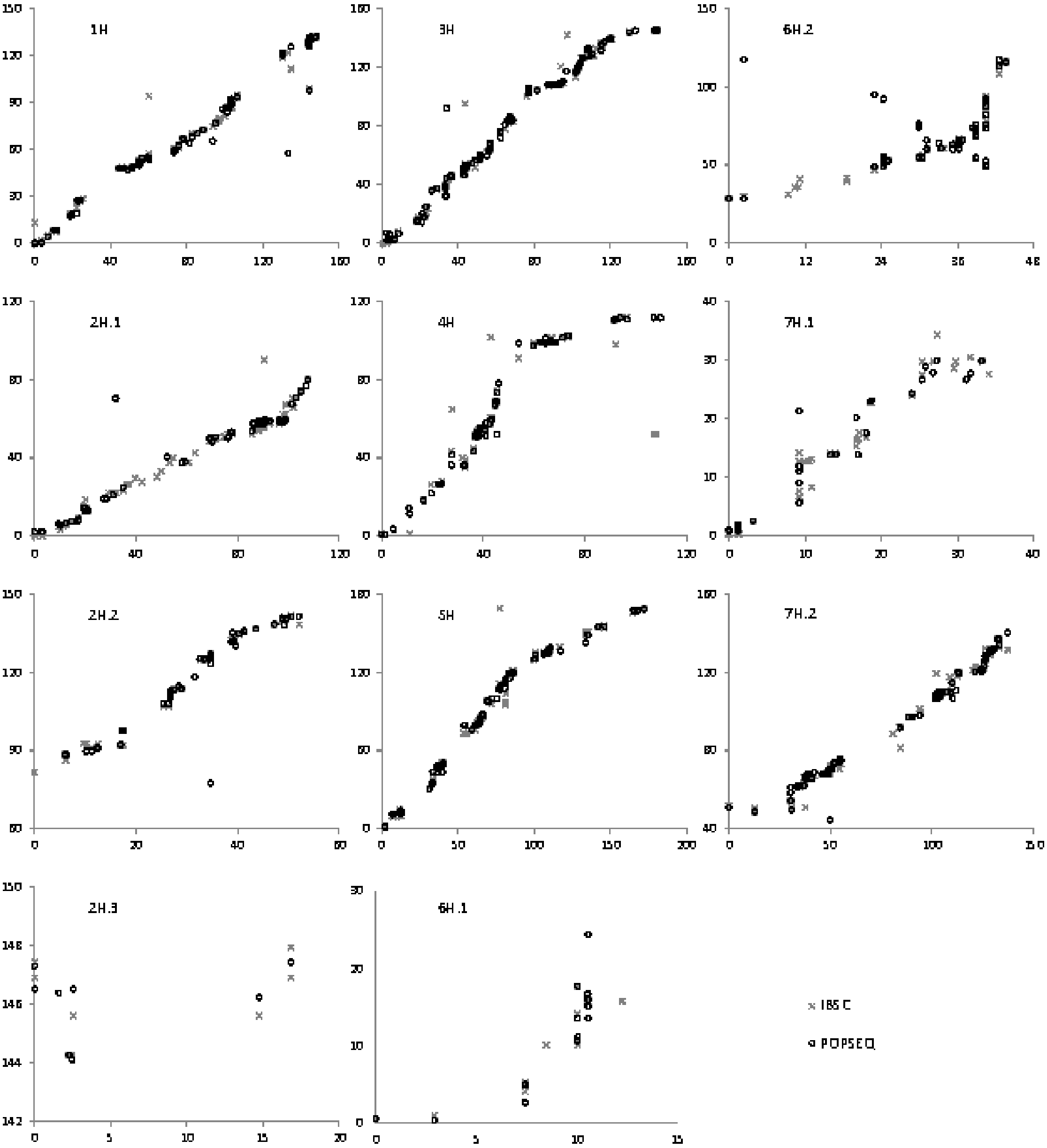


Figure 3.3. Comparison of BARLEYMAP positions and genetic map. 2D scatter plots comparing the RIL population map (X axis) against the IBSC and POPSEQ in-silico maps (Y axis). Positions of marker loci in cM. The positions of the IBSC genetic/physical map (grey crosses) and the POPSEQ map (black circles) were obtained using the hierarchical method of BARLEYMAP [Find markers].

3.4. Discussion

Plant breeders have relied upon large numbers of de novo genetic maps and consensus maps to deduce information about the relative position of their markers in relation to others. The lack of common markers between maps has hindered the progress towards the identification of genes or QTL underlying relevant traits for breeding. The era of abundant sequence data is providing the opportunity to identify numerous new markers, which are implemented in relatively cheap and high-throughput platforms, widely used by the community. This is the case of GBS protocols or array genotyping systems based on data from SNP calling pipelines.

In addition, such diversity of markers makes it possible to construct high-resolution genetic maps, which, within genome sequencing projects, are used in conjunction with physical maps to anchor sequences from shotgun or BAC sequencing. These resources may not constitute a complete genome, but often contain a high proportion of the genes of an organism, correctly placed in linear order. Many of the absent assembled contigs come from highly repetitive, less gene abundant regions (Mayer et al., 2012). Thus, exploiting such sequence-enriched maps can be of help when locating genetic markers, when relating and comparing different maps to each other, or in map-based cloning. This must be done with caution, since the actual genotype or population under analysis could be more or less closely related to the sequence references or even it could bear local rearrangements (Farré et al., 2012). Moreover, these sequence-enriched maps tend to have specific features for different species, since each genome project may opt to use one or several genotypes as references, or could use different sequencing technologies and sources. For these reasons, it would be helpful to have tools flexible enough to help fill the gap between specific genomic databases and the data used by plant breeders.

General resources, such as Ensembl Plants (Kersey et al., 2014), or more specific ones, as the IPK Barley server (http://webblast.ipk-gatersleben.de/barley/viroblast.php), can certainly be of help for these tasks. However, they are purely sequence-based and do not make explicit use of the genetic maps underlying the physical assembly. Therefore, they do not filter alignment matches in order to summarize mapping results, thus not considering possible redundant positions as well as those with non-consistent locations along the genome, originated from subtle differences among data sources. In addition, the choice of BLAST as the only search engine complicates mapping transcripts, as introns frequently interrupt the matching regions and produce short local alignments, confounding query coverage. Finally, these resources fail to include collections of genetic markers routinely used by breeders for genotyping their plant materials. On the other hand, HarvEST (Close et al., 2007), another important barley resource, does include SNP markers and IBSC positions of Morex genes and homologs in other grasses, but cannot be used to interactively map selected DNA sequences within the genome.

A unique feature of BARLEYMAP is the integration of alignment to sequence references and mapping to genetic and physical frameworks. The combined use of Blastn and GMAP allows BARLEYMAP to align transcripts, and markers derived from them, as demonstrated here by aligning flcDNAs, ESTs, and several genetic marker collections. Moreover, the use of a hierarchical method for alignment provides a reasonable compromise between the use of a single DB and the direct merging of results from the independent alignment to several DBs. In the first case, a number of queries may be absent, depending on the completeness of the assembly or presence-absence polymorphisms. For instance, cultivar Morex, as a spring cultivar, lacks the *VrnH2* gene (von Zitzewitz et al., 2005). Being an incomplete reference, other genes might only be found in alternative datasets, as the subset of flcDNAs (21%) that cannot be confidently aligned to Morex but is found in other references. The second approach, the alignment of every sequence to every reference, in addition to being a time-consuming process, produces queries with multiple targets and redundancy, both difficult to identify and fix, and can significantly reduce the number of useful markers associated to a single, unambiguous map location. The hierarchical method reduces computing time by aligning only the remaining unaligned sequences. In addition, queries with multiple mappings will arise only when the different locations are found in the same DB. As a drawback, the hierarchical method could be masking true multiple alignments (for example copy-number variation polymorphisms) in the case of markers for which different targets are found in different DBs. However, most of those multiple positions seem to be very close to each other and are almost completely removed when using the hierarchical method. This suggests that such multiple positions are mostly artificial, generated by the independent mapping to different assemblies and sources. For efficiency and to ease downstream analysis, the web application uses only the hierarchical method when querying several DBs. The standalone application gives the user full control on using or not the hierarchical method.

Therefore, BARLEYMAP allows barley geneticists and breeders to exploit their new and existing genotyping data in an accessible and time-saving manner, by integrating different marker types and flexible annotation retrieval in a single framework. It does so efficiently, as demonstrated by the good agreement between the orders of a purpose-built genetic map and the positions derived from BARLEYMAP. According to these observations it would be tempting to skip the mapping step altogether for any new population under study, and to proceed for further analyses using directly the positions derived from sequences-enriched genetic/physical maps. This benchmark suggests that analyses based on positions such as those produced by BARLEYMAP from currently available barley resources would produce reasonable results. However, the differences obtained by aligning the GBS markers to the two main genomic resources (IBSC and POPSEQ) advise against using such information as the gold standard for position, at least until the accuracy of barley references improves, and even then maybe only for genotypes close enough to the existing references.

A similar statement can be said for fine mapping purposes. Despite the fact that it can be of great help the use of knowledge about surrounding genes and markers provided by BARLEYMAP, when working with a marker defined interval, the positions and relative order of such features should be assessed carefully due to the technical and biological variability that might exist in the reference data (Hofmann et al., 2013; Liu et al., 2014).

Finally, BARLEYMAP allows research groups to use custom databases, maps and pre-computed datasets of markers, so that they may work with their own data and share it in a light-weight manner. Therefore, it provides a framework that ranges from a ready-to-work application for the retrieval of positional data from barley resources, up to a customizable pipeline that allows working with sequence-based positional data, if available, from any organism.

3.5. References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25**,** 3389-402.

Boller, B. (2013). Interview with Beat Boller, President of EUCARPIA, the European Association for Research on Plant Breeding. *International Innovation (Environment)***,** 42-43.

Close, T.J., Wanamaker, S., Roose, M.L., and Lyon, M. (2007). "HarvEST: an EST database and viewing software", in *Plant bioinformatics: methods and protocols,* ed. D. Edwards, (Totowa, New Jersey: Humana Press), 161-77. doi: 10.1007/978-1-59745-535-0_7

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., et al. (2012). Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44**,** 1388-92. doi: 10.1038/ng.2447

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79**,** 7696-701. doi: 10.1128/AEM.02411-13

Farré, A., Cuadrado, A., Lacasa-Benito, I., Cistué, L., Schubert, I., Comadran, J., et al. (2012). Genetic characterization of a reciprocal translocation present in a widely grown barley variety. *Mol Breed* 30**,** 1109-1119. doi: 10.1007/s11032-011-9698-z

Hofmann, K., Silvar, C., Casas, A.M., Herz, M., Buttner, B., Gracia, M.P., et al. (2013). Fine mapping of the Rrs1 resistance locus against scald in two large populations derived from Spanish barley landraces. *Theor Appl Genet* 126**,** 3091-102. doi: 10.1007/s00122-013-2196-4

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463**,** 763-8. doi: 10.1038/nature08747

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449**,** 463-7. doi: 10.1038/nature06148

Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., et al. (2013). Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496**,** 91-5. doi: 10.1038/nature12028

Kersey, P.J., Allen, J.E., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., et al. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res* 42**,** D546-52. doi: 10.1093/nar/gkt979

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 888**,** 67-89. doi: 10.1007/978-1-61779-870-2_5

Liu, H., Bayer, M., Druka, A., Russell, J.R., Hackett, C.A., Poland, J., et al. (2014). An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. *BMC Genomics* 15**,** 104. doi: 10.1186/1471-2164-15-104

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., et al. (2013). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* 76**,** 718-27. doi: 10.1111/tpj.12319

Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., et al. (2011). Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.* 156**,** 20-8. doi: 10.1104/pp.110.171579

Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 23**,** 1249-63. doi: 10.1105/tpc.110.082537

Mayer, K.F.X., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491**,** 711-6. doi: 10.1038/nature11543

Muñoz-Amatriain, M., Eichten, S.R., Wicker, T., Richmond, T.A., Mascher, M., Steuernagel, B., et al. (2013). Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol* 14**,** R58. doi: 10.1186/gb-2013-14-6-r58

Ooijen, J.W.v. (2006). "JoinMap 4, software for the calculation of genetics linkage maps in experimental populations".

Paux, E., Sourdille, P., Mackay, I., and Feuillet, C. (2012). Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnol Adv* 30**,** 1071-88. doi: 10.1016/j.biotechadv.2011.09.015

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., et al. (2008). A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322**,** 101-4. doi: 10.1126/science.1161847

Payne, R.W., Murray, D.A., Harding, S.A., Baird, D.B. & Soutar, D.M. (2009). "GenStat for Windows (12th Edition) Introduction".

Poland, J.A., Brown, P.J., Sorrells, M.E., and Jannink, J.L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7**,** e32253. doi: 10.1371/journal.pone.0032253

Poursarebani, N., Ariyadasa, R., Zhou, R., Schulte, D., Steuernagel, B., Martis, M.M., et al. (2013). Conserved synteny-based anchoring of the barley genome physical map. *Funct Integr Genomics* 13**,** 339-50. doi: 10.1007/s10142-013-0327-2

Sharma, S.K., Bolser, D., de Boer, J., Sonderkaer, M., Amoros, W., Carboni, M.F., et al. (2013). Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 (Bethesda)* 3**,** 2031-47. doi: 10.1534/g3.113.007153

von Zitzewitz, J., Szucs, P., Dubcovsky, J., Yan, L., Francia, E., Pecchioni, N., et al. (2005). Molecular and structural characterization of barley vernalization genes. *Plant Mol Biol* 59**,** 449-67. doi: 10.1007/s11103-005-0351-2

Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E., et al. (2006). A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* 7**,** 206. doi: 10.1186/1471-2164-7-206

Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21**,** 1859-75. doi: 10.1093/bioinformatics/bti310

Wu, Y., Bhat, P.R., Close, T.J., and Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4**,** e1000212. doi: 10.1371/journal.pgen.1000212

Yahiaoui, S., Cuesta-Marcos, A., Gracia, M.P., Medina, B., Lasa, J.M., Casas, A.M., et al. (2014). Spanish barley landraces outperform modern cultivars at low-productivity sites. *Plant Breeding* 133**,** 218–226. doi: 10.1111/pbr.12148

Yu, J.Z., Young, C.J.L., Pepper, A.E., Li, F., Yu, S., Buyyarapu, R., et al. (Year). "Toward Cotton Molecular Breeding: Challenges and Opportunities", in: *International Plant & Animal Genome XXII*), W604.

4. A cluster of NBS-LRR genes resides in a barley powdery mildew resistance QTL on 7HL

4.1. Introduction

Blumeria graminis is an obligate biotrophic fungal ectoparasite of grasses. It colonizes the surface of leaves, feeding from the epidermal cells by means of specialized organs called haustoria (Jørgensen, 1988). The *forma specialis hordei* causes powdery mildew in barley (*Hordeum vulgare* L.), which leads to severe losses in yield and grain quality in temperate latitudes worldwide (Zhang et al., 2005; Ames et al., 2015). This results in a significant economic impact since barley is one of the most widely grown crops (for a recent review, see Verstegen et al., 2014). Consequently, the interaction of barley and powdery mildew has been extensively studied (for a recent review, see Schweizer, 2014) and many resistance genes known as mildew genes (*Ml* genes) have been described (Friedt and Ordon, 2007).

However, most of them are still molecularly uncharacterized. Among cloned genes, the recessive *mlo* stands out; providing durable resistance (Jorgensen, 1992) which has remained effective for over 30 years and copes with a broad spectrum of pathogen isolates (Büschges et al., 1997). The other major powdery mildew resistance genes cloned so far are located at the *Mla* locus, which consists of a cluster of genes encoding for related proteins (Wei et al., 1999). Several *Mla* alleles have been cloned (Halterman et al., 2001; Zhou et al., 2001) out of the many resistance specificities described for this locus (Jørgensen and Wolfe, 1994).

Cloning of *mlo* and *Mla* involved long and laborious efforts. Specifically, fine-mapping of these genes consisted in recurrent steps of marker development, polymorphism detection and genotyping, looking for recombinants. This was done to narrow down the respective genetic intervals until an affordable physical size of the region was achieved, and subsequently resolved by chromosome walking or sequencing of subclones developed using yeast or bacterial artificial chromosome (BAC) clones. This cumbersome procedure was most challenging for species like barley due to the lack of genomic resources and its large and highly repetitive genome (Krattinger et al., 2009). However, the recent advent of high-throughput sequencing, by means of NGS technologies, has accelerated the development of synteny resources (Mayer et al., 2011), sequenced enriched physical-maps (Mayer et al., 2012; Mascher et al., 2013a; Ariyadasa et al., 2014; Muñoz-Amatriaín et al., 2015), genotyping (Comadran et al., 2012; Poland et al., 2012) and sequence capture platforms (Mascher et al., 2013b). In consequence, gene cloning now benefits from the easier and faster genotyping of high-resolution mapping populations, high-throughput polymorphism detection in parental lines, and new fine mapping approaches, such as mapping-by-sequencing (Mascher et al., 2014).

Typical disease resistance genes from plant innate immunity encode receptors usually activated through recognition of molecules from the pathogen (Flor, 1971). These receptors are usually subdivided in two classes. Transmembrane pattern-recognition receptors represent the first active line of defense at the plant cell surface (Jones and Dangl, 2006). They enable the recognition of microbe-associated molecular patterns and induce pattern-triggered immunity. In contrast, a second class of resistance proteins induces elicitor-triggered immunity, detecting either the action or the structure of pathogen molecules inside host cells. These receptors are polymorphic, defining a repertoire for the detection of distinct pathogen effectors (Maekawa et al., 2011). Most genes in this second class encode proteins of the NBS-LRR family (McHale et al., 2006).

NBS-LRRs are abundant in plant genomes (Yue et al., 2013) and are encoded by genes often located in clusters of closely related members (Michelmore and Meyers, 1998). These evolve through rapid expansion and contraction of gene families (Meyers et al., 2003; Monosi et al., 2004; Zhou et al., 2004). In barley, an example of an NBS-LRR cluster is that residing in the *Mla* locus (Seeholzer et al., 2010). NBS-LRR genes encode two protein domains. The nucleotide-binding site (NBS) domain bears a string of motifs largely conserved in plants, both in sequence and in order (Marone et al., 2013). NBS domains are followed by a leucine-rich repeat (LRR) domain, which is generally more variable, often associated with direct or indirect non-self-recognition (Spoel and Dong, 2012). Besides *Mla* genes, many other disease resistance genes have been associated to NBS-LRR loci in plants (reviewed in Marone et al., 2013). For instance, in barley Rpg5/rpg4 confers resistance to *Puccinia graminis* (Brueggeman et al., 2008), and Rdg2a to *Drechslera graminea* (Bulgarelli et al., 2010). Additional NBS-LRR genes have been cloned in wheat and its wild relatives (discussed in Gu et al., 2015).

This study took advantage of the sequencing-based genomic resources available for barley to fine map a powdery mildew resistance QTL. A high-resolution mapping population was developed to narrow down the QTL interval, followed by exome sequencing of recombinant lines with contrasting resistance phenotypes. The results revealed that genes located in the physical region corresponding to the genetic interval where the QTL is placed, formed a cluster of closely related NBS-LRRs, of which the resistant lines have unique haplotypes.

4.2. Materials and methods

4.2.1. Plant material and mapping population

A BC_1F_2 population was obtained from the cross Plaisant x RIL151. Recombinant inbred line (RIL) 151 derives from the SBCC097 x Plaisant population (Silvar et al., 2010). This line has only one of the two resistance QTL identified in the original donor landrace, on 7HL (Silvar et al., 2012). BC_1F_2 seeds were planted in 96-well trays and sampled 10 days after sowing. For each individual BC_1F_2 plant, a 0.6 cm leaf disk was cut. DNA extraction and amplification was carried out with the Extract-N-Amp Plant PCR kit (Sigma, USA). A cleaved amplified polymorphic sequence (CAPS) marker, QBS58, and a microsatellite, EBmac0755, were used as flanking markers to delimit the QTL interval. Restriction digestion of PCR products was carried out in a 20 µl volume using 1.5 U of the respective restriction endonuclease (NEB, Fermentas). Plants were selected if they showed recombination between both markers. Data from another 4 markers (QBS52, QBS46, QBS44 and QBS36) were used to perform linkage analysis with JoinMap 4.0 (Ooijen, 2006), using Kosambi's map function. Selected plants were vernalized for 6 weeks at 3-8°C, 8 h light, then transplanted to pots and transferred to a growth chamber, where the plants were grown under long-day conditions (16 h light, 250 µmol m-2 s-1, 20°C, 60% relative humidity/8h dark, 16°C, 65% relative humidity). Plants were bagged before seed setting.

To select homozygous recombinants in the BC_1F_3 generation, 20 progeny plants of each selected BC_1F_2 plant were screened as explained above. Additional CAPS and pyrosequencing markers were incorporated at this stage. To verify the genotype of the BC_1F_4 recombinant lines, genomic DNA was isolated from frozen leaves using the NucleoSpin Plant II kit (Macherey-Nagel, Germany).

4.2.2. *Pathogen isolates and disease assessment*

Four isolates of *B. graminis* f. sp. *hordei* (R79, R126, R164 and R225) were used to score resistance/susceptibility in the parents and BC_1F_4 recombinant lines. These isolates were propagated on plants of the susceptible cv. Igri. The seedlings were grown under mildew-free conditions at 20°C with 60-70% relative humidity and a 16 h light/8 h dark photoperiod. Ten days after sowing, when the first leaf was fully expanded, five plants per line were inoculated with the different isolates by brushing them with powdery mildew spores. Inoculated plants were maintained under the same conditions described above. The infection types were recorded on a scale of 0–4 (including intertypes) 10 days after inoculation, following the procedure of Torp et al. (1978) and Jensen et al. (1992). Plants with infection scores <2 were classified as resistant, otherwise were labelled as susceptible. Pictures were also taken 10 days after infection.

4.2.3. *Exome sequencing*

Genomic DNA from three BC_1F_4 lines (1476, 1766 and 2085) was extracted from leaf tissue using the NucleoSpin Plant II XL kit from Macherey-Nagel. Exome capture and DNA sequencing was performed at CNAG (Centro Nacional de Análisis Genómico, Barcelona). DNA capture was performed in a single reaction with the Roche Nimblegene SeqCap EZ Developer kit (Mascher et al., 2013b), following the instructions from the manufacturer. DNA was barcoded with TruSeq adapters and pooled before hybridization to the exome probes. DNA fragmentation and size selection was performed to produce 2x101 bp paired-end reads with average insert size of 150 bp. Sample preparation followed standard Illumina TruSeq procedures. Sequencing was performed in two separate runs of an Illumina HiSeq2000, each in a single lane.

Reads were aligned to the Morex whole genome sequencing (WGS) assembly (Mayer et al., 2012) with BWA MEM (Li and Durbin, 2009) with default parameters. Read duplicates were tagged by means of MarkDuplicates from picard-tools-1.113 (http://broadinstitute.github.io/picard). Variant detection was performed combining SAMtools (Li et al., 2009) and GATK (McKenna et al., 2010). Variants were filtered out, requiring a minimum depth of 10 and a minimum quality of 30 in each genotyped line. Polymorphic variants were obtained comparing the data of the BC_1F_4 lines with variants for SBCC097 and Plaisant from another exome capture essay (unpublished).

To look for the recombination points in the sequences of the three BC_1F_4 lines, a score was assigned to each variant identified after the exome capture. If a variant was like Plaisant, the score was increased by 1. If the variant was like SBCC097, the score was decreased by 1 instead. If it was different to the parents, the score remained unchanged. Therefore, the variants in which the three lines were Plaisant-like received a score of +3 in that position in the genome. On the contrary, if all three lines were like SBCC097, the score was -3. This was repeated for every variant. The scores of the variants lying on a single Morex WGS contig were averaged to obtain a single contig score.

4.2.4. Identification and annotation of the BACs located within the QTL region

Contigs of each BAC associated to finger-printed contig (FPC) 591, from IBGSC (Mayer et al., 2012) and University of California Riverside (UCR BACs, hereafter; Muñoz-Amatriaín et al., 2015), were concatenated to build up BAC pseudoscaffolds. Gene annotations were obtained from IBGSC data, by alignment of the associated Morex WGS contigs to Uniref90 and UniprotKB (blastx, maximum e-value 1e-50) and by identification and annotation of open reading frames (ORFs) with 'getorf' (Rice et al., 2000; -minsize 90) and the script 'run_predict.sh' from CPC (Coding Potential Calculator, version 0.9-r2; Kong et al., 2007). Searches of NBS and LRR motifs (taken from Table 1 in Jupe et al. (2012)) were performed with MAST (MEME suite 4.10.1; Bailey and Gribskov (1998)). Structure of the NBS-LRR genes was obtained after alignment of the predicted proteins to NCBI 'nr' protein database. Multiple alignments of the proteins were performed with Clustal Omega (Sievers et al., 2011).

4.2.5. Finding and assembling heterozygous mapping regions

Although the lines used for this study should all be homozygous in the QTL region, a number of sites with heterozygous variants were found after aligning exome sequences to the reference. To systematically locate these regions, an analysis of the number of different k-mers mapping to the pseudoscaffolds was carried out. Read mappings from exome sequencing were surveyed to quantify each different 50-mer aligning to each position in the reference, considering only those sampled at least 4 times. Sets of reads from the segments with more than one kind of k-mer (therefore annotated as "heterozygous mappings", HMs) and mapping to disease resistance proteins were assembled with Trinity (Grabherr et al., 2011). The sequence contigs obtained for the different BC_1F_4 lines were compared and clustered. A representative sequence was chosen from each cluster and a genotype was assigned to it based on its presence-absence pattern across BC_1F_4 lines. Several overlapping contigs, which showed the same PAV in the lines, were assembled together.

4.2.6. Validation of the genotypes found with the exome capture by PCR

The genotypes of the parents and the recombinant lines were checked for those Morex WGS contigs which had polymorphisms associated with the resistance/susceptibility phenotype. These included contigs 1622651, 167712, 211721, and 50573. Amplicons were used to validate the genotypes of the lines corresponding to sequences present in BACs M01 and D03 from FPC 591. In addition, the PAV polymorphism of the lines was assessed for the 2 largest new assembled sequence contigs (ELOC1 and ELOC2), including cultivar Morex. Primers were designed with Primer 3 (Untergasser et al., 2012) and validated by running isPCR (https://genome.ucsc.edu/cgi-bin/hgPcr) against the WGS assemblies from IBGSC data. In addition, primers were designed to amplify the unknown fragments between Morex WGS contig 50573 and both ELOC1 and Morex WGS contig 44875, by Long Range PCR.

4.2.7. Characterization of the new assembled sequence contigs

Putative ORFs encompassing the assembled ELOCs were searched with ORF Finder. In addition, CPC was conducted to evaluate their protein-coding potential. The resulting DNA sequences were searched for in the Uniprot Plants and NCBI 'nr' databases. Both sequences were also compared against the IBGSC databases and Haruna Nijo flcDNAs (Matsumoto et al., 2011) with Barleymap (Cantalapiedra et al., 2015). The predicted aminoacid sequences coded by those ORFs were compared to each other with blastp.

4.2.8. Real-Time PCR of the assembled sequence contigs

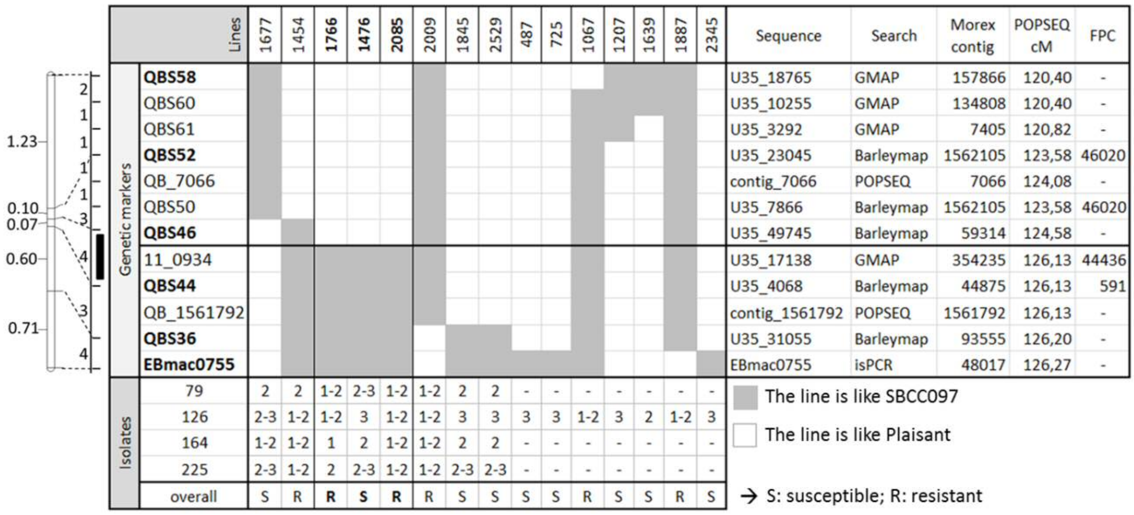
For Real-Time quantitative PCR (RTq-PCR) experiments, 7-day-old plants were inoculated with powdery mildew isolate R79 in the greenhouse. Two samples per line were collected at 12, 24, 48 and 72 h after infection. Each sample consisted of the pooled leaf tissue of two plants.

Total RNA was extracted from frozen samples using the Aurum TM Total RNA Mini Kit (BioRad, USA) following the manufacturer's instructions. First-strand cDNA was synthesized from 100 ng of total RNA by using the iScript cDNA Synthesis Kit (BioRad). Primers were designed with Primer Express 3.0 (Applied Biosystems, USA). RTq-PCR was performed in 50 µl of reaction mixture made up of 2.5 µl of cDNA, 1 × iQ SYBR Green Supermix (BioRad) and 0.3 µM of each specific primer. The Actin gene was used as a constitutively expressed reference gene to normalize expression as in Trevaskis et al. (2006).

4.3. Results

4.3.1. Fine mapping of the resistance locus

To fine map the resistance QTL identified on 7HL in the SBCC097 x Plaisant population (Silvar et al., 2010), a RIL containing only this QTL (RIL151, Silvar et al., 2012) was backcrossed to Plaisant. A large BC_1F_2 population was obtained, and tested for recombination between markers QBS58 and EBmac0755, flanking the 7HL QTL. Out of 2,899 BC_1F_2 plants tested, 152 recombinants were identified and grown until maturity. Twenty five BC_1F_3 families were then screened to identify homozygous recombinants, which were further tested with the markers obtained in previous studies, exploiting synteny and physical information (Silvar et al., 2012; Silvar et al., 2013b). This procedure identified 15 BC_1F_4 plants covering the whole region (Figure 4.1). A genetic map of the region was constructed with the information of the entire BC_1F_2 generation and allowed narrowing the position of the QTL down to a 0.07 cM interval between markers QBS46 and QBS44. Furthermore, three BC_1F_4 lines, one susceptible (1476) and two resistant (1766 and 2085), showed the same genotype flanking the QTL but different phenotype (Figure 4.1). Therefore, the gene or genes responsible for the resistance lay within the interval between QBS46 and QBS44.



Lines	1677	1454	**1766**	**1476**	**2085**	2009	1845	2529	487	725	1067	1207	1639	1887	2345	Sequence	Search	Morex contig	POPSEQ cM	FPC
Genetic markers																				
QBS58	▨					▨						▨	▨	▨		U35_18765	GMAP	157866	120,40	-
QBS60	▨					▨					▨	▨	▨	▨		U35_10255	GMAP	134808	120,40	-
QBS61	▨					▨					▨	▨		▨		U35_3292	GMAP	7405	120,82	-
QBS52	▨					▨					▨			▨		U35_23045	Barleymap	1562105	123,58	46020
QB_7066	▨					▨					▨			▨		contig_7066	POPSEQ	7066	124,08	-
QBS50	▨					▨					▨			▨		U35_7866	Barleymap	1562105	123,58	46020
QBS46		▨				▨					▨			▨		U35_49745	Barleymap	59314	124,58	-
11_0934		▨	▨	▨	▨	▨					▨			▨		U35_17138	GMAP	354235	126,13	44436
QBS44		▨	▨	▨	▨	▨					▨			▨		U35_4068	Barleymap	44875	126,13	591
QB_1561792		▨	▨	▨	▨	▨					▨			▨		contig_1561792	POPSEQ	1561792	126,13	-
QBS36		▨	▨	▨	▨		▨	▨			▨			▨		U35_31055	Barleymap	93555	126,20	-
EBmac0755		▨	▨	▨	▨		▨	▨	▨	▨	▨				▨	EBmac0755	isPCR	48017	126,27	-
Isolates																				
79	2	2	1-2	2-3	1-2	1-2	2	2	-	-	-	-	-	-	-	▨ The line is like SBCC097				
126	2-3	1-2	1-2	3	1-2	1-2	3	3	3	3	1-2	3	2	1-2	3					
164	1-2	1-2	1	2	1-2	1-2	2	2	-	-	-	-	-	-	-	□ The line is like Plaisant				
225	2-3	1-2	2	2-3	1-2	1-2	2-3	2-3	-	-	-	-	-	-	-					
overall	S	R	**R**	**S**	**R**	R	S	S	S	S	R	S	S	R	S	→ S: susceptible; R: resistant				

Figure 4.1. Fine mapping of the 7HL QTL. Left: genetic map of BC_1F_2 mapping population (distances in cM) showing a schematic distribution of the recombinants found in the BC_1F_3 by marker interval. The black vertical bar indicates the position of the QTL. Center: graphical genotypes of the 15 BC_1F_4 lines. Markers assayed in the BC_1F_2 are highlighted in bold type. The lines sequenced in this study (1766, 1476, 2085) are separated from the others by thick vertical lines. The thick horizontal line between QBS46 and 11_0934 markers the most likely position of the resistance gene. The bottom table summarizes the evaluation of the lines for resistance to four different powdery mildew isolates. Right: Table showing the sequences used to locate the genetic markers in the barley genome, and the sources (POPSEQ) or search methods used, Barleymap or GMAP (Wu and Watanabe, 2005). The target WGS contigs are shown ("Morex contig" column) along with their position in chromosome 7H ("POPSEQ cM" column), as well as the physical contigs ("FPC" column) associated to them.

4.3.2. *Analysis of exome sequencing polymorphisms*

Exome sequencing of the parents and the three BC_1F_4 lines was performed in order to identify the differences between the resistant and the susceptible plants. Analysis of the read data from exome sequencing involves a mapping step using a reference, the Morex WGS assembly (Mayer et al., 2012) in this case. However, the region associated to the resistance was majorly of interest here. Therefore, the genetic markers from the previous section were located in the POPSEQ map (Mascher et al., 2013a) and the identified positions (Figure 4.1) were used to anchor available genomic resources to the region. This yielded a set of 973 Morex WGS contigs associated to 17 FPCs, which are contigs with assigned physical positions. Comparing the variants between the parents, 1,037 polymorphisms were identified, corresponding to 120 Morex WGS contigs (out of the 973 just described). The genotypes of the BC_1F_4 lines were checked, looking for variants consistent with the phenotypic profile of the lines (1476 like the susceptible parent, Plaisant; the other two like the resistant parent, SBCC097), as those would be the most informative towards finding candidate genes. Only one of the Morex sequences, contig 50573, presented haplotypes fully in agreement with the phenotypic profile of the lines. This contig has a single annotated gene, a "Pentatricopeptide repeat-containing protein" (MLOC_65722 in IBGSC data). A CAPS marker designed for this gene was assayed on all 15 BC_1F_4 lines, and its position within the QTL region was confirmed.

4.3.3. Physical localization of the resistance locus

From the previous analysis, only Morex contig 50573 was unambiguously located within the QTL interval. However, although its genetic POPSEQ map position was known, it could not be found in the IBGSC physical map, hindering its direct physical localization. Nonetheless, most of the variants in the remaining Morex WGS contigs were clearly located on either side of the candidate region (i.e., the three lines had the same genotype). Looking at the genotypes of the lines from exome data, the position and order of Morex WGS contigs was not always in agreement with the POPSEQ map. If only Morex WGS contigs with known physical position were considered, the genotypes of the recombinant lines indicated the likely physical location of the recombination breakpoints within FPC 591, more specifically, between contigs 167712 and 44875 (Figure 4.2A). The position of yet another Morex WGS contig, 211721, was ambiguous. The genotypes of the lines for these contigs were confirmed by PCR assays.

To further delimit the physical position of the resistance locus, the BACs associated to FPC 591 in the IBGSC physical map were retrieved (Figure 4.2B). Among BACs with available sequence data, HVVMRXALLmA0204M01 (M01 hereafter) spans a central segment of FPC 591. Among the Morex WGS contigs aligning to M01, 167712 and 211721 were identified ~2.5 kb apart. Moreover, Morex contig 44875 was associated to BAC HVVMRXALLEA0187D03 (D03 from now on), both from IBGSC anchoring data and by our homology searches (identity 99.75 %, full target coverage, bitscore 1448; to D03 BES MRX2BAD187D03T71). D03 covers the right half of FPC 591, but it has not been fully sequenced yet. No other BACs providing new data within the QTL interval were identified. Candidate genes should thus be placed within the minimum tiling path (MTP) defined by BACs M01 and D03.

During the progress of this work, a new assembly of BACs (UCR BACs) was published. In this assembly, two extra BACs were associated to FPC 591 (Figure 4.2B): 0139I11 and 0758B20 (I11 and B20 from now on). BAC I11 was compared to M01. Most of the I11 sequences are already present in M01, but with a different arrangement. In contrast, the comparison of B20 and M01 pseudoscaffolds showed that they are mostly different, with only a few related regions. Among the Morex WGS contigs which aligned to B20, contigs 50573 and 44875 were found, separated by 4,234 bases. Note that Morex WGS contig 50573 is the only one with a haplotype in agreement with the phenotypes of the lines, hence supporting the position of the resistance locus within FPC 591.

4.3.4. Searching for candidate genes in the reference cultivar Morex

Candidate genes were searched for in the annotated Morex genome. Alignments of Morex WGS contigs, anchored to BAC M01, against IBGSC and Uniref90 sequences revealed eight gene annotations: five "Disease resistance protein RPM1", two transposon-related and one "Putative disease resistance protein RGA4". In-house annotation of the ORFs identified in the M01 pseudoscaffold (see Materials and methods) confirmed the presence of the RPM1- and transposon-related sequences, including loci not associated to Morex WGS contigs and, therefore, lacking exome capture probes. When the whole pseudoscaffold was self-aligned, the ORFs annotated as RPM1 proteins appeared to be related to each other. Since RPM1 belongs to the NBS-LRR family of resistance-genes, motifs which are known to be conserved in domains of NBS-LRR genes (Jupe et al., 2012) were searched for in the region using the software MAST. Most RPM1-related loci were also confirmed by the MAST scan (Figure 4.2C). Overall, nine segments were identified with highly significant motifs from the N-terminal, NBS and linker domains; three of them with LRR motifs. The same analysis was applied to BAC I11, which showed almost the same features as M01, as expected.

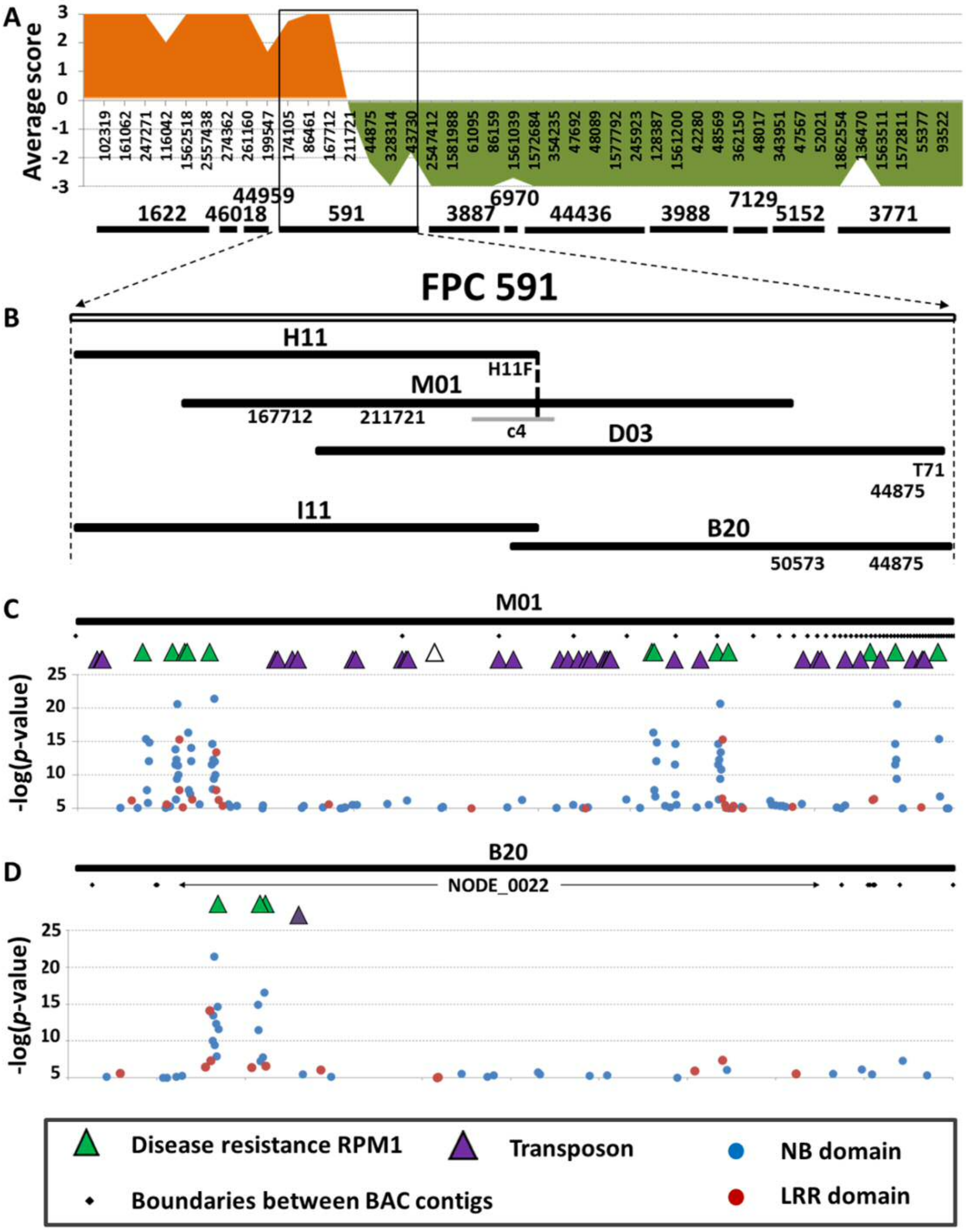


Figure 4.2. Analysis of BACs in MTP of FPC 591. A: average scores of the Morex WGS contigs considering the genotypes of the BC_1F_4 lines in relation to the parents. Orange: positive score, more lines are like Plaisant; green: negative score, more lines are like SBCC097. Contigs are sorted by increasing FPC cM position, and by POPSEQ position to resolve coincidences. FPCs are shown as black horizontal bars. B: IBGSC (H11, M01 and D03) and UCR (I11 and B20) BACs in FPC 591. Morex WGS contigs 167712 and 211721, and BES H11F and BAC contig c4, are anchored to M01. Morex WGS contigs 44875 and 50573 are anchored to B20. C: analysis of the pseudoscaffold of BAC M01. Triangles of different colors are ORFs of genes (see legend; white triangle: RGA4). The scatterplot shows the -log10(P-value) of the NBS and LRR motifs identified throughout the pseudoscaffold (blue dots: NBS domains; red dots: LRR domains). D: analysis of the pseudoscaffold of BAC B20. NODE_0022 is the longest contig in the BAC.

On the other hand, IBGSC annotation of the Morex WGS contigs associated to UCR BAC B20 showed up 2 genes: a "Pentatricopeptide repeat-containing protein" in contig 50573, mentioned earlier, and a "WD-repeat protein 57 IPR015943" in contig 44875. Both results were confirmed with alignments to Uniref90. In addition, another 3 Uniref90 hits to the left of contig 50573 were obtained; all labeled as "Disease resistance protein RPM1", both using raw Morex WGS contigs and in silico identified ORFs as queries. Again MAST scans of NBS-LRR motifs confirmed these results (Figure 4.2D) and, as with M01, several hits related to transposons were obtained close to them.

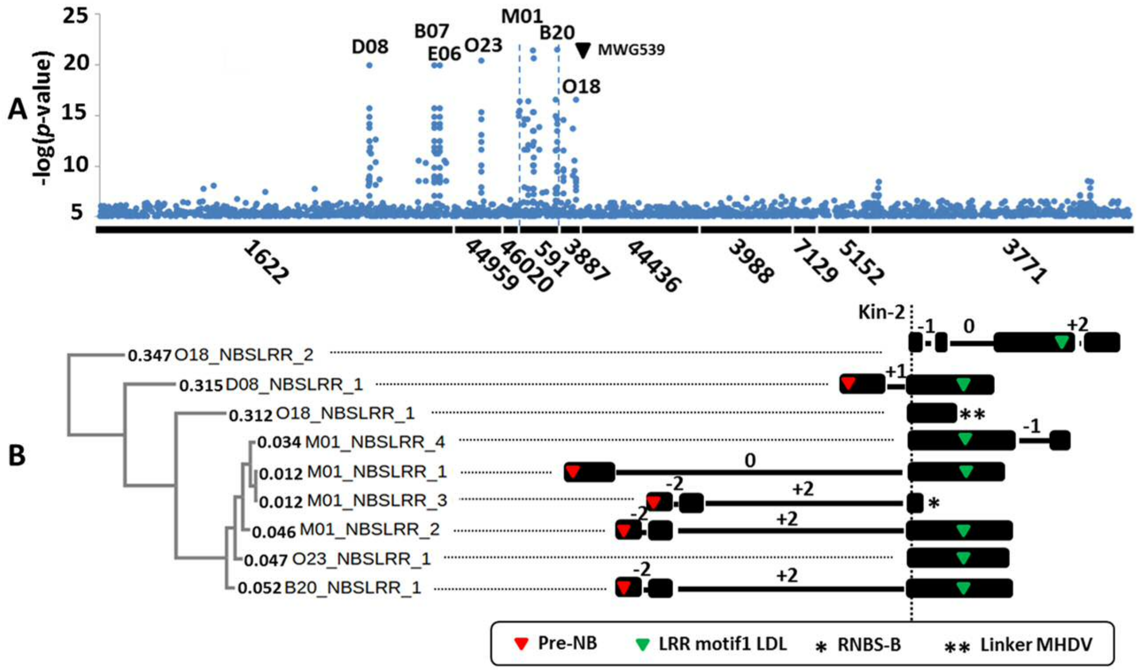


Figure 4.3. NBS and LRR motifs found in the region of FPC 591. A: Significance of the motifs found in the whole region (of about 5.6 Mb). Vertical dashed blue lines demarcate the motifs found within FPC 591. A black triangle indicates the physical position of RFLP marker MWG539, close to the Mlf locus (Schönfeld et al., 1996). B: UPGMA clustering of the predicted proteins containing NBS-LRR motifs. Protein names are prefixed with their respective BAC codes. Distances obtained from the multiple alignment are shown to the left of each protein name. Inferred gene structures are shown to the right (black boxes: exons; black horizontal lines: introns). The number on each intron shows the frame change from one exon to the next. Motifs shown on gene structures are named after Table 1 in Jupe et al. (2012). A vertical dashed line shows the position of the Kinase-2 motif, to which the structures of genes have been aligned. Asterisks indicate the presence of a specific motif at the end of the available sequence of the corresponding gene.

Analysis of NBS-LRR motifs in a wide physical region around FPC 591 (55 UCR BACs, spanning 5.6 Mb) revealed that the cluster is mostly circumscribed to the resistance locus (Figure 4.3A). A few other NBS-LRR genes were detected outside the locus, but these were unrelated both in terms of sequence and gene structure (Figure 4.3B).

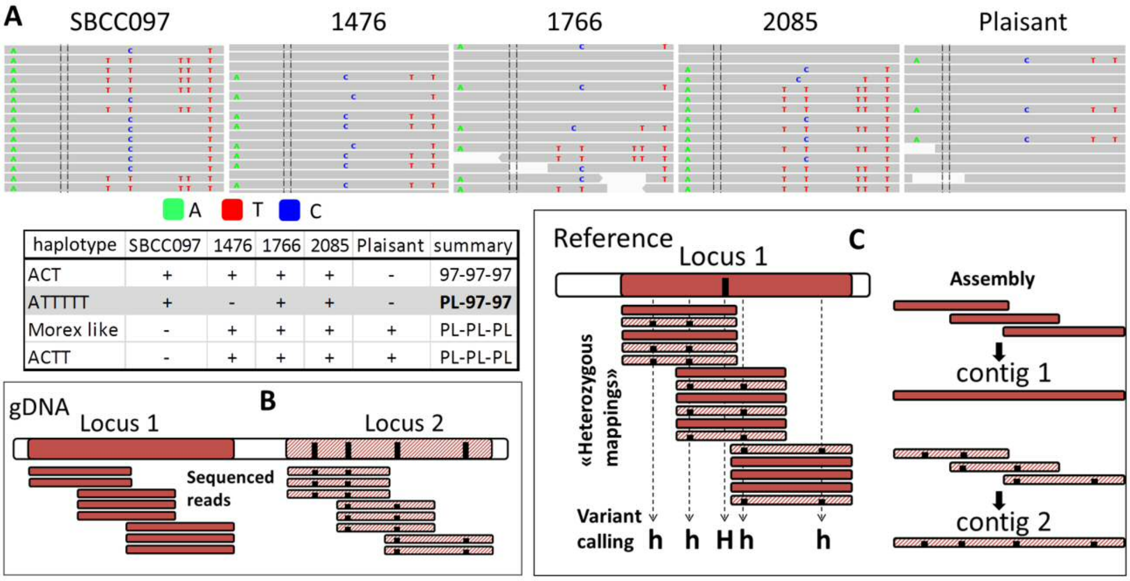
Therefore, besides a Pentatricopeptide repeat-containing protein and a WD-repeat protein, the MTP spanning the resistance locus in Morex is rich in transposons and contains a cluster of closely related NBS-LRR genes.

4.3.5. Analysis of "heterozygous mappings" in Morex

As shown above, only Morex WGS contig 50573 had a haplotype consistent with being within the resistance locus. However, there were other Morex WGS contigs for which some variants were consistent but others were not. Many of the variants in those contigs were apparently heterozygous. This was highly unlikely, as the parents were homozygous, the BC_1F_4 plants were selected to be homozygous for the interval of interest and the possibility of having double recombinants within such a small region was negligible. In fact, visual inspection of the mappings producing those variants revealed different populations of reads stacking to the same locus (Figure 4.4A), in contrast with the mappings from contig 50573, which produced unambiguous homozygous single-nucleotide polymorphisms (SNPs). The apparent heterozygous genotypes were confirmed through PCR amplification of CAPS markers. Note that these variants were abundant and linked in recurrent groups, as independent haplotypes, instead of being spread out randomly among the reads. Thus, it is unlikely that they are the result of sequencing errors. Instead, these mappings could have been produced by piling up closely related sequences (repeats, paralogous genes) which were captured by the exome baits (Jupe et al., 2013; Mascher et al., 2013b), but for which the original locus would not be present in the reference. Since they affect variant calling, producing apparent heterozygous variants, from now on this kind of mappings will be referred to as "heterozygous mappings" (HMs) (Figure 4.4, B and C). Almost all Morex WGS contigs with HMs, whose variants had genotypes in agreement with the phenotypic profile of the lines, could be annotated as homologs to "Disease resistance protein RPM1" or "Disease resistance protein RPP13", after alignment to the Uniprot Plants database (http://www.uniprot.org/blast/). Some of those contigs are the ones located within or close to FPC 591. Taken together, these results suggest that there are sequences related to disease resistance proteins, which are not present in the Morex reference but are likely within the resistance locus in the genomes of SBCC097 or Plaisant.

In this study, the distribution and abundance of HMs in the resistance locus region was analyzed in more detail to i) assess whether the differences between the recombinant lines were likely to be related with the disease resistance, ii) verify whether the presence of HMs was a feature exclusive of the sequences related to NBS-LRR genes in the region of interest, and to iii) identify and demarcate the segments of the reference in which they occur. This last objective would allow obtaining the reads which produce the HMs and assembling them into sequence contigs (Figure 4.4C).

Therefore, we analyzed the number of different 50-mers, fragments of reads of 50 bases, mapping to each position of Morex WGS contigs anchored to BACs M01 and B20 in the three BC_1F_4 lines. Note that the reads from our sequencing data are 101-mers, but to be able to capture diversity in a given position a smaller k-mer size had to be chosen, since mapping duplicates were removed in a previous step. Wherever several 50-mers mapped to the same position, HMs would be likely found; each 50-mer being possibly derived from a different genomic locus. Notably, we found different 50-mers mapping to most of the loci related with NBS-LRR genes, although not all the mapped loci belonged to that class. Out of the covered positions, 74.4 and 89.5% had a single 50-mer in M01 and in B20, respectively. Interestingly, differences among the lines seemed to be associated mostly to disease resistance loci. First, the resistant lines had a larger percentage of positions with several 50-mers (i.e. with HMs) in M01, although not in B20. Furthermore, taking into account only the reference positions within annotated NBS-LRR genes, the difference between the resistant lines and the susceptible one increased in both BACs. Therefore, the differences between the two BACs can to a large extent be explained by the greater abundance of NBS-LRR related sequences in M01 and B20 (49.6 and 11.7% of the mapped bases, respectively).



haplotype	SBCC097	1476	1766	2085	Plaisant	summary
ACT	+	+	+	+	-	97-97-97
ATTTTT	+	-	+	+	-	**PL-97-97**
Morex like	-	+	+	+	+	PL-PL-PL
ACTT	-	+	+	+	+	PL-PL-PL

Figure 4.4. Heterozygous mappings (HMs). A: images captured from Integrative Genomics Viewer (IGV), showing reads (gray horizontal bars) mapping to a specific interval of Morex WGS contig 1622651. Colored characters show the variants detected for each genotype in relation to the Morex reference. The table summarizes the haplotypes identified, along with their presence-absence type ("+" or "-") in the lines. Genotypes of the three BC_1F_4 recombinant lines relative to the parents are shown in the "summary" column. One group of variants (ATTTTT, light gray background) is consistent with the phenotypic resistance profile of the lines ("PL-97-97" or susceptible-resistant-resistant). B: schematic representation of the reads that would be obtained after sequencing two closely related loci. The two loci are represented by horizontal bars (red background; plain for Locus 1, striped for Locus 2), with a few hypothetical differences (black vertical bars). C: reads from B are mapped back to the reference. In the example shown, the reference lacks one locus (Locus 2), and all sequenced reads hit the existing one (Locus 1), producing apparent HMs. As a result, variant calling yields heterozygous calls ("h") and homozygous calls ("H") intermixed. A new assembly could solve this region, yielding independent contigs resembling the original loci, due to the presence of the four genotypic variants between the two loci.

4.3.6. De-novo *assembly of exome sequence reads spanning the resistance locus*

Analysis of HMs pointed towards the presence of NBS-LRR related sequences within the resistance locus, absent from the Morex reference. In light of this, a template-guided assembly of reads producing HMs was performed. Firstly, Morex WGS contig fragments located within FPC 591, related to disease resistance genes and producing HMs were chosen

(11 loci). Secondly, six further Morex WGS contig fragments with HMs and variants in agreement with the phenotypes of the lines were selected. Finally, Morex WGS contig 50573, harboring the "Pentatricopeptide repeat-containing protein", was included as a control. Read subsets mapping to the 18 selected segments were retrieved, and an independent assembly for each genotype was performed (for both parents and the three BC_1F_4 lines). These operations yielded 203 sequence contigs, with an average of almost 41 contigs per line. These new contigs were clustered, and a representative sequence per cluster was selected, yielding 31 representative sequences. Based on the presence or absence of those sequences, PAV genotypes for each cluster were assigned to each line. Representative sequences showing the same PAV genotypic profiles were then compared to each other, leading to the assembly of 5 of them into a contig of 981 nucleotides (ELOC1), and another 4 into a contig of 787 bases (ELOC2). Therefore, the final set comprised 24 sequence contigs, for which the lines had different PAV genotypes. ELOC1 and ELOC2 were the largest assembled contigs. ELOC1 was absent in Plaisant and 1476, while ELOC2 was only present in SBCC097 and 1766. The absence of ELOC2 from the resistant line 2085 was in agreement with the fewer number of 50-mers identified in this line in comparison with 1766, and it suggested that 2085 and 1476 contained the smallest interval flanking the resistance locus.

4.3.7. *Validation and characterization of the new assembled sequence contigs*

We designed primers to perform PCR amplification of ELOC1 and ELOC2. The PCRs confirmed the PAV genotypes of the 15 BC_1F_4 lines and the parents (Figure 4.5). In addition, the absence of both sequences in cultivar Morex was verified (data not shown). To check whether this result was a consequence of polymorphism on the primers, the reads from the exome capture of SBCC097, Plaisant, Morex (from the same exome capture experiment), and lines 1476, 1766 and 2085 were re-aligned to the new contigs. This confirmed the PAV variation found on them. Moreover, the products of amplification of the lines SBCC097 and 1766 were Sanger-sequenced and further validated.

In silico ORF calling was performed with both ELOCs, obtaining two partial ORFs of 322 and 252 amino acids for ELOC1 and ELOC2, respectively. In addition, their protein-coding potential was checked, with log-odds scores of 82.73 and 57.46 for ELOC1 and ELOC2, respectively. The percentage of identity between the two amino acid sequences was 92%, and their alignment covered most of ELOC2. Looking for similar proteins in Uniprot Plants and NCBI 'nr' databases, results were found within the range of identities obtained when comparing the NBS-LRR proteins in the QTL region in Morex, and comparable with paralogous genes found in other NBS-LRR clusters (Wei et al., 1999; Kuang et al., 2004; Bulgarelli et al., 2010). Moreover, the ELOCs were aligned against the Morex NBS-LRR predicted proteins of the region. The best hits had almost full coverage and 87.9 and 91.6% identity, for ELOC1 and ELOC2, respectively. Alignment of DNA sequences of the ELOCs to the IBGSC databases produced similar results. Also, these alignments revealed that the contigs contained only the LRR domain, lacking the NBS one.

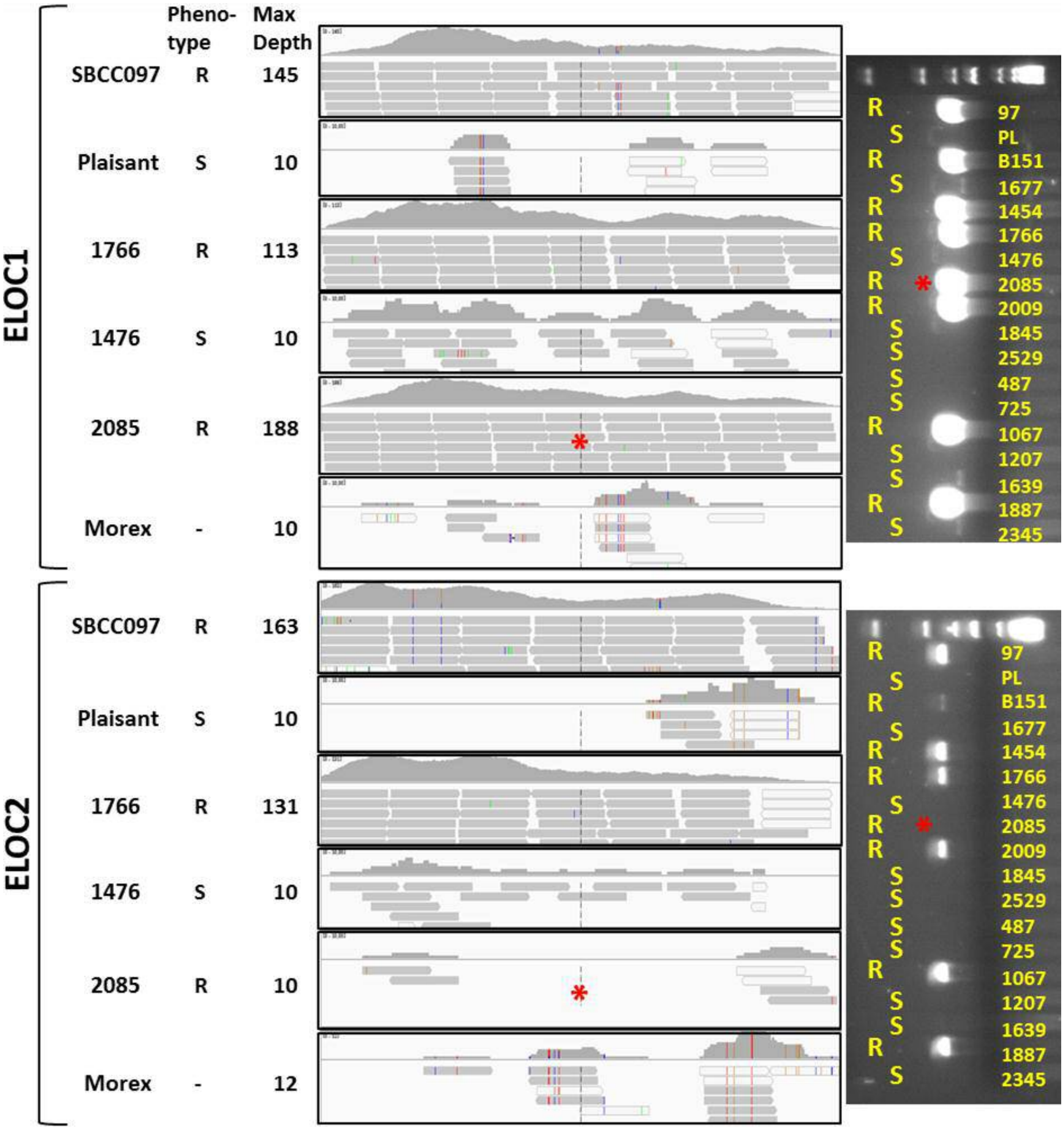


Figure 4.5. Presence-absence genotypes for ELOC1 and ELOC2. Left: phenotypes of the two parents, the three sequenced lines and Morex, along with the maximum depth of coverage (“Max Depth”) obtained after mapping the exome sequencing reads to the new assembled contigs, ELOC1 (top) and ELOC2 (bottom). Center: images captured from IGV, showing the profile of depth of coverage throughout the contigs (top) and individual reads mapped (bottom). Resistant lines have large depths of coverage and similar profiles, covering the whole contigs, with the exception of 2085 in ELOC2 (red asterisk). Susceptible lines have low depth of coverage and irregular, incomplete mapping profiles. Right: gel electrophoresis of PCR amplicons of ELOC1 and ELOC2 for the two parents, the resistant line RIL151 and the fifteen BC_1F_4 lines, along with their phenotypes. Resistant lines have presence genotypes whereas susceptible lines have absence genotypes, with the exception of 2085 in ELOC2 (red asterisk). R: resistant. S: susceptible.

RTq-PCR was used to check the expression of both new contigs. No specific amplicon was obtained for ELOC2 and, therefore, it could either be a pseudogene (Kuang et al., 2004) or be expressed in another tissue or developmental stage (Tan et al., 2007). Nonetheless, amplification was positive for ELOC1, confirming its transcription in leaves of SBCC097 and the two resistant BC_1F_4 lines, although this is not a definitive evidence of the gene being functional (Wei et al., 2002; Monosi et al., 2004). The RTq-PCR was performed for SBCC097 at different time points, spanning 72 h after infection. Apparently, there was no change in ELOC1 expression in response to the infection, although this is not irreconcilable with being involved in the resistance or even being regulated at another stage than transcription (Tan et al., 2007).

4.4. Discussion

Barley research has been accelerated by the availability of abundant genomic resources published over the last years. In some cases, this has led to faster gene cloning, like cloning of *HvCEN* by Comadran et al. (2012). However, other barley genes have not been cloned yet despite their known phenotypic effect and genetic localization, partly due to the lack of such resources until recently. The continuous improvement of barley physical resources (Mayer et al., 2012; Mascher et al., 2013a; Ariyadasa et al., 2014; Muñoz-Amatriaín et al., 2015) allows the adoption of more efficient methodologies for genetic studies involving high-throughput genotyping, marker development, gene discovery, expression analysis, synteny and genome comparative studies. The exome capture probe set developed by Mascher et al. (2013b) for barley is already being used for gene cloning purposes. Mascher et al. (2014) used it to identify *HvMND*, a gene that regulates the rate of leaf initiation, and Pankin et al. (2014) to identify a candidate for *HvPHYC*. In both cases, exome capture was performed on bulked plants with extreme phenotypes from BC_1F_2 populations between mutants and the wild type.

In this work, the same exome capture probe set was used to sequence three recombinant lines for a powdery mildew resistance QTL. The resistance allele was contributed by a Spanish landrace, showing a wide resistance profile (resistance to 23 out of 27 isolates tested) after a thorough disease survey (Silvar et al., 2011) with the accessions from the SBCC (Igartua et al., 1998). Such line had two QTL conferring race-specific resistances on chromosome 7H (Silvar et al., 2010). The mechanism of resistance of this line was classified as consistent with "intermediate-acting" genes, governing resistance mainly at the post-penetration stage (Silvar et al., 2013a). Genomic approaches allowed the development of new markers to narrow down the QTL intervals (Silvar et al., 2012; Silvar et al., 2013b), but were insufficient to definitely locate a manageable physical location or a set of candidate genes for the stronger QTL on 7HL, which is the subject of this work.

From that point, a large F_2 population was created and screened with markers from those previous studies, aiming to identify recombinant lines to further narrow down the QTL interval. The final interval, just 0.07 cM wide, was apparently small enough to land on potential candidates, as this size is comparable with other intervals used in successful gene cloning attempts in barley (reviewed in Krattinger et al., 2009). Again, the analysis of available genomic resources was insufficient to locate candidate genes or to delimit the resistance to a single physical contig. Although the markers were found in the Morex WGS assembly and a POPSEQ map position could be assigned to them, many other Morex WGS contigs with positions within the QTL interval were identified, leading to a large list of annotated genes. Moreover, since the current barley maps are incomplete, additional contigs could have gone unnoticed. Finally, since not all the contigs to which the markers hit were anchored to physical contigs, the physical localization of the QTL remained unknown. An additional challenge was the search of genetic markers from previous studies in the reference. Several of the markers were only found through the analysis of chimeras from GMAP alignments, likely due to the fragmented nature of the Morex WGS assembly.

Exome sequencing of the parents and three recombinant lines allowed the identification of abundant polymorphic variants. This is a faster and more powerful alternative to the search of markers by in-silico comparison of genomic resources from different genotypes or by extrapolation of markers from other populations, since many of these are not necessarily polymorphic between the parental lines of the population under study. However, in this work, most of the homozygous SNPs were located outside the QTL. Only a single Pentatricopeptide-repeat containing protein was easily identified within the QTL region, and its corresponding Morex WGS contig lacked physical anchoring. Despite that, the analysis of the profile of variants along the physical contigs in the region was enough to point towards a single FPC which could contain entirely the QTL. This highlights the usefulness of exome sequencing for fine mapping purposes. However, this work demonstrates the technical challenges encountered. Some positions of Morex WGS contigs were not in agreement with the genotypes of our lines. Differences in collinearity between several genetic maps and the POPSEQ reference have been already described (Cantalapiedra et al., 2015; Silvar et al., 2015). These incongruences are important for fine mapping purposes. A single physical contig holding the resistance locus was identified only after removing the Morex WGS contigs not associated to physical positions and using a score to average together the genotypes of the variants within each Morex WGS contig.

Despite the scarcity of homozygous SNPs found within the QTL region, we observed abundant heterozygous SNPs which were polymorphic between the parents as PAV. Although the work with SNPs and small indels is rather straightforward, working with other kinds of variation such as copy-number variation (CNV) or PAV requires using alternative approaches, for example analyzing mapping depth (Mascher et al., 2014). In this work, heterozygous mappings (HMs) are defined as those producing heterozygous variants probably due to the collapse of reads from paralogous genes absent in the reference genome. This phenomenon has been recently described among homoeologous genes in an exome sequencing experiment in wheat (King et al., 2015). In studies focused on variant discovery, HMs can confound the discrimination of true variants at a given locus. However, this study used HMs to identify the regions with polymorphic HMs, through k-mer analysis, to further assemble different paralogous genes and assess their expression. Though this approach aimed to locate regions with HMs, k-mer abundance could be directly used for genotyping purposes. As with CNV, analysis of HMs is related to the number of copies of a given sequence. However, the analysis of CNV through mapping depth should cope with the different efficiencies in the hybridization and PCR amplification steps during exome sequencing when the sequences are different. In contrast, the analysis of k-mer abundance has the drawback of being unable to differentiate the copies when they are identical to each other. In addition, analysis of HMs could provide insights into the loci and gene families for which the reference genome is incomplete or shows larger variation between different genotypes. Finally, we genotyped the HMs as PAV polymorphisms by means of template-guided assembly and clustering of the resulting sequence contigs. An alternative approach would be to directly compare the presence or absence of the individual k-mers mapping to a given position in the genotypes, although this would not provide assembled contigs. In both cases, the main difficulty resides in differentiating between orthologous and paralogous genes, allelic variants and isoforms (Kuang et al., 2004; Seeholzer et al., 2010), either when clustering the contigs from the assembly or when considering that all orthologous k-mers from the different genotypes are mapping to the same reference locus, and not to another closely related one. In any case, the methods used in this study were implemented from standard tools which were combined to accomplish our specific goals, and thus could be further developed and optimized to cope with peculiarities of HMs.

Both the analysis of the sequenced BACs and the genotyping of HMs pointed towards a cluster of related NBS-LRR genes in the resistance locus. These are good candidates for a resistance gene, although we have to be aware that the sequences captured are limited by the baits used and it cannot be ruled out that the actual resistance gene is absent from the capture reactions and/or from the reference genome. NBS-LRR genes are abundant in many plant genomes and are often organized in clusters of one or more groups of related paralogous genes (Michelmore and Meyers, 1998), which makes their assembly difficult. This problem was evident in this study as revealed by the huge difference in size, number and composition of contigs in equivalent sequenced BACs from independent assemblies (e.g. M01 from IBGSC and I11 from UCR). In addition, a common trend observed in NBS-LRR genes in grasses is the rapid expansion and loss of members from those groups (Li et al., 2010; Yang et al., 2013), leading to PAV and CNV between genotypes. Genes found in that region in Morex were poorly annotated and most of them were split into different WGS contigs. Therefore, the exact number and structure of the genes in this cluster remains unknown both in cultivar Morex and in the resistant line SBCC097. In our assembly, the NBS-LRR genes were incomplete, lacking the NBS domains. We do not know whether these genes are actually incomplete or the NBS domains do exist but were not captured. Lack of exome capture reads covering the genes completely, for instance due to the presence of large introns in them, could lead to incomplete assemblies. Nonetheless, the NBS domains are usually more conserved than the LRR ones (Meyers et al., 1999; Pan et al., 2000; Seeholzer et al., 2010), and this could hinder the independent assembly of the different paralogous genes.

This study made extensive use of state-of-the-art genomic resources available for barley. Several aspects which could be considered when working with these resources arise from our analysis. We have already mentioned some of them, like the lack of position of many Morex WGS contigs or the incomplete annotation of genes in the region. Regarding contig positions, we describe the combined use of both POPSEQ map of Morex WGS contigs and their anchoring to BACs to obtain as many sequences as possible close to our resistance locus. Additional information from the recent publication of sequenced BACs from UCR, a different assembly to that of IBGSC, allowed to complete the MTP of the region and confirmed the features identified using IBGSC data. Furthermore, it highlighted the discrepancies between assemblies, even when corresponding to the same barley genotype, at least in regions with repetitive sequences like the clustered NBS-LRR genes and transposons found in our region.

Finally, identification of the full sequence at these loci would require obtaining BAC libraries and the use of long-read sequencing technologies. Sequencing the whole region could reveal candidate genes which have gone unnoticed, and it could contribute to the understanding of structure and diversification of NBS-LRR genes. Furthermore, sequencing the region, which is rich in resistance genes in barley, could help identifying other resistances. For example, *Mlf* (Schönfeld et al., 1996), which has been associated to this region previously (Backes et al., 2003), given the close physical location of its linked RFLP probe to our QTL. Although BAC libraries are available for cultivar Morex and a few more accessions, this is still not the case for most barley genotypes. Until those resources are available, the exploitation of exome capture to assemble reads from HMs was used in this study to identify candidates not present in the reference or in the exome capture target space, through similarity with closely related genes.

4.5. References

Ames, N., Dreiseitl, A., Steffenson, B.J., and Muehlbauer, G.J. (2015). Mining wild barley for powdery mildew resistance. *Plant Pathol.* 64**,** 1396-1406.

Ariyadasa, R., Mascher, M., Nussbaumer, T., Schulte, D., Frenkel, Z., Poursarebani, N., et al. (2014). A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. *Plant Physiol.* 164**,** 412-423. doi: 10.1104/pp.113.228213

Backes, G., Madsen, L.H., Jaiser, H., Stougaard, J., Herz, M., Mohler, V., et al. (2003). Localisation of genes for resistance against Blumeria graminis f.sp. hordei and Puccinia graminis in a cross between a barley cultivar and a wild barley (Hordeum vulgare ssp. spontaneum) line. *Theor. Appl. Genet.* 106**,** 353-362.

Bailey, T.L., and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14**,** 48-54.

Brueggeman, R., Druka, A., Nirmala, J., Cavileer, T., Drader, T., Rostoks, N., et al. (2008). The stem rust resistance gene Rpg5 encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. *Proc. Natl. Acad. Sci. U.S.A.* 105**,** 14970-14975.

Bulgarelli, D., Biselli, C., Collins, N.C., Consonni, G., Stanca, A.M., Schulze-Lefert, P., et al. (2010). The CC-NB-LRR-type Rdg2a resistance gene confers immunity to the seed-borne barley leaf stripe pathogen in the absence of hypersensitive cell death. *PLoS One* 5**,** e12599.

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., et al. (1997). The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell* 88**,** 695-705.

Cantalapiedra, C.P., Boudiar, R., Casas, A.M., Igartua, E., and Contreras-Moreira, B. (2015). BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol. Breeding* 15. doi: 10.1007/s11032-015-0253-1

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., et al. (2012). Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44**,** 1388-92. doi: 10.1038/ng.2447

Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9**,** 275-296.

Friedt, W., and Ordon, F. (2007). "Molecular markers for gene pyramiding and disease resistance breeding in barley", in *Genomics assisted crop improvement,* eds. R.K. Varshney & R. Tuberosa, (Dordrecht, The Netherlands: Springer), 81-101.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29**,** 644-52. doi: 10.1038/nbt.1883

Gu, L., Si, W., Zhao, L., Yang, S., and Zhang, X. (2015). Dynamic evolution of NBS-LRR genes in bread wheat and its progenitors. *Mol. Genet. Genomics* 290**,** 727-738.

Halterman, D., Zhou, F.S., Wei, F.S., Wise, R.P., and Schulze-Lefert, P. (2001). The MLA6 coiled-coil NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. *Plant J.* 25**,** 335-348.

Igartua, E., Gracia, M.P., Lasa, J.M., Medina, B., Molina-Cano, J.L., Montoya, J.L., et al. (1998). The Spanish barley core collection. *Genet Res Crop Evol* 45**,** 475-481. doi: 10.1023/A:1008662515059

Jensen, H.P., Christensen, E., and Jorgensen, J.H. (1992). Powdery mildew resistance genes in 127 northwest European spring barley varieties. *Plant Breed.* 108**,** 210-228.

Jones, D.G.J., and Dangl, J.L. (2006). The plant immune system. *Nature* 444**,** 323-329.

Jorgensen, J.H. (1992). Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica* 63**,** 141-152.

Jørgensen, J.H. (1988). "*Erysiphe graminis,* powdery mildew of cereals and grasses", in *Genetics of plant pathogenic fungi,* ed. G.S. Sidhu, (London: Academic Press Limited), 137-157.

Jørgensen, J.H., and Wolfe, M. (1994). Genetics of powdery mildew resistance in barley. *Crit. Rev. Plant Sci.* 13**,** 97-119.

Jupe, F., Pritchard, L., Etherington, G.J., Mackenzie, K., Cock, P.J., Wright, F., et al. (2012). Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* 13**,** 75.

Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G.J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J* 76**,** 530-44. doi: 10.1111/tpj.12307

King, R., Bird, N., Ramirez-Gonzalez, R., Coqhill, J.A., Patil, A., Hassani-Pak, K., et al. (2015). Mutation scanning in wheat by exon capture and next-generation sequencing. *PLoS One* 10**,** e0137549.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35**,** W345-9. doi: 10.1093/nar/gkm391

Krattinger, S., Wicker, T., and Keller, B. (2009). "Map-based cloning of genes in Triticeae (wheat and barley)", in *Genetics and genomics of the Triticeae,* eds. C. Feuillet & G. Muehlbauer, (Dordrecht, The Netherlands: Springer), 337-357.

Kuang, H., Woo, S., Meyers, B.C., Nevo, E., and Michelmore, R.W. (2004). Multiple genetic processes result in heterogeneus rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16**,** 2870-2894.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25**,** 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9. doi: 10.1093/bioinformatics/btp352

Li, J., Ding, J., Zhang, W., Zhang, Y., Tang, P., Chen, J.Q., et al. (2010). Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* 283, 427-438.

Maekawa, T., Kufer, T.A., and Schulze-Lefert, P. (2011). NLR functions in plant and animal immune systems: so far and yet so close. *Nat. Immunology* 12, 817-826.

Marone, D., Russo, M., Laidò, G., De Leonardis, A., and Mastrangelo, A. (2013). Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *Int. J. Mol. Sci.* 14, 7302-7326.

Mascher, M., Jost, M., Kuon, J., Himmelbach, A., Abfalg, A., Beier, S., et al. (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15, R78. doi: 10.1186/gb-2014-15-6-r78

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., et al. (2013a). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* 76, 718-27. doi: 10.1111/tpj.12319

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013b). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J.* 76, 494-505. doi: 10.1111/tpj.12294.

Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., et al. (2011). Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.* 156, 20-8. doi: 10.1104/pp.110.171579

Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 23, 1249-63. doi: 10.1105/tpc.110.082537

Mayer, K.F.X., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711-6. doi: 10.1038/nature11543

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303.

McHale, L., Tan, X., Koehl, P., and Michelmore, R. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7, 212.

Meyers, B.C., Dickerman, A.W., Michelmore, R.W., Sivaramakrishnan, S., Sobral, B.W., and Young, N.D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317-332.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell* 15, 809-834.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8**,** 1113-1130.

Monosi, B., Wisser, R.J., Pennill, L., and Hulbert, S.H. (2004). Full-genome analysis of resistance gene homologues in rice. *Theor. Appl. Genet.* 109**,** 1434-1447.

Muñoz-Amatriaín, M., Lonardi, S., Luo, M., Madishetty, K., Svensson, J.T., Moscou, M.J., et al. (2015). Sequencing of 15622 gene-bearing BACs clarifies the gene-dense regions of the barley genome. *Plant J.* 84**,** 216-227. doi: 10.1111/tpj.12959

Ooijen, J.W.v. (2006). "JoinMap 4, software for the calculation of genetics linkage maps in experimental populations".

Pan, Q., Wendel, J., and Fluhr, R. (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50**,** 203-213.

Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., et al. (2014). Mapping-by-sequencing identifies HvPhytochrome C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics* 198**,** 383-396. doi: 10.1534/genetics.114.165613

Poland, J.A., Brown, P.J., Sorrells, M.E., and Jannink, J.L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7**,** e32253. doi: 10.1371/journal.pone.0032253

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16**,** 276-277.

Schönfeld, M., Ragni, A., Fischbeck, G., and Jahoor, A. (1996). RFLP mapping of three new loci for resistance genes to powdery mildew (*Erysiphe graminis* f. sp. *hordei*) in barley. *Theor. Appl. Genet.* 93**,** 48-56.

Schweizer, P. (2014). "Host and nonhost response to attack by fungal pathogens", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 197-235.

Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., et al. (2010). Diversity at the Mla powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol Plant Microbe Interact* 23**,** 497-509. doi: 10.1094/MPMI-23-4-0497

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7**,** 539. doi: 10.1038/msb.2011.75

Silvar, C., Dhif, H., Igartua, E., Kopahnke, D., Gracia, M.P., Lasa, J.M., et al. (2010). Identification of quantitative trait loci for resistance to powdery mildew in a Spanish barley landrace. *Mol. Breed.* 25**,** 581-592.

Silvar, C., Flath, K., Kopahnke, D., Gracia, M.P., Lasa, J.M., Casas, A.M., et al. (2011). Analysis of powdery mildew resistance in the Spanish barley core collection. *Plant Breed.* 130**,** 195-202.

Silvar, C., Kopahnke, D., Flath, K., Serfling, A., Perovic, D., Casas, A.M., et al. (2013a). Resistance to powdery mildew in one Spanish barley landrace hardly resembles other previously identified wild barley resistance. *Eur. J. Plant Pathol.* 136, 459-468.

Silvar, C., Martis, M.M., Nussbaumer, T., Haag, N., Rauser, R., Keilwagen, J., et al. (2015). Assessing the barley genome zipper and genomic resources for breeding purposes. *Plant Genome* 8, 1-14.

Silvar, C., Perovic, D., Nussbaumer, T., Spannagl, M., Usadel, B., Casas, A., et al. (2013b). Towards positional isolation of three quantitative trait loci conferring resistance to powdery mildew in two Spanish barley landraces. *PLoS One* 8, e67336. doi: 10.1371/journal.pone.0067336

Silvar, C., Perovic, D., Scholz, U., Casas, A.M., Igartua, E., and Ordon, F. (2012). Fine mapping and comparative genomics integration of two quantitative trait loci controlling resistance to powdery mildew in a Spanish barley landrace. *Theor. Appl. Genet.* 124, 49-62.

Spoel, S.H., and Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.* 12, 89-100.

Tan, X., Meyers, B.C., Kozik, A., West, M.A.L., Morgante, M., St. Clair, D.A., et al. (2007). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. *BMC Plant Biol.* 7, 56.

Torp, J., Jensen, H.P., and Jorgensen, H.J. (1978). Powdery mildew resistance genes in 106 northwest European spring barley varieties. *Kgl. Vet. Landbohojsk. Årsskr.* 1, 75-102.

Trevaskis, B., Hemming, M.N., Peacock, W.J., and Dennis, E.S. (2006). HvVRN2 responds to daylength, whereas HvVRN1 is regulated by vernalization and developmental status. *Plant Physiology* 140, 1397-1405. doi: 10.1104/pp.105.073486

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al. (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.

Verstegen, H., Köneke, O., Korzun, V., and Broock, R.v. (2014). "The world importance of barley and challenges to further improvements", in *Biotechnological approaches to barley improvement,* eds. J. Kumlehn & N. Stein, (Berlin Heidelberg: Springer-Verlag), 3-19.

Wei, F., Gobelman-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., et al. (1999). The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 153, 1929-1948.

Wei, F., Wing, R.A., and Wise, R.P. (2002). Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell* 14, 1903-1917.

Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J., et al. (2013). Repidly evolving *R* genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18572-18577.

Yue, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J., et al. (2013). Rapidly evolving *R* genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. U.S.A.* 110**,** 18572-18577.

Zhang, Z., Henderson, C., Perfect, E., Carver, T.L.W., Thomas, B.J., Skamnioti, P., et al. (2005). Of genes and genomes, needles and haystacks: Blumeria graminis and functionality. *Mol. Plant Pathol.* 6**,** 561-575.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Valè, G., Yahiaoui, N., et al. (2001). Cell-autonomous expression of barley Mla1 confers race-specific resistance to the powdery mildew fungus via a Rar1-independent signaling pathway. *Plant Cell* 13**,** 337-350.

Zhou, T., Wang, Y., Chen, J., Araki, H., Jing, Z., Jiang, K., et al. (2004). Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* 271**,** 402-415.

5. Large differences in gene expression between elite barley cultivar Scarlett and a Spanish landrace under drought and heat stress

5.1. *Introduction*

Barley (*Hordeum vulgare* L.) is the fourth cereal crop in relevance worldwide. Like most crops, its production is affected by environmental stresses, drought being the most important among them (Cattivelli et al., 2008). Drought is already prominent at several major agricultural areas throughout the world (Luck et al., 2015), and its effects are predicted to worsen due to growing water demand, shrinking water supply and increased seasonal variability (Barnabas et al., 2008; Luck et al., 2015). An increment of overall temperature is also expected (Barnabas et al., 2008; IPCC, 2014). Actually, many stresses often occur in combination, as is the case of drought and heat, thus being more harmful (Challinor et al., 2014; Mickelbart et al., 2015). However, modern breeding has been directed mainly towards increasing yield, without considering yield stability as a major goal (Mittler, 2006). Therefore, attention is growing towards minimizing the gap between yields under optimal and stress conditions (Cattivelli et al., 2008), to cope with current yield variability (Keating et al., 2010), and to contribute to adaptation to global change (Challinor et al., 2014).

An appropriate strategy to achieve this goal is the exploitation of genetic diversity not yet incorporated into elite cultivars (Dwivedi et al., 2016). As in other crops, current barley cultivars exhibit a narrower genetic basis than wild progenitors (*Hordeum vulgare* ssp. *spontaneum*) and landraces, which are the primary source of useful genes for breeding programs (Fischbeck, 2003; Dawson et al., 2015). Furthermore, in environments with low productivity, landraces and old cultivars often outperform modern genotypes (Ceccarelli et al., 1998; Pswarayi et al., 2008; Yahiaoui et al., 2014). In comparison with wheat, barley has been grown in a wider range of environmental conditions, and is the predominant crop in marginal areas with little precipitation. Accordingly, it is sown in large expanses of the Mediterranean-climatic regions (Ceccarelli, 1994; Ryan et al., 2009), where drought can occur at any moment during the life cycle of crops, being particularly frequent during the terminal stages (Turner, 2004), when different components of grain yield can be largely influenced (Fischer and Turner, 1978; Saini and Westgate, 1999; Araus et al., 2002). Therefore, barley landraces adapted to such conditions could bear genes useful for breeding programs aiming to obtain better yields under drought.

Technical advances in the last decade have potential to improve crop breeding processes (Rivers et al., 2015). High-throughput sequencing technologies are providing new powerful tools to study the association between plant genotypic and phenotypic variation (Varshney et al., 2014; Dawson et al., 2015). One of these, RNAseq (Mortazavi et al., 2008), is currently employed with different aims in crop genetics, like polymorphism detection and transcript profiling (Varshney et al., 2009). The latter can be used to analyze gene expression networks involved in different processes; for example, those related with resistance to abiotic stresses. However, analyses of *cis*-regulatory elements of transcription factors (TFs) and of promoters of genes involved in a given response have been rare in barley, likely due to the absence of adequate genomics resources.

In this work, two contrasting barley genotypes were subjected to prolonged water deficit, either alone or combined with heat. Spanish barley landrace SBCC073 was the best yielding genotype, among 159 landraces and 25 old and modern cultivars, in field trials in Spain in which average yield was below 3 t ha^{-1} (Yahiaoui et al., 2014). Here, it was compared to a modern cultivar, Scarlett, sensitive to water stress (Sayed et al., 2012). *De novo* assemblies of transcriptomes of both genotypes were obtained and gene expression changes evaluated both in developing inflorescences and leaves. Metabolic pathways, biological processes, molecular functions, co-expression clusters and *cis*-regulatory elements of drought-modulated genes are reported.

5.2. *Materials and methods*

5.2.1. *Plant material and drought experiments*

Seeds of Spanish barley landrace SBCC073 (http://www.eead.csic.es/EEAD/barley/core.php?var=73) and of cultivar Scarlett were sown. Seedlings were allowed to grow for one week and then were vernalized for 24 days, in order to synchronize flowering. At the end of the vernalization period, plants at the 3-leaf stage were transferred to 28.0 x 20.8 cm (height x diameter) black plastic pots (one seedling per pot) with standard substrate made of peat, fine sand and perlite Europerl B-10 (Europerlita Española SA, Barcelona, Spain), from a mix with 46 kg, 150 kg and 50 L, respectively. Two series of pots were placed in a greenhouse (natural photoperiod, controlled maximum temperature 28ºC, average daily temperature 25±2°C during the day and 21±3°C at night) and in a growth chamber (16h light / 8h dark, 21 ºC daytime / 18 ºC night temperature). Additional pots filled only with substrate were used to estimate dry weight and field capacity (FC). Soluble fertilizer was provided with irrigation. Plants were treated with fungicide (Triadimenol 25%) to prevent powdery mildew build-up.

Drought treatments started 30 days after transplant at the end of the vernalization period. Water application was not interrupted abruptly. Instead, it was gradually reduced to resemble a slow drying soil, based on weight of each pot relative to the estimated FC. Pots were weighted, watered, rotated and their positions swapped every two days. Once the target fraction of FC was reached, the pots were watered to keep such weight constant. Treatment levels in the growth chamber were 70% and 20% FC, whereas an intermediate level of 50% FC was applied in the greenhouse. At the sampling date, all plants in the water-stress treatments had been at the target fraction of FC for at least 14 days. Temperature and relative humidity in the greenhouse were automatically recorded.

5.2.2. *Measurement of phenotypic traits*

Several traits were recorded 60 days after transplant. Leaf water potential (LWP) in leaves was measured at noon using a Scholander chamber (SF-PRES-70, Solfranc Tecnologías SL, Vila-Seca, Spain). Stomatal conductance (SCo) was measured, starting at 9 am, using a leaf porometer (Decagon Devices, Pullman, WA, USA). Relative water content (RWC) was also estimated, as described in Talame et al. (2007). For each plant, three independent measurements were taken for LWP, SCo and RWC. In addition, tiller number (TN) and number of tillers reaching at least Zadoks stage 49 (Zadoks et al., 1974), i.e., visibly emerging spikes (VSN) were counted. All measures were taken at two biological replicates.

5.2.3. *RNA extraction and transcriptome sequencing*

Two tissues, young inflorescences and leaves (including last expanded leaves and flag leaves), were sampled at 60 days after transplant. Fresh material was harvested and frozen in liquid N_2 before RNA extraction with the NucleoSpin® RNA Plant kit (Macherey-Nagel, Düren, Germany). RNA quality was assessed with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and with Bioanalyzer 2100 hardware (Agilent, Santa Clara, CA, USA; average RIN: 6.7 for leaves, 8.1 for flowers). Barcoded cDNA libraries were prepared at CNAG (Barcelona, Spain) following Illumina TruSeq standard procedures, and eventually sequenced in an Illumina HiSeq2000 sequencer, using a full flow-cell, 4 samples per lane, to produce 2x101 bp paired-end reads. The whole dataset consisted of 2 biological replicates from greenhouse plants (2 tissues x 2 replicates x 2 genotypes), 2 biological replicates of developing inflorescences and 3 biological replicates of leaves from plants subjected to drought and well irrigated plants in the growth chamber (5 x 2 genotypes x 2 treatments).

5.2.4. *RNAseq data preprocessing and transcriptome assembly*

Raw reads were sequentially processed with FASTQC v0.10.0 (Andrews, 2010) and Trimmomatic v0.22 (Bolger et al., 2014), discarding stretches of mean Phred score <28 and cropping the first nucleotides to ensure a per-position A, C, G, T frequency near 0.25. Only reads of length ≥ 80 nucleotides were kept for further analysis. Surviving reads were error-corrected with Musket v1.0.6 (Liu et al., 2013) and default parameters. Then, reads were assembled following two different procedures, *de novo* and reference-guided.

De novo assemblies were obtained using Trinity r2013-02-25 recommended procedures (Haas et al., 2013). First, reads from sample replicates were pooled together and *in silico* normalized, to a maximum coverage of 30. This procedure was repeated with the resulting read sets to obtain, for each genotype, a final set of normalized reads. These were used for *de novo* assembly of SBCC073 and Scarlett transcriptomes.

A reference-guided assembly (RGA) was generated with the Tuxedo pipeline (Trapnell et al., 2012). First, clean reads were mapped to the IBGSC cv. Morex assembly (Mayer et al., 2012) with Tophat2 (v2.0.9; --b2-very-sensitive, --b2-scor-min C,-28,0 –read-mismatches 4 –read-gap-length 12 –read-edit-dist 12 -G 21Aug12_Transcript_and_CDS_structure.gff). This mapping procedure was performed in two steps, a first one to exclude reads with multiple mappings to the whole reference assembly (-M, -g 1, --no-discordant) and a second one to identify reads mapping unambiguously to gene coding loci (-g 2, --no-discordant, --no-mixed). Mappings were used as input for Cufflinks (v2.2.1). Individual assemblies were merged with the reference Morex assembly with Cuffmerge.

5.2.5. *Correction, validation and annotation of de novo transcriptomes*

Clean reads were mapped back to the *de novo* transcriptomes using Trinity script *alignReads.pl* with Bowtie (Langmead et al., 2009). In addition, the newly assembled isoforms were mapped to Morex, Bowman, Barke WGS (Whole Genome Shotgun) assemblies (Mayer et al., 2012) and Haruna Nijo flcDNAs (Matsumoto et al., 2011) with the script *bmaux_align_fasta* from the Barleymap package (Cantalapiedra et al., 2015) (hierarchical=yes query-mode=cdna thres-id=98 thres-cov=10), keeping together sequences matching the same reference sequence. Sequences in each of these groups were clustered with WCD-express v0.6.3 (Hazelhurst and Liptak, 2011) using threshold=24, which is equivalent to a 98% identity cut-off.

Presence of these isoforms in existing references was further confirmed by aligning them iteratively to additional sequence repositories. These were the Haruna Nijo genome assembly (Sato et al., 2016), genome contigs of Chinese Spring wheat (Mayer et al., 2014), barley ESTs from HarvEST assembly 36 (Close et al., 2007), the MIPS repeat database (Nussbaumer et al., 2013), and sequences from *Hordeum*, *Brachypodium, Triticum, Oryza* or *Aegilops* in the nt NCBI database (ftp.ncbi.nlm.nih.gov/blast/db). Alignment to Morex, Bowman and Barke WGS assemblies, and to Haruna Nijo genome and flcDNAs was repeated with a more stringent coverage threshold (thres-cov=80). Finally, transcripts were scanned for the presence of sequencing vectors by comparison with the EMVec database (ftp://ftp.ebi.ac.uk/pub/databases/emvec/) and as a result 64 sequences were removed.

Gene annotation of assembled contigs was performed with the script *transcripts2cdsCPP.pl* (-n 50) from GET_HOMOLOGUES-EST (v 04052016, https://github.com/eead-csic-compbio/get_homologues), which uses Transdecoder (https://transdecoder.github.io/) and blastx alignments to SwissProt proteins to define CDS sequences. Clusters obtained with GET_HOMOLOGUES-EST (get_homologues-est.pl -t 0 -M -S 96 -A –L), requiring percentage sequence identity > 96, were used to obtain reciprocal correspondences between transcripts from SBCC073 and Scarlett assemblies. PFAM domains in translated CDS sequences were also annotated (*get_homologues-est.pl* –D).

5.2.6. *Analysis of gene expression*

Differential expression contrasts were performed for each genotype, tissue and treatment; both for isoforms and genes. For this purpose, we compared three different pipelines.

For the first one, estimation of expression levels of isoforms and genes was done with RSEM v.1.2.11 (Li and Dewey, 2011), using Bowtie2 (Langmead and Salzberg, 2012) and otherwise default parameters. RSEM 'expected counts' were used as input for differential expression analyses with the 'glm' functions of the R (R Development Core Team, 2008) Bioconductor package edgeR v3.8.6 (Robinson et al., 2010) (false discovery rate function "BH" set to 0.001). A minimum CPM (counts per million) of 0.4, equivalent to around 10 RSEM 'expected counts' based on a linear regression (R-square = 1, intercept ~ 0, slope = 25), was required in at least half of the samples to include an isoform or a gene in the analysis.

A second method relied on kallisto v0.42.5 (Bray et al., 2016) to obtain 'expected counts' and to generate 100 bootstrap samples for each replicate, followed by test for differential expression with sleuth v.0.28.0 Wald test (Pimentel et al., 2016), using the previously generated bootstrap samples.

For the third method, Cuffquant and Cuffdiff v.2.2.1 (Trapnell et al., 2013) were used to test differential expression, with FDR 0.05, on the RGA transcripts.

Principal component analyses (PCA) of the resulting expression estimates from kallisto were done with the function PCA from R package FactoMineR 1.29 (Lê et al., 2008). Correlation analysis was performed using the R package corrplot 0.73 (Wei and Simko, 2014).

5.2.7. RT-qPCR validation

Reference genes for calculating relative expression were either searched in the literature or selected from our RNAseq data. The latter were those with the smallest coefficient of variation of expression values across samples, among isoforms not reported as differentially expressed (DE) by edgeR. DE isoforms to be checked with RT-qPCR were chosen randomly from bins covering the range of edgeR logFC. All the selected DE isoforms had TPM (transcripts per million) greater than 1. Primers for both reference genes and DE isoforms were designed with Primer Express 3.0 (Applied Biosystems). Conservation of the target sequences was checked in both SBCC73 and Scarlett isoforms. Whenever possible, one of the primers of the pair was set over an exon-exon junction and towards the 3' end.

The same DNase I-treated RNA samples used for RNAseq were utilized for the RT-qPCR assays. First strand cDNA synthesis was made from 2 µg of total RNA to a final volume of 40 µl containing oligo(dT)20 for priming and SuperScript III Reverse Transcriptase (Invitrogen, Cat.No. 18080-044). All the RT-qPCR reactions were performed in an ABI7500 (Applied Biosystems, Foster City, CA, USA) with the following PCR profile: 95ºC 10 min pre-denaturation step; 95ºC 15 sec denaturation and 60ºC 50 sec annealing (40 cycles), followed by a melting curve 60ºC-95ºC default ramp rate. The efficiency of primers was obtained from calibration curves with 1:5 dilution series and at least 4 points fitted in a linear regression with R-square over 0.99. We used NormFinder (Andersen et al., 2004) to analyze the stability value of the reference genes. Relative change of expression was calculated according to Pfaffl (2001), but using the geometric mean of three reference genes as normalization factor (Vandesompele et al., 2002).

5.2.8. Functional annotation of differentially expressed isoforms

Software CPC (Kong et al., 2007) was used to tag DE isoforms as coding or non-coding, and to obtain Uniref90 best hits. In addition, contained CDS sequences were deduced and PFAM protein domains annotated, as explained earlier for all the isoforms of each transcriptome. GO terms for each DE isoform were obtained with in-house script barleyGO (http://www.eead.csic.es/compbio/soft/barleyGO.tgz). Enrichment tests for PFAM domains and GO terms were performed in R using the Fisher exact test (p-value < 0.01). For the GO terms, we used the R package topGO (Alexa and Rahnenfuhrer, 2016).

DE isoforms were searched in metabolic pathways databases, including KEGG (Kanehisa et al., 2016), PlantReactome (Tello-Ruiz et al., 2016) and PlantCyc (Plant Metabolic Network, 2016). For KEGG, we obtained the list of genes of *Oryza sativa* ("osa"), from which we retrieved Orthology identifiers and pathways. DE isoforms were aligned to those genes with blastn (-perc_identity 75 –num_alignments 1), discarding hits with low query coverage in the alignment ('qcovs' < 70). PlantReactome (file "gene_ids_by_pathway_and_species.tab") was explored with Morex gene identifiers to obtain the pathways involved in differential expression. The gene identifiers were derived from mappings of *de novo* assemblies to the Morex reference genome from the validation step using the Barleymap package, as explained above. In the case of PlantCyc, we obtained the blast set "plantcyc.fasta" and enzymes annotation ("PMN11_June2016/plantcyc_pathways.20160601"), and used a custom script to match annotated enzymes with blastx (-evalue 0.00001 –num_alignments 1), filtering hits with percentage identity ≥ 75. Enzymes and pathways were grouped in broader categories manually, by merging their textual descriptions in KEGG and PlantCyc.

5.2.9. *Comparison with related studies*

The literature was surveyed to obtain protein and transcript sequences which had been previously associated with response to water deprivation in barley. These drought-related sequences were aligned with Blast[p|x] to genes from the Haruna Nijo genome assembly, which allowed mapping them to their corresponding DE isoforms from this study.

5.2.10. *Clustering and identification of cis-regulatory elements of co-expressed genes*

DE isoforms were clustered based on their TPM values (from kallisto). Distance between each pair of isoforms was calculated with Pearson correlation. This metric was weighted with Euclidean distance, under the hypothesis that isoforms sharing their expression pattern, but differing in magnitude, might have promoters which could be overlooked when clustered together with Pearson correlation only. These distances were used to perform hierarchical clustering (R package hclust, method="complete"). To declare the final number of clusters, the dendrogram was pruned when 95% of clusters had an internal average distance below 0.001% of the initial average distance of all DE isoforms.

The following procedure was used to recover promoter sequences corresponding to the genes present in the expression clusters. DE isoforms from each cluster were mapped to transcripts from the Morex WGS assembly (Blastn -perc_identity 98). For each cluster containing 10 or more genes, repeat-masked promoter sequences (-1000, +200 nucleotides around TSS) were retrieved from the RSAT::Plants server (http://plants.rsat.eu, version Hordeum_vulgare.082214v1.29) (Medina-Rivera et al., 2015). As negative controls, promoter sequences were retrieved from randomly generated gene clusters of the same size. Enrichment in GO terms and motif discovery with oligo-analysis and dyad-analysis were performed following the protocol of (Contreras-Moreira et al., 2016). Motif scores within upstream regions of co-expressed genes and their orthologous genes in *Brachypodium distachyon* reference (v1.0.29) (International Brachypodium Initiative, 2010), were obtained with the program matrix-scan from RSAT::Plants. These scores were also calculated for motifs generated by permutation of the bases of each discovered motif. Therefore, two types of evidences were used to assess the reliability of discovered motifs: i) their statistical significance compared to the negative controls, and ii) their matrix-scan scores compared to the scores of permuted motifs. Discovered motifs were annotated by comparison to plant regulatory motifs in the footprintDB repository (Sebastian and Contreras-Moreira, 2014). The highest scoring motif, in terms of footprintDB 'Ncor' score, was selected as the best hit. The full report on the promoter analysis, including source code, is available at http://floresta.eead.csic.es/rsat/data/barley_drought_clusters.

Finally, deduced peptide sequences of DE isoforms annotated as transcription factors with iTAK (http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi), were used to predict their putative DNA-binding motifs with footprintDB.

5.3. Results

5.3.1. Growth of Scarlett and SBCC073 plants subjected to drought

Two different experiments were set up, in which plants were placed in a growth chamber or in a greenhouse. The growth chamber was kept at strictly controlled environmental conditions, whereas the greenhouse underwent a natural photoperiod (August - September, 2012, starting with 14 h 23 min and ending with 11 h 46 min daylight, http://www.fomento.gob.es/salidapuestasol/2012/Zaragoza-2012.txt) and controlled, but more variable, temperature and humidity. Both daytime and night temperatures in the greenhouse were higher than in the growth chamber, whereas relative humidity was similar on average. In both settings, water stress was imposed after initiation of the stem elongation stage. Growth chamber plants were watered in order to conserve 70% field capacity (FC) (controls, C), or instead subjected to reduced irrigation, up to 20% FC (drought, D). Greenhouse plants were irrigated to an intermediate 50% FC (mild drought and heat, MDH). These experiments are outlined in Figure 5.1.

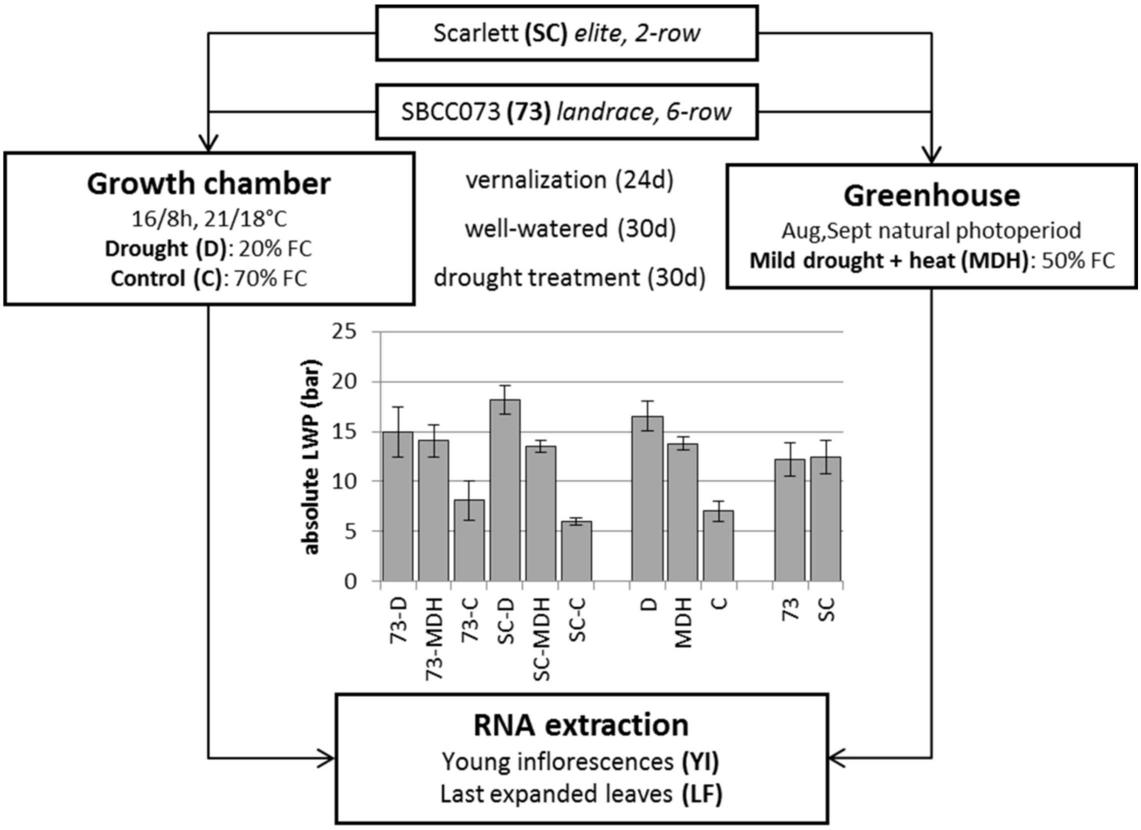


Figure 5.1. Design of stress treatments, and leaf water potential patterns. SBCC073 (73) and Scarlett (SC) plants were placed in a growth chamber and in a greenhouse. Growth chamber plants were either watered to 70% FC (control, C) or instead 20% FC (drought, D). Greenhouse plants were subjected to combined mild drought (50% FC) and heat stress (MDH). Drought treatments lasted 30 days (30d), after 24d of vernalization and 30d of normal irrigation. The bar plot shows average ± SEM absolute leaf water potential (LWP).

Daily loss of water, based on the weights of pots, was largest in C plants, intermediate under MDH and lowest under D. The same trend was observed for leaf water potential (LWP), summarized in Figure 5.1. LWP was proportional to the three imposed water regimes, with plants subjected to drought (D and MDH) showing larger absolute LWP that those well-watered. The largest value corresponded to Scarlett plants under D, in which SBCC073 plants had values comparable to those of both SBCC073 and Scarlett plants under MDH. Likewise, minimum values for stomatal conductance (SCo) were recorded for plants under D (Table 5.1). However, the largest SCo was found under MDH. Relative water content (RWC) was lowest for plants under D, in both genotypes, whereas under MDH, it was closer to that of C plants in SBCC073, and closer to that of plants under D in Scarlett. Tiller number (TN) and visible spike number (VSN) were also affected by water deprivation, being larger in C than under D, both in SBCC073 and Scarlett. Under MDH, similarly to the RWC observations, TN was less affected in SBCC073 than in Scarlett.

Table 5.1. Physiological measurements of plants in the drought experiments. Treatments corresponded to control (C) and drought (D) in the growth chamber, at 70% and 20% field capacity (FC), respectively; whereas greenhouse plants were kept at mild drought and heat (MDH, 50% FC). Physiological and morphological measurements were absolute leaf water potential (LWP), stomatal conductance (SCo), relative water content (RWC) of leaves, tiller number (TN) and visible spike number (VSN).

Treatment	LWP (bar)	SCo (mmol/m2s)	RWC	TN	VSN
	------------------------------- SBCC073 -------------------------------				
C	8.09	33.57	0.94	13	4
MDH	14.10	40.93	0.97	11	1
D	14.95	23.02	0.82	8	3
	------------------------------- Scarlett -------------------------------				
C	6.00	12.45	0.92	16	2
MDH	13.47	39.00	0.85	5	0
D	18.15	0.25	0.87	11	2

5.3.2. *Assembly and validation of Scarlett and SBCC073 transcriptomes*

Sequencing of cDNA libraries, derived from leaf (LF) and young inflorescence (YI) transcripts, yielded 1.18 billion paired-end sequence reads. From this dataset, we assembled separate *de novo* transcriptomes for Scarlett and SBCC073, as well as a reference-guided assembly (RGA).

The *de novo* assemblies yielded similar numbers and lengths of isoforms for both genotypes (Table 5.2). These sets, with 103,623 genes in SBCC073 and 113,962 in Scarlett, were comparable but larger than the annotated gene sets for the Morex cultivar (Mayer et al., 2012), with 75,258 high and low confidence genes, and with the results from the RGA (75,204 genes). Validation and correction of the *de novo* isoforms was performed in three stages. First, the clean reads were mapped back to the assembled transcripts, to compute the fraction of well aligned pairs of reads (both reads mapped, correct orientation and insert size), which was near 83% for both cultivars. Second, *de novo* subcomponents were revised for re-clustering. This requires some explanation. Whereas RGA contigs are isoforms associated to known genes from the reference, *de novo* assembly generates contigs which are isoforms clustered in so called subcomponents. In some cases, these subcomponents accumulate closely related sequences, for instance from paralogous genes or expressed pseudogenes, which should be separated. Therefore, this second step consisted in re-clustering isoforms from subcomponents to genes, by alignment to annotated references (see Methods), and assigning them to different loci when appropriate. The final number of genes in the *de novo* assemblies was 112,923 in SBCC073 and 123,582 in Scarlett. Third, the isoforms were matched to a variety of genomic and transcriptomic sequence repositories of barley, wheat and other grasses. In total, 93% of SBCC073 and 87% of Scarlett genes could be confirmed.

These sequence comparisons are further illustrated in Figure 5.2. Note that at least 10% alignment coverage was required in all cases. Further, the alignment against Morex, Barke, Bowman and Haruna Nijo was repeated, with a more strict minimum coverage of 80%. This test confirmed that 88,293 (78% of SBCC073) and 92,713 (75% of Scarlett) genes map with high confidence to previously reported barley sequences.

5.3.3. *Analysis of gene expression*

Clean paired-end reads were mapped back to SBCC073 and Scarlett assemblies, to estimate expression counts for each transcript. These estimates were subsequently used to identify DE tags (genes and isoforms) between stressed treatments and C, for each tissue and genotype. For this purpose, we compared three different pipelines, which rely on different software for each of the two steps: RSEM-edgeR, kallisto-sleuth and Cuffquant-Cuffdiff. In addition, a set of isoforms from YI were randomly chosen to test their expression by RT-qPCR, using genes selected from the literature and from our RNAseq expression data as references.

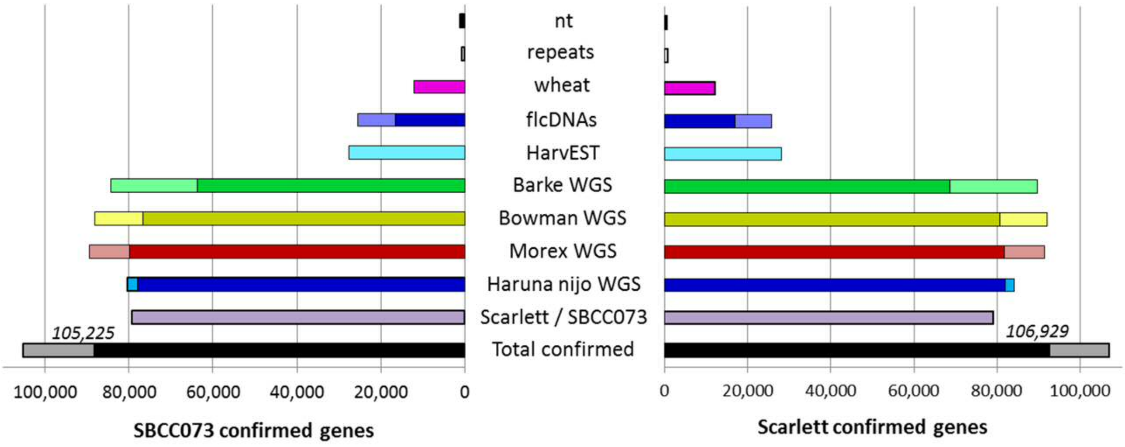
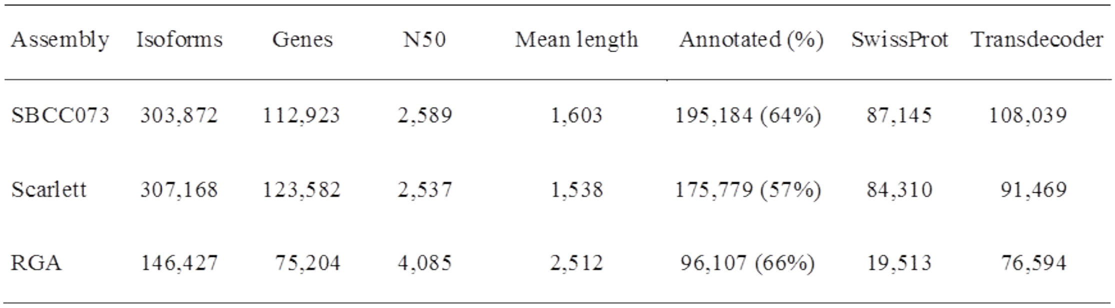


Figure 5.2. De novo assembled genes confirmed in existing barley references. Bars indicate the number of assembled genes of landrace SBCC073 (left) and cultivar Scarlett (right) which were confirmed by alignment to each other, and to several sequence repositories of barley and wheat (for list, see text). The total number of genes confirmed for each of the two assemblies is also shown (bottom black/grey bars). The alignments required 98% identity and a minimum alignment query coverage of either 10% (whole bars) or 80% (fraction of bars filled with a darker color).

The results of differential expression computed with kallisto-sleuth had the best agreement with those of RT-qPCR (Figure 5.3). The outcome of the RSEM-edgeR pipeline was comparable to kallisto-sleuth after discarding a few outliers. Moreover, PCA and clustering of samples, using expression estimates from kallisto, showed good correlation between replicates. When the expression estimates, obtained with the three methods, were directly compared, RSEM-edge and kallisto-sleuth showed the best agreement. In order to reduce false positives, final DE tags were obtained from the intersection between those two methods.

Overall, the response differed between genotypes in YI, and between treatments in LF (Figure 5.4). Under D, we found almost no response in SBCC073, either in YI or LF samples, whereas in Scarlett, YI samples had many DE tags. On the contrary, abundant changes in gene expression were observed under MDH, with the exception of YI from SBCC073, which remained mostly unaltered. Regarding the proportion of up-regulated tags over total DE tags, in LF under MDH it was close to 50%, in both genotypes, whereas in YI from Scarlett plants it increased under D (62.6% in isoforms, 61.4% in genes) and decreased dramatically under MDH. There was high agreement between DE genes and DE isoforms in all contrasts, aalthough some DE genes were different to those found when analyzing isoforms. On the other hand, common DE tags between different contrasts were negligible, with the exception of LF under MDH, in which Scarlett and SBCC073 shared a low but sizable fraction.

Table 5.2. Statistics of *de novo* and reference-guided assemblies. Rows correspond to either *de novo* assemblies (SBCC073 and Scarlett) or reference-guided assembly (RGA). The upper part of the table shows the number of isoforms and genes, as obtained from the assembler, along with statistics on length of isoforms (N50 and mean length). The bottom half shows the number and percentage of annotated isoforms, and whether this annotation was obtained from alignment to SwissProt database or by CDS *de novo* prediction with Transdecoder.



Assembly	Isoforms	Genes	N50	Mean length	Annotated (%)	SwissProt	Transdecoder
SBCC073	303,872	112,923	2,589	1,603	195,184 (64%)	87,145	108,039
Scarlett	307,168	123,582	2,537	1,538	175,779 (57%)	84,310	91,469
RGA	146,427	75,204	4,085	2,512	96,107 (66%)	19,513	76,594

Finally, overall gene expression changes (number of DE tags and cumulative logFC from each contrast) were compared with the physiological measurements. Some large correlations were obtained, although these results must be considered with care due to the small sample size. For LWP, we found a positive correlation with YI overall logFC of isoforms (r 0.97, p-value 0.03) and number of DE tags (r 0.99, p-value 0.01). SCo exhibited strong positive correlation with gene expression changes in LF (ranges: r 0.95 - 0.98, p-values 0.05 - 0.02), to which VSN showed strong negative correlation (ranges: r -0.91 - -0.96, p-values 0.04 - 0.09).

DE isoforms were annotated combining different strategies, as described in Materials and Methods. The main annotation results are detailed in the following sections.

5.3.4. Differentially expressed isoforms in leaves under drought

As explained in the previous section, just a few isoforms were DE in LF under D. In both genotypes, we found an up-regulated isoform encoding a polyamine oxidase, involved in spermine and spermidine degradation. In addition, an isoform corresponding to a chlorophyll apoprotein from photosystem II was down-regulated in Scarlett. However, this change was not observed in SBCC073, which instead showed induction of transcripts of three proteins, namely ABA/WDS (abscisic acid / water deficit stress) induced protein, ribonuclease T2 and calcineurin-like phosphoesterase. Other DE isoforms were annotated as non-coding or of unknown function.

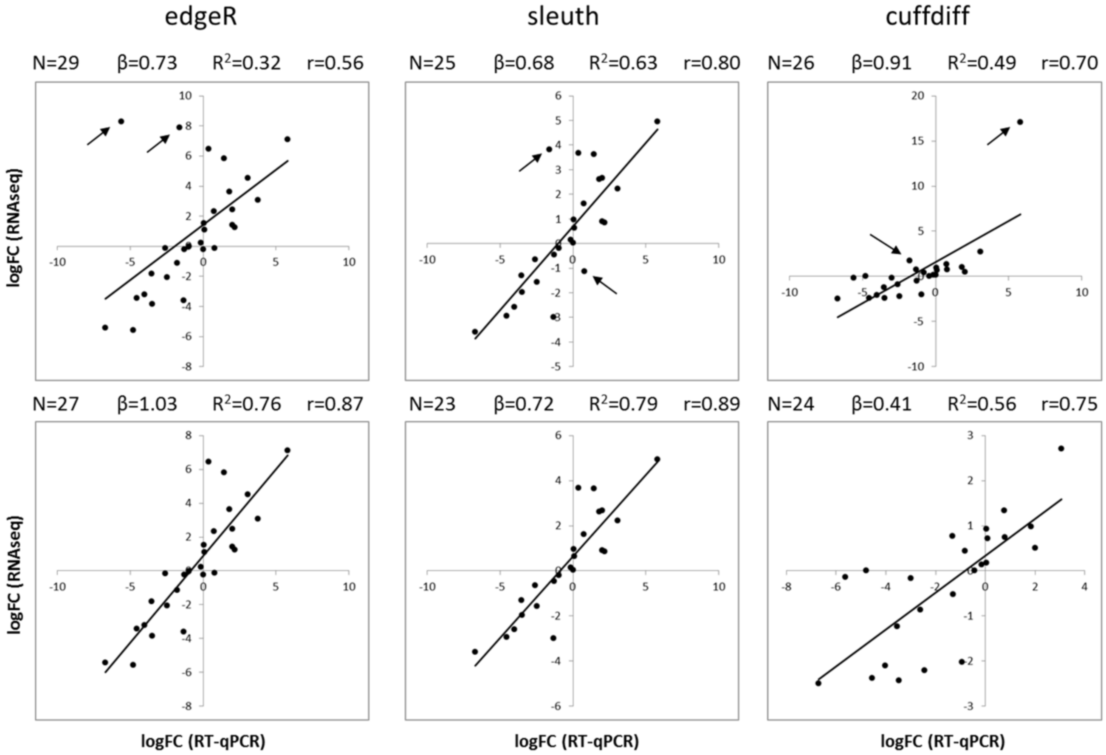


Figure 5.3. Comparison of RT-qPCR and RNAseq gene expression results. Scatterplots show the logFC of isoforms obtained with RT-qPCR (horizontal axis) and with RNAseq (vertical axis). LogFC values from RNAseq were obtained with three different analysis methods: edgeR (left), sleuth (center) and Cuffdiff (right). Plots on the top show all available data, whereas plots on the bottom show data after removing the two most scattered data points (black arrows). Black lines correspond to a linear regression. N: number of data points; β: slope of regression; R2: coefficient of determination; r: Pearson correlation coefficient.

5.3.5. Differentially expressed isoforms in leaves under mild drought and heat

There were more DE tags in LF under MDH, and involved a more diverse array of gene functions than under D. The same polyamine oxidase induced in LF under D was also observed up-regulated in Scarlett under MDH. Intriguingly, in SBCC073 we found up-regulated a transcript encoding a spermidine synthase.

Some GO terms were enriched in both genotypes, including "phosphorelay signal transduction system", "pyrimidine-containing compound biosynthesis process", "response to temperature stimulus", "response to water deprivation" and "thiamine biosynthetic process". Other pathways and cellular processes involved in the responses of both genotypes were starch phosphorylation, chorismate biosynthesis, L-ascorbate biosynthesis and recycling, DMNT biosynthesis (a volatile homoterpene), and other proteins involved in protein folding, proteolysis and defense response (Figure 5.5). We also found in both genotypes up-regulation of isoforms annotated as CCA1/LHY MYB-related TF. Moreover, we found another DE gene annotated as MYB-related TF in both genotypes, which is similar to *Arabidopsis thaliana* TCL2, and an additional uncharacterized MYB-related TF in SBCC073 only. At the same time, down-regulation of other genes related with circadian rhythm was detected, like adagio-like protein 3 and a PRR1 (HvTOC1) transcription regulator. In SBCC073, we found also down-regulation of another circadian clock related gene, annotated as APRR3. Another gene up-regulated in both genotypes was annotated as protein kinase CIPK9. Regarding transporters, repressed transcripts encoding aquaporins were noticed in both genotypes. There were a few other protein domains regulated in both genotypes, most of them repressed.

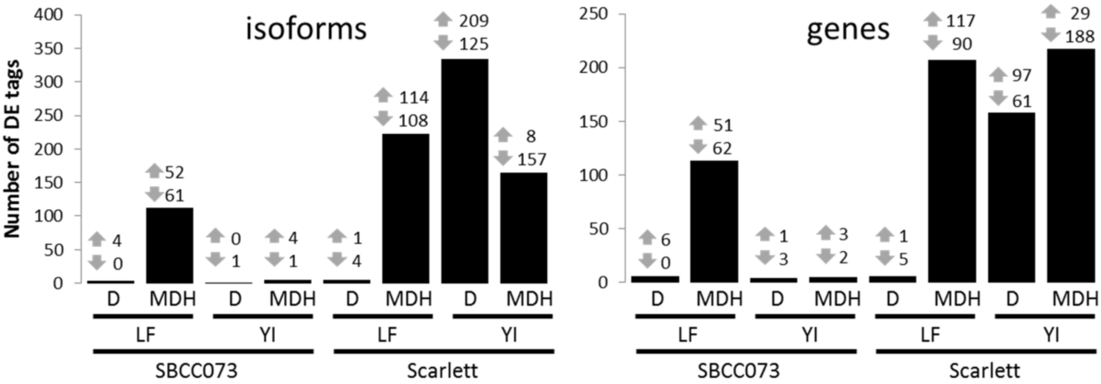


Figure 5.4. Number of differentially expressed isoforms and genes. Number of up-regulated (up arrows) and down-regulated (down arrows) differentially expressed tags (isoforms, left; genes, right), for each contrast. Bars show the sum of both induced and repressed tags. LF: leaves. YI: young inflorescences. D: drought treatment. MDH: mild drought and heat treatment.

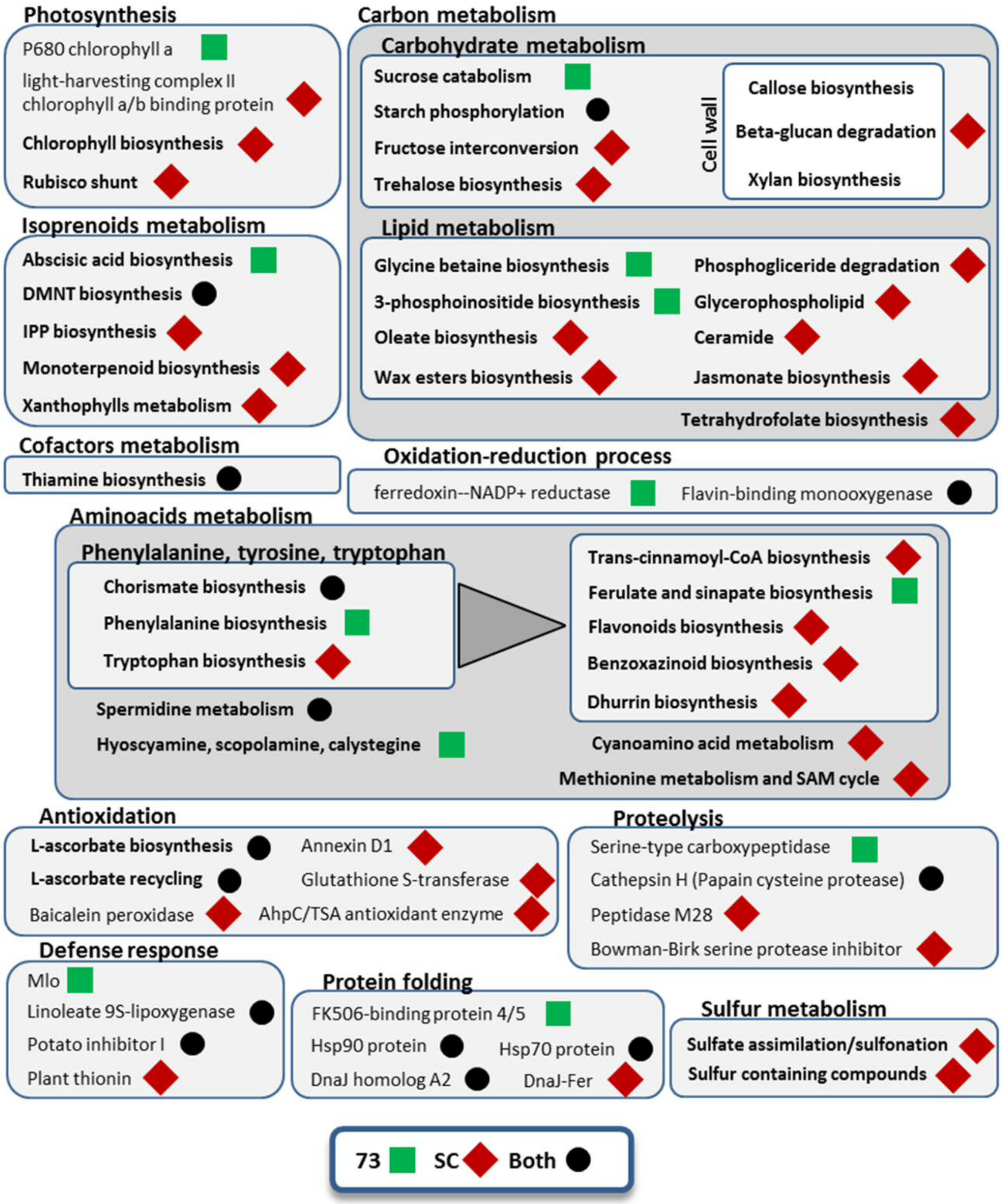

Photosynthesis
P680 chlorophyll a 🟩
light-harvesting complex II chlorophyll a/b binding protein ◆
Chlorophyll biosynthesis ◆
Rubisco shunt ◆
Isoprenoids metabolism
Abscisic acid biosynthesis 🟩
DMNT biosynthesis ●
IPP biosynthesis ◆
Monoterpenoid biosynthesis ◆
Xanthophylls metabolism ◆
Cofactors metabolism
Thiamine biosynthesis ●
Carbon metabolism
Carbohydrate metabolism
Sucrose catabolism 🟩
Starch phosphorylation ●
Fructose interconversion ◆
Trehalose biosynthesis ◆
Cell wall
Callose biosynthesis
Beta-glucan degradation ◆
Xylan biosynthesis
Lipid metabolism
Glycine betaine biosynthesis 🟩
3-phosphoinositide biosynthesis 🟩
Oleate biosynthesis ◆
Wax esters biosynthesis ◆
Phosphogliceride degradation ◆
Glycerophospholipid ◆
Ceramide ◆
Jasmonate biosynthesis ◆
Tetrahydrofolate biosynthesis ◆
Oxidation-reduction process
ferredoxin--NADP+ reductase 🟩
Flavin-binding monooxygenase ●
Aminoacids metabolism
Phenylalanine, tyrosine, tryptophan
Chorismate biosynthesis ●
Phenylalanine biosynthesis 🟩
Tryptophan biosynthesis ◆
Trans-cinnamoyl-CoA biosynthesis ◆
Ferulate and sinapate biosynthesis 🟩
Flavonoids biosynthesis ◆
Benzoxazinoid biosynthesis ◆
Dhurrin biosynthesis ◆
Spermidine metabolism ●
Hyoscyamine, scopolamine, calystegine 🟩
Cyanoamino acid metabolism ◆
Methionine metabolism and SAM cycle ◆
Antioxidation
L-ascorbate biosynthesis ●
L-ascorbate recycling ●
Baicalein peroxidase ◆
Annexin D1 ◆
Glutathione S-transferase ◆
AhpC/TSA antioxidant enzyme ◆
Proteolysis
Serine-type carboxypeptidase 🟩
Cathepsin H (Papain cysteine protease) ●
Peptidase M28 ◆
Bowman-Birk serine protease inhibitor ◆
Defense response
Mlo 🟩
Linoleate 9S-lipoxygenase ●
Potato inhibitor I ●
Plant thionin ◆
Protein folding
FK506-binding protein 4/5 🟩
Hsp90 protein ●
Hsp70 protein ●
DnaJ homolog A2 ●
DnaJ-Fer ◆
Sulfur metabolism
Sulfate assimilation/sulfonation ◆
Sulfur containing compounds ◆
73 🟩 SC ◆ Both ●

Figure 5.5. Metabolic pathways and cellular processes with differentially expressed isoforms from leaves under mild drought and heat. Metabolic pathways, cellular processes and proteins with differentially expressed isoforms are grouped into more general processes, within boxes. Bold categories include several differentially expressed isoforms from a given pathway or process, whereas non-bold names are from specific proteins. Green squares represent processes affected only in SBCC073 (73) plants, whereas red diamonds indicate those altered only in Scarlett (SC). Processes and proteins with changes in gene expression in both genotypes are marked with a black circle. A triangle links the metabolism of aromatic amino acids with downstream pathways of secondary metabolites obtained from them.

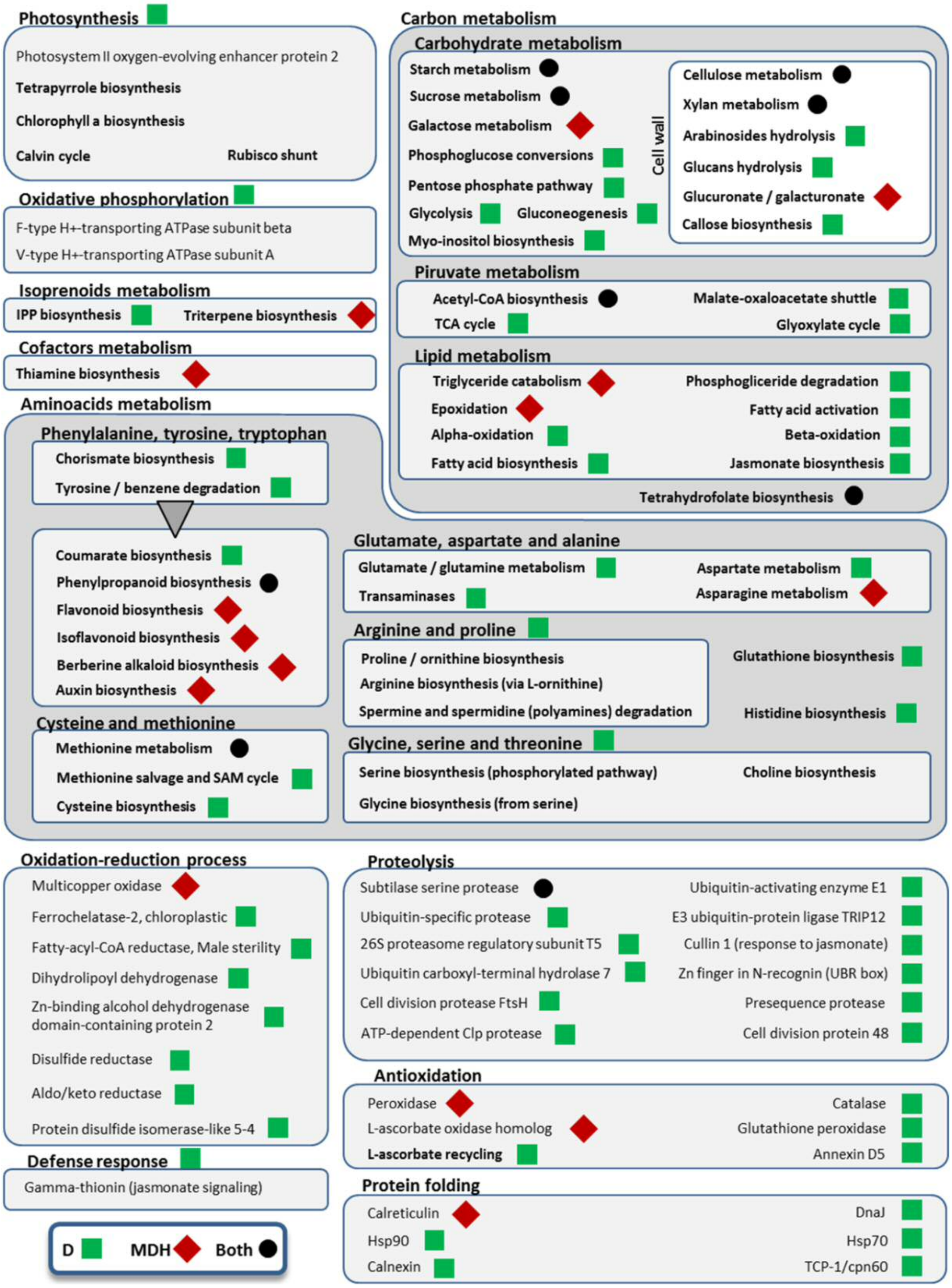

Photosynthesis 🟩
Photosystem II oxygen-evolving enhancer protein 2
Tetrapyrrole biosynthesis
Chlorophyll a biosynthesis
Calvin cycle
Rubisco shunt
Oxidative phosphorylation 🟩
F-type H+-transporting ATPase subunit beta
V-type H+-transporting ATPase subunit A
Isoprenoids metabolism
IPP biosynthesis 🟩
Triterpene biosynthesis ◆
Cofactors metabolism
Thiamine biosynthesis ◆
Aminoacids metabolism
Phenylalanine, tyrosine, tryptophan
Chorismate biosynthesis 🟩
Tyrosine / benzene degradation 🟩
Coumarate biosynthesis 🟩
Phenylpropanoid biosynthesis ●
Flavonoid biosynthesis ◆
Isoflavonoid biosynthesis ◆
Berberine alkaloid biosynthesis ◆
Auxin biosynthesis ◆
Cysteine and methionine
Methionine metabolism ●
Methionine salvage and SAM cycle 🟩
Cysteine biosynthesis 🟩
Carbon metabolism
Carbohydrate metabolism
Starch metabolism ●
Sucrose metabolism ●
Galactose metabolism ◆
Phosphoglucose conversions 🟩
Pentose phosphate pathway 🟩
Glycolysis 🟩
Gluconeogenesis 🟩
Myo-inositol biosynthesis 🟩
Cell wall
Cellulose metabolism ●
Xylan metabolism ●
Arabinosides hydrolysis 🟩
Glucans hydrolysis 🟩
Glucuronate / galacturonate ◆
Callose biosynthesis 🟩
Piruvate metabolism
Acetyl-CoA biosynthesis ●
TCA cycle 🟩
Malate-oxaloacetate shuttle 🟩
Glyoxylate cycle 🟩
Lipid metabolism
Triglyceride catabolism ◆
Epoxidation ◆
Alpha-oxidation 🟩
Fatty acid biosynthesis 🟩
Phosphogliceride degradation 🟩
Fatty acid activation 🟩
Beta-oxidation 🟩
Jasmonate biosynthesis 🟩
Tetrahydrofolate biosynthesis ●
Glutamate, aspartate and alanine
Glutamate / glutamine metabolism 🟩
Transaminases 🟩
Aspartate metabolism 🟩
Asparagine metabolism ◆
Arginine and proline 🟩
Proline / ornithine biosynthesis
Arginine biosynthesis (via L-ornithine)
Spermine and spermidine (polyamines) degradation
Glutathione biosynthesis 🟩
Histidine biosynthesis 🟩
Glycine, serine and threonine 🟩
Serine biosynthesis (phosphorylated pathway)
Glycine biosynthesis (from serine)
Choline biosynthesis
Oxidation-reduction process
Multicopper oxidase ◆
Ferrochelatase-2, chloroplastic 🟩
Fatty-acyl-CoA reductase, Male sterility 🟩
Dihydrolipoyl dehydrogenase 🟩
Zn-binding alcohol dehydrogenase domain-containing protein 2 🟩
Disulfide reductase 🟩
Aldo/keto reductase 🟩
Protein disulfide isomerase-like 5-4 🟩
Defense response 🟩
Gamma-thionin (jasmonate signaling)
D 🟩 MDH ◆ Both ●
Proteolysis
Subtilase serine protease ●
Ubiquitin-specific protease 🟩
26S proteasome regulatory subunit T5 🟩
Ubiquitin carboxyl-terminal hydrolase 7 🟩
Cell division protease FtsH 🟩
ATP-dependent Clp protease 🟩
Ubiquitin-activating enzyme E1 🟩
E3 ubiquitin-protein ligase TRIP12 🟩
Cullin 1 (response to jasmonate) 🟩
Zn finger in N-recognin (UBR box) 🟩
Presequence protease 🟩
Cell division protein 48 🟩
Antioxidation
Peroxidase ◆
L-ascorbate oxidase homolog ◆
L-ascorbate recycling 🟩
Catalase 🟩
Glutathione peroxidase 🟩
Annexin D5 🟩
Protein folding
Calreticulin ◆
Hsp90 🟩
Calnexin 🟩
DnaJ 🟩
Hsp70 🟩
TCP-1/cpn60 🟩

Figure 5.6. Metabolic pathways and cellular processes with differentially expressed isoforms from Scarlett young inflorescences. Metabolic pathways, cellular processes and proteins with differentially expressed isoforms are grouped into more general processes, within boxes. Bold categories include several differentially expressed isoforms from a given pathway or process, whereas non-bold names are from specific proteins. Green squares point out processes altered only under drought (D), whereas red diamonds indicate processes affected only in the mild drought and heat experiment (MDH). Processes and proteins with changes in gene expression in both treatments are marked with a black circle. A triangle links the metabolism of aromatic amino acids with downstream pathways of secondary metabolites obtained from them.

Differences between genotypes were also seen among DE transcripts in LF under MDH. For instance, in SBCC073 there was enrichment of terms such as "actin filament-based movement", "ammonium ion metabolic process" and "defense response by cell wall thickening", while in Scarlett a greater variety of response-related terms were obtained, such as "response to abscisic acid", "response to bacterium", "response to ethylene", "response to hydrogen peroxide" or "response to wounding". Also, DE isoforms related to glycine betaine biosynthesis and to abscisic acid (ABA) biosynthesis were seen in SBCC073, whereas trehalose biosynthesis was involved in the response of Scarlett LF to MDH (Figure 5.5). Moreover, isoforms involved in cell wall, epidermis (wax esters) and membrane lipids (glycerophospholipids, ceramide) metabolism were up-regulated in Scarlett but not present among SBCC073 DE isoforms. This was also the case of some defense response metabolic pathways (benzoxazinoids and dhurrin biosynthesis), xanthophylls metabolism, several antioxidation related proteins (like baicalein peroxidase or glutathione S-transferase) or sulphur metabolism related proteins. We also found differences among TFs and protein kinases (PKs). For instance, CIPK17 and a C2C2-Dof TF, whose best SwissProt hit is Arabidopsis protein CDF2, were up-regulated, and an AP2/ERF-AP2 TF (related to *Brassica napus* BBM2) down-regulated, all in SBCC073. Instead, repression of a TUB TF, similar to *O. sativa* subsp. *japonica* TULP7, and induction of both a bZIP TF and a jasmonate ZIM TIFY TF, the latter related to *O. sativa* subsp. *japonica* TIFY6B, was noticed in Scarlett. Besides aquaporins, already mentioned, DE isoforms related to transport processes were different between genotypes, being more abundant in Scarlett. These included lipid transfer proteins, phosphate, potassium, triose-phosphate, adenine, vacuolar amino acid and ABC transporters, and a repressed NUCLEAR FUSION DEFECTIVE 4 (NFD4) protein.

5.3.6. Differentially expressed isoforms in young inflorescences in SBCC073

In YI, the transcriptional responses were markedly different between genotypes, with only minor responses in plants of genotype SBCC073 under both treatments. Indeed, a single down-regulated transcript was identified in SBCC073 under D, annotated as Pollen Ole e 1 allergen/extension. Under MDH, a repressed isoform was annotated as "non-coding", whereas four up-regulated isoforms corresponded to CCA1/LHY.

5.3.7. Differentially expressed isoforms in young inflorescences in Scarlett

In contrast with what was seen in SBCC073, YI from Scarlett showed abundant gene expression changes. Enriched GO terms found both under D and under MDH were scarce (Table 5.3), including cell wall-related processes "beta-glucan biosynthetic process", "lignin metabolic process", "phenylpropanoid metabolic process", and "cell wall organization or biogenesis", and others like "response to carbon dioxide" and "sucrose metabolic process". Other shared DE tags included isoforms involved in tetrahydrofolate biosynthesis and a subtilase serine protease (Figure 5.6). Among DE TFs in YI, we found B3-ARF isoforms (Auxin response factors with B3 and PB1 domains) induced under both treatments. However, reciprocal alignment revealed that they belong to different genes (blastn, alignment coverage 48% and percentage of identity 63%). The most similar protein of the isoform in the D treatment was ARF21, also known as OsARF7b, whereas the closest homologue of the isoform found under MDH was ARF11.

Besides B3-ARF TFs, only a few other isoforms were up-regulated in Scarlett YI under MDH, corresponding to an elongation factor EF-1, a DNA topoisomerase, a kinesin motor domain, CCA1/LHY, and a condensing complex subunit protein. All the others were down-regulated, whose enriched GO terms included "cellulose biosynthetic process", "xylan biosynthetic process", "flavonoid biosynthetic process", "mitotic chromosome condensation", "plasmodesmata-mediated intercellular transport" and "mucilage extrusion from seed coat" (Table 5.3). Other differences with respect to the D treatment were the involvement of enzymes from thiamine biosynthesis, triglyceride catabolism, epoxidation, berberine alkaloid biosynthesis or auxin biosynthesis (Figure 5.6). Among repressed isoforms related with transporters, we found sugar and lysine-histidine transporters, a PRA1 family protein B2 (a protein family related to regulation of vesicle trafficking, (Kamei et al., 2008), and several ABC transporters. Other proteins (and protein domains) which were found DE only under MDH included an expansin-B3, a putative cell wall protein, a PMR5/Cas1p, and several germin-like proteins.

Table 5.3. Gene Ontology terms enriched in Scarlett young inflorescences. The upper left section shows the GO terms enriched in both experiments (MDH: mild drought and heat; D: drought). The upper right section shows the GO terms enriched only under MDH. The bottom section shows the GO terms enriched only among differentially expressed isoforms under D.

MDH and D	MDH only
Beta-glucan biosynthetic process	Cellulose biosynthetic process
Lignin metabolic process	Xylan biosynthetic process
Phenylpropanoid metabolic process	Plasmodesmata-mediated intercellular transport
Response to carbon dioxide	Mucilage extrusion from seed coat
Sucrose metabolic process	Flavonoid biosynthetic process
Cell wall organization or biogenesis	Mitotic chromosome condensation
D only	
ARF protein signal transduction	Growth
Aspartate family amino acid biosynthetic process	Hydrogen peroxide catabolic process
ATP generation from ADP	L-alanine catabolic process, by transamination
ATP hydrolysis coupled proton transport	L-phenylalanine catabolic process
Callose deposition in cell wall	Methionine biosynthetic process
Carbohydrate catabolic process	NADP metabolic process
Cell wall thickening	ncRNA transcription
Cellular response to starvation	Pentose-phosphate shunt
De-etiolation	Polycistronic mRNA processing
Embryo development ending in seed dormancy	Positive regulation of embryonic development
Ethylene biosynthetic process	Positive regulation of ribosome biogenesis
Glucose metabolic process	Primary root development
Glycerol catabolic process	Protein import into chloroplast stroma
Pyruvate metabolic process	Starch metabolic process
Response to metal iron	Sulfur amino acid biosynthetic process
Response to hormone	Translation elongation
Response to osmotic stress	Tricarboxylic acid metabolic process
S-adenosylmethionine biosynthetic process	Triglyceride mobilization
Seed development	Wax biosynthetic process

Under D, Scarlett YI showed almost twice as many induced than repressed isoforms. The number of enriched GO terms was greater than for all the other contrasts, including numerous enriched processes (Table 5.3) and metabolic pathways (Figure 5.6), related with responses to abiotic stress (cell wall thickening, biosynthesis of wax, triglyceride mobilization, expansin-A7), development (seed, embryo and root development), central metabolism (starch, glucose, pyruvate, many amino acids, fatty acids biosynthesis, activation and beta-oxidation), hormones (ethylene, jasmonate), energy (ATP and NADP metabolism related proteins, F and V-type H+-transporting ATPases), nucleic acids and proteins metabolism, antioxidation, proteolysis, protein folding, numerous proteins involved in transport and vesicle trafficking, tRNA synthetases, an up-regulated MADS-MIKC TF whose best hit in SwissProt is *O. sativa* subsp. *japonica* MADS6, several PKs (like CIPK30) and phosphatases (like phosphoinositide phosphatase SAC7), proteins involved in interactions and signal transduction (SNF2, ASPR1 topless-related protein 1, 14-3-3 protein epsilon, CypP450), cytoskeleton proteins (tubulin, myosin, fimbrin and villin domains), and even processes related with photosynthetic tissues, like biosynthesis of chlorophyll a or tetrapyrrole, or induction of a Rubisco activase.

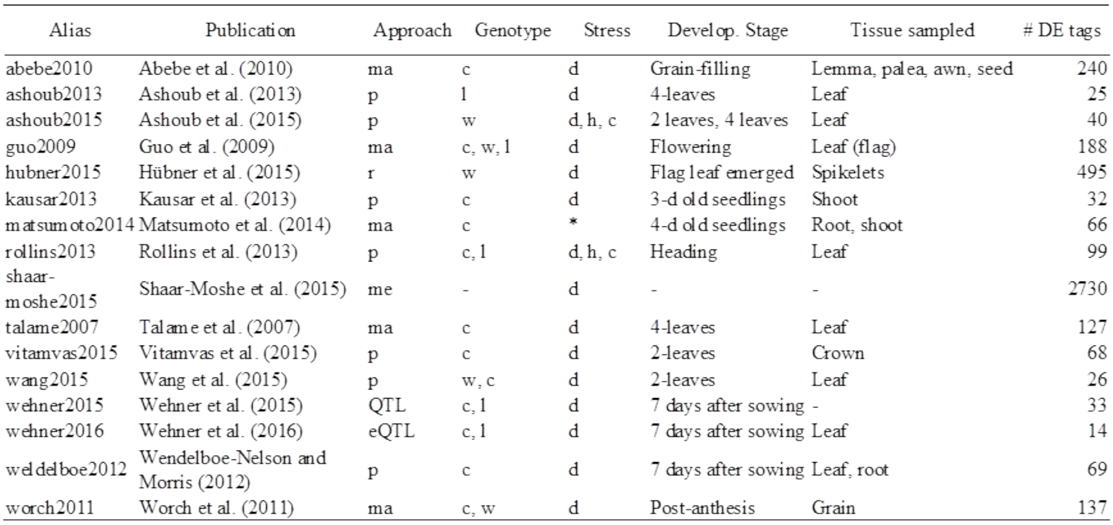
All these evidences indicate that responses to D and MDH of Scarlett YI were different, and that reproductive tissues were undergoing large gene expression changes, especially under D.

5.3.8. Comparison with related studies

We surveyed the literature reporting genes and proteins expressed in barley in response to water deprivation. The goal was to compare those sequences to the DE transcripts identified in this work. The studies listed in Table 5.4 include 5 microarray experiments, 7 based on proteomics, 1 RNAseq study, 1 QTL work, 1 surveying expression QTL and 1 meta-analysis. Most of them focused on barley plants under drought, with a few exceptions. The work "matsumoto2014" surveyed responses to desiccation, salt stress and ABA. In addition, both "ashoub2015" and "rollins2013" combined drought and heat stress. The meta-analysis "shaar-moshe2015" compared drought related genes from different plant species. Although many of these works (9) sampled leaves, other tissues were also analyzed in some of them (mainly shoots, roots, spikes and grain).

Out of 4389 DE tags (proteins, genes and transcripts) reported overall in the studies above, more than half (2730) were barley genes included in the meta-analysis "shaar-moshe2015" and, indeed, that study matches the largest number of DE tags of the current work. However, in relative terms, the most similar were those of "ashoub2013", "ashoub2015", "vitamvas2015", "wang2015", "kausar2013" and "rollins2013", in decreasing order, whose DE tags were also found in the present study in proportions ranging from 52% to 32% (see white bars in Figure 5.7). Interestingly, these are all proteomics studies. DE transcripts from Scarlett YI under D matched the largest percentage of DE tags from the surveyed studies.

Table 5.4. Studies from the literature assessing protein or transcript expression changes in response to drought in barley. An alias was assigned to each study, to facilitate referring to them. There are different approaches in the comparison dataset, including microarrays (ma), proteomics (p), RNAseq (r), meta-analysis (me), a QTL study and one based on eQTLs. The genotypes used for the experiments involve barley cultivars (c), landraces (l) or wild barley (w). The type of stress applied was drought (d), heat (h), drought and heat combined (c), or dessication, salt and ABA in the case of "matsumoto2014" (*). Stresses were applied during different developmental stages, and the tissue sampled was varied also. Finally, the number of differentially expressed tags (transcripts, genes, proteins) included in the comparison dataset is shown (# DE tags).



Alias	Publication	Approach	Genotype	Stress	Develop. Stage	Tissue sampled	# DE tags
abebe2010	Abebe et al. (2010)	ma	c	d	Grain-filling	Lemma, palea, awn, seed	240
ashoub2013	Ashoub et al. (2013)	p	l	d	4-leaves	Leaf	25
ashoub2015	Ashoub et al. (2015)	p	w	d, h, c	2 leaves, 4 leaves	Leaf	40
guo2009	Guo et al. (2009)	ma	c, w, l	d	Flowering	Leaf (flag)	188
hubner2015	Hübner et al. (2015)	r	w	d	Flag leaf emerged	Spikelets	495
kausar2013	Kausar et al. (2013)	p	c	d	3-d old seedlings	Shoot	32
matsumoto2014	Matsumoto et al. (2014)	ma	c	*	4-d old seedlings	Root, shoot	66
rollins2013	Rollins et al. (2013)	p	c, l	d, h, c	Heading	Leaf	99
shaar-moshe2015	Shaar-Moshe et al. (2015)	me	-	d	-	-	2730
talame2007	Talame et al. (2007)	ma	c	d	4-leaves	Leaf	127
vitamvas2015	Vitamvas et al. (2015)	p	c	d	2-leaves	Crown	68
wang2015	Wang et al. (2015)	p	w, c	d	2-leaves	Leaf	26
wehner2015	Wehner et al. (2015)	QTL	c, l	d	7 days after sowing	-	33
wehner2016	Wehner et al. (2016)	eQTL	c, l	d	7 days after sowing	Leaf	14
weldelboe2012	Wendelboe-Nelson and Morris (2012)	p	c	d	7 days after sowing	Leaf, root	69
worch2011	Worch et al. (2011)	ma	c, w	d	Post-anthesis	Grain	137

We also recorded the number of DE tags found in individual contrasts of our study, which had already been identified in previous studies. These figures for the four main contrasts of our study, Scarlett YI under D, Scarlett YI under MDH, SBCC073 LF under MDH and Scarlett LF under MDH, were 44%, 30%, 56% and 52%, respectively. The largest figures found for the leaf contrasts likely reflect the prevalence of studies which sampled LF tissues.

A total 470 DE isoforms were not found in previous studies, whereas 160 were in just one study and 47 in two. Only 19 DE isoforms were in common in three or more studies. These DE isoforms included several 70kDa and 90kDa heat shock proteins, a S-methyltransferase from S-adenosyl-L-methionine cycle and an N-methyltransferase involved in choline biosynthesis, transcripts related with photosynthesis and carbon fixation, a sucrose synthase, a phosphoglycerate mutase and a triose-phosphate isomerase, a glutathione peroxidase, an ATP synthase and a V-type H+-transporting ATPase subunit, an aspartate kinase, a protein with Potato inhibitor I family domain and a spermidine synthase (Table 5.5).

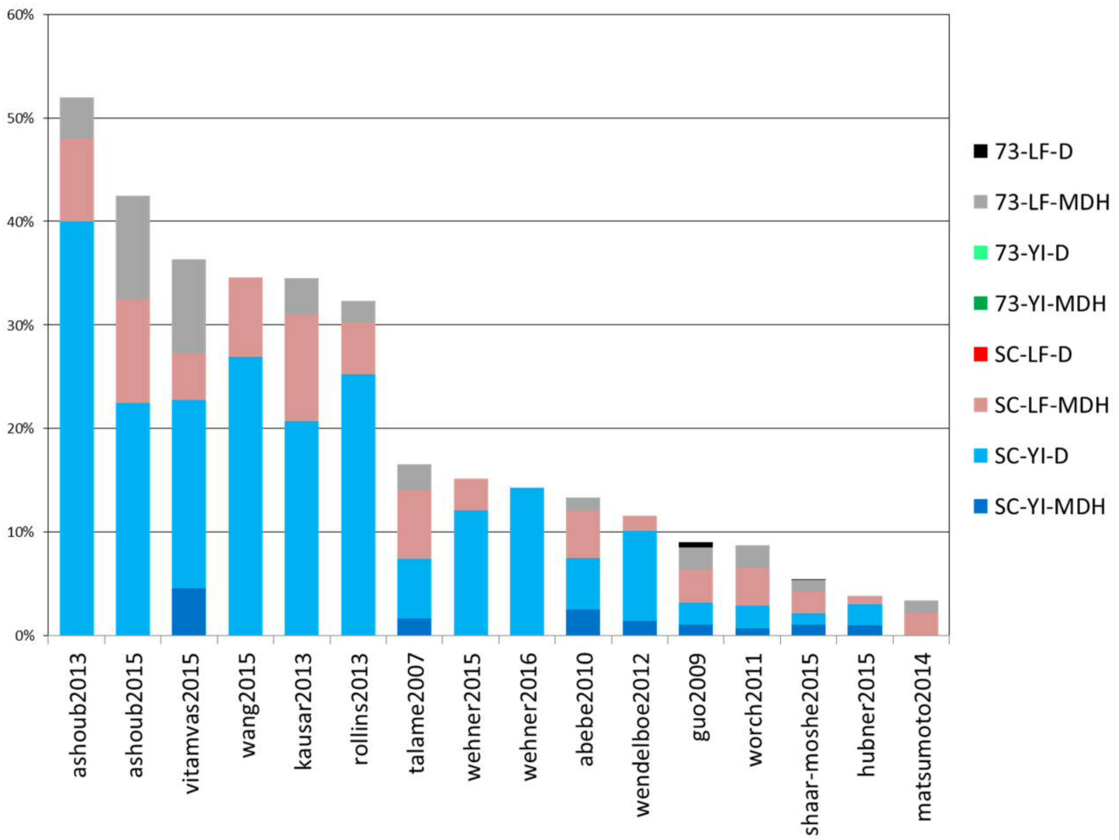


Figure 5.7. Percentage of differentially expressed tags from other studies which were identified in the present work. Bars indicate the percentage of differentially expressed tags (proteins, genes or isoforms) from other studies which were identified in this work. Each color represents the contribution of each contrast.

5.3.9. *Analysis of co-expressed genes*

DE isoforms were clustered based on their expression patterns across samples, with the aim of identifying shared regulatory motifs in their upstream genomic regions. We obtained 23 clusters, 14 of them with more than 10 isoforms. Several clusters contained mostly isoforms from a given contrast while others had mixed DE tags from different treatments.

In order to validate the expression-based gene clusters, they were tested for Gene Ontology (GO) enrichment. Moreover, to test the hypothesis that co-expressed genes might share *cis*-regulatory sequences, their upstream sequences were subjected to motif discovery algorithms and the DNA motifs found were annotated. Finally, the resulting regulatory motifs were compared to the binding predictions of DE expressed TFs identified in this work, trying to link these TFs to clusters of DE tags.

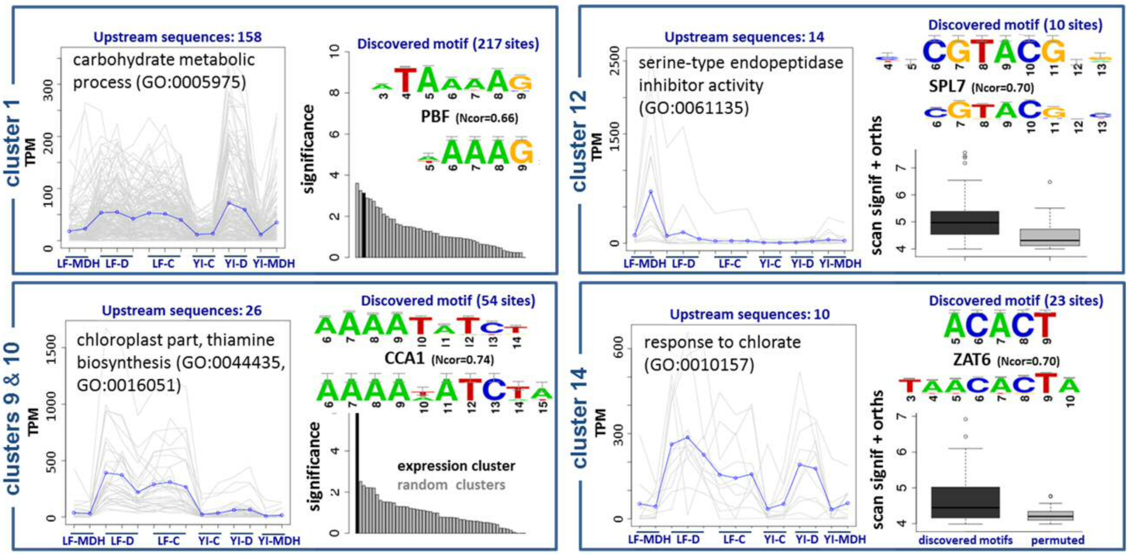


Figure 5.8. Enriched DNA motifs in promoters of differentially co-expressed isoforms. Gene Ontology enrichment and regulatory motifs discovered in 5 clusters of co-expressed isoforms. For each cluster, a plot is shown on the left with the expression profile, where LF and YI correspond to leaf and young inflorescence tissues, and G, D and C to greenhouse, chamber and control replicates, respectively. Regulatory motifs are shown on the right side of each cluster box, with the discovered consensus sequence on top and the most similar motif in footprintDB aligned below. Cluster 10 was found to be very similar to cluster 9, and thus is not shown. The evidence supporting the motifs of clusters 1, 9 and 10 is their significance (black bars) when compared to negative controls (grey bars). Motifs of clusters 12 and 14 (dark boxplots) have higher scores than their shuffled motifs (grey boxplots) when scanned along the cluster upstream sequences and their Brachypodium distachyon orthologues.

The results are summarized in Figure 5.8. Upstream sequences of genes from cluster 1, with functional annotations related to the metabolism of carbohydrates, contain a wtATAAAAGw site, which is similar to motifs of TATA-binding proteins and Dof TFs (Yanagisawa, 2002). We observed a C2C2-Dof TF up-regulated in SBCC073 LF under MDH (see previous sections), although we were not able to identify DNA-binding domains associated to it. Therefore, we cannot confirm whether or not C2C2-Dof protein binds to this motif to regulate genes in cluster 1, but the possibility deserves further investigation. Promoter sequences of genes in clusters 9 and 10, which group mostly transcripts down-regulated in LF under MDH, contain sites identical to the consensus of CCA1/LHY, which belongs to the MYB/SANT family (Green and Tobin, 1999). These sites were independently predicted by oligo-analysis (AAAATATCTy) and dyad-analysis (aAAAkaTCTw), indicating that they are high-confidence predictions. Genes of this cluster are annotated as components of thiamine biosynthesis in the chloroplast. Accordingly, CCA1/LHY, which was up-regulated in SBCC073 and Scarlett samples under MDH, binds to the same motif (aAAATATCTkY). Cluster 12 had predicted yaCGTACGtr *cis*-elements. Genes in this cluster were induced in LF under MDH, and are annotated as heat shock proteins. Finally, genes in cluster 14 are annotated as components of salinity response, and share *cis*-elements of consensus smACACTbm.

Table 5.5. Differentially expressed isoforms found in three or more previous studies. Each row corresponds to a differentially expressed (DE) isoform which was observed in three or more previous studies. Gray-filled cells indicate the contrast in which it was declared as DE (73: SBCC073, SC: Scarlett, YI: young inflorescences, LF: leaves, D: severe drought treatment, MDH: mild drought and heat treatment). The presence of the DE isoform in a given study is highlighted with grey background. Functional annotation: 00425: 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase; 01438: HSP 70kDa; 30291: PSII; 03771: Rubisco activase; 23857: phosphoetanolamine N-methyltransferase; 15018: HSP 70kDa; 46536: sucrose synthase; 49313: Rubisco; 46824: 2,3-bisphosphoglycerate-independent phosphoglycerate mutase; 22980: HSP 90kDa; 03577: glutathione peroxidase; 43420: V-type H+-transporting ATPase; 19971: ATP synthase; 20214: triose-phosphate isomerase; 18227: spermidine synthase; 49597: potato inhibitor I family; 33995: unknown; 01544: HSP 70kDa; 15965: aspartate kinase.

DE isoform	73-LF-MDH	SC-LF-MDH	SC-YI-D	abebe2010	ashoub2013	ashoub2015	guo2009	hubner2015	kausar2013	matsumoto2014	rollins2013	shaar-moshe2015	talame2007	vitamvas2015	wang2015	wehner2015	wehner2016	wendelboe2012	worch2011
00425			▨		▨	▨			▨		▨		▨	▨				▨	
01438		▨		▨	▨	▨	▨					▨	▨						▨
30291			▨	▨					▨		▨			▨				▨	
03771		▨	▨			▨			▨		▨		▨		▨				
23857	▨						▨					▨	▨	▨					▨
15018	▨		▨			▨			▨		▨			▨					
46536			▨		▨									▨		▨			▨
49313		▨			▨						▨				▨			▨	
46824			▨			▨					▨			▨					
22980			▨								▨	▨			▨				
03577			▨					▨						▨		▨			
43420			▨		▨						▨							▨	
19971			▨			▨			▨		▨								
20214			▨	▨					▨		▨								
18227	▨			▨			▨					▨							
49597	▨	▨		▨								▨	▨						
33995		▨		▨			▨					▨							
01544		▨						▨	▨					▨					
15965		▨		▨								▨	▨						

Out of 11 DE TFs, 7 were associated with DNA-binding domains (Table 5.6), including CCA1/LHY (see above), the MYB-related TF of unknown function DE in SBCC073 LF under MDH, the MADS-MIKC up-regulated in Scarlett YI under D (AwRGaAAaww), the B3-ARF TFs induced in Scarlett YI either under D or MDH (yTTGTCtC), the bZIP up-regulated in Scarlett LF under MDH (cayrACACGTgkt) and the AP2/ERF-AP2 down-regulated in SBCC073 LF under MDH (CACrrwTCCCrAkG). It is possible that these genes were in part regulating the changes in gene expression in response to the treatments. However, these could not be linked to the motifs identified in promoters.

Table 5.6. Predicted DNA motifs for differentially expressed transcription factors. DE isoforms which were annotated as TFs in all the contrasts (73: SBCC073, SC: Scarlett, YI: young inflorescences, LF: leaves, D: severe drought treatment, MDH: mild drought and heat treatment) are shown along with their iTAK-annotated Pfam domains, whether they were induced (up) or repressed (dn), the BLASTP E-value of homologous TFs, the sequence motif predicted by footprintDB and the best SwissProt hit, along with its gene name prefixed with acronym of the organism (At: *Arabidopsis thaliana*; Bn: *Brassica napus*; Os: *Oryza sativa* subsp. *japonica*).

Isoform	Pfam	Contrast	Up/Down-regulated	E-value	DNA motif	SwissProt
comp690102_c3	AP2/ERF-AP2	73-LF-MDH	dn	7.00E-79	CACrrwTCCCrAkG	Q8LSN2-BnBBM2
comp700847_c0	B3-ARF	SC-YI-D	up	7.00E-150	yTTGTCtC	Q6YZW0-OsARF21
comp61422_c0	B3-ARF	SC-YI-MDH	up	1.00E-98	yTTGTCtC	Q85983-OsARF11
comp59053_c0	bZIP	SC-LF-MDH	up	7.00E-42	cayrACACGTgkt	-
comp688195_c0	C2C2-Dof	73-LF-MDH	up	-	-	Q93ZL5-AtCDF2
comp67310_c0	CCA1/LHY	SC-YI/LF-MDH	up	0.00E+00	waGATAttt	Q6R0H1-AtLHY
comp53438_c1	CCA1/LHY	73-YI/LF-MDH	up	0.00E+00	waGATAttt	Q6R0H1-AtLHY
comp51250_c2	MYB-related	73-LF-MDH	up	5.00E-46	waGATwttww	-
comp61039_c0	MADS-MIKC	SC-YI-D	up	8.00E-61	AwRGaAAaww	Q6EU39-OsMADS6
comp689206_c7	MYB-related	73-LF-MDH	up	-	-	B3H4X8-AtTCL2
comp66417_c0	MYB-related	SC-LF-MDH	up	-	-	B3H4X8-AtTCL2
comp64196_c0	TIFY	SC-LF-MDH	up	-	-	Q6ES51-OsTIFY6B
comp702448_c0	TUB	SC-LF-MDH	dn	-	-	Q7XSV4-OsTULP7

5.4. Discussion

In this work, ***de novo* assemblies** of Spanish landrace SBCC073 and elite cultivar Scarlett were generated. These assemblies had a larger number of isoforms and genes than current barley references. This could be an effect of sequencing errors and non-coding sequences being expressed, but also of absence of actual transcripts from the references. Nonetheless, the use of all available reference sequences (Morex, Barke, Bowman, Haruna Nijo) led to the confirmation of a substantial percentage of those isoforms, allowing the identification of more assembled isoforms than using any of them separately. This highlights the variability in gene content between genome references, which poses a problem when working with non-reference genotypes as in the present study. In light of this, an advantage of *de novo* assemblies resides in recovering genotype-specific transcripts and in reducing mapping errors produced by polymorphisms. Therefore, using them as reference, as we have done in this study, allows diminishing the proportion of unmapped reads and increasing mapping accuracy, which is essential for gene expression assays. Moreover, we tested three different pipelines for differential expression, and those based on *de novo* assemblies had a better agreement with RT-qPCR results.

Plants from Scarlett and SBCC073 were subjected to severe drought and mild drought combined with heat, during the reproductive stage, and **physiological responses** were measured. Water-stressed plants showed reduced daily loss of water, increased absolute leaf water potential, changes in stomatal conductance, reduced tiller number and reduced spike number, at the end of the experiment. However, there were also differences between the genotypes, indicating different strategies of adaptation to stress. Absolute leaf water potential under severe drought was higher in Scarlett than in SBCC073. Moreover, under combined mild drought and heat, Scarlett exhibited the lowest tiller number, with relative water content comparable to plants under severe drought. In comparison, both measurements were close to that of well-watered plants in SBCC073, under the combined stress. Taken together, these results indicate that Scarlett was more susceptible to mild drought and heat than SBCC073. Experiments carried out in pots, like this, have the disadvantage of not mimicking natural conditions perfectly. On the other hand, experiments in controlled settings actually help to limit variation due to interaction with environment. For instance, rooting depth is kept out of the equation as, although the pots were large, the roots readily explored all soil volume. Hence, potential genotypic differences in soil exploring capacity cannot be held responsible for the genotypic disparities in physiological measurements. Given that soil conditions and water availability were similar for the two genotypes, it can be concluded that SBCC073 was more drought tolerant than Scarlett.

Regarding **gene expression**, the responses to the stresses were specific of each tissue and genotype. Drought almost did not impact SBCC073, whereas the combination of mild drought and heat only affected its leaves. In contrast, gene expression in both Scarlett tissues was strongly altered in the greenhouse, whereas severe drought alone impacted young inflorescences only.

Overall, we found few changes in **leaves** under severe drought stress. Although related studies found more differences in gene expression in leaves, most of them studied early responses and only a few addressed prolonged stresses, as in the present study. Processes involved in plant responses to water deficit are different depending on the temporal scale, being those related with drought resistance and grain production, like phenology adjustment, acclimation, fertility and harvest index, affected by medium- to long-term water scarcity (Passioura, 2004). Severe brief stresses, which are rare in the field, are more related with plant survival (Passioura, 2002). Nonetheless, another study focused on long-lasting water and heat stress (Ashoub et al., 2015) reported many gene expression changes. However, that study involved wild barley seedlings starting at the stage of two leaves. Leaves from adult plants, like the ones in our study, are expected to show different responses to drought than those of seedlings (Blum, 2005). Mature flowering plants could have a more limited transcriptional response to prolonged drought stress due to acclimation or enhanced tolerance, which could be achieved, for example, through selective senescence of older leaves or the development of a deep root system (Blum, 2005; 2009). Studies similar to ours, in which the stress conditions were maintained for a long period, and samples were taken from adult plants, have provided contrasting results. The closest result to ours was found by Rollins et al. (2013), who reported no changes in leaf proteome of mature barley plants under drought stress, but apparent changes due to heat. Others, however, did find differentially expressed genes in flag leaves of adult barley plants (Guo et al., 2009) or changes in protein expression in mature leaves of wheat drought tolerant genotypes (Ford et al., 2011).

In contrast with the drought treatment, we found numerous differentially expressed transcripts in leaves under combined drought and heat stress. There is scarce information about the optimum temperature for barley growth. We can assume that it is close to the one reported for wheat, whose optimum range is between 18 and 23 ºC (Slafer and Rawson, 1995; Porter and Gawith, 1999). A previous study showed that high temperature (25ºC) resulted in rapid progression through reproductive development in long days (Hemming et al., 2012). The temperatures in the greenhouse clearly exceeded that range for most of the experimental period and, therefore, experienced a combination of heat and drought stress, together with a wider range of variation for other environmental factors than control plants, such as a mild powdery mildew infection, presence of phytophagous insects, and variable natural photoperiod.

In such conditions, there were several DE isoforms in common in both genotypes. For example, transcription of CCA1/LHY was induced in Scarlett and SBCC073, in both leaves and young inflorescences. The observed changes in expression of CCA1/LHY might be related to photoperiod rather than to tolerance to stress, given that CCA1/LHY is a component of the circadian clock (Campoli et al., 2012; Deng et al., 2015), and other genes related with circadian clock were also differentially expressed in leaves under mild drought and heat, like HvPRR1/TOC1 (Campoli et al., 2012) and an homolog of Arabidopsis adagio-like protein 3. Even so, CCA1/LHY has been shown to be controlled by heat (Karayekov et al., 2013) and reported to play a key role in abiotic stress (Grundy et al., 2015) in other species. Also, among differentially expressed transcripts in leaves, the most recurrent were those related with polyamines (like spermine and spermidine), which were identified in leaves from both genotypes, under severe drought alone and under drought combined with heat. These are small aliphatic amines which have been associated to numerous stresses in plants, including osmotic stress and heat (Bouchereau et al., 1999), and their knock-out mutants in Arabidopsis show increased susceptibility to drought stress (Yamaguchi et al., 2007). However, their specific roles in drought stressed plants remain obscure (Capell et al., 2004; Do et al., 2013).

Besides that, Scarlett leaves displayed more numerous and functionally diverse differentially expressed transcripts than SBCC073, under mild drought and heat. Despite presenting comparable stomatal conductance to SBCC073, Scarlett showed increased responses in genes related to photosynthesis and carbon fixation metabolism, as well as antioxidant enzymes. Also, this genotype seems to react more actively to pathogen attack under MDH, as seen by the increased biosynthesis of molecules related to defense responses. Another interesting genotypic difference was that glycine betaine biosynthesis was induced in SBCC073, whereas in Scarlett trehalose biosynthesis was induced instead. These two compounds have an alleged osmoprotectant function in organisms. While glycine-betaine is well known in plants, including cereals (Ashraf and Foolad, 2007), trehalose is not common in plants (Majumder et al., 2009). These results point towards the presence of effects on different pathways, and different genotypic strategies to cope with the combination of stresses encountered in the greenhouse treatment.

In **young inflorescences**, there were noticeable changes in gene expression in Scarlett, but almost none in SBCC073, in both stress treatments. As in leaves, this could indicate that Scarlett inflorescences were suffering more from stress than those of SBCC073. A similar interpretation was made by (Hübner et al., 2015), who found a larger proportion of differentially expressed genes for this plant organ in response to stress in sensitive genotypes of wild barley. It is intriguing that inflorescences from Scarlett in the greenhouse showed primarily repressed transcripts, most of them related with metabolism of carbohydrates, reorganization of cell wall and biosynthesis of secondary metabolites. Also, two transcripts involved in indole-3-acetic acid (IAA) biosynthesis were repressed: an L-tryptophan transaminase, which catalyzes the conversion of tryptophan to indole-3-pyruvate, and an indole-3-pyruvate monooxygenase, which yields IAA. This is a key auxin, a phytohormone which regulates many critical developmental processes (Woodward and Bartel, 2005). Barley developing inflorescences are a source of IAA (Wolbang et al., 2004), involved in modulation of stem growth and of floret primordia development (Leopold and Thimann, 1949). We could speculate that this could be an attempt to delay spike development in the face of severe stress.

Differentially expressed transcripts were compared with those from **related studies**. Disparities with other studies partly reflect differences in experimental set up and vegetal material assessed, but other causes are also possible. Interestingly, agreement was better with works based on proteomics than on transcriptomics. This may reflect a statistical bias, due to the choice of strict significance thresholds in our case and in proteomics studies. In fact, the number of differentially expressed proteins reported from proteomics studies was low, which could explain in part the large percentage of coincidences. On the other hand, RNAseq sampling and expression range is different from that of microarrays (Ozturk et al., 2002), which predominated in the gene expression datasets used for comparison, which could favor obtaining results closer to those of proteomics. Actually, there was only one study using RNAseq in the comparison dataset (Hübner et al., 2015), but similarities with it were also scarce. These authors sequenced transcripts from barley immature spikelets subjected to prolonged water stress, which is rather similar to our experiment. However, they worked with wild barley, whereas this study employed a landrace and an elite cultivar. Wild barley holds much more diversity than cultivated types, with considerable variation in physiological and phenotypic characteristics, and presents specific environmental adaptations to stress like temperature and rainfall (Ellis et al., 2000; Hübner et al., 2013). Therefore, it is feasible that the responses to abiotic stresses of wild barley are different to those of cultivated genotypes. In addition, the methodology in that study, an approach based on RGA, was also different from the one adopted here. As mentioned above, we show that such method produced different outcomes than *de novo* assemblies.

Overall agreement between studies was limited, as seen by the few DE isoforms found in common in three or more studies. A previous meta-analysis of gene expression in response to drought (Shaar-Moshe et al., 2015) also detected few common differentially expressed transcripts between studies, although in this case the comparison involved different plant families. This notwithstanding, some processes are recurrently found in drought studies in barley, including ours, independently of the diversity of genotypes and environmental conditions employed. Hence, these could play central roles in the response of barley to abiotic stress. Many of these have already been discussed and reviewed, like the role of polyamines (see above) (Guo et al., 2009; Abebe et al., 2010; Ashoub et al., 2013), proteases (Ford et al., 2011; Ashoub et al., 2013), glycine betaine and other osmoprotectants (Abebe et al., 2010; Ashoub et al., 2013; Ashoub et al., 2015), ascorbic acid (Guo et al., 2009; Wendelboe-Nelson and Morris, 2012; Wang et al., 2015), lipoxygenases (Wendelboe-Nelson and Morris, 2012; Ashoub et al., 2015), aldehyde dehydrogenase (Guo et al., 2009), and also components of photosystem II, carbohydrates metabolism, heat shock proteins, methionine metabolism, or antioxidant enzymes like catalases, which are well known to be involved in stress responses in plants (Krasensky and Jonak, 2012; Marco et al., 2015).

In order to understand the role of differentially expressed genes, it is important to analyze how these genes are orchestrated. Here, this was accomplished by discovering potential *cis*-elements within upstream promoter sequences. Indeed, this study shows that RNAseq can be exploited to obtain biologically relevant conclusions from **co-expressed genes** using currently available barley genomic resources. As a proof of concept, the CCA1/LHY TF, up-regulated in leaves under mild drought and heat, was associated to two clusters of repressed transcripts, which harbor high-confidence CCA1 binding sites in their promoter sequences. Genes in those clusters were related to thiamine biosynthesis in the chloroplast, an early response to stress known to be linked to the circadian clock (Bocobza et al., 2013; Wang et al., 2016). Transcripts from thiamine biosynthesis were repressed in another study assessing barley under drought (Talame et al., 2007), indicating that thiamine could play an important role in drought response, maybe regulating function of enzymes for which it is a cofactor, enhancing tolerance to oxidative damage, or as a signaling molecule in adaptation mechanisms to abiotic stress (Tunc-Ozdemir et al., 2009; Goyer, 2010). Therefore, we were able to associate gene regulation apparently elicited by CCA1/LHY with a previously known stress response linked to regulation of thiamine biosynthesis, through analysis of DNA-binding motifs.

Besides CCA1/LHY, we were able to identify other promoters and DNA-binding affinities of TFs. A motif involved in the regulation of heat shock proteins matches a SBP zinc-finger protein SPL7, which has been described as a TF related to heat stress in rice (Yamanouchi et al., 2002). Genes from another cluster shared a motif whose best hits were Arabidopsis ZAT6, belonging to a family of zinc-finger repressors involved in responses to salt stress (Ciftci-Yilmaz et al., 2007), and AZF2, a C2/H2 zinc-finger which negatively regulates abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions (Kodaira et al., 2011). Moreover, among hundreds of differentially expressed transcripts, only 11 TFs were found in this study (including CCA1/LHY). As an example, we found differential expression of transcripts of a MYB-related protein, whose closest SwissProt homologues are single-repeat R3 MYB TFs from Arabidopsis. These are involved in epidermal cell fate specification, more specifically in regulation of trichome development (Gan et al., 2011). Therefore, this MYB-related protein could have a similar role of that of GT factors in wheat, which ahev been related to drought tolerance and trichome development (Zheng et al., 2016). Some of the TFs identified here have already been associated with abiotic stress in rice or Arabidopsis. In example, we found a bZIP TF whose DNA-binding motif corresponds to that of ABRE (ABA-responsive element) *cis*-element, and thus could be regulating ABA-responsive genes (Nakashima et al., 2014). We also found an AP2/ERF-AP2 TF differentially expressed in SBCC073 leaves. The AP2/ERF is a large family of plant-specific TFs, which includes dehydration-responsive element-binding (DREB) proteins, involved in the activation of drought responsive genes (Mizoi et al., 2012). However, the TF reported here was similar to BABY BOOM genes from *Brassica napus*, in which they promote embryo development (Boutilier et al., 2002). We also found differentially expressed transcripts related to a MADS-MIKC homologue of OsMADS6, related with floral organ and meristem identities in rice (Li et al., 2010), up-regulated in Scarlett developing inflorescences under drought; an uncharacterized MYB-related TF, in SBCC073 leaves only; a C2C2-Dof, similar to Arabidopsis CDF2, which regulates miRNAs involved in control of flowering time (Sun et al., 2015); a TF of the TIFY family, whose members are responsive jasmonic acid and to abiotic stresses (Ye et al., 2009); a TUBBY-like protein (TULP), which have been associated to sensitivity to ABA in Arabidopsis (Lai et al., 2004); and two transcripts matching different B3-ARF (auxin responding factor with B3 domains) from Arabidopsis. Therefore, the responses observed here seem to have only partial overlap with those already described in other plants. For example, NAC TFs (Nakashima et al., 2012) have not been found in this study. Taking advantage of DNA-binding motifs allows linking TFs and groups of co-expressed genes through their common interface, and provides an additional layer of insight on the dynamics of stress responses in plants. Signaling pathways in response to drought in barley, especially depending on type of stress, development stage, tissue and genotype, remain to be deciphered (Gürel et al., 2016), although it is expected that different responses and strategies will be favored in different agronomic contexts.

Well-adapted accession SBCC073 is currently being tested under water stress field conditions in populations derived from crosses, to search for QTL that control agronomic traits. The catalog of sequence transcripts and expression profiles from the current study will complement this population-based approach to unravel the genetic control of drought responses which impact grain yield.

5.5. References

Abebe, T., Melmaiee, K., Berg, V., and Wise, R.P. (2010). Drought response in the spikes of barley: gene expression in the lemma, palea, awn, and seed. *Funct. Integr. Genomics* 10, 191-205. doi: 10.1007/s10142-009-0149-4

Alexa, A., and Rahnenfuhrer, J. (2016). "topGO: enrichment analysis for Gene Ontology. R package version 2.24.0". Available online at: https://bioconductor.org/packages/release/bioc/html/topGO.html (Accessed June 1, 2016)

Andersen, C.L., Ledet-Jensen, J., and Ørntoft, T.F. (2004). Normalization of real-time quantitative RT-PCR data: a model based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res.* 64, 5245-5250. doi: 10.1158/0008-5472.can-04-0496

Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data". Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed April 24, 2014)

Araus, J.L., Slafer, G.A., Reynolds, M.P., and Royo, C. (2002). Plant breeding and drought in C3 cereals: what should we breed for? *Ann. Bot.* 89, 925-940. doi: 10.1093/aob/mcf049

Ashoub, A., Baeumlisberger, M., Neupaertl, M., Karas, M., and Bruggemann, W. (2015). Characterization of common and distinctive adjustments of wild barley leaf proteome under drought acclimation, heat stress and their combination. *Plant Mol. Biol.* 87, 459-71. doi: 10.1007/s11103-015-0291-4

Ashoub, A., Beckhaus, T., Berberich, T., Karas, M., and Bruggemann, W. (2013). Comparative analysis of barley leaf proteome as affected by drought stress. *Planta* 237, 771-81. doi: 10.1007/s00425-012-1798-4

Ashraf, M., and Foolad, M.R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environ. Exper. Bot.* 59, 206-216. doi: 10.1016/j.envexpbot.2005.12.006

Barnabas, B., Jager, K., and Feher, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 31, 11-38. doi: 10.1111/j.1365-3040.2007.01727.x

Blum, A. (2005). Drought resistance, water-use efficiency, and yield potential - are they compatible, dissonant, or mutually exclusive? *Aust. J. Agric. Res.* 56, 1159-118. doi: 10.1071/AR05069

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res.* 112, 119-123. doi: 10.1016/j.fcr.2009.03.009

Bocobza, S.E., Malitsky, S., Araújo, W.L., Nunes-Nesi, A., Meir, S., Shapira, M., et al. (2013). Orchestration of thiamin biosynthesis and central metabolism by combined action of the thiamin pyrophosphate riboswitch and the circadian clock in Arabidopsis. *Plant Cell* 25, 288-307. doi: 10.1105/tpc.112.106385

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-20. doi: 10.1093/bioinformatics/btu170

Bouchereau, A., Aziz, A., Larher, F., and Martin-Tanguy, J. (1999). Polyamines and environmental challenges: recent development. *Plant Sci.* 140**,** 103-125. doi: 10.1016/S0168-9452(98)00218-0

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14**,** 1737-1749. doi: 10.1105/tpc.001941

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnol.* 34**,** 525-527. doi: 10.1038/nbt.3519

Campoli, C., Shtaya, M., Davis, S.J., and Von Korff, M. (2012). Expression conservation within the circadian clock of a monocot: natural variation at barley *Ppd-H1* affects circadian expression of flowering time genes, but not clock orthologs. *BMC Plant Biol.* 12. doi: 10.1186/1471-2229-12-97

Cantalapiedra, C.P., Boudiar, R., Casas, A.M., Igartua, E., and Contreras-Moreira, B. (2015). BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol. Breeding* 15. doi: 10.1007/s11032-015-0253-1

Capell, T., Bassie, L., and Christou, P. (2004). Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proc. Natl. Acad. Sci. U.S.A.* 101**,** 9909-9914. doi: 10.1073/pnas.0306974101

Cattivelli, L., Rizza, F., Badeck, F.-W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., et al. (2008). Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Res.* 105**,** 1-14. doi: 10.1016/j.fcr.2007.07.004

Ceccarelli, S. (1994). Specific adaptation and breeding for marginal conditions. *Euphytica* 77**,** 205-219. doi: 10.1007/BF02262633

Ceccarelli, S., Grando, S., and Impiglia, A. (1998). Choice of selection strategy in breeding barley for stress environments. *Euphytica* 103**,** 307-318. doi: 10.1023/A:1018647001429

Ciftci-Yilmaz, S., Morsy, M.R., Song, L., Coutu, A., Krizek, B.A., Lewis, M.W., et al. (2007). The EAR-motif of the Cys2/His2-type zinc finger protein Zat7 plays a key role in the defense response of Arabidopsis to salinity stress. *J. Biol. Chem.* 282**,** 9260-9268. doi: 10.1074/jbc.M611093200

Close, T.J., Wanamaker, S., Roose, M.L., and Lyon, M. (2007). "HarvEST: an EST database and viewing software", in *Plant bioinformatics: methods and protocols,* ed. D. Edwards, (Totowa, New Jersey: Humana Press), 161-77. doi: 10.1007/978-1-59745-535-0_7

Contreras-Moreira, B., Castro-Mondragon, J.A., Rioualen, C., Cantalapiedra, C.P., and Van Helden, J. (2016). "RSAT::Plants: Motif discovery within clusters of upstream sequences in plant genomes", in *Plant synthetic promoters: methods and protocols,* ed. R. Hehl, accepted.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Chang.* 4**,** 287-291. doi: 10.1038/NCLIMATE2153

Dawson, I.K., Russell, J., Powell, W., Steffenson, B., Thomas, W.T.B., and Waugh, R. (2015). Barley: a translational model for adaptation to climate change. *New Phytol.* 206**,** 913-31. doi: 10.1111/nph.13266

Deng, W., Clausen, J., Boden, S., Oliver, S.N., Casao, M.C., Ford, B., et al. (2015). Dawn and dusk set states of the circadian oscillator in sprouting barley (Hordeum vulgare) seedlings. *PLoS ONE* 10**,** e0129781. doi: 10.1371/journal.pone.0129781

Do, P.T., Degenkolbe, T., Erban, A., Heyer, A.G., Kopka, J., Köhl, K.I., et al. (2013). Dissecting rice polyamine metabolism under controlled long-term drought stress. *PLoS ONE* 8**,** e60325. doi: 10.1371/journal.pone.0060325

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K., and Ortiz, R. (2016). Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci.* 21**,** 31-42. doi: 10.1016/j.tplants.2015.10.012

Ellis, R.P., Forster, B.P., Robinson, D., Handley, L.L., Gordon, D.C., Russell, J.R., et al. (2000). Wild barley: a source of genes for crop improvement in the 21st century? *J. Exp. Bot.* 51**,** 9-17. doi: 10.1093/jexbot/51.342.9

Fischbeck, G. (2003). "Diversification through breeding", in *Diversity in barley (Hordeum vulgare),* eds. R. Von Bothmer, T. Van Hintum, H. Knüpffer & K. Sato, (Amsterdam: Elsevier Science B.V.), 29-52.

Fischer, R.A., and Turner, N.C. (1978). Plant productivity in the arid and semiarid zones. *Annu. Rev. Plant Physiol.* 29**,** 277-317. doi: 10.1146/annurev.pp.29.060178.001425

Ford, K.L., Cassin, A., and Bacic, A. (2011). Quantitative proteomic analysis of wheat cultivars with differing drought stress tolerance. *Front. Plant Sci.* 2. doi: 10.3389/fpls.2011.00044

Gan, L., Xia, K., chen, J., and Shucai, W. (2011). Functional characterization of TRICHOMELESS2, a new single-repeat R3 MYB transcription factor in the regulation of trichome patterning in Arabidopsis. *BMC Plant Biol.* 11**,** 176. doi: 10.1186/1471-2229-11-176

Goyer, A. (2010). Thiamine in plants: aspects of its metabolism and functions. *Phytochemistry* 71**,** 1615-24. doi: 10.1016/j.phytochem.2010.06.022

Green, R.M., and Tobin, E.M. (1999). Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 96**,** 4176-4179. doi: 10.1073/pnas.96.7.4176

Grundy, J., Stoker, C., and Carre, I.A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00648

Guo, P., Baum, M., Grando, S., Ceccarelli, S., Bai, G., Li, R., et al. (2009). Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *J. Exp. Bot.* 60**,** 3531-44. doi: 10.1093/jxb/erp194

Gürel, F., Özturk, Z.N., Uçarli, C., and Rosellini, D. (2016). Barley genes as tools to confer abiotic stress tolerance in crops. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01137

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8**,** 1494-512. doi: 10.1038/nprot.2013.084

Hazelhurst, S., and Liptak, Z. (2011). KABOOM! A new suffix array based algorithm for clustering expression data. *Bioinformatics* 27**,** 3348-55. doi: 10.1093/bioinformatics/btr560

Hemming, M.N., Walford, S.A., Fieg, S., Dennis, E.S., and Trevaskis, B. (2012). Identification of high-temperature-responsive genes in cereals. *Plant Physiol.* 158**,** 1439-1450. doi: 10.1104/ pp.111.192013

Hübner, S., Bdolach, E., Ein-Gedy, S., Schmid, K.J., Korol, A., and Fridman, E. (2013). Phenotypic landscapes: phenological patters in wild and cultivated barley. *J. Evol. Biol.* 26**,** 163-174. doi: 10.1111/jeb.12043

Hübner, S., Korol, A.B., and Schmid, K.J. (2015). RNA-Seq analysis identifies genes associated with differential reproductive success under drought-stress in accessions of wild barley *Hordeum spontaneum*. *BMC Plant Biol.* 15**,** 134. doi: 10.1186/s12870-015-0528-z

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463**,** 763-8. doi: 10.1038/nature08747

Intergovernmental Panel on Climate Change (IPCC), (2014). "Climate change 2014 synthesis report". Available online at: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_All_Topics.pdf (Accessed July 12, 2016)

Kamei, C.L.A., Boruc, J., Vandepoele, K., Van den Daele, H., Maes, S., Russinova, E., et al. (2008). The PRA1 gene family in Arabidopsis. *Plant Physiol.* 147**,** 1735-1749. doi: 10.1104/pp. 108.122226

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44**,** D457-D462. doi: 10.1093/nar/gkv1070

Karayekov, E., Sellaro, R., Legris, M., Yanovsky, M.J., and Casal, J.J. (2013). Heat shock-induced fluctuations in clock and light signaling enhance phytochrome B-mediated Arabidopsis deetiolation. *Plant Cell* 25**,** 2892-2906. doi: 10.1105/tpc.113.114306

Keating, B.A., Carberry, P.S., Bindraban, P.S., Asseng, S., Meinke, H., and Dixon, J. (2010). Eco-efficient agriculture: concepts, challenges and opportunities. *Crop Sci.* 50**,** S-109-S-119. doi: 10.2135/cropsci2009.10.0594

Kodaira, K.S., Qin, F., Tran, L.S., Maruyama, K., Kidokoro, S., Fujita, Y., et al. (2011). Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant Physiol.* 157**,** 742-756. doi: 10.1104/pp.111.182683

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35**,** W345-9. doi: 10.1093/nar/gkm391

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63**,** 1593-608. doi: 10.1093/jxb/err460

Lai, C., Lee, C., Chen, P., Wu, S., Yang, C., and Shaw, J. (2004). Molecular analyses of the Arabidopsis TUBBY-like protein gene family. *Plant Physiol.* 134**,** 1586-1597. doi: 10.1104/pp. 103.037820

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9**,** 357-359. doi: 10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10**,** R25. doi: 10.1186/gb-2009-10-3-r25

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* 25**,** 1-18. doi: 10.18637/jss.v025.i01

Leopold, A.C., and Thimann, K.V. (1949). The effect of auxin on flower initiation. *Am. J. Bot.* 36**,** 342-347.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12**,** 323. doi: 10.1186/1471-2105-12-323

Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H., et al. (2010). The AGL6-like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Res.* 20**,** 299-313. doi: 10.1038/cr.2009.143

Liu, Y., Schroder, J., and Schmidt, B. (2013). Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics* 29**,** 308-15. doi: 10.1093/bioinformatics/bts690

Luck, M., Landis, M., and Gassert, F. 2015. "Aqueduct water stress projections: decadal projections of water supply and demand using CMIP5 GCMs". Technical Note. Washington, D. C.: World Resources Institute. Available: wri.org/publication/aqueduct-water-stress-projections (Accessed July 4, 2016)

Majumder, A.L., Sengupta, S., and Goswami, L. (2009). "Osmolyte regulation in abiotic stress.", in *Abiotic stress adaptation in plants: physiological, molecular and genomic foundation.,* eds. A. Pareek, S.K. Sopory & H.J. Bohnert, (The Netherlands: Springer), 349-370.

Marco, F., Bitrián, M., Carrasco, P., Rajam, M.V., Alcázar, R., and Tiburcio, A.F. (2015). "Genetic Engineering Strategies for Abiotic Stress Tolerance in Plants", in *Plant biology and biotechnology,* eds. B. Bahadur, L. Sahijram, M.V. Rajam & K.V. Krishnamurthy, (India: Springer), 579-609. doi: 10.1007/978-81-322-2283-5_29

Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., et al. (2011). Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.* 156, 20-8. doi: 10.1104/pp.110.171579

Mayer, K.F.X., Rogers, J., Doležel, J., Pozniak, C., Eversole, K., Feuillet, C., et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* 345. doi: 10.1126/science.1251788

Mayer, K.F.X., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711-6. doi: 10.1038/nature11543

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., et al. (2015). RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res.* 43, W50-W56. doi: 10.1093/nar/gkv362

Mickelbart, M.V., Hasegawa, P.M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Rev. Genet.* 16, 237-251. doi: 10.1038/nrg3901

Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15-19. doi: 10.1016/j.tplants.2005.11.002

Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819, 86-96. doi: 10.1016/j.bbagrm.2011.08.004

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621-628. doi: 10.1038/nmeth.1226

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819, 97-103. doi: 10.1016/j.bbagrm.2011.10.005

Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00170

Nussbaumer, T., Martis, M.M., Roessner, S.K., Pfeifer, M., Bader, K.C., Sharma, S., et al. (2013). MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* 41, D1144-51. doi: 10.1093/nar/gks1153

Ozturk, Z.N., Talamé, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., et al. (2002). Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol Biol* 48, 551-573. doi: 10.1023/A:1014875215580

Passioura, J.B. (2002). Environmental biology and crop improvement. *Funct. Plant Biol.* 29, 537-546. doi: 10.1071/FP02020

Passioura, J.B. (2004). "Increasing crop productivity when water is scarce - from breeding to field management.", in *New directions for a diverse planet,* eds. R.A. Fischer, N. Turner, J.

Angus, L. Mcintyre, M. Robertson, A. Borrell & D. Lloyd, (Brisbane, Australia: Proc. 4th Int. Crop Science Congress).

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29**,** e45. doi: 10.1093/nar/29.9.e45

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2016). Differential analysis of RNA-Seq incorporating quantification uncertainty. *BioRxiv*. doi: 10.1101/058164

Plant Metabolic Network (2016). "PlantCyc database". Available online at: http://www.plantcyc.org (Accessed June 20, 2016)

Porter, J.R., and Gawith, M. (1999). Temperatures and the growth and development of wheat: a review. *Eur. J. Agron.* 10**,** 23-36. doi: 10.1016/S1161-0301(98)00047-1

Pswarayi, A., van Eeuwijk, F.A., Ceccarelli, S., Grando, S., Comadran, J., Russell, J., et al. (2008). Barley adaptation and improvement in the Mediterranean basin. *Plant Breeding* 127**,** 554-560. doi: 10.1111/j.1439-0523.2008.01522.x

R Development Core Team (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rivers, J., Warthmann, N., Pogson, B.J., and Borevitz, J.O. (2015). Genomic breeding for food, environment and livelihoods. *Food Secur.* 7**,** 375-382. doi: 10.1007/s12571-015-0431-3

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26**,** 139-40. doi: 10.1093/bioinformatics/btp616

Rollins, J.A., Habte, E., Templer, S.E., Colby, T., Schmidt, J., and von Korff, M. (2013). Leaf proteome alterations in the context of physiological and morphological responses to drought and heat stress in barley (Hordeum vulgare L.). *J. Exp. Bot.* 64**,** 3201-12. doi: 10.1093/jxb/ert158

Ryan, J., Abdel Monem, M., and Amri, A. (2009). Nitrogen fertilizer response of some barley varieties in semi-arid conditions in Morocco. *J. Agricult. Sci. Technol.* 11**,** 227-236.

Saini, H.S., and Westgate, M. (1999). Reproductive development in grain crops during drought. *Adv. Agron.* 68**,** 59-96. doi: 10.1016/S0065-2113(08)60843-3

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J., and Itoh, T. (2016). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Res.* 23**,** 21-28. doi: 10.1093/dnares/dsv033

Sayed, M.A., Schumann, H., Pillen, K., Naz, A.A., and Léon, J. (2012). AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (*Hordeum vulgare L.*). *BMC Genet.* 13. doi: 10.1186/1471-2156-13-61

Sebastian, A., and Contreras-Moreira, B. (2014). footprintDB: a database of transcription factors with annotated cis elements and binding interfaces. *Bioinformatics* 30**,** 258-65. doi: 10.1093/bioinformatics/btt663

Shaar-Moshe, L., Hubner, S., and Peleg, Z. (2015). Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biol.* 15, 111. doi: 10.1186/s12870-015-0493-6

Slafer, G.A., and Rawson, H.M. (1995). Base and optimum temperatures vary with genotype and stage of development in wheat. *Plant Cell Environ.* 18, 671-679. doi: 10.1111/j.1365-3040.1995.tb00568.x

Sun, Z., Guo, T., Liu, Y., Liu, Q., and Fang, Y. (2015). The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS Genet.* 11, e1005598. doi: 10.1371/journal.pgen.1005598

Talame, V., Ozturk, N.Z., Bohnert, H.J., and Tuberosa, R. (2007). Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis. *J. Exp. Bot.* 58, 229-40. doi: 10.1093/jxb/erl163

Tello-Ruiz, M., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., et al. (2016). Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 44, D1133-D1140. doi: 10.1093/nar/gkv1179

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46-53. doi: 10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562-578. doi: 10.1038/nprot.2012.016

Tunc-Ozdemir, M., Miller, G., Song, L., Kim, J., Sodek, A., Koussevitzky, S., et al. (2009). Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. *Plant Physiol.* 151, 421-32. doi: 10.1104/pp.109.140046

Turner, N.C. (2004). Sustainable production of crops and pastures under drought in a Mediterranean environment. *Ann. Appl. Biol.* 144, 139-147.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, research0034.1-research0034.11.

Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522-530. doi: 10.1016/j.tibtech.2009.05.006

Varshney, R.K., Terauchi, R., and McCough, S.R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12, e1001883. doi: 10.1371/journal.pbio.1001883

Wang, L., Ye, X., Liu, H., Liu, X., Wei, C., Huang, Y., et al. (2016). Both overexpression and suppression of an *Oryza sativa* NB-LRR-like gene OsLSR result in autoactivation of immune response and thiamine accumulaion. *Sci. Rep.* 6. doi: 10.1038/srep24079

Wang, N., Zhao, J., He, X., Sun, H., Zhang, G., and Wu, F. (2015). Comparative proteomic analysis of drought tolerance in the two contrasting Tibetan wild genotypes and cultivated genotype. *BMC Genomics* 16**,** 432. doi: 10.1186/s12864-015-1657-3

Wei, T., and Simko, V. (2014). "Corrplot: visualization of a correlation matrix". Available online at: http://CRAN.R-project.org/package=corrplot (Accessed September 24, 2014)

Wendelboe-Nelson, C., and Morris, P.C. (2012). Proteins linked to drought tolerance revealed by DIGE analysis of drought resistant and susceptible barley varieties. *Proteomics* 12**,** 3374-85. doi: 10.1002/pmic.201200154

Wolbang, C.M., Chandler, P.M., Smith, J.J., and Ross, J.J. (2004). Auxin from the developing inflorescence is required for the biosynthesis of active gibberellins in barley stems. *Plant Physiol.* 134**,** 769-776. doi: 10.1104/pp.103.030460

Woodward, A.W., and Bartel, B. (2005). Auxin: regulation, action, and interaction. *Ann. Bot.* 95**,** 707-735. doi: 10.1093/aob/mci083

Yahiaoui, S., Cuesta-Marcos, A., Gracia, M.P., Medina, B., Lasa, J.M., Casas, A.M., et al. (2014). Spanish barley landraces outperform modern cultivars at low-productivity sites. *Plant Breeding* 133**,** 218–226. doi: 10.1111/pbr.12148

Yamaguchi, K., Takahashi, Y., Berberich, T., Imai, A., Takahashi, T., Michael, A.J., et al. (2007). A protective role for the polyamine spermine against drought stress in Arabidopsis. *Biochem. Biophys. Res. Commun.* 352**,** 486-490. doi: 10.1016/j.bbrc.2006.11.041

Yamanouchi, U., Yano, M., Lin, H., Ashikari, M., and Yamada, K. (2002). A rice spotted leaf gene, Spl7, encodes a heat stress transcription factor protein. *Proc. Natl. Acad. Sci. U.S.A.* 99**,** 7530-7535. doi: 10.1073/pnas.112209199

Yanagisawa, S. (2002). The Dof family of plant transcription factors. *Trends Plant Sci.* 7**,** 555-560. doi: 10.1016/S1360-1385(02)02362-2

Ye, H., Du, H., Tang, N., Li, X., and Xiong, L. (2009). Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. *Plant Mol. Biol.* 71**,** 291-305. doi: 10.1007/s11103-009-9524-8

Zadoks, J.C., Chang, T.T., and Konzak, C.F. (1974). A decimal code for the growth stages of cereals. *Weed Res* 14**,** 415-421. doi: 10.1111/j.1365-3180.1974.tb01084.x

Zheng, X., Liu, H., Ji, H., Wang, Y., Dong, B., Qiao, Y., et al. (2016). The wheat GT factor TaGT2L1D negatively regulates drought tolerance and plant development. *Sci. Rep.* 6. doi: 10.1038/srep27042

6. General discussion

The specific research objectives have been discussed in the previous chapters. This chapter, instead, is a personal reflection on aspects related to the main objective of the thesis, from a general perspective, with examples from my work. In particular, here I discuss briefly about the adoption of high-throughput sequencing (HTS) and bioinformatics, by a research group focused on plant genetic variability and breeding.

A requirement to work with HTS data is the incorporation of new computational infrastructures, and human resources to i) store, ii) handle, and iii) process the data (Marx, 2013). HTS technologies yield enormous amounts of data (Schadt et al., 2010). This fact, by itself, poses a challenge to research groups willing to take advantage of them. Storing massive amounts of data, obtained with substantial effort and cost, also involves securing its integrity, while providing access to it to researchers working concurrently, through computer networks. Therefore, the knowledge to set up, manage, and process files stored in computer network architectures is a requirement to work with HTS data. Also, analyzing such amounts of data in a reasonable time requires powerful computational equipment, including high-performance multi-core processors, complemented with fast and large read-access memory (RAM) modules. To take advantage of this hardware, programming skills are important, as is the ability to test, choose and run the appropriate tools for each specific analysis. These are the reasons why many laboratories are investing in dedicated computational infrastructures, and increasingly demanding professionals with the right skills for the computational analysis of HTS data. This is the case of the research institutes I visited during this PhD project, including the Bioinformatics Unit at NIAS (Tsukuba, Japan), the Bioinformatics group at IPK (Gatersleben, Germany), and several groups at CNAG (Barcelona). In our local group at EEAD-CSIC (Zaragoza), the adoption of HTS methods has propelled the acquisition of new computing servers and workstations, network-attached storage (NAS) hard drives, and also Web servers to provide access, to the research community, to the results of our research. As an example of the volume of data which is produced by HTS approaches, taking into account only the experiments described in this work, 1.32 billion reads were produced, to a total of 267 Giga bases sequenced, and a raw data load of 200 Gigabytes (compressed). Note that downstream analyses of these data increase significantly the total size of files to be stored. Regarding data processing capabilities, mapping of RNAseq reads, with standard HTS software, requires processing them in parallel, in multi-core processors with large memory availability, and *de novo* assembly of barley transcriptomes can take from hours to days. In summary, **HTS data requires incorporating IT infrastructure and resources** suitable to locate, integrate, and provide proper access to it (Howe et al., 2008).

Results from analyses of HTS data are hardly interpretable without a basic knowledge about the underlying algorithms of bioinformatics tools. This includes an understanding of the purpose of software parameters, and the outcomes obtained by modifying them. As an example, in the exome capture data analyzed in this work, tuning of the parameters of the Trinity *de novo* assembler was critical to disentangle heterozygous mappings into independent contigs. This led afterwards to the identification of candidate genes for the disease resistance under study. This requirement, of using specific strategies to achieve particular goals, is the reason why general purpose tools are so scarce, and are usually nested in purpose-specific pipelines, built by bioinformatics professionals (Chang, 2015). In this regard, this PhD thesis highlights, in every chapter, that **taking the most of each HTS dataset requires adapting the tools and approaches to the specific questions being addressed.**

Moreover, the analysis and interpretation of genomic data require enlarging and strengthening the collaborative approaches that are already commonplace in Biology (Ward et al., 2012). In this PhD thesis, taking advantage of the sequence-enriched physical and genetic maps of barley required thorough study of all the associated resources, through continuous crosstalk with geneticists and breeders. This cooperative effort drove the development of the integrated tool BARLEYMAP, and more generally, has been essential to this work, and to the growing expertise of the group in the use of HTS data. For example, it was indispensable for me in order to learn about basic terms used in crop breeding and molecular plant biology, like yield, drought tolerance, field capacity, plant fungal pathogens, disease resistance genes, mapping populations, linkage or association mapping, genetic markers, PCR techniques, RNA isolation and quality validation, spike development, flowering time, vernalization, to mention a few. In the opposite direction, it also was important to be able to communicate the necessary bioinformatics terms, like read mapping, *de novo* assembly, reference sequence, Web server, linux scripting, programming languages, and many more. From this point of view, this work reports also about the progress made by bidirectional interactions **with researchers from different, but complementary, backgrounds, as a necessary and beneficial requirement to take advantage of HTS data**.

Complex, tailor-made bioinformatics analyses also represent a challenge when publishing results. Reproducing such analyses and pipelines should be straightforward to any laboratory with the proper computational infrastructures, even to those without extensive bioinformatics know-how, given that suitable scripts and computer programs are provided, along with detailed protocols, software versions and specific parameters used. While there is progress in this direction (Ince et al., 2012; Macdonald and Boutros, 2016), this is not always a priority neither for publishers and reviewers nor for the scientific community foreign to bioinformatics. Here, we made public the code implementing BARLEYMAP, besides providing access to the software through both a Web and a standalone application. Also, the implementation and server deployment was described in the paper, and further documentation explaining the operation and configuration of the application was distributed along the code. Regarding the exome capture experiment, the whole bioinformatics procedures were included as supplementary material of the main paper, to ensure that all the relevant information, needed to reproduce the experiments, was published. Also, scripts with specific pieces of code were published. Raw reads from both the exome sequencing and from the RNAseq experiment were uploaded to public repositories. Moreover, *de novo* assembled transcriptomes and detailed analysis pipelines will be made accessible to the research community once the RNAseq manuscript is accepted. In conclusion, both publication boards and **the scientific community should keep pushing forward towards comprehensively documented, and reproducible, Biology science**.

In summary, HTS has brought many new possibilities, but also new challenges, to plant research. To take full advantage of them, appropriate computational infrastructures and human abilities must be incorporated to research groups. Bioinformatics plays an essential role to provide appropriate solutions to specific experimental designs and research goals in Biology. To live up to these high expectations, crosstalk between bioinformatics and genetics must be fostered, adapting the new genomic resources to specific needs. Due to this specificity, publishing in detail the methods and approaches carried out is critical to ensure the reproducibility and validity of HTS results, and, in general, of current biological science.

References

Chang, J. (2015). Reward bioinformaticians. *Nature* 520, 151-152.

Howe, D., Constanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., et al. (2008). The future of biocuration. *Nature* 455, 47-50.

Ince, D.C., Hatton, L., and Graham-Cumming, J. (2012). The case for open computer programs. *Nature* 482, 485-488.

Macdonald, J.M., and Boutros, P.C. (2016). Log::ProgramInfo: a Perl module to collect and log data for bioinformatics pipelines. *Source Code Biol. Med.* 11.

Marx, V. (2013). The big challenges of Big Data. *Nature* 498, 255-260.

Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., and Nolan, G.P. (2010). Computational solutions to large-scale data management and analysis. *Nat Rev Genet* 11, 647-57. doi: 10.1038/nrg2857

Ward, R.M., Schmieder, R., Highnam, G., and Mittelman, D. (2012). Big data challenges and opportunities in high-throughput sequencing. *Syst Biomed* 1, 29-34.

7. Conclusions

The conclusions obtained from the development of a software tool, BARLEYMAP, to address the first objective of this work, are:

1) The implementation of BARLEYMAP, a software tool which combines several alignment algorithms, all the available reference sequences of barley, and results from several barley sequence-enriched maps, provides position for a larger number of genetic markers than using those resources separately.

2) The accuracy of the positions of genetic markers obtained with BARLEYMAP is comparable to that of barley genetic maps, suggesting that such positions could be used without the requirement of calculating a genetic map for subsequent analyses.

3) The integration of the resources published alongside the sequence-enriched physical and genetic maps of barley into a single bioinformatics resource, BARLEYMAP, provides easy access to them.

In relation with the second objective of this work, using exome capture and sequencing to accelerate gene cloning, we conclude that:

4) Exome capture and sequencing provides additional information for fine mapping. In the case of the powdery mildew resistance QTL studied here, sequencing of just three informative recombinant lines allowed increasing the density of markers within the QTL.

5) Fine mapping through existing sequencing methods, including exome capture, is hampered by the current state of barley reference sequences, fragmented and poorly annotated, especially in regions containing repetitive sequences, and clusters of closely related genes and pseudogenes.

6) The combination of the different genomic sequence resources available for barley provides a more comprehensive reference than using them separately, improving the possibility of success of approaches like fine mapping.

7) Heterozygous mappings are apparently produced by erroneous mapping of reads to paralogous loci, collapsing into the same place in the reference sequence. These features indicate the presence of polymorphic members of gene families, and their polymorphisms can be disentangled to identify candidate genes, including those absent in the reference sequence.

8) A cluster of closely related genes, encoding NBS-LRR proteins, is co-located with the powdery mildew resistance QTL from Spanish barley landrace SBCC097, and includes a candidate gene for the resistance, absent from cultivars Morex and Plaisant, and expressed in Spanish barley landrace SBCC097.

Finally, several conclusions can be obtained from the abiotic stress experiment, carried out to address the third objective:

9) *De novo* assembled transcriptomes can be used as valid reference sequences. In our work, annotated transcriptomes of landrace SBCC073 and cultivar Scarlett, were effectively used to calculate gene expression, and could be incorporated to the pool of reference sequences available for barley.

10) Cultivar Scarlett is more susceptible to drought and heat stress than landrace SBCC073, as indicated by physiological and agronomical measurements, and also by the degree of changes in gene expression observed in adult leaves and developing inflorescences.

11) Common biological processes are found in response to drought across experiments and genotypes from different studies. In contrast, particular genes with altered gene expression are rarely conserved, due to differences in experimental setups, biological material used, or noise. Further studies would benefit from focusing on processes rather than on particular genes.

12) Expression patterns of genes are correlated with biological processes. Analysis of promoter sequences of co-expressed genes can be performed with currently available barley genomic resources, and lead to the discovery of shared regulatory elements. Some of these cis-elements can be linked to drought-responsive candidate transcription factors.

8. Annexes

- **Cantalapiedra CP**, Boudiar R, Casas AM, Igartua E, Contreras-Moreira B. 2015. BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol. Breeding* 35:13.

- **Cantalapiedra CP**, Contreras-Moreira B, Silvar C, Perovic D, Ordon F, Gracia MP, Igartua E, Casas AM. 2016. A cluster of Nucleotide-Binding Site - Leucine-Rich Repeat genes resides in a barley powdery mildew resistance Quantitative Trait Loci on 7HL. *Plant Genome*. 9(2):14.

- **Cantalapiedra CP**, García-Pereira MJ, Gracia MP, Igartua E, Casas AM, Contreras-Moreira B. 2016. Large differences in gene expression between elite barley cultivar Scarlett and a Spanish landrace under drought and heat stress. *Front. Plant Sci.* (under review)

-

Mol Breeding (2015) 35:13
DOI 10.1007/s11032-015-0253-1

BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley

Carlos P. Cantalapiedra · Ridha Boudiar · Ana M. Casas · Ernesto Igartua · Bruno Contreras-Moreira

Received: 12 April 2014 / Accepted: 4 October 2014 / Published online: 20 January 2015

Abstract The BARLEYMAP pipeline was designed to map both genomic sequences and transcripts against sequence-enriched genetic/physical frameworks, with plant breeders as the main target users. It reports the most probable genomic locations of queries after merging results from different resources so that diversity obtained from re-sequencing experiments can be exploited. In addition, the application lists surrounding annotated genes and markers, facilitating downstream analyses. Pre-computed marker datasets can also be created and browsed to facilitate searches and cross referencing. Performance is evaluated by mapping two sets of long transcripts and by locating the physical and genetic positions of four marker collections widely used for high-throughput genotyping of barley cultivars. In addition, genome positions retrieved by BARLEYMAP are compared to positions within a conventional genetic map for a population of recombinant inbred lines, yielding a gene-order accuracy of 96 %. These results reveal advantages and drawbacks of current *in silico* approaches for barley genomics. A web application to make use of barley data is available at http://floresta.eead.csic.es/barleymap. The pipeline can be set up for any species with similar sequence resources, for which a fully functional standalone version is available for download.

Keywords Barley · Marker · Genetic and physical maps · Genotyping-by-sequencing · Gene annotation · Sequence mapping

Introduction

The main challenge for users of genomic data for applied purposes is the efficient use of the enormous amount of data generated by sequencing (Boller 2013). To aid geneticists and breeders of the *Triticeae* crops, some of the most important species for food security, several tools and data repositories have been developed recently, including HarvEST (Close et al. 2007), the T3 toolbox (http://triticeaetoolbox.org) or the Genome Zippers (Mayer et al. 2011).

Electronic supplementary material The online version of this article (doi:10.1007/s11032-015-0253-1) contains supplementary material, which is available to authorized users.

C. P. Cantalapiedra (✉) · R. Boudiar · A. M. Casas · E. Igartua · B. Contreras-Moreira (✉)
Estación Experimental de Aula Dei (EEAD-CSIC), Avda. Montañana, 1005, 50059 Zaragoza, Spain
e-mail: cpcantalapiedra@eead.csic.es

B. Contreras-Moreira
e-mail: bcontreras@eead.csic.es

C. P. Cantalapiedra
Plant Biology and Biotechnology PhD Program, Universitat Autònoma de Barcelona, Barcelona, Spain

B. Contreras-Moreira
Fundación ARAID, calle María de Luna 11, 50018 Zaragoza, Spain

The public release of the sequence-enriched genetic and physical map of barley (*Hordeum vulgare* L.) is being exploited for different purposes and already benefits breeding programs and companies worldwide, which previously had to rely solely on genetic maps and synteny-driven predictions. However, the current genomic assemblies are highly fragmented, as barley contains a major fraction of repeated sequences that hinder the assembly process (International Barley Genome Sequence Consortium 2012) (IBSC). Moreover, the anchored sequences come from different cultivars and sequencing methods, increasing the richness as well as the complexity of the reference map. In addition, another sequence-enriched map, based on one of the previous assemblies, has been published recently (POPSEQ, Mascher et al. 2013).

Due to that complexity, it can be a daunting task for plant breeders to place arbitrary nucleotide sequences within the barley genome and to identify nearby genes and genetic markers, useful for tasks such as genetic map assessment or map-based cloning. Furthermore, it is expected that some sequences will have multiple matches due to the presence of putative duplicated chromosome segments, paralogs and pseudogenes, as well as possible inconsistencies in the assembly (Muñoz-Amatriain et al. 2013; Poursarebani et al. 2013).

The described genomic patchwork is not exclusive to barley, as genomes from other species have been and are currently being assembled with the aid of sequence-enriched maps, especially since the advent of next generation sequencing methods and when dealing with highly repetitive genomes. Examples of the last are some species related to barley: *Brachypodium distachyon* (International Brachypodium Initiative 2010), *Aegilops tauschii* (Jia et al. 2013) and hexaploid wheat (*Triticum aestivum* L., Paux et al. 2008, 2012). Among dicots, examples include grapevine (*Vitis vinifera* L., Jaillon et al. 2007), potato (*Solanum tuberosum* L., Sharma et al. 2013) or allotetraploid cotton (*Gossypium hirsutum* L., Yu et al. 2014).

Here we present a generic software platform designed to exploit genetic and physical information from sequence-enriched maps. As such, it can be configured to work with different sequence databases and maps, and thus it may take advantage of resequencing data. The application can be used with two types of input:

1. DNA sequences, which are aligned to genome assemblies to estimate their likely genomic positions. Two strategies are supported, allowing users to map either: (1) arbitrary genomic sequences and/or (2) transcripts or expressed sequence tags (ESTs), allowing for possible introns in the alignment.
2. Standard marker identifiers so that users can have immediate access to pre-computed positions of markers. For example, those widely used in high-throughput genotyping experiments for a given species.

The BARLEYMAP pipeline, available at http://floresta.eead.csic.es/barleymap, provides researchers a simple mapping report with details on genetic and physical position of markers, as well as additional results with surrounding genes and known markers from other datasets. Here it is benchmarked and implemented as a web tool with barley data, although its use can be extended, with the standalone version, to any other species with similar genomic resources available.

Materials and methods

Pipeline outline

The BARLEYMAP pipeline (Fig. 1a) was mainly implemented in Python 2.6 and includes SplitBlast, a Perl script for distributing BLAST jobs (Contreras-Moreira and Vinuesa 2013). It has two main commands: (Align sequences) and (Find markers). The first one uses a batch of FASTA-formatted DNA sequences as input, which are aligned by means of Blastn:Megablast from the BLAST package (Altschul et al. 1997), GMAP (Wu and Watanabe 2005) or both. The “auto” mode calls both programs sequentially: input sequences are first aligned by Blastn, and those which do not yield alignments over customizable sequence identity and query coverage thresholds (default: 98 and 95 %, respectively) are then passed to GMAP. Results from both programs are filtered. In the case of Blastn, only the alignments with the best bit score are kept. Lacking bit scores, GMAP results are filtered by defining bad hits as those with both identity and coverage worse than those of other hits, as well as those marked as chimeras. The alignment step is performed against one or more sequence databases

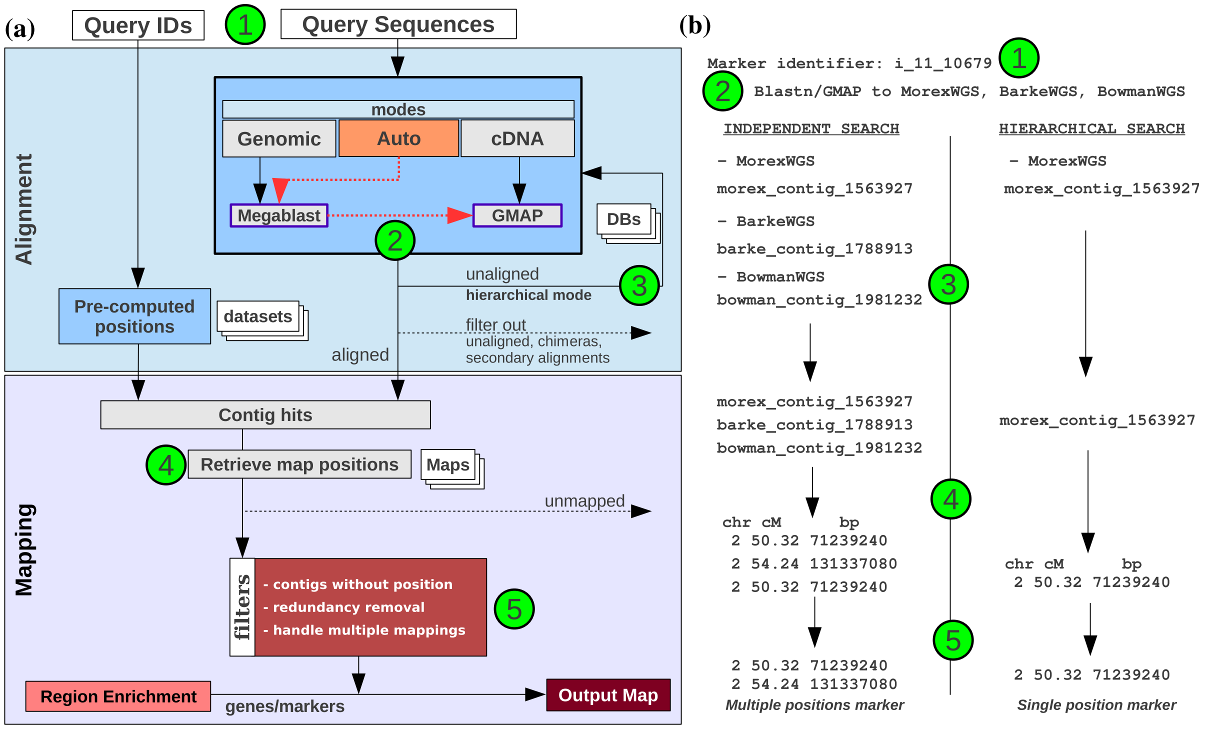


Fig. 1 BARLEYMAP pipeline. **a** Two types of input can be queried: identifiers (query IDs) or FASTA sequences. The alignment modes allow to query for genomic and/or transcript sequences. The "auto" mode uses both Blastn:Megablast and GMAP (*dotted arrows* inside "modes" box). This will be repeated for each sequence reference (DB), independently, unless the hierarchical search is specified, in which case only unaligned queries will be searched in the remaining DBs. If those do not align against any DB, they will be discarded, along with secondary alignments, alignments without position (unmapped) and GMAP chimeras (*dotted arrows*). Alternatively, alignment targets can be recovered from pre-computed data. Map positions of the targets will be associated with the queries, and after several filtering steps, enrichment with surrounding genes and markers will be performed. Finally, annotation of genes may be appended to the results. **b** An example with marker i_11_10679, from the Infinium dataset. First, it is searched by means of sequence alignments against the barley shotgun assemblies. With the hierarchical search (*right track*), the marker is found in the Morex assembly, so no other DBs are queried. The position (chr: chromosome; cM: genetic position in centimorgan; bp: physical position in base pairs) of the Morex contig, which is the target of the alignment, is retrieved from the IBSC map and finally reported. If DBs are queried independently (*left track*), all the results are kept and the position of such contigs is retrieved. Finally, as the redundancy filter cannot distinguish between actual different positions and erroneous differences, it reports a marker with multiple positions. *Circled numbers* are used to relate the different steps from **a**, **b** flowcharts

(DBs in Fig. 1a). These can be queried independently, merging the results afterwards, or by using a hierarchical strategy, in which only those queries not found in one DB are searched in the next ones (Fig. 1b). The (Find markers) command instead takes a list of query identifiers as input and retrieves their alignment targets from pre-computed datasets.

For the mapping step, the positions of targets in one or more genetic/physical maps are looked up and transferred to the initial queries. Results that provide the same location for a given query are merged into a single record. Once map positions have been compiled, the output report is augmented with genes or genetic markers anchored to those genome regions. Finally, the user has toggle controls to append to the results of the functional annotation of those genes, as well as the genes to which the additional markers hit.

Barley data configuration and application distribution

BARLEYMAP was originally configured to work with barley data. Whole genome shotgun (WGS) assemblies of cultivars Morex, Barke and Bowman, as well as Morex bacterial artificial chromosome (BAC) contigs and BAC-end sequences (BES) from the IBSC (2012), are employed as DBs. Genetic positions are retrieved separately from two recently published maps: the genetic/physical framework from the IBSC and the POPSEQ map of Morex contigs (Mascher et al. 2013). For the first one, mapping positions were obtained from the AC datasets and assigned to the DBs depending on the original source of the anchored sequence. As pre-calculated datasets, several collections of genetic markers were compiled: (1) Infinium® iSelect 9K (Comadran et al. 2012), (2) DArTs™ (Wenzl et al. 2006), (3) DArTseq™ (Diversity Arrays Technology, Australia; Kilian et al. 2012) and (4) a set of SNPs generated via genotyping-by-sequencing (GBS) for the Oregon Wolfe Barley (OWB) population (Poland et al. 2012). All of them were aligned to the DBs by means of BARLEYMAP (Align sequences). Cultivar Haruna Nijo full-length cDNAs (flcDNAs, Matsumoto et al. 2011) and HarvEST assembly #36 cDNA sequences (Close et al. 2007), including 32,331 unigenes and 37,817 singletons, were aligned to the DBs as well. The default values of identity and coverage described above were used as thresholds for the alignments in all cases, performing both Blastn and GMAP steps for aligning against every DB independently. For comparison purposes, the previous datasets were also located using the hierarchical search with BARLEYMAP (Find markers) over the WGS assemblies (Morex, Barke and Bowman), BACs and BES references, in that order.

Finally, barley genes, including introns and up to 5,000 bp upstream of each transcript, were extracted from the Morex assembly, by means of custom scripts using the GTF data for high-confidence (HC) and low-confidence (LC) genes from the MIPS FTP site (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data). Those two gene sets were used as targets for matching of all the markers from the pre-computed datasets. The same thresholds described above to align markers to the reference DBs were applied, using the hierarchical search to prioritize hits on the HC dataset. Functional annotations were also downloaded from the MIPS FTP site.

The standalone version of BARLEYMAP is distributed with the pre-computed barley datasets to support the (Find markers) mode without further requirements (the total package is ~15 MB). The attached documentation explains the configuration required to run the (Align sequences) mode and to add custom DBs, maps or datasets, including those from any other organism for which similar sequence-based mapping resources are available. The BARLEYMAP web application relies on a CherryPy web server to handle client requests, and enables the user to query all the barley resources described above. When several DBs are chosen by the user, the web application runs the hierarchical search by querying the WGS assemblies of cultivars—Morex, Bowman and Barke—Morex BAC contigs and BES, in that order.

Genetic map construction

The performance of BARLEYMAP was benchmarked against a newly developed genetic map for the barley population SBCC073 × Orria. SBCC073 is a Spanish landrace-derived inbred line (from Archidona, Málaga, Spain), with high yield under drought (Yahiaoui et al. 2014). Orria [(((Api × Kristina) × M66.85) × Sigfrido's) × 79W40762] is a semi-dwarf cultivar selected in Spain from a CIMMYT nursery, which is highly productive across most Spanish regions. This cross was carried out within the Spanish National Breeding Program. This is a population of 101 BC1F5 lines, originally developed to carry out quantitative trait locus (QTL) studies, which was genotyped with a DArTseq™ GBS assay. One BC1F5 line was discarded on the basis of high percentages of heterozygous data. Therefore, the final mapping population comprised 100 lines. A genetic map was constructed in a two-step process, using first Joinmap 4 (Van Ooijen 2006) and then MSTMap (Wu et al. 2008). Resulting linkage groups were assigned to barley chromosomes based on the genomic positions assigned by BARLEYMAP.

The same polymorphic SNP markers were also queried by means of BARLEYMAP (Find markers) to both IBSC and POPSEQ maps, in hierarchical mode, to obtain *in silico* maps. Spearman rank correlations were calculated between positions in the resulting genetic map and positions in the genetic/physical maps of IBSC and POPSEQ, using GenStat 16 (Payne et al. 2009).

Results

Alignment of barley transcripts

To test the alignment step of BARLEYMAP (Fig. 1a), the "auto" mode was selected to match long transcripts against the WGS assemblies of cultivars Morex, Barke and Bowman, as well as against the BAC contigs and BES from the IBSC, in that order by means of the hierarchical search. Of the 28,620 flcDNAs from cultivar Haruna Nijo (Matsumoto et al. 2011), 60 % were successfully aligned, with 68.5 % of the alignments obtained by GMAP (Fig. 2). Applying the same method, at least one hit was found for 59 % out of 70,148 HarvEST cDNA sequences, with almost 60 % of them aligned by Blastn. 79 and 86 % of the previous hits were matched against the first queried database, the WGS assembly of cultivar Morex. The rest, 3,578 and 5,725 queries, respectively, could only be matched in the remaining references.

Alignment of barley markers

A second benchmark consisted of mapping diverse collections of genetic markers, described in "Materials and Methods" section, which are widely used by geneticists and breeders:

1. 7,864 Infinium® iSelect SNPs.
2. 2,000 Diversity Array Technology presence–absence (PAV) markers (DArTs™).
3. 24,061 GBS markers, including both SNP and PAV markers (DArTseq™)
4. 34,396 GBS SNP markers from the OWB population.

As observed for transcripts, a significant number of Infinium (30 %) and DArT (16 %) markers could only be confidently aligned with GMAP (Fig. 2). However, this proportion was tiny for GBS markers, especially for DArTseq SNPs, which were mostly aligned by Blastn. Nonetheless, around 1,400 OWB GBS markers were aligned by GMAP.

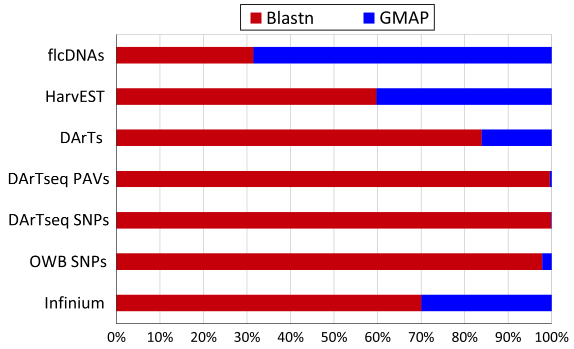


Fig. 2 Percentage of sequences found by either Blastn or GMAP, using the hierarchical method to align every dataset to barley sequence references

Although these markers are short DNA sequences, their alignments produced mostly single hits (over 98 %) when searched independently in the WGS assemblies of cultivars Morex, Barke and Bowman. However, such percentage was smaller for BAC contigs and BES references (64 and 88 %, respectively). Using the hierarchical method, this percentage was near 99 % for every marker dataset (Table 1).

The databases yielding the highest number of aligned markers were the WGS assemblies (Online Resource 1, Figure S1), with those from cultivars Morex and Bowman being slightly more informative than the one from cultivar Barke. The number of markers aligned to BAC contigs and BES references was smaller in comparison. In all cases, the use of the hierarchical search method resulted in a larger number of markers available for position retrieval.

Mapping of aligned markers to barley genetic/ physical maps

Markers aligned to sequence DBs (Table 1) were then assigned genetic positions retrieved from the IBSC and POPSEQ sequence-enriched maps (Online Resource 2). While POPSEQ comprises only contigs from the Morex assembly, IBSC map positions can be retrieved for contigs from up to five different DBs. Thus, in the latter case, marker positions were obtained either (1) by merging the results from their alignment to each DB independently or (2) from the hits obtained with the hierarchical method (see "Materials and Methods" section). As summarized in Table 2, the highest number of markers was mapped to the IBSC map, with 59 % of them having a single map position. In contrast, the POPSEQ results had the least number of mapped markers, but 99 % of them had a single map position. Regarding the hierarchical search, it misses ~4,300 marker positions with respect to IBSC, but a large majority of the sequences mapped (99 %) had a single map position, just as observed for POPSEQ.

A significant fraction of all mapped markers lie on identical genetic positions and do not contribute to effectively resolve genomic intervals. Thus, considering only unique genetic locations, the hierarchical search method yields the maximum number of landmarks, with 6,908. This advantage of the hierarchical

Table 1 Genetic markers aligned by BARLEYMAP to barley sequence references, using the hierarchical search method

Marker sets	Markers	Aligned (%)	Single target (%)
DArTs	2,000	1,340 (67.0)	1,334 (99.6)
DArTseq PAVs	15,526	7,498 (48.3)	7,456 (99.4)
DArTseq SNPs	8,535	6,876 (80.6)	6,832 (99.4)
OWB SNPs	34,396	22,992 (66.8)	22,731 (98.9)
Infinium	7,864	7,304 (92.9)	7,291 (99.8)
Total	68,321	46,010 (67.3)	45,644 (99.2)

The proportion of matched queries with a single alignment hit is shown as well

method when compared to the IBSC results comes at the cost of masking markers with multiple positions in different DBs. However, the information lost is mostly redundant, as revealed by the analysis of the positions of markers: for markers with multiple locations in the same DB reported by both search methods, 102 out of 140 (73 %) lay in different chromosomes; for those removed by the hierarchical method (15,493), only 8 % are in different chromosomes and most of the remaining are <5 cM apart, as shown in Online Resource 1, Figure S2.

Matching of genetic markers to barley genes

By taking the IBSC gene annotations, the sequences of genes, including introns and up to 5,000 bp upstream of each transcript, were obtained from the WGS assembly of cultivar Morex, yielding 62,426 HC and 69,299 LC sequences. A total of 68,321 markers from the datasets in Table 1 were matched to these gene sequences with the (Align sequences) command, hierarchical search and default parameters, as explained in "Materials and Methods" section. Of these, 39.23 % matched currently annotated genes, with 68 % being HC genes.

Validating genetic maps of barley populations

The population SBCC073 × Orria yielded 2,483 polymorphic SNPs. These were filtered according to the presence of missing data (<10 %), heterozygotes (<10 %) or allelic frequency of the donor parent (SBCC073) over 75 %. After filtering, 1,227 SNPs were used to construct a genetic map. In a first step, linkage groups were created with software Joinmap using the maximum likelihood algorithm. Then, in a second step, the distances between markers were recalculated based on the Kosambi's mapping function using MSTMap, which works more efficiently when the number of markers is large. A total of 11 linkage groups were thus identified, representing 4 whole chromosomes (1H, 3H, 4H and 5H) and 3 fragmented ones (chromosome 2H in 3 groups, chromosomes 6H and 7H in 2 groups each). Linkage groups were assigned to chromosomes, and the resulting genetic positions of the 1,227 SNP markers were compared to the positions assigned to them by BARLEYMAP by hierarchically searching against either POPSEQ or IBSC references. Correlation analyses, summarized in Fig. 3 and Online Resource 1, Table S1, reveal that locus order in the genetic map derived from the population is largely similar to the implicit ordering of positions automatically assigned by the (Find markers) command. The weighted averages obtained across linkage groups for POPSEQ and IBSC were 0.92 and 0.96, respectively. There were nonetheless three exceptions: (1) a small linkage group made of 10 markers for which the genetic map is necessarily less consistent than for larger groups; (2) linkage group 4H and; (3) linkage group 6H.2. For these last

Table 2 Result of mapping all the 68,321 markers from Table 1 to the IBSC and POPSEQ maps

Map/search type	Markers with map position	Markers with single position	Unique genetic positions
IBSC/independent	38,528	22,891	5,675
POPSEQ/Morex assembly	30,330	30,232	2,721
IBSC/hierarchical	34,203	34,063	6,908

For IBSC, results obtained by the independent and hierarchical search strategies are shown

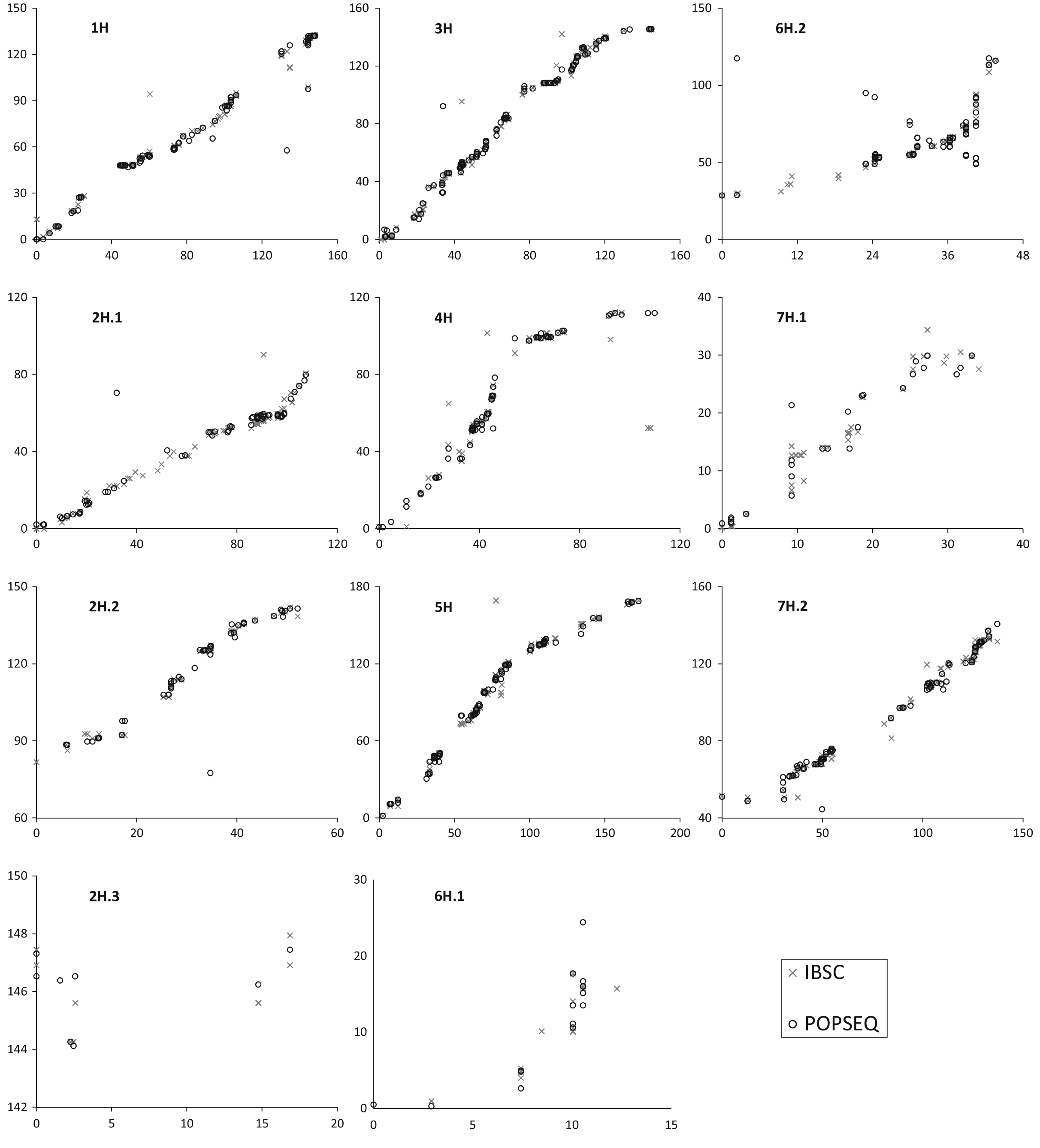


Fig. 3 2D scatter plots comparing the RIL population map (*X* axis) against the IBSC and POPSEQ *in silico* maps (*Y* axis). Positions of marker loci in cM. The positions of the IBSC genetic/physical map (*grey crosses*) and the POPSEQ map (*black circles*) were obtained using the hierarchical method of BARLEYMAP (Find markers)

two groups, there was good agreement with only one of the two physical maps used, pointing to local discrepancies between the data from IBSC and POPSEQ (see Fig. 3).

Discussion

Plant breeders have relied upon large numbers of de novo genetic maps and consensus maps to deduce information about the relative position of their markers in relation to others. The lack of common markers between maps has hindered the progress towards the identification of genes or QTL underlying relevant traits for breeding. The era of abundant sequence data is providing the opportunity to identify numerous new markers, which are implemented in relatively cheap and high-throughput platforms, widely used by the community. This is the case of GBS protocols or array genotyping systems based on data from SNP calling pipelines.

In addition, such diversity of markers makes it possible to construct high-resolution genetic maps that, within genome sequencing projects, are used in conjunction with physical maps to anchor sequences from shotgun or BAC sequencing. These resources may not constitute a complete genome, but often contain a high proportion of the genes of an organism, correctly placed in linear order. Many of the absent assembled contigs come from highly repetitive, less gene abundant regions (International Barley Genome Sequence Consortium 2012). Thus, exploiting such sequence-enriched maps can be of help when locating genetic markers, when relating and comparing different maps to each other, or in map-based cloning. This must be done with caution, since the actual genotype or population under analysis could be more or less closely related to the sequence references or could even bear local rearrangements (Farré et al. 2012). Moreover, these sequence-enriched maps tend to have specific features for different species, since each genome project may opt to use one or several genotypes as references or could use different sequencing technologies and sources. For these reasons, it would be helpful to have tools flexible enough to help fill the gap between specific genomic databases and the data used by plant breeders.

General resources, such as Ensembl Plants (Kersey et al. 2014), or more specific ones, as the IPK Barley server (http://webblast.ipk-gatersleben.de/barley/viroblast.php), can certainly be of help for these tasks. However, they are purely sequence-based and do not make explicit use of the genetic maps underlying the physical assembly. Therefore, they do not filter alignment matches in order to summarize mapping results, thus not considering possible redundant positions as well as those with non-consistent locations along the genome, originated from subtle differences among data sources. In addition, the choice of BLAST as the only search engine complicates mapping transcripts. While BLAST is able to generate local alignments that may be used to reconstruct a complete spliced alignment, there is an extensive literature reporting the importance of using specialized algorithms for performing spliced alignments. The reason is not only for the convenience of obtaining directly a full-length alignment, including its overall statistics, but furthermore to consider micro-exons, large introns, donor/acceptor splice sites and other features related to spliced sequences that could facilitate its correct identification. This is especially important in the presence of paralogs, pseudogenes and segmental duplications in the entire genome, which can hinder joining together local alignments, and can be addressed better with programs which perform both the mapping and alignment steps in a single job (see Gotoh 2008 and references therein). Finally, these resources fail to include collections of genetic markers routinely used by breeders for genotyping their plant materials. On the other hand, HarvEST (Close et al. 2007), another important barley resource, does include SNP markers and IBSC positions of Morex genes and homologs in other grasses, but cannot be used to interactively map selected DNA sequences within the genome.

A unique feature of BARLEYMAP is the integration of alignment to sequence references and mapping to genetic and physical frameworks. Being designed to facilitate the access to positional information, BARLEYMAP concentrates in hiding the underlying redundancy and complexity by means of a series of filters. First, it allows the user to directly filter alignment results by percent identity and query coverage. Then, it considers that the user should be typically interested in the best alignment result, which is automatically selected by the BARLEYMAP web server (behaviour that may be disabled in the standalone application). Moreover, it provides an explicit control on the presence of results from multiple mapping queries in the final report, avoiding redundant results both from the alignment and the mapping steps. In the first case, different hits to the same contig will share the same genetic and physical anchored position. In the second one, different contigs may be anchored to the same position, therefore yielding redundant results. Additionally, it facilitates the interpretation of unmapped queries, by separating those with alignment hit from those without it. The combined use of Blastn and GMAP allows BARLEYMAP to align transcripts, and markers derived from them, as demonstrated here by aligning flcDNAs, ESTs and several genetic marker collections. Moreover, the use of a hierarchical method for alignment provides a reasonable compromise between the use of a single DB and the direct merging of results from the independent alignment to several DBs. In the first case, a number of queries may be absent, depending on the completeness of the assembly or presence–absence polymorphisms. For instance, cultivar Morex, as a spring cultivar, lacks the *VrnH2* gene (von Zitzewitz et al. 2005). Being an incomplete reference, other genes might only be found in alternative datasets, as the subset of flcDNAs (21 %) that cannot be confidently aligned to Morex but are found in other references. The second approach, the alignment of every sequence to every reference, in addition to being a time-consuming process, produces queries with multiple targets and redundancy, both difficult to identify and fix, and can significantly reduce the number of useful markers associated with a single, unambiguous map location. The hierarchical method reduces computing time by aligning only the remaining unaligned sequences. In addition, queries with multiple mappings will arise only when the different locations are found in the same DB. As a drawback, the hierarchical method could be masking true multiple alignments (for example, copy number variation polymorphisms) in the case of markers for which different targets are found in different DBs. However, most of those multiple positions seem to be very close to each other and are almost completely removed when using the hierarchical method. This suggests that such multiple positions are mostly artificial, generated by the independent mapping to different assemblies and sources. For efficiency and to ease downstream analysis, the web application uses only the hierarchical method when querying several DBs. The standalone application gives the user full control on using or not the hierarchical method.

BARLEYMAP allows barley geneticists and breeders to exploit their new and existing genotyping data in an accessible and time-saving manner, by integrating different marker types and flexible annotation retrieval in a single framework. It does so efficiently, as demonstrated by the good agreement between the orders of a purpose-built genetic map and the positions derived from BARLEYMAP (Online Resource 1, Table S1). According to these observations, it would be tempting to skip the mapping step altogether for any new population under study and to proceed for further analyses using directly the positions derived from sequence-enriched genetic/physical maps. This benchmark suggests that analyses based on positions such as those produced by BARLEYMAP from currently available barley resources would produce reasonable results. However, the different outcome obtained by aligning the GBS markers to the two main genomic resources (IBSC and POPSEQ) advise against using such information as the gold standard for position, at least until the accuracy of barley references improves, and even then maybe only for genotypes close enough to the existing references.

A similar statement can be made for fine mapping purposes. Despite the fact that it can be of great help to use knowledge about surrounding genes and markers provided by BARLEYMAP, when working with a marker defined interval, the positions and relative order of such features should be assessed carefully due to the technical and biological variability that might exist in the reference data (Hofmann et al. 2013; Liu et al. 2014).

Finally, BARLEYMAP allows research groups to use custom databases, maps and pre-computed datasets of markers so that they may work with their own data and share it in a light-weight manner. Therefore, it provides a framework that ranges from a ready-to-work application for the retrieval of positional data from barley resources, up to a customizable pipeline that allows working with sequence-based positional data, if available, from any organism.

Acknowledgments We thank Andrzej Kilian for his help with DArT markers. This work was funded by DGA—Obra Social La Caixa (Grant Number GA-LC-059-2011) and by the Spanish Ministry of Science and Innovation (Projects AGL2010-21929 and RTA2009-00006-C04-02). Carlos P. Cantalapiedra is funded by Spanish Ministry of Science and Innovation (Grant Number BES-2011-045905 linked to Project AGL2010-21929). Ridha Boudiar was supported by a Master's fellowship from IAMZ-CIHEAM.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Boller B (2013) Interview with Beat Boller, President of EUCARPIA. the European Association for Research on Plant Breeding, International Innovation (Environment), pp 42–43

Close TJ, Wanamaker S, Roose ML, Lyon M (2007) HarvEST. Methods Mol Biol 406:161–177

Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, Shaw P, Bayer M, Thomas W, Marshall D, Hedley P, Tondelli A, Pecchioni N, Francia E, Korzun V, Walther A, Waugh R (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44:1388–1392

Contreras-Moreira B, Vinuesa P (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696–7701

Farré A, Cuadrado A, Lacasa-Benito I, Cistué L, Schubert I, Comadran J, Jansen J, Romagosa I (2012) Genetic characterization of a reciprocal translocation present in a widely grown barley variety. Mol Breed 30:1109–1119

Gotoh O (2008) A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res 36:2630–2638

Hofmann K, Silvar C, Casas AM, Herz M, Buttner B, Gracia MP, Contreras-Moreira B, Wallwork H, Igartua E, Schweizer G (2013) Fine mapping of the Rrs1 resistance locus against scald in two large populations derived from Spanish barley landraces. Theor Appl Genet 126:3091–3102

International Barley Genome Sequence Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian Public Consortium for Grapevine Genome C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KF, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Yang H, Liu X, He Z, Mao L (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95

Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, Hughes DS, Humphrey J, Kerhornou A, Khobova J, Langridge N, McDowall MD, Maheswari U, Maslen G, Nuhn M, Ong CK, Paulini M, Pedro H, Toneva I, Tuli MA, Walts B, Williams G, Wilson D, Youens-Clark K, Monaco MK, Stein J, Wei X, Ware D, Bolser DM, Howe KL, Kulesha E, Lawson D, Staines DM (2014) Ensembl genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res 42:D546–D552

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol Biol 888:67–89

Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, Ramsay L, Hedley PE, Waugh R (2014) An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genom 15:104

Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Munoz-Amatriain M, Close TJ, Wise RP, Schulman AH, Himmelbach A, Mayer KF, Scholz U, Poland JA, Stein N, Waugh R (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J 76:718–727

Matsumoto T, Tanaka T, Sakai H, Amano N, Kanamori H, Kurita K, Kikuta A, Kamiya K, Yamamoto M, Ikawa H, Fujii N, Hori K, Itoh T, Sato K (2011) Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol 156:20–28

Mayer KF, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalakova M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Muñoz-Amatriain M, Eichten SR, Wicker T, Richmond TA, Mascher M, Steuernagel B, Scholz U, Ariyadasa R, Spannagl M, Nussbaumer T, Mayer KF, Taudien S, Platzer M, Jeddeloh JA, Springer NM, Muehlbauer GJ, Stein N (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. Genome Biol 14:R58

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Paux E, Sourdille P, Mackay I, Feuillet C (2012) Sequence-based marker development in wheat: advances and applications to breeding. Biotechnol Adv 30:1071–1088

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2009) GenStat for windows (12th edn) introduction. VSN International, Hemel Hempstead

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Poursarebani N, Ariyadasa R, Zhou R, Schulte D, Steuernagel B, Martis MM, Graner A, Schweizer P, Scholz U, Mayer K, Stein N (2013) Conserved synteny-based anchoring of the barley genome physical map. Funct Integr Genomics 13:339–350

Sharma SK, Bolser D, de Boer J, Sonderkaer M, Amoros W, Carboni MF, D'Ambrosio JM, de la Cruz G, Di Genova A, Douches DS, Eguiluz M, Guo X, Guzman F, Hackett CA, Hamilton JP, Li G, Li Y, Lozano R, Maass A, Marshall D, Martinez D, McLean K, Mejia N, Milne L, Munive S, Nagy I, Ponce O, Ramirez M, Simon R, Thomson SJ, Torres Y, Waugh R, Zhang Z, Huang S, Visser RG, Bachem CW, Sagredo B, Feingold SE, Orjeda G, Veilleux RE, Bonierbale M, Jacobs JM, Milbourne D, Martin DM, Bryan GJ (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 3:2031–2047

Van Ooijen JW (2006) JoinMap 4, software for the calculation of genetics linkage maps in experimental populations. Kyazma B.V, Wageningen

von Zitzewitz J, Szucs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen TH, Hayes PM, Skinner JS (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59:449–467

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genom 7:206

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4:e1000212

Yahiaoui S, Cuesta-Marcos A, Gracia MP, Medina B, Lasa JM, Casas AM, Ciudad FJ, Montoya JL, Moralejo M, Molina-Cano JL, Igartua E (2014) Spanish barley landraces outperform modern cultivars at low-productivity sites. Plant Breeding 133:218–226

Yu JZ, Young CJL, Pepper AE, Li F, Yu S, Buyyarapu R, Sharma GC, Hinze LL, Percy RG (2014) Toward cotton molecular breeding: challenges and opportunities. In: International plant and animal genome XXII, San Diego, CA, USA, p W604

Published May 27, 2016

A Cluster of Nucleotide-Binding Site–Leucine-Rich Repeat Genes Resides in a Barley Powdery Mildew Resistance Quantitative Trait Loci on 7HL

Carlos P. Cantalapiedra, Bruno Contreras-Moreira, Cristina Silvar, Dragan Perovic, Frank Ordon, María Pilar Gracia, Ernesto Igartua, and Ana M. Casas*

Abstract

Powdery mildew causes severe yield losses in barley production worldwide. Although many resistance genes have been described, only a few have already been cloned. A strong QTL (quantitative trait locus) conferring resistance to a wide array of powdery mildew isolates was identified in a Spanish barley landrace on the long arm of chromosome 7H. Previous studies narrowed down the QTL position, but were unable to identify candidate genes or physically locate the resistance. In this study, the exome of three recombinant lines from a high-resolution mapping population was sequenced and analyzed, narrowing the position of the resistance down to a single physical contig. Closer inspection of the region revealed a cluster of closely related NBS-LRR (nucleotide-binding site–leucine-rich repeat containing protein) genes. Large differences were found between the resistant lines and the reference genome of cultivar Morex, in the form of PAV (presence-absence variation) in the composition of the NBS-LRR cluster. Finally, a template-guided assembly was performed and subsequent expression analysis revealed that one of the new assembled candidate genes is transcribed. In summary, the results suggest that NBS-LRR genes, absent from the reference and the susceptible genotypes, could be functional and responsible for the powdery mildew resistance. The procedure followed is an example of the use of NGS (next-generation sequencing) tools to tackle the challenges of gene cloning when the target gene is absent from the reference genome.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.10.0101

Powdery mildew (*Blumeria graminis*) is an obligate biotrophic fungal ectoparasite of grasses. It colonizes the surface of leaves, feeding from the epidermal cells by means of specialized organs called haustoria (Jørgensen, 1988). The forma specialis *hordei* causes powdery mildew in barley (*Hordeum vulgare* L.), which leads to severe losses in yield and grain quality in temperate latitudes worldwide (Ames et al., 2015; Zhang et al., 2005). This results in a significant economic impact since barley is one of the most widely grown crops (Verstegen et al., 2014). Consequently, the interaction of barley and powdery mildew has been extensively studied (for a recent review, see Schweizer, 2014) and many resistance genes known as mildew genes (*Ml* genes) have been described (Friedt and Ordon, 2007).

However, most of them are still molecularly uncharacterized. Among cloned genes, the recessive *mlo* stands out; providing durable resistance (Jørgensen, 1992) which has

C.P. Cantalapiedra, B. Contreras-Moreira, M.P. Gracia, E. Igartua, and A.M. Casas, Dep. of Genetics and Plant Production, Estación Experimental de Aula Dei, EEAD-CSIC, Avda. Montañana 1005, 50059 Zaragoza, Spain; C.P. Cantalapiedra, Plant Biology and Biotechnology PhD Program, Universitat Autònoma de Barcelona, Spain; B. Contreras-Moreira, Fundación ARAID, Spain; C. Silvar, Universidade da Coruna, Spain; F. Ordon, D. Perovic, Julius Kühn-Institute JKI, Quedlinburg, Germany. Received 19 Oct. 2015. Accepted 24 Feb. 2016. *Corresponding author (acasas@eead.csic.es).

Abbreviations: BAC, bacterial artificial chromosome; BES, BAC-End sequence; CAPS, cleaved amplified polymorphic sequence; CPC, Coding Potential Calculator; CNV, copy-number variation; FPC, finger-printed contig; HM, heterozygous mapping; IBGSC, International Barley Genome Sequencing Consortium; LRR, leucine-rich repeat; MTP, minimum tiling path; NBS, nucleotide-binding site; NGS, next-generation sequencing; ORF, open reading frame; PAV, presence-absence variation; PCR, polymerase chain reaction; QTL, quantitative trait locus; RIL, recombinant inbred line; RTq, real-time quantitative; SBCC, Spanish Barley Core Collection; SNP, single-nucleotide polymorphism; UCR, University of California Riverside; WGS, whole genome sequencing.

remained effective for over 30 yr and copes with a broad spectrum of pathogen isolates (Büschges et al., 1997). The other major powdery mildew resistance genes cloned so far are located at the *Mla* locus, which consists of a cluster of genes encoding for related proteins (Wei et al., 1999). Several *Mla* alleles have been cloned (Zhou et al., 2001; Halterman et al., 2001) out of the many resistance specificities described for this locus (Jørgensen and Wolfe, 1994).

Cloning of *mlo* and *Mla* involved long and laborious efforts. Specifically, fine-mapping of these genes consisted in recurrent steps of marker development, polymorphism detection and genotyping, looking for recombinants. This was done to narrow down the respective genetic intervals until an affordable physical size of the region was achieved, and subsequently resolved by chromosome walking or sequencing of subclones developed using yeast or bacterial artificial chromosome (BAC) clones. This cumbersome procedure was most challenging for species like barley due to the lack of genomic resources and its large and highly repetitive genome (Krattinger et al., 2009). However, the recent advent of high-throughput sequencing, by means of NGS technologies, has accelerated the development of synteny resources (Mayer et al., 2011), sequenced enriched physical maps (Ariyadasa et al., 2014; International Barley Genome Sequencing Consortium [IBGSC], 2012; Mascher et al., 2013b; Muñoz-Amatriaín et al., 2015), genotyping (Comadran et al., 2012; Poland et al., 2012), and sequence capture platforms (Mascher et al., 2013a). In consequence, gene cloning now benefits from the easier and faster genotyping of high-resolution mapping populations, high-throughput polymorphism detection in parental lines, and new fine mapping approaches, such as mapping-by-sequencing (Mascher et al., 2014).

Typical disease resistance genes from plant innate immunity encode receptors usually activated through recognition of molecules from the pathogen (Flor, 1971). These receptors are usually subdivided in two classes. Transmembrane pattern-recognition receptors represent the first active line of defense at the plant cell surface (Jones and Dangl, 2006). They enable the recognition of microbe-associated molecular patterns and induce pattern-triggered immunity. In contrast, a second class of resistance proteins induces elicitor-triggered immunity, detecting either the action or the structure of pathogen molecules inside host cells. These receptors are polymorphic, defining a repertoire for the detection of distinct pathogen effectors (Maekawa et al., 2011). Most genes in this second class encode proteins of the NBS-LRR family (McHale et al., 2006).

NBS-LRRs are abundant in plant genomes (Yue et al., 2012) and are encoded by genes often located in clusters of closely related members (Michelmore and Meyers, 1998). These evolve through rapid expansion and contraction of gene families (Meyers et al., 2003; Monosi et al., 2004; Zhou et al., 2004). In barley, an example of an NBS-LRR cluster is that residing in the *Mla* locus (Seeholzer et al., 2010). NBS-LRR genes encode two protein domains. The NBS domain bears a string of motifs largely conserved in plants, both in sequence and in order (Marone et al., 2013). NBS domains are followed by a LRR domain, which is generally more variable, often associated with direct or indirect non-self-recognition (Spoel and Dong, 2012). Besides *Mla* genes, many other disease resistance genes have been associated to NBS-LRR loci in plants (reviewed in Marone et al., 2013). For instance, in barley *Rpg5*/*rpg4* confers resistance to *Puccinia graminis* (Brueggeman et al., 2008), and *Rdg2a* to *Drechslera graminea* (Bulgarelli et al., 2010). Additional NBS-LRR genes have been cloned in wheat and its wild relatives (discussed in Gu et al., 2015).

This study took advantage of the sequencing-based genomic resources available for barley to fine map a powdery mildew resistance QTL. A high-resolution mapping population was developed to narrow down the QTL interval, followed by exome sequencing of recombinant lines with contrasting resistance phenotypes. The results revealed that genes located in the physical region corresponding to the genetic interval where the QTL is placed, formed a cluster of closely related NBS-LRRs, of which the resistant lines have unique haplotypes.

Materials and Methods

Plant Material and Mapping Population

A BC_1F_2 population was obtained from the cross Plaisant × RIL151. Recombinant inbred line (RIL) 151 derives from the SBCC097 × Plaisant population (SBCC, Spanish Barley Core Collection; Silvar et al., 2010). This line has only one of the two resistance QTL identified in the original donor landrace, on 7HL (Silvar et al., 2012). BC_1F_2 seeds were planted in 96-well trays and sampled 10 d after sowing. For each individual BC_1F_2 plant, a 0.6 cm leaf disk was cut. DNA extraction and amplification was performed with the Extract-N-Amp Plant polymerase chain reaction (PCR) kit (Sigma, San Antonio, TX). A cleaved amplified polymorphic sequence (CAPS) marker, QBS58, and a microsatellite, EBmac0755, were used as flanking markers to delimit the QTL interval. Restriction digestion of PCR products was performed in a 20 μL volume using 1.5 U of the respective restriction endonuclease (Fermentas). Plants were selected if they showed recombination between both markers. Data from another four markers (QBS52, QBS46, QBS44, and QBS36) were used to perform linkage analysis with JoinMap 4.0 (van Ooijen, 2006), using Kosambi's map function. Selected plants were vernalized for 6 wk at 3 to 8°C, 8 h light, then transplanted to pots and transferred to a growth chamber, where the plants were grown under long-day conditions (16 h light, 250 μmol m^{-2} s^{-1}, 20°C, 60% relative humidity; 8h dark, 16°C, 65% relative humidity). Plants were bagged before seed setting.

To select homozygous recombinants in the BC_1F_3 generation, 20 progeny plants of each selected BC_1F_2 plant were screened as explained above. Additional CAPS and pyrosequencing markers were incorporated at this stage. To verify the genotype of the BC_1F_4 recombinant lines, genomic DNA was isolated from frozen leaves using the NucleoSpin Plant II kit (Macherey-Nagel, Germany). The complete set of markers used can be found in Supplemental File 1.

Pathogen Isolates and Disease Assessment

Four isolates of *B. graminis* f. sp. *hordei* (R79, R126, R164, and R225) were used to score resistance and susceptibility in the parents and BC_1F_4 recombinant lines. These isolates were propagated on plants of the susceptible cultivar Igri. The seedlings were grown under mildew-free conditions, at 20°C with 60 to 70% relative humidity, and a 16 h light/8 h dark photoperiod. Ten days after sowing, when the first leaf was fully expanded, five plants per line were inoculated with the different isolates by brushing them with powdery mildew spores. Inoculated plants were maintained under the same conditions described above. The infection types were recorded on a scale of 0 to 4 (including intertypes) 10 d after inoculation, following the procedure of Torp et al. (1978) and Jensen et al. (1992). Plants with infection scores < 2 were classified as resistant, otherwise were labeled as susceptible. Pictures were also taken 10 d after infection.

Exome Sequencing

Genomic DNA from three BC_1F_4 lines (1476, 1766, and 2085) was extracted from leaf tissue using the NucleoSpin Plant II XL kit from Macherey-Nagel. Exome capture and DNA sequencing was performed at CNAG (Centro Nacional de Análisis Genómico, Barcelona). DNA capture was performed in a single reaction with the Roche Nimblegene SeqCap EZ Developer kit (Mascher et al., 2013a), following the instructions from the manufacturer. DNA was barcoded with TruSeq adapters and pooled before hybridization to the exome probes. DNA fragmentation and size selection was performed to produce 2 × 101 bp paired-end reads with average insert size of 150 bp. Sample preparation followed standard Illumina TruSeq procedures. Sequencing was performed in two separate runs of an Illumina HiSeq2000, each in a single lane.

Reads were aligned to the Morex whole genome sequencing (WGS) assembly (IBGSC, 2012) with BWA MEM (Li and Durbin, 2009) with default parameters. Read duplicates were tagged by means of MarkDuplicates from picard-tools-1.113 (http://broadinstitute.github.io/picard). Variant detection was performed combining SAMtools (Li et al., 2009) and GATK (McKenna et al., 2010) (see Supplemental Materials and Methods). Variants were filtered out, requiring a minimum depth of 10 and a minimum quality of 30 in each genotyped line. Polymorphic variants were obtained comparing the data of the BC_1F_4 lines with variants for SBCC097 and Plaisant from another exome capture essay (Cantalapiedra, Contreras-Moreira, Gracia, Igartua, and Casas, unpublished data, 2014).

To look for the recombination points in the sequences of the three BC_1F_4 lines, a score was assigned to each variant identified after the exome capture. If a variant was like Plaisant, the score was increased by 1. If the variant was like SBCC097, the score was decreased by 1 instead. If it was different to the parents, the score remained unchanged. Therefore, the variants in which the three lines were Plaisant-like received a score of +3 in that position in the genome. On the contrary, if all three lines were like SBCC097, the score was –3. This was repeated for every variant. The scores of the variants lying on a single Morex WGS contig were averaged to obtain a single contig score.

Identification and Annotation of the BACs Located within the QTL Region

Contigs of each BAC associated to finger-printed contig (FPC) 591, from IBGSC (2012) and University of California Riverside (UCR BACs, hereafter; Muñoz-Amatriaín et al., 2015), were concatenated to build up BAC pseudoscaffolds. Gene annotations were obtained from IBGSC data, by alignment of the associated Morex WGS contigs to Uniref90 and UniprotKB (blastx, maximum e-value $1e^{-50}$) and by identification and annotation of open reading frames (ORFs) with getorf (Rice et al., 2000; -minsize 90) and the script run_predict.sh from CPC (Coding Potential Calculator, v.0.9-r2; Kong et al., 2007). Searches of NBS and LRR motifs (taken from Table 1 in Jupe et al., 2012) were performed with MAST (MEME suite 4.10.1; Bailey and Gribskov, 1998). Structure of the NBS-LRR genes was obtained after alignment of the predicted proteins to NCBI nr protein database (see Supplemental Materials and Methods). Multiple alignments of the proteins were performed with Clustal Omega (Sievers et al., 2011).

Finding and Assembling Heterozygous Mapping Regions

Although the lines used for this study should all be homozygous in the QTL region, a number of sites with heterozygous variants were found after aligning exome sequences to the reference. To systematically locate these regions, an analysis of the number of different k-mers mapping to the pseudoscaffolds was performed. Read mappings from exome sequencing were surveyed to quantify each different 50-mer aligning to each position in the reference, considering only those sampled at least four times. The scripts used for k-mer analysis are available in Supplemental File 2. Sets of reads from the segments with more than one kind of k-mer (therefore annotated as heterozygous mappings, HMs) and mapping to disease resistance proteins were assembled with Trinity (Grabherr et al., 2011; parameters located in Supplemental Materials and Methods). The sequence contigs obtained for the different BC_1F_4 lines were compared and clustered. A representative sequence was chosen from each cluster and a genotype was assigned to it based on its presence-absence pattern across BC_1F_4 lines. Several overlapping contigs, which showed the same PAV in the lines, were assembled together.

Validation of the Genotypes Found with the Exome Capture by PCR

The genotypes of the parents and the recombinant lines were checked for those Morex WGS contigs which had polymorphisms associated with the resistance or susceptibility phenotype. These included contigs 1622651, 167712, 211721, and 50573. Amplicons were used to validate the genotypes of the lines corresponding to sequences present in BACs M01 and D03 from FPC 591. In addition, the PAV polymorphism of the lines was assessed for the two largest new assembled sequence contigs (ELOC1 and ELOC2), including cultivar Morex. Primers were designed with Primer 3 (Untergasser et al., 2012) and validated by running isPCR (https://genome.ucsc.edu/cgi-bin/hgPcr, verified 22 Apr. 2016) against the WGS assemblies from IBGSC data. In addition, primers were designed to amplify the unknown fragments between Morex WGS contig 50573 and both ELOC1 and Morex WGS contig 44875, by Long Range PCR. The primers and their respective PCR conditions can be found in Supplemental File 1.

Characterization of the New Assembled Sequence Contigs

Putative ORFs encompassing the assembled ELOCs were searched with ORF Finder. In addition, CPC was conducted to evaluate their protein-coding potential. The resulting DNA sequences were searched for in the Uniprot Plants and NCBI nr databases. Both sequences were also compared against the IBGSC databases and Haruna Nijo flcDNAs (Matsumoto et al., 2011) with Barleymap (Cantalapiedra et al., 2015). The predicted aminoacid sequences coded by those ORFs were compared to each other with blastp.

Real-Time PCR of the Assembled Sequence Contigs

For Real-Time quantitative polymerase chain reaction (RTq-PCR) experiments, 7-d-old plants were inoculated with powdery mildew isolate R79 in the greenhouse. Two samples per line were collected at 12, 24, 48, and 72 h after infection. Each sample consisted of the pooled leaf tissue of two plants.

Total RNA was extracted from frozen samples using the Aurum TM Total RNA Mini Kit (BioRad, Hercules, CA) following the manufacturer's instructions. First-strand cDNA was synthesized from 100 ng of total RNA by using the iScript cDNA Synthesis Kit (BioRad). Primers were designed with Primer Express 3.0 (Applied Biosystems, Carlsbad, CA). RTq-PCR was performed in 50 μL of reaction mixture made up of 2.5 μL of cDNA, 1 × iQ SYBR Green Supermix (BioRad) and 0.3 μM of each specific primer. Primers and PCR conditions can be found in Supplemental File 1. The Actin gene was used as a constitutively expressed reference gene to normalize expression as in Trevaskis et al. (2006).

Results

Fine Mapping of the Resistance Locus

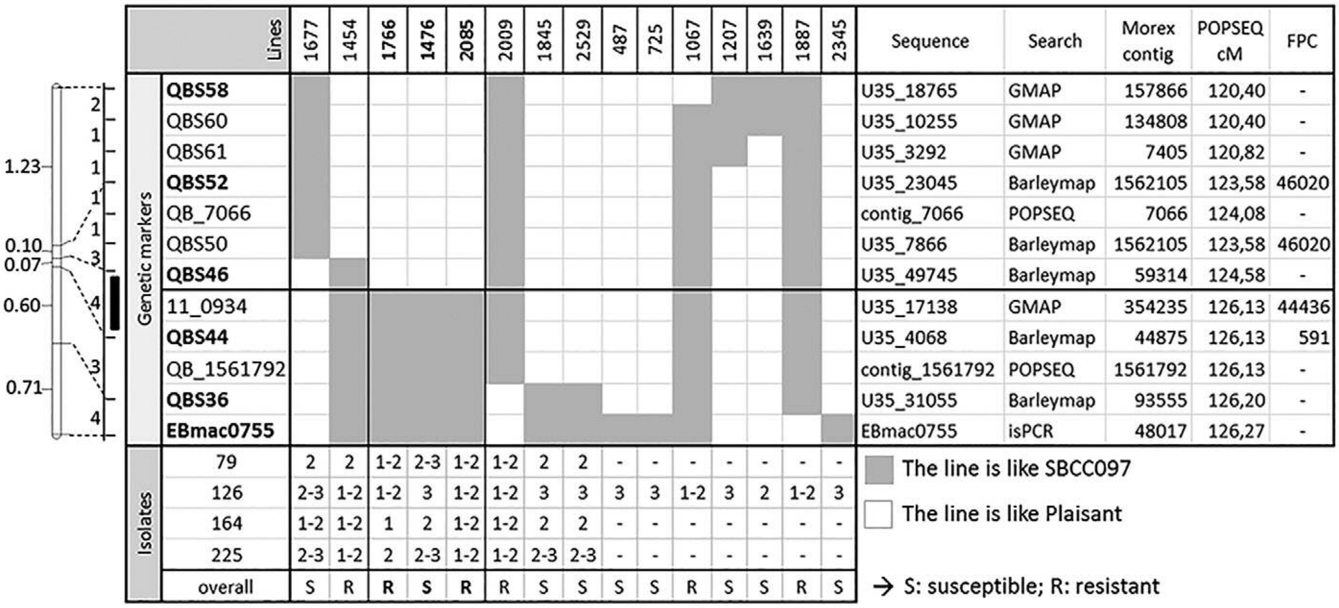
To fine map the resistance QTL identified on 7HL in the SBCC097 × Plaisant population (Silvar et al., 2010), a RIL containing only this QTL (RIL151, Silvar et al., 2012) was backcrossed to Plaisant. A large BC_1F_2 population was obtained, and tested for recombination between markers QBS58 and EBmac0755, flanking the 7HL QTL. Out of 2899 BC_1F_2 plants tested, 152 recombinants were identified and grown until maturity. Twenty-five BC_1F_3 families were then screened to identify homozygous recombinants, which were further tested with the markers obtained in previous studies, exploiting synteny and physical information (Silvar et al., 2012; 2013b). This procedure identified 15 BC_1F_4 plants covering the whole region (Fig. 1). A genetic map of the region was constructed with the information of the entire BC_1F_2 generation and allowed narrowing the position of the QTL down to a 0.07 cM interval between markers QBS46 and QBS44. Furthermore, three BC_1F_4 lines, one susceptible (1476) and two resistant (1766 and 2085), showed the same genotype flanking the QTL but different phenotype (Fig. 1, Supplemental Fig. S1). Therefore, the gene or genes responsible for the resistance lay within the interval between QBS46 and QBS44.

Analysis of Exome Sequencing Polymorphisms

Exome sequencing of the parents and the three BC_1F_4 lines was performed in order to identify the differences between the resistant and the susceptible plants (Supplemental Table S2). Analysis of the read data from exome sequencing involves a mapping step using a reference, the Morex WGS assembly (IBGSC, 2012) in this case. However, the region associated to the resistance was majorly of interest here. Therefore, the genetic markers from the previous section were located in the POPSEQ map (Mascher et al., 2013b) and the identified positions (Fig. 1) were used to anchor available genomic resources to the region (Supplemental Materials and Methods, Supplemental Fig. S2). This yielded a set of 973 Morex WGS contigs (Supplemental File 3) associated to 17 FPCs, which are contigs with assigned physical positions (Supplemental Table S1). Comparing the variants between the parents, 1037 polymorphisms were identified, corresponding to 120 Morex WGS contigs (out of the 973 just described). The genotypes of the BC_1F_4 lines were checked, looking for variants consistent with the phenotypic profile of the lines (1476 like the susceptible parent, Plaisant; the other two like the resistant parent, SBCC097), as those would be the most informative toward finding candidate genes. Only one of the Morex sequences, contig 50573, presented haplotypes fully in agreement with the phenotypic profile of the lines. This contig has a single annotated gene, a "Pentatricopeptide repeat-containing protein" (MLOC_65722 in IBGSC data). A CAPS marker designed for this gene was assayed on all 15 BC_1F_4 lines, and its position within the QTL region was confirmed.

Physical Localization of the Resistance Locus

From the previous analysis, only Morex contig 50573 was unambiguously located within the QTL interval. However, although its genetic POPSEQ map position was known, it



Isolates	1677	1454	1766	1476	2085	2009	1845	2529	487	725	1067	1207	1639	1887	2345
79	2	2	1-2	2-3	1-2	1-2	2	2	-	-	-	-	-	-	-
126	2-3	1-2	1-2	3	1-2	1-2	3	3	3	3	1-2	3	2	1-2	3
164	1-2	1-2	1	2	1-2	1-2	2	2	-	-	-	-	-	-	-
225	2-3	1-2	2	2-3	1-2	1-2	2-3	2-3	-	-	-	-	-	-	-
overall	S	R	R	S	R	R	S	S	S	S	R	S	S	R	S

Sequence	Search	Morex contig	POPSEQ cM	FPC
U35_18765	GMAP	157866	120,40	-
U35_10255	GMAP	134808	120,40	-
U35_3292	GMAP	7405	120,82	-
U35_23045	Barleymap	1562105	123,58	46020
contig_7066	POPSEQ	7066	124,08	-
U35_7866	Barleymap	1562105	123,58	46020
U35_49745	Barleymap	59314	124,58	-
U35_17138	GMAP	354235	126,13	44436
U35_4068	Barleymap	44875	126,13	591
contig_1561792	POPSEQ	1561792	126,13	-
U35_31055	Barleymap	93555	126,20	-
EBmac0755	isPCR	48017	126,27	-

Fig. 1. Fine mapping of the 7HL quantitative trait loci (QTL). Left: Genetic map of BC_1F_2 mapping population (distances in cM) showing a schematic distribution of the recombinants found in the BC_1F_3 by marker interval. The black vertical bar indicates the position of the QTL. Center: Graphical genotypes of the 15 BC_1F_4 lines. Markers assayed in the BC_1F_2 are highlighted in bold type. The lines sequenced in this study (1766, 1476, 2085) are separated from the others by thick vertical lines. The thick horizontal line between QBS46 and 11_0934 marks the most likely position of the resistance gene. The bottom table summarizes the evaluation of the lines for resistance to four different powdery mildew isolates. Right: Table showing the sequences used to locate the genetic markers in the barley genome, and the sources (POPSEQ) or search methods used, Barleymap or GMAP (Wu and Watanabe, 2005). The target whole genome sequencing contigs are shown ("Morex contig" column) along with their position in chromosome 7H ("POPSEQ cM" column), as well as the physical contigs ("FPC" column) associated to them.

could not be found in the IBGSC physical map, hindering its direct physical localization. Nonetheless, most of the variants in the remaining Morex WGS contigs were clearly located on either side of the candidate region (i.e., the three lines had the same genotype). Looking at the genotypes of the lines from exome data, the position and order of Morex WGS contigs was not always in agreement with the POPSEQ map (Supplemental Fig. S3). If only Morex WGS contigs with known physical position were considered, the genotypes of the recombinant lines indicated the likely physical location of the recombination breakpoints within FPC 591, more specifically, between contigs 167712 and 44875 (Fig. 2A). The position of yet another Morex WGS contig, 211721, was ambiguous. The genotypes of the lines for these contigs were confirmed by PCR assays.

To further delimit the physical position of the resistance locus, the BACs associated to FPC 591 in the IBGSC physical map were retrieved (Fig. 2B). Among BACs with available sequence data, HVVMRXALLmA0204M01 (M01 hereafter) spans a central segment of FPC 591. Among the Morex WGS contigs aligning to M01 (Supplemental File 4), 167712 and 211721 were identified ~2.5 kb apart. Moreover, Morex contig 44875 was associated to BAC HVVMRXALLEA0187D03 (D03 from now on), both from IBGSC anchoring data and by our homology searches (identity 99.75%, full target coverage, bitscore 1448; to D03 BES MRX2BAD187D03T71). D03 covers the right half of FPC 591, but it has not been fully sequenced yet. No other BACs providing new data within the QTL interval were identified. Candidate genes should thus be placed within the minimum tiling path (MTP) defined by BACs M01 and D03.

During the progress of this work, a new assembly of BACs (UCR BACs) was published. In this assembly, two extra BACs were associated to FPC 591 (Fig. 2B): 0139I11 and 0758B20 (I11 and B20 from now on). BAC I11 (Supplemental File 5) was compared to M01 (Supplemental Fig. S4A). Most of the I11 sequences are already present in M01, but with a different arrangement. In contrast, the comparison of B20 and M01 pseudoscaffolds (Supplemental Fig. S4B) showed that they are mostly different, with only a few related regions. Among the Morex WGS contigs which aligned to B20 (Supplemental File 6), contigs 50573 and 44875 were found, separated by 4234 bases. Note that Morex WGS contig 50573 is the only one with a haplotype in agreement with the phenotypes of the lines, hence supporting the position of the resistance locus within FPC 591.

Searching for Candidate Genes in the Reference Cultivar Morex

Candidate genes were searched for in the annotated Morex genome. Alignments of Morex WGS contigs, anchored to BAC M01, against IBGSC and Uniref90 sequences, revealed eight gene annotations: five "Disease resistance protein RPM1," two transposon-related and one "Putative disease resistance protein RGA4." In-house annotation of the ORFs identified in the M01 pseudoscaffold (see Materials and Methods) confirmed the presence of the RPM1- and

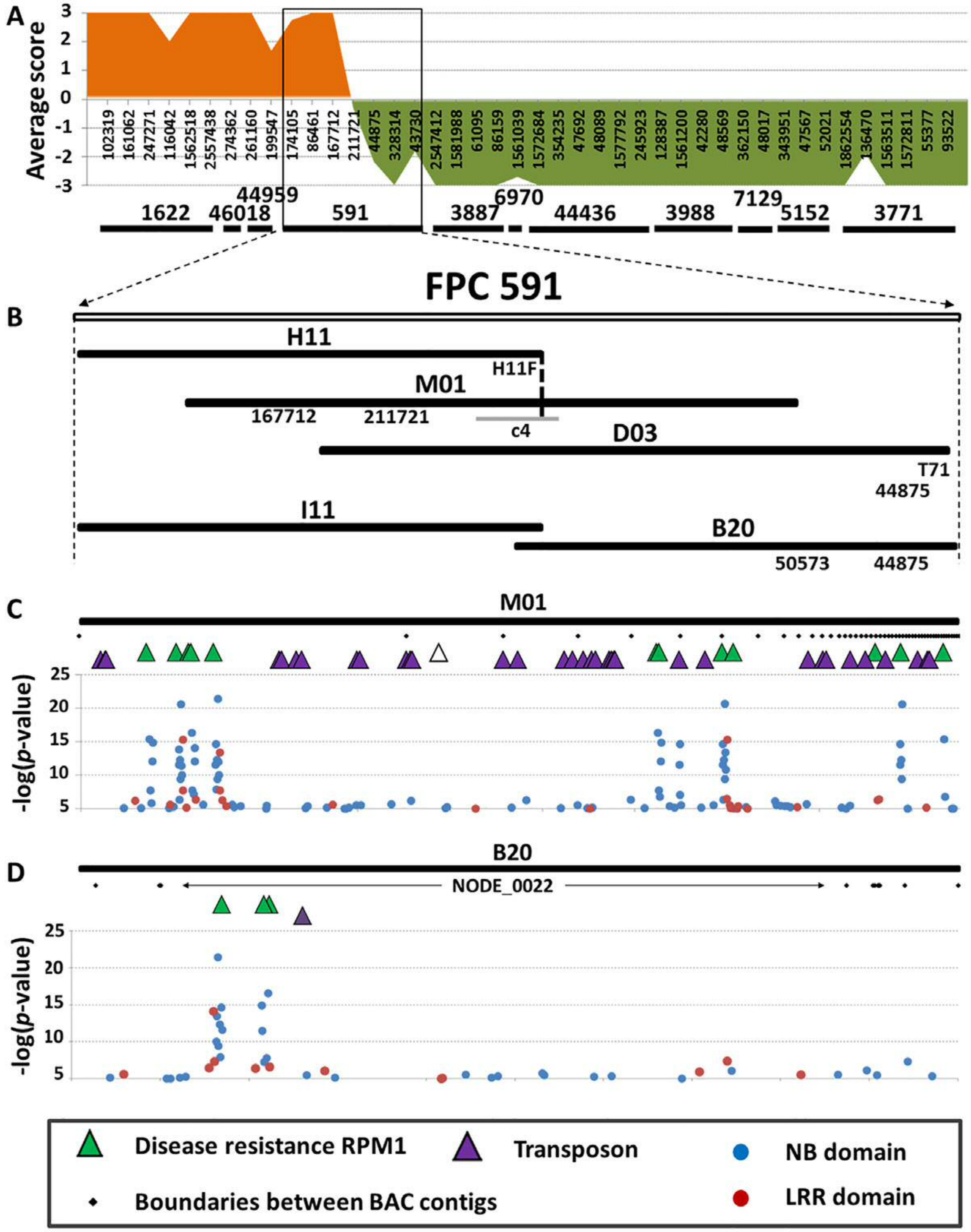


Fig. 2. Analysis of bacterial artificial chromosomes (BACs) in minimum tiling path (MTP) of finger-printed contig (FPC) 591. (A) Average scores of the Morex whole genome sequencing (WGS) contigs considering the genotypes of the BC_1F_4 lines in relation to the parents. Orange: positive score, more lines are like Plaisant; green: negative score, more lines are like SBCC097. Contigs are sorted by increasing FPC cM position, and by POPSEQ position to resolve coincidences, from left (120.4 cM) to right (126.6 cM). FPCs are shown as black horizontal bars. (B) IBGSC (H11, M01, and D03) and UCR (I11 and B20) BACs covering FPC 591. Morex WGS contigs 167712 and 211721 are anchored to M01. BAC-End sequence (BES) H11F and BAC contig c4 of M01 match by sequence alignment (vertical dashed line). BES T71 and Morex WGS contig 44875 align to each other. Morex WGS contigs 44875 and 50573 are anchored to B20. (C) Analysis of the pseudoscaffold of BAC M01, represented as a black horizontal bar. Green triangles are ORFs annotated as RPM1 by alignment to Uniref90. A white triangle shows an ORF annotated as RGA4, which seems to be related to transposons. Purple triangles show the position of ORFs annotated as transposons. The scatterplot shows the-$\log_{10}$(p-value) of the NBS and LRR motifs identified throughout the pseudoscaffold (blue dots, NBS domains; red dots, LRR domains). (D) Analysis of the pseudoscaffold of BAC B20. Note that NODE_0022 is highlighted as the longest contig in the BAC.

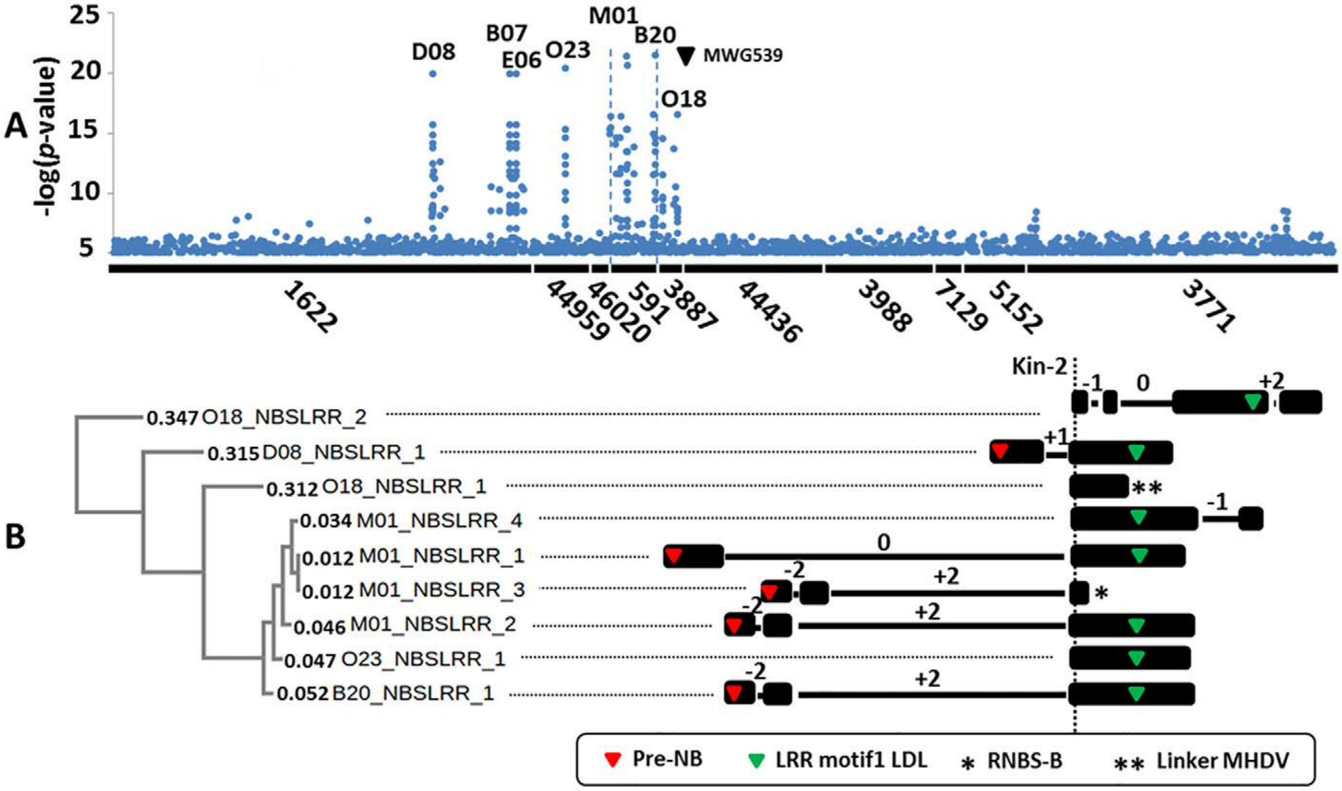


Fig. 3. Nucleotide-binding site (NBS) and leucine-rich repeat (LRR) motifs found in the region of FPC 591. (A) Significance of the motifs found in the whole region (of about 5.6 Mb). Vertical dashed blue lines demarcate the motifs found within FPC 591. A black triangle indicates the physical position of RFLP marker MWG539, close to the *Mlf* locus (Schönfeld et al., 1996). (B) UPGMA clustering of the predicted proteins containing NBS-LRR motifs. Protein names are prefixed with their respective BAC codes. Distances obtained from the multiple alignment are shown to the left of each protein name. Inferred gene structures are shown to the right (black boxes: exons; black horizontal lines: introns). The number on each intron shows the frame change from one exon to the next. Motifs shown on gene structures are named after Table 1 in Jupe et al. (2012). A vertical dashed line shows the position of the Kinase-2 motif, to which the structures of genes have been aligned. Asterisks indicate the presence of a specific motif at the end of the available sequence of the corresponding gene.

transposon-related sequences, including loci not associated to Morex WGS contigs and, therefore, lacking exome capture probes. When the whole pseudoscaffold was self-aligned, the ORFs annotated as RPM1 proteins appeared to be related to each other (Supplemental Fig. S5). Since RPM1 belongs to the NBS-LRR family of resistance-genes, motifs which are known to be conserved in domains of NBS-LRR genes (Jupe et al., 2012) were searched for in the region using the software MAST. Most RPM1-related loci were also confirmed by the MAST scan (Fig. 2C). Overall, nine segments were identified with highly significant motifs from the N-terminal, NBS and linker domains; three of them with LRR motifs (Supplemental File 4). The same analysis was applied to BAC I11, which showed almost the same features as M01, as expected (Supplemental File 5).

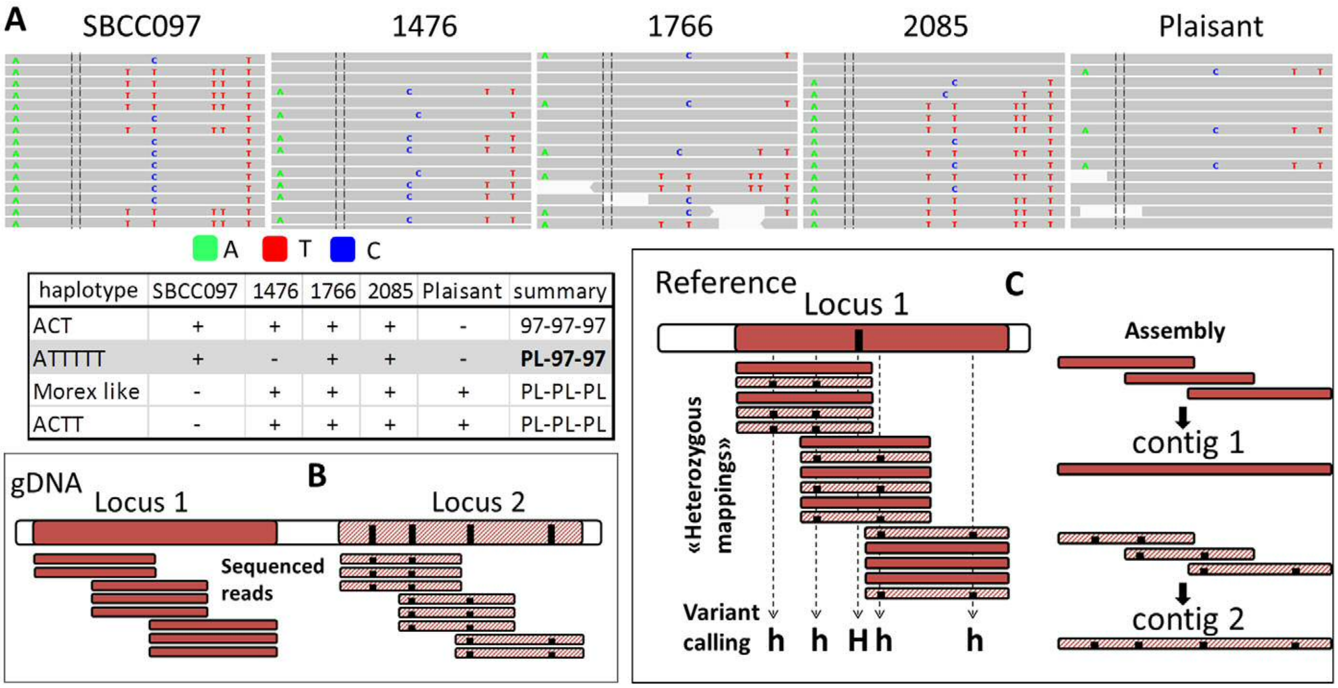
On the other hand, IBGSC annotation of the Morex WGS contigs associated to UCR BAC B20 showed up 2 genes: a "Pentatricopeptide repeat-containing protein" in contig 50573, mentioned earlier, and a "WD-repeat protein 57 IPR015943" in contig 44875. Both results were confirmed with alignments to Uniref90. In addition, another 3 Uniref90 hits to the left of contig 50573 were obtained; all labeled as "Disease resistance protein RPM1," both using raw Morex WGS contigs and in silico identified ORFs as queries. Again MAST scans of NBS-LRR motifs confirmed these results (Fig. 2D) and, as with M01, several hits related to transposons were obtained close to them (Supplemental File 6).

Analysis of NBS-LRR motifs in a wide physical region around FPC 591 (55 UCR BACs, spanning 5.6 Mb) revealed that the cluster is mostly circumscribed to the resistance locus (Fig. 3A). A few other NBS-LRR genes were detected outside the locus, but these were unrelated both in terms of sequence and gene structure (Fig. 3B, Supplemental Files 7 and 8).

Therefore, besides a Pentatricopeptide repeat-containing protein and a WD-repeat protein, the MTP spanning the resistance locus in Morex is rich in transposons and contains a cluster of closely related NBS-LRR genes.

Analysis of Heterozygous Mappings in Morex

As shown above, only Morex WGS contig 50573 had a haplotype consistent with being within the resistance locus. However, there were other Morex WGS contigs for which some variants were consistent, but others were not. Many of the variants in those contigs were apparently heterozygous. This was highly unlikely, as the



haplotype	SBCC097	1476	1766	2085	Plaisant	summary
ACT	+	+	+	+	-	97-97-97
ATTTTT	+	-	+	+	-	**PL-97-97**
Morex like	-	+	+	+	+	PL-PL-PL
ACTT	-	+	+	+	+	PL-PL-PL

Fig. 4. Heterozygous mappings (HMs). (A) Images captured from Integrative Genomics Viewer (Integrative Genomics Viewer [IGV]; Robinson et al., 2011), showing reads (gray horizontal bars) mapping to a specific interval of Morex whole genome sequencing contig 1622651. Colored characters show the variants detected for each genotype in relation to the Morex reference. The table summarizes the haplotypes identified, along with their presence-absence type (+ or –) in the lines. Genotypes of the three BC_1F_4 recombinant lines relative to the parents are shown in the "summary" column. One group of variants (ATTTTT, light gray background) is consistent with the phenotypic resistance profile of the lines (PL-97-97 or susceptible-resistant-resistant). (B) schematic representation of the reads that would be obtained after sequencing two closely related loci. The two loci are represented by horizontal bars (red background; plain for Locus 1, striped for Locus 2), with a few hypothetical differences (black vertical bars). (C) Reads from (B) are mapped back to the reference. In the example shown, the reference lacks one locus (Locus 2), and all sequenced reads hit the existing one (Locus 1), producing apparent HMs. As a result, variant calling yields heterozygous calls (h) and homozygous calls (H) intermixed. A new assembly could solve this region, yielding independent contigs resembling the original loci, due to the presence of the four genotypic variants between the two loci.

parents were homozygous, the BC_1F_4 plants were selected to be homozygous for the interval of interest and the possibility of having double recombinants within such a small region was negligible. In fact, visual inspection of the mappings producing those variants revealed different populations of reads stacking to the same locus (Fig. 4A), in contrast with the mappings from contig 50573, which produced unambiguous homozygous single-nucleotide polymorphisms (SNPs; Supplemental Fig. S6). The apparent heterozygous genotypes were confirmed through PCR amplification of CAPS markers (Supplemental Fig. S7). Note that these variants were abundant and linked in recurrent groups, as independent haplotypes, instead of being spread out randomly among the reads. Thus, it is unlikely that they are the result of sequencing errors. Instead, these mappings could have been produced by piling up closely related sequences (repeats, paralogous genes) which were captured by the exome baits (Mascher et al., 2013a; Jupe et al., 2013), but for which the original locus would not be present in the reference. Since they affect variant calling, producing apparent heterozygous variants, from now on this kind of mappings will be referred to as HMs (Fig. 4B and 4C). Almost all Morex WGS contigs with HMs, whose variants had genotypes in agreement with the phenotypic profile of the lines, could be annotated as homologs to "Disease resistance protein RPM1" or "Disease resistance protein RPP13" (Supplemental Table S3), after alignment to the Uniprot Plants database (http://www.uniprot.org/blast/, verified 22 Apr. 2016). Some of those contigs are the ones located within or close to FPC 591 (Supplemental Fig. S8–S11). Taken together, these results suggest that there are sequences related to disease resistance proteins, which are not present in the Morex reference, but are likely within the resistance locus in the genomes of SBCC097 or Plaisant.

In this study, the distribution and abundance of HMs in the resistance locus region was analyzed in more detail to (i) assess whether the differences between the recombinant lines were likely to be related with the disease resistance, (ii) verify whether the presence of HMs was a feature exclusive of the sequences related to NBS-LRR genes in the region of interest, and to (iii) identify and demarcate the segments of the reference in which they occur. This last objective would allow obtaining the reads which produce the HMs and assembling them into sequence contigs (Fig. 4C).

Therefore, we analyzed the number of different 50-mers, fragments of reads of 50 bases, mapping to each position of Morex WGS contigs anchored to BACs M01 and B20 in the three BC_1F_4 lines. Note that the reads from our sequencing data are 101-mers, but to be able to capture diversity in a given position a smaller k-mer size had to be chosen, since mapping duplicates were removed in a previous step. Wherever several 50-mers mapped to the same position, HMs would be likely found; each 50-mer being possibly derived from a different genomic locus. Notably, we found different 50-mers mapping to most of the loci related with NBS-LRR genes, although not all the mapped loci belonged to that class (Supplemental Fig. S12). Out of the covered positions, 74.4 and 89.5% had a single 50-mer in M01 and in B20, respectively. Interestingly, differences among the lines seemed to be associated mostly to disease resistance loci. First, the resistant lines had a larger percentage of positions with several 50-mers (i.e., with HMs) in M01 (Supplemental Fig. S13A), although not in B20. Furthermore, taking into account only the reference positions within annotated NBS-LRR genes, the difference between the resistant lines and the susceptible one increased in both BACs (Supplemental Fig. S13B). Therefore, the differences between the two BACs can to a large extent be explained by the greater abundance of NBS-LRR related sequences in M01 and B20 (49.6 and 11.7% of the mapped bases, respectively).

De Novo assembly of Exome Sequence Reads Spanning the Resistance Locus

Analysis of HMs pointed toward the presence of NBS-LRR related sequences within the resistance locus, absent from the Morex reference. In light of this, a template-guided assembly of reads producing HMs was performed. First, Morex WGS contig fragments located within FPC 591, related to disease resistance genes and producing HMs were chosen (11 loci). Second, six further Morex WGS contig fragments with HMs and variants in agreement with the phenotypes of the lines were selected. Finally, Morex WGS contig 50573, harboring the "Pentatricopeptide repeat-containing protein," was included as a control. Read subsets mapping to the 18 selected segments were retrieved, and an independent assembly for each genotype was performed (for both parents and the three BC_1F_4 lines, Supplemental File 9). These operations yielded 203 sequence contigs, with an average of almost 41 contigs per line. These new contigs were clustered, and a representative sequence per cluster was selected (see Supplemental materials and methods and Supplemental Fig. S14), yielding 31 representative sequences. Based on the presence or absence of those sequences, PAV genotypes for each cluster were assigned to each line. Representative sequences showing the same PAV genotypic profiles were then compared with each other, leading to the assembly of five of them into a contig of 981 nucleotides (ELOC1), and another four into a contig of 787 bases (ELOC2). Therefore, the final set comprised 24 sequence contigs, for which the lines had different PAV genotypes (Supplemental File 9). ELOC1 and ELOC2 were the largest assembled contigs. ELOC1 was absent in Plaisant and 1476, while ELOC2 was only present in SBCC097 and 1766. The absence of ELOC2 from the resistant line 2085 was in agreement with the fewer number of 50-mers identified in this line in comparison with 1766, and it suggested that 2085 and 1476 contained the smallest interval flanking the resistance locus.

Validation and Characterization of the New Assembled Sequence Contigs

We designed primers to perform PCR amplification of ELOC1 and ELOC2. The PCRs confirmed the PAV genotypes of the 15 BC_1F_4 lines and the parents (Fig. 5). In addition, the absence of both sequences in cultivar Morex was verified (data not shown). To check whether this result was a consequence of polymorphism on the primers, the reads from the exome capture of SBCC097, Plaisant, Morex (from the same exome capture experiment), and lines 1476, 1766, and 2085 were realigned to the new contigs. This confirmed the PAV variation found on them. Moreover, the products of amplification of the lines SBCC097 and 1766 were Sanger-sequenced and further validated.

In silico ORF calling was performed with both ELOCs, obtaining two partial ORFs of 322 and 252 amino acids for ELOC1 and ELOC2, respectively. In addition, their protein-coding potential was checked, with log-odds scores of 82.73 and 57.46 for ELOC1 and ELOC2, respectively. The percentage of identity between the two amino acid sequences was 92%, and their alignment covered most of ELOC2. Looking for similar proteins in Uniprot Plants and NCBI nr databases, results were found (Supplemental File 10) within the range of identities obtained when comparing the NBS-LRR proteins in the QTL region in Morex (Supplemental File 8), and comparable with paralogous genes found in other NBS-LRR clusters (Bulgarelli et al., 2010; Kuang et al., 2004; Wei et al., 1999). Moreover, the ELOCs were aligned against the Morex NBS-LRR predicted proteins of the region. The best hits had almost full coverage and 87.9 and 91.6% identity, for ELOC1 and ELOC2, respectively. Alignment of DNA sequences of the ELOCs to the IBGSC databases produced similar results. Also, these alignments revealed that the contigs contained only the LRR domain, lacking the NBS one.

RTq-PCR was used to check the expression of both new contigs. No specific amplicon was obtained for ELOC2 and, therefore, it could either be a pseudogene (Kuang et al., 2004) or be expressed in another tissue or developmental stage (Tan et al., 2007). Nonetheless, amplification was positive for ELOC1, confirming its transcription in leaves of SBCC097 and the two resistant BC_1F_4 lines, although this is not a definitive evidence of the gene being functional (Monosi et al., 2004; Wei et al., 2002). The RTq-PCR was performed for SBCC097 at different time points, spanning 72 h after infection. Apparently, there was no change in ELOC1 expression in response to the infection, although this is not irreconcilable with being involved in the resistance or even being regulated at another stage than transcription (Tan et al., 2007).

Discussion

Barley research has been accelerated by the availability of abundant genomic resources published over the last years. In some cases, this has led to faster gene cloning, like

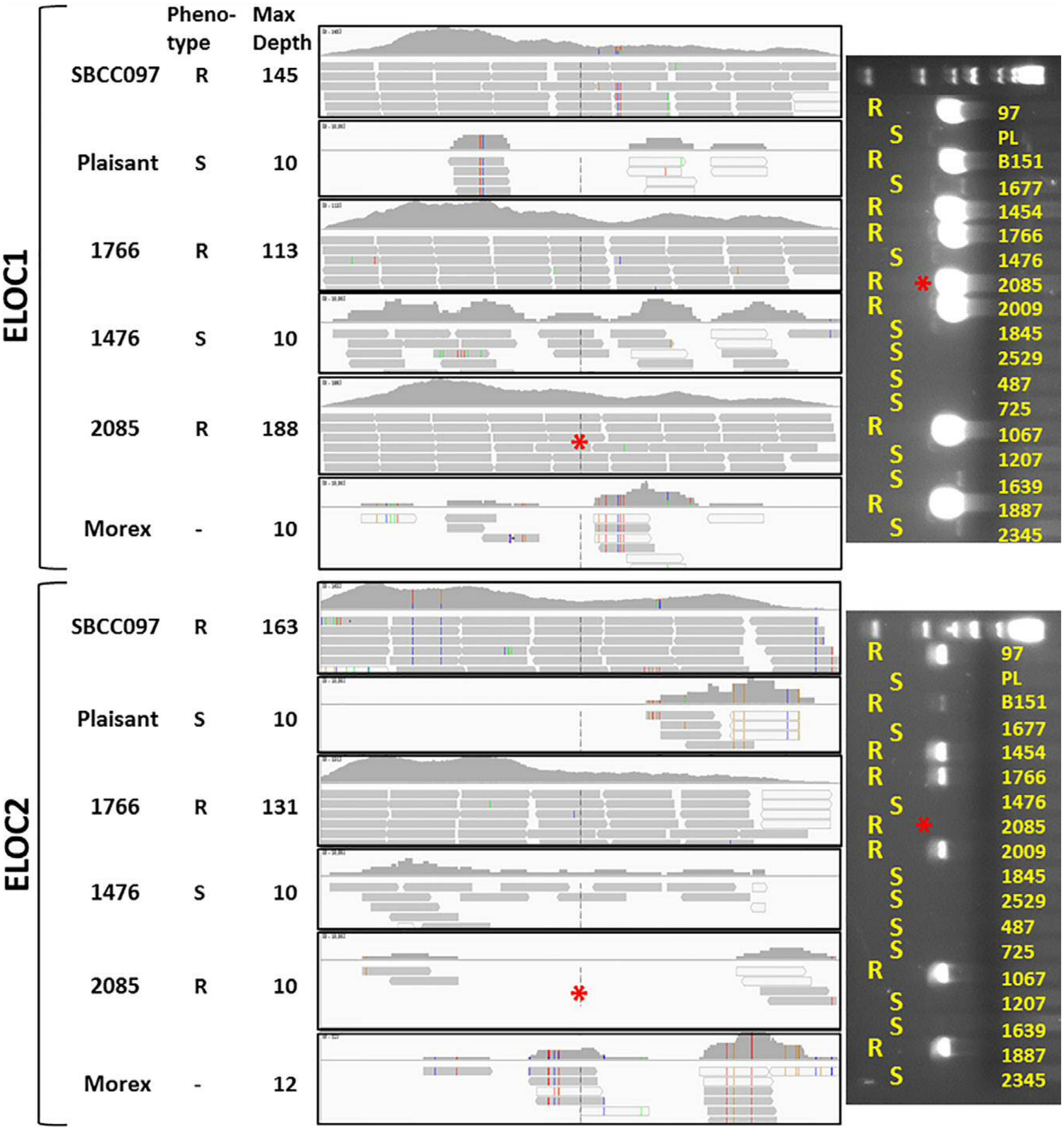


Fig. 5. Presence-absence genotypes for ELOC1 (top) and ELOC2 (bottom). Left: Phenotypes of the two parents, the three sequenced lines, and Morex, along with the maximum depth of coverage (Max Depth) obtained after mapping the exome sequencing reads to ELOC1 and ELOC2 (the two new assembled contigs). Center: Images captured from Integrative Genomics Viewer (IGV), showing the profile of depth of coverage throughout the contigs (top) and individual reads mapped (bottom). Resistant lines have large depths of coverage and similar profiles, covering the whole contigs, with the exception of 2085 in ELOC2 (red asterisk). Susceptible lines have low depth of coverage and irregular, incomplete mapping profiles. Right: Gel electrophoresis of polymerase chain reaction (PCR) amplicons of ELOC1 and ELOC2 for the two parents, the resistant line RIL151 and the 15 BC_1F_4 lines, along with their phenotypes. Resistant lines have presence genotypes whereas susceptible lines have absence genotypes, with the exception of 2085 in ELOC2 (red asterisk). R, resistant; S, susceptible.

cloning of *HvCEN* by Comadran et al. (2012). However, other barley genes have not been cloned yet despite their known phenotypic effect and genetic localization, partly due to the lack of such resources until recently. The continuous improvement of barley physical resources (Ariyadasa et al., 2014; IBGSC, 2012; Mascher et al., 2013b; Muñoz-Amatriaín et al., 2015) allows the adoption of more efficient methodologies for genetic studies involving high-throughput genotyping, marker development, gene discovery, expression analysis, synteny and genome comparative studies. The exome capture probe set developed by Mascher et al. (2013a) for barley is already being used for gene cloning purposes. Mascher et al. (2014) used it to identify *HvMND*, a gene that regulates the rate of leaf initiation, and Pankin et al. (2014) to identify a candidate for *HvPHYC*. In both cases, exome capture was performed on bulked plants with extreme phenotypes from BC_1F_2 populations between mutants and the wild-type.

In this work, the same exome capture probe set was used to sequence three recombinant lines for a powdery mildew resistance QTL. The resistance allele was contributed by a Spanish landrace, showing a wide resistance profile (resistance to 23 out of 27 isolates tested) after a thorough disease survey (Silvar et al., 2011) with the accessions from the SBCC (Igartua et al., 1998). Such line had two QTL conferring race-specific resistances on chromosome 7H (Silvar et al., 2010). The mechanism of resistance of this line was classified as consistent with "intermediate-acting" genes, governing resistance mainly at the postpenetration stage (Silvar et al., 2013a). Genomic approaches allowed the development of new markers to narrow down the QTL intervals (Silvar et al., 2012, 2013b), but were insufficient to definitely locate a manageable physical location or a set of candidate genes for the stronger QTL on 7HL, which is the subject of this work.

From that point, a large F_2 population was created and screened with markers from those previous studies, aiming to identify recombinant lines to further narrow down the QTL interval. The final interval, just 0.07 cM wide, was apparently small enough to land on potential candidates, as this size is comparable with other intervals used in successful gene cloning attempts in barley (reviewed in Krattinger et al., 2009). Again, the analysis of available genomic resources was insufficient to locate candidate genes or to delimit the resistance to a single physical contig. Although the markers were found in the Morex WGS assembly and a POPSEQ map position could be assigned to them, many other Morex WGS contigs with positions within the QTL interval were identified, leading to a large list of annotated genes. Moreover, since the current barley maps are incomplete, additional contigs could have gone unnoticed. Finally, since not all the contigs to which the markers hit were anchored to physical contigs, the physical localization of the QTL remained unknown. An additional challenge was the search of genetic markers from previous studies in the reference. Several of the markers were only found through the analysis of chimeras from GMAP alignments, likely due to the fragmented nature of the Morex WGS assembly.

Exome sequencing of the parents and three recombinant lines allowed the identification of abundant polymorphic variants. This is a faster and more powerful alternative to the search of markers by in silico comparison of genomic resources from different genotypes or by extrapolation of markers from other populations, since many of these are not necessarily polymorphic between the parental lines of the population under study. However, in this work, most of the homozygous SNPs were located outside the QTL. Only a single Pentatricopeptide-repeat containing protein was easily identified within the QTL region, and its corresponding Morex WGS contig lacked physical anchoring. Despite that, the analysis of the profile of variants along the physical contigs in the region was enough to point toward a single FPC which could contain entirely the QTL. This highlights the usefulness of exome sequencing for fine mapping purposes. However, this work demonstrates the technical challenges encountered. Some positions of Morex WGS contigs were not in agreement with the genotypes of our lines. Differences in collinearity between several genetic maps and the POPSEQ reference have been already described (Cantalapiedra et al., 2015; Silvar et al., 2015). These incongruences are important for fine mapping purposes. A single physical contig holding the resistance locus was identified only after removing the Morex WGS contigs not associated to physical positions and using a score to average together the genotypes of the variants within each Morex WGS contig.

Despite the scarcity of homozygous SNPs found within the QTL region, we observed abundant heterozygous SNPs which were polymorphic between the parents as PAV. Although the work with SNPs and small indels is rather straightforward, working with other kinds of variation such as copy-number variation (CNV) or PAV requires using alternative approaches, for example analyzing mapping depth (Mascher et al., 2014). In this work, HMs are defined as those producing heterozygous variants probably due to the collapse of reads from paralogous genes absent in the reference genome. This phenomenon has been recently described among homeologous genes in an exome sequencing experiment in wheat (King et al., 2015). In studies focused on variant discovery, HMs can confound the discrimination of true variants at a given locus. However, this study used HMs to identify the regions with polymorphic HMs, through k-mer analysis, to further assemble different paralogous genes and assess their expression. Though this approach aimed to locate regions with HMs, k-mer abundance could be directly used for genotyping purposes. As with CNV, analysis of HMs is related to the number of copies of a given sequence. However, the analysis of CNV through mapping depth should cope with the different efficiencies in the hybridization and PCR amplification steps during exome sequencing when the sequences are different. In contrast, the analysis of k-mer abundance has the drawback of being unable to differentiate the copies when they are identical to each other. In addition, analysis of HMs could provide insights into the loci and gene families for which the reference genome is incomplete or shows larger variation between different genotypes. Finally, we genotyped the HMs as PAV polymorphisms by means of template-guided assembly and clustering of the resulting sequence contigs. An alternative approach would be to directly compare the presence or absence of the individual k-mers mapping to a given position in the genotypes, although this would not provide assembled contigs. In both cases, the main difficulty resides in differentiating between orthologous and paralogous genes, allelic variants and isoforms (Kuang et al., 2004; Seeholzer et al., 2010), either when clustering the contigs from the assembly or when considering that all orthologous k-mers from the different genotypes are mapping to the same reference locus, and not to another closely related one. In any case, the methods used in this study were implemented from standard tools which were combined to accomplish our specific goals, and thus could be further developed and optimized to cope with peculiarities of HMs.

Both the analysis of the sequenced BACs and the genotyping of HMs pointed toward a cluster of related NBS-LRR genes in the resistance locus. These are good candidates for a resistance gene, although we have to be aware that the sequences captured are limited by the baits used and it cannot be ruled out that the actual resistance gene is absent from the capture reactions and/or from the reference genome. NBS-LRR genes are abundant in many plant genomes and are often organized in clusters of one or more groups of related paralogous genes (Michelmore and Meyers, 1998), which makes their assembly difficult. This problem was evident in this study as revealed by the huge difference in size, number and composition of contigs in equivalent sequenced BACs from independent assemblies (e.g., M01 from IBGSC and I11 from UCR). In addition, a common trend observed in NBS-LRR genes in grasses is the rapid expansion and loss of members from those groups (Li et al., 2010; Yang et al., 2013), leading to PAV and CNV between genotypes. Genes found in that region in Morex were poorly annotated and most of them were split into different WGS contigs. Therefore, the exact number and structure of the genes in this cluster remains unknown both in cultivar Morex and in the resistant line SBCC097. In our assembly, the NBS-LRR genes were incomplete, lacking the NBS domains. We do not know whether these genes are actually incomplete or the NBS domains do exist but were not captured. Lack of exome capture reads covering the genes completely, for instance due to the presence of large introns in them, could lead to incomplete assemblies. Nonetheless, the NBS domains are usually more conserved than the LRR ones (Meyers et al., 1999; Pan et al., 2000; Seeholzer et al., 2010), and this could hinder the independent assembly of the different paralogous genes.

This study made extensive use of state-of-the-art genomic resources available for barley. Several aspects which could be considered when working with these resources arise from our analysis. We have already mentioned some of them, like the lack of position of many Morex WGS contigs or the incomplete annotation of genes in the region. Regarding contig positions, we describe the combined use of both POPSEQ map of Morex WGS contigs and their anchoring to BACs to obtain as many sequences as possible close to our resistance locus. Additional information from the recent publication of sequenced BACs from UCR, a different assembly to that of IBGSC, allowed to complete the MTP of the region and confirmed the features identified using IBGSC data. Furthermore, it highlighted the discrepancies between assemblies, even when corresponding to the same barley genotype, at least in regions with repetitive sequences like the clustered NBS-LRR genes and transposons found in our region.

Finally, identification of the full sequence at these loci would require obtaining BAC libraries and the use of long-read sequencing technologies. Sequencing the whole region could reveal candidate genes which have gone unnoticed, and it could contribute to the understanding of structure and diversification of NBS-LRR genes. Furthermore, sequencing the region, which is rich in resistance genes in barley, could help identifying other resistances. For example, *Mlf* (Schönfeld et al., 1996), which has been associated to this region previously (Backes et al., 2003), given the close physical location of its linked RFLP probe to our QTL. Although BAC libraries are available for cultivar Morex and a few more accessions, this is still not the case for most barley genotypes. Until those resources are available, the exploitation of exome capture to assemble reads from HMs was used in this study to identify candidates not present in the reference or in the exome capture target space, through similarity with closely related genes.

Accession Numbers

The NGS data for both parents and recombinant lines are accessible at European Nucleotide Archive under project no. PRJEB11739.

Supplemental Information Available

Supplemental information is included with this article. Supplement Part 1: Supplemental Materials and Methods, Supplemental Tables S1 to S3, and Supplemental Figures S1 to S13. Supplement Part 2: Spreadsheets, multiple alignments, scripts.

Acknowledgments

The authors would like to thank Dr. Nils Stein, Dr. Robbie Waugh, and IBGSC members for allowing early use of barley exome capture platform. We are also grateful to Dr. Nils Stein and Dr. Uwe Scholz for allowing the use of computational infrastructure at IPK, Dr. María Muñoz-Amatriaín for her hints to work with UCR BACs, and Drs. Kerstin Flath and Doris Kopahnke for providing and multiplying mildew isolates. This work was funded by Spanish projects AGL2010-21929, AGL2013-48756, RFP2012-00015-00-00, and EUI2009-04075 (national code for Plant-KBBE project ExpResBar). CPC was funded by Spanish MINECO (Grant no. BES-2011-045905 linked to project AGL2010-21929).

References

Ames, N., A. Dreiseitl, B.J. Steffenson, and G.J. Muehlbauer. 2015. Mining wild barley for powdery mildew resistance. Plant Pathol. 64(6):1396–1406. doi:10.1111/ppa.12384

Ariyadasa, R., M. Mascher, T. Nussbaumer, D. Schulte, Z. Frenkel, N. Poursarebani, et al. 2014. A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. Plant Physiol. 164:412–423. doi:10.1104/pp.113.228213

Backes, G., L.H. Madsen, H. Jaiser, J. Stougaard, M. Herz, V. Mohler, et al. 2003. Localisation of genes for resistance against *Blumeria graminis* f.sp. *hordei* and *Puccinia graminis* in a cross between a barley cultivar and a wild barley (*Hordeum vulgare* ssp. *spontaneum*) line. Theor. Appl. Genet. 106:353–362. doi:10.1007/s00122-002-1148-1

Bailey, T.L., and M. Gribskov. 1998. Combining evidence using p-values: Application to sequence homology searches. Bioinformatics 14:48–54. doi:10.1093/bioinformatics/14.1.48

Brueggeman, R., A. Druka, J. Nirmala, T. Cavileer, T. Drader, N. Rostoks, et al. 2008. The stem rust resistance gene *Rpg5* encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. Proc. Natl. Acad. Sci. USA 105:14,970–14,975. doi:10.1073/pnas.0807270105

Bulgarelli, D., C. Biselli, N.C. Collins, G. Consonni, A.M. Stanca, P. Schulze-Lefert, et al. 2010. The CC-NB-LRR-Type *Rdg2a* resistance gene confers immunity to the seed-borne barley leaf stripe pathogen in the absence of hypersensitive cell death. PLoS ONE 5(9):e12599. doi:10.1371/journal.pone.0012599

Büschges, R., K. Hollrichner, R. Panstruga, G. Simons, M. Wolter, et al. 1997. The barley *Mlo* gene: A novel control element of plant pathogen resistance. Cell 88:695–705. doi:10.1016/S0092-8674(00)81912-1

Cantalapiedra, C.P., R. Boudiar, A.M. Casas, E. Igartua, and B. Contreras-Moreira. 2015. BARLEYMAP: Physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. Mol. Breed. 35:13. doi:10.1007/s11032-015-0253-1

Comadran, J., B. Kilian, J. Russell, L. Ramsay, N. Stein, M. Ganal, et al. 2012. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. Nat. Genet. 44:1388–1392. doi:10.1038/ng.2447

Flor, H.H. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9:275–296. doi:10.1146/annurev.py.09.090171.001423

Friedt, W., and F. Ordon. 2007. Molecular markers for gene pyramiding and disease resistance breeding in barley. In: R.K. Varshney and R. Tuberosa, editors, Genomics assisted crop improvement: 2. Genomics applications in crops. Springer, Dordrecht, the Netherlands. p. 81–101.

Grabherr, M., B.J. Hass, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652. doi:10.1038/nbt.1883

Gu, L., W. Si, L. Zhao, S. Yang, and X. Zhang. 2015. Dynamic evolution of NBS-LRR genes in bread wheat and its progenitors. Mol. Genet. Genomics 290:727–738. doi:10.1007/s00438-014-0948-8

Halterman, D., F.S. Zhou, F.S. Wei, R.P. Wise, and P. Schulze-Lefert. 2001. The MLA6 coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J. 25:335–348. doi:10.1046/j.1365-313x.2001.00982.x

Igartua, E., M.P. Gracia, J.M. Lasa, B. Medina, J.L. Molina-Cano, J.L. Montoya, et al. 1998. The Spanish barley core collection. Genet. Resour. Crop Evol. 45:475–481. doi:10.1023/A:1008662515059

International Barley Genome Sequencing Consortium (IBGSC). 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi:10.1038/nature11543

Jensen, H.P., E. Christensen, and J.H. Jørgensen. 1992. Powdery mildew resistance genes in 127 Northwest European spring barley varieties. Plant Breed. 108:210–228. doi:10.1111/j.1439-0523.1992.tb00122.x

Jones, D.G.J., and J.L. Dangl. 2006. The plant immune system. Nature 444:323–329. doi:10.1038/nature05286

Jørgensen, J.H. 1988. *Erysiphe graminis*, powdery mildew of cereals and grasses. In: G.S. Sidhu, editor, Genetics of plant pathogenic fungi. Adv. Plant Pathol. 6:137–157.

Jørgensen, J.H. 1992. Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. Euphytica 63:141–152. doi:10.1007/BF00023919

Jørgensen, J.H., and M. Wolfe. 1994. Genetics of powdery mildew resistance in barley. Crit. Rev. Plant Sci. 13:97–119. doi:10.1080/07352689409701910

Jupe, F., L. Pritchard, G.J. Etherington, K. MacKenzie, P.J. Cock, F. Wright, et al. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics 13:75. doi:10.1186/1471-2164-13-75

Jupe, F., K. Witek, W. Verweij, J. Sliwka, L. Pritchard, G.J. Etherington, et al. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J. 76:530–544. doi:10.1111/tpj.12307

King, R., N. Bird, R. Ramirez-Gonzalez, J.A. Coghill, A. Patil, K. Hassani-Pak, et al. 2015. Mutation scanning in wheat by exon capture and next-generation sequencing. PLoS ONE 10(9):e0137549. doi:10.1371/journal.pone.0137549

Kong, L., Y. Zhang, Z.Q. Ye, X.Q. Liu, S.Q. Zhao, L. Wei, et al. 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35:W345–W349. doi:10.1093/nar/gkm391

Krattinger, S., T. Wicker, and B. Keller. 2009. Map-based cloning of genes in Triticeae (wheat and barley). Chapter 12. In: C. Feuillet and G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer, Dordrecht, the Netherlands. p. 337–357. doi:10.1007/978-0-387-77489-3_12

Kuang, H., S. Woo, B.C. Meyers, E. Nevo, and R.W. Michelmore. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–2894. doi:10.1105/tpc.104.025502

Li, J., J. Ding, W. Zhang, Y. Zhang, P. Tang, J. Chen, et al. 2010. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. Mol. Genet. Genomics 283:427–438. doi:10.1007/s00438-010-0527-6

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Maekawa, T., T.A. Kufer, and P. Schulze-Lefert. 2011. NLR functions in plant and animal immune systems: So far and yet so close. Nat. Immunol. 12:817–826. doi:10.1038/ni.2083

Marone, D., M. Russo, G. Laidò, A. De Leonardis, and A. Mastrangelo. 2013. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. Int. J. Mol. Sci. 14:7302–7326. doi:10.3390/ijms14047302

Mascher, M., M. Jost, J.E. Kuon, A. Himmelbach, A. Aßfalg, S. Beier, et al. 2014. Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol. 15:R78. doi:10.1186/gb-2014-15-6-r78

Mascher, M., G.J. Muehlbauer, D.S. Rokhsar, J. Chapman, J. Schmutz, K. Barry, et al. 2013b. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). Plant J. 76:718–727. doi:10.1111/tpj.12319

Mascher, M., T.A. Richmond, D.J. Gerhardt, A. Himmelbach, L. Clissold, D. Sampath, et al. 2013a. Barley whole exome capture: A tool for genomic research in the genus *Hordeum* and beyond. Plant J. 76:494–505. doi:10.1111/tpj.12294

Matsumoto, T., T. Tanaka, H. Sakai, N. Amano, H. Kanamori, K. Kurita, et al. 2011. Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol. 156:20–28. doi:10.1104/pp.110.171579

Mayer, K.F.X., M. Martis, P.E. Hedley, H. Simková, H. Liu, J.A. Morris, et al. 2011. Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263. doi:10.1105/tpc.110.082537

McHale, L., X. Tan, P. Koehl, and R. Michelmore. 2006. Plant NBS-LRR proteins: Adaptable guards. Genome Biol. 7:212. doi:10.1186/gb-2006-7-4-212

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110

Meyers, B.C., A.W. Dickerman, R.W. Michelmore, S. Sivaramakrishnan, B.W. Sobral, and N.D. Young. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. 20:317–332. doi:10.1046/j.1365-313X.1999.t01-1-00606.x

Meyers, B.C., A. Kozik, A. Griego, H. Kuang, and R.W. Michelmore. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15:809–834. doi:10.1105/tpc.009308

Michelmore, R.W., and B.C. Meyers. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113–1130. doi:10.1101/gr.8.11.1113

Monosi, B., R.J. Wisser, L. Pennill, and S.H. Hulbert. 2004. Full-genome analysis of resistance gene homologues in rice. Theor. Appl. Genet. 109:1434–1447. doi:10.1007/s00122-004-1758-x

Muñoz-Amatriaín, M., S. Lonardi, M.C. Luo, K. Madishetty, J.T. Svensson, M.J. Moscou, et al. 2015. Sequencing of 15,622 gene-bearing

BACs clarifies the gene-dense regions of the barley genome. Plant J. 84:216–227. doi:10.1111/tpj.12959

Pan, Q., J. Wendel, and R. Fluhr. 2000. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J. Mol. Evol. 50:203–213. doi:10.1007/s002399910023.

Pankin, A., C. Campoli, X. Dong, B. Kilian, R. Sharma, A. Himmelbach, et al. 2014. Mapping-by-sequencing identifies *HvPHYTOCHROME C* as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. Genetics 198:383–396. doi:10.1534/genetics.114.165613

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotype-by-sequencing approach. PLoS ONE 7(2):E32253. doi:10.1371/journal.pone.0032253

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277. doi:10.1016/S0168-9525(00)02024-2

Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz et al. 2011. Integrative genomics viewer. Nat. Biotech. 29:24–26. doi:10.1038/nbt.1754

Schönfeld, M., A. Ragni, G. Fischbeck, and A. Jahoor. 1996. RFLP mapping of three new loci for resistance genes to powdery mildew (*Erysiphe graminis* f. sp. *hordei*) in barley. Theor. Appl. Genet. 93:48–56. doi:10.1007/BF00225726

Schweizer, P. 2014. Host and nonhost response to attack by fungal pathogens. Chapter 11. In: J. Kumlehn and N. Stein, editors, Biotechnological approaches to barley improvement. Springer, Berlin, Germany. p. 197–235. doi:10.1007/978-3-662-44406-1_11

Seeholzer, S., T. Tsuchimatsu, T. Jordan, S. Bieri, S. Pajonk, W. Yang, et al. 2010. Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. Mol. Plant Microbe Interact. 23:497–509. doi:10.1094/MPMI-23-4-0497

Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539. doi:10.1038/msb.2011.75

Silvar, C., H. Dhif, E. Igartua, D. Kopahnke, M.P. Gracia, J.M. Lasa, et al. 2010. Identification of quantitative trait loci for resistance to powdery mildew in a Spanish barley landrace. Mol. Breed. 25:581–592. doi:10.1007/s11032-009-9354-z

Silvar, C., K. Flath, D. Kopahnke, M.P. Gracia, J.M. Lasa, A.M. Casas, et al. 2011. Analysis of powdery mildew resistance in the Spanish barley core collection. Plant Breed. 130:195–202. doi:10.1111/j.1439-0523.2010.01843.x

Silvar, C., D. Kopahnke, K. Flath, A. Serfling, D. Perovic, A.M. Casas, et al. 2013a. Resistance to powdery mildew in one Spanish barley landrace hardly resembles other previously identified wild barley resistances. Eur. J. Plant Pathol. 136:459–468. doi:10.1007/s10658-013-0178-7

Silvar, C., M.M. Martis, T. Nussbaumer, N. Haag, R. Rauser, J. Keilwagen, et al. 2015. Assessing the barley genome zipper and the genomic resources for breeding purposes. Plant Gen. 8(3) doi:10.3835/plantgenome2015.06.0045

Silvar, C., D. Perovic, T. Nussbaumer, M. Spannagl, B. Usadel, A. Casas, et al. 2013b. Towards positional isolation of three quantitative trait loci conferring resistance to powdery mildew in two Spanish barley landraces. PLoS ONE 8(6):e67336. doi:10.1371/journal.pone.0067336

Silvar, C., D. Perovic, U. Scholz, A.M. Casas, E. Igartua, and F. Ordon. 2012. Fine mapping and comparative genomics integration of two quantitative trait loci controlling resistance to powdery mildew in a Spanish barley landrace. Theor. Appl. Genet. 124:49–62. doi:10.1007/s00122-011-1686-5

Spoel, S.H., and X. Dong. 2012. How do plants achieve immunity? Defence without specialized immune cells. Nat. Rev. Immunol. 12:89–100. doi:10.1038/nri3141

Tan, X., B.C. Meyers, A. Kozik, M.A. West, M. Morgante, et al. 2007. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol. 7:56. doi:10.1186/1471-2229-7-56

Torp, J., H.P. Jensen, and H.J. Jorgensen. 1978. Powdery mildew resistance genes in 106 northwest European spring barley varieties. In: Yearbook. Royal Veterinary and Agricultural University, Copenhagen, Denmark. p. 75–102.

Trevaskis, B., M.N. Hemming, W.J. Peacock, and E.S. Dennis. 2006. *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. Plant Physiol. 140:1397–1405. doi:10.1104/pp.105.073486

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, et al. 2012. Primer3—New capabilities and interfaces. Nucleic Acids Res. 40(15):e115. doi:10.1093/nar/gks596

van Ooijen, J.W. 2006. JoinMap®4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.

Verstegen, H., O. Köneke, V. Korzun, and R. von Broock. 2014. The world importance of barley and challenges to further improvements. In: Kumlehn, J., and N. Stein, editors, Biotechnological approaches to barley improvement. Springer, Berlin, Germany. p. 3–19. doi:10.1007/978-3-662-44406-1

Wei, F., K. Gobelman-Werner, S.M. Morroll, J. Kurth, L. Mao, R. Wing, et al. 1999. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929–1948.

Wei, F., R.A. Wing, and R.P. Wise. 2002. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14:1903–1917. doi:10.1105/tpc.002238

Wu, T.D., and C.K. Watanabe. 2005. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875. doi:10.1093/bioinformatics/bti310

Yang, S., J. Li, X. Zhang, Q. Zhang, J. Huang, J.Q. Chen, et al. 2013. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc. Natl. Acad. Sci. USA 110:18572–18577. doi:10.1073/pnas.1318211110

Yue, J., B.C. Meyers, J. Chen, D. Tian, and S. Yang. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol. 193:1049–1063. doi:10.1111/j.1469-8137.2011.04006.x

Zhang, Z., C. Henderson, E. Perfect, T.L.W. Carver, B.J. Thomas, P. Skamnioti, et al. 2005. Of genes and genomes, needles and haystacks: *Blumeria graminis* and functionality. Mol. Plant Pathol. 6:561–575. doi:10.1111/j.1364-3703.2005.00303.x

Zhou, F.S., J.C. Kurth, F.S. Wei, C. Elliott, G. Vale, N. Yahiaoui, et al. 2001. Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 13:337–350. doi:10.1105/tpc.13.2.337

Zhou, T., Y. Wang, J. Chen, H. Araki, Z. Jing, K. Jiang, et al. 2004. Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol. Genet. Genomics 271:402–415. doi:10.1007/s00438-004-0990-z

Large Differences in Gene Expression between Elite Barley Cultivar Scarlett and a Spanish Landrace under Drought and Heat Stress

Carlos P. Cantalapiedra[1], María J. García-Pereira[1], María P. Gracia[1], Ernesto Igartua[1], Ana M. Casas[1], Bruno Contreras-Moreira[1, 2*]

[1]Estación Experimental de Aula Dei, CSIC, Spain, [2]Fundación ARAID, Spain

Submitted to Journal:
Frontiers in Plant Science

Specialty Section:
Crop Science and Horticulture

Article type:
Original Research Article

Manuscript ID:
225909

Received on:
17 Aug 2016

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

EI, PG, AC, conceived the experiment and designed the greenhouse and growth chamber experiment. CC, AC, and BC designed the sequencing experiments. CC and AC grew and dissected the plants, made physiological measurements and extracted RNA. CC, MG and BC analyzed RNAseq data. CC and AC performed RT-qPCR experiments. All authors read and approved the final manuscript.

Keywords

barley, Landrace, drought, heat, transcriptome profiling, Gene Expression, RNAseq

Abstract

Word count: 221

Drought causes important losses in crop production every season. Improvement for drought tolerance could take advantage of the diversity held in germplasm collections, much of which has not been incorporated yet into modern breeding. Spanish landraces constitute a promising resource for barley breeding, as they were widely grown until last century and still show good yielding ability under stress. Here, we study the transcriptome expression landscape two genotypes, an outstanding Spanish landrace-derived inbred line (SBCC073) and a modern cultivar (Scarlett). Gene expression of adult plants after prolonged stresses, either drought or drought combined with heat, was monitored. Transcriptome of mature leaves presented little changes under severe drought, whereas abundant gene expression changes were observed under combined mild drought and heat. Developing inflorescences of SBCC073 exhibited mostly unaltered gene expression, whereas numerous changes were found in the same tissues for Scarlett. Genotypic differences in physiological traits and gene expression patterns confirmed the superior behavior of landrace SBCC073 under abiotic stress. A comparison with related studies in barley, addressing gene expression responses to drought, revealed common biological processes, but moderate agreement regarding individual differentially expressed transcripts. Special emphasis was put in the search of co-expressed genes and underlying common regulatory motifs. Overall, 11 transcription factors were identified, and one of them matched cis-regulatory motifs discovered upstream of co-expressed genes involved in those responses.

Funding statement

This work was funded by DGA - Obra Social La Caixa [grant number GA-LC-059-2011] and by the Spanish Ministry of Economy and Competitiveness [projects AGL2010-21929, RFP-2012-00015-00-00 and AGL2013-48756-R]. Carlos P. Cantalapiedra is funded by [grant BES-2011-045905 linked to project AGL2010-21929].

Ethics statement

(Authors are required to state the ethical considerations of their study in the manuscript including for cases where the study was exempt from ethical approval procedures.)

Did the study presented in the manuscript involve human or animal subjects: No

Large Differences in Gene Expression between Elite Barley Cultivar Scarlett and a Spanish Landrace under Drought and Heat Stress

Carlos P Cantalapiedra[1†], María J García-Pereira[1†], M Pilar Gracia[1], Ernesto Igartua[1], Ana M Casas[1] and Bruno Contreras-Moreira[1,2*]

[1] Department of Genetics and Plant Production, Estación Experimental de Aula Dei (EEAD-CSIC), Zaragoza, Spain
[2] Fundación ARAID, Zaragoza, Spain

* Correspondence:
Bruno Contreras-Moreira
bcontreras@eead.csic.es

Keywords: barley, landrace, drought, heat, transcriptome profiling, gene expression, RNAseq

Total words: 9071

Total figures: 8

Total tables: 6

Abstract

Drought causes important losses in crop production every season. Improvement for drought tolerance could take advantage of the diversity held in germplasm collections, much of which has not been incorporated yet into modern breeding. Spanish landraces constitute a promising resource for barley breeding, as they were widely grown until last century and still show good yielding ability under stress. Here, we study the transcriptome expression landscape two genotypes, an outstanding Spanish landrace–derived inbred line (SBCC073) and a modern cultivar (Scarlett). Gene expression of adult plants after prolonged stresses, either drought or drought combined with heat, was monitored. Transcriptome of mature leaves presented little changes under severe drought, whereas abundant gene expression changes were observed under combined mild drought and heat. Developing inflorescences of SBCC073 exhibited mostly unaltered gene expression, whereas numerous changes were found in the same tissues for Scarlett. Genotypic differences in physiological traits and gene expression patterns confirmed the superior behavior of landrace SBCC073 under abiotic stress. A comparison with related studies in barley, addressing gene expression responses to drought, revealed common biological processes, but moderate agreement regarding individual differentially expressed transcripts. Special emphasis was put in the search of co–expressed genes and underlying common regulatory motifs. Overall, 11 transcription factors were identified, and one of them matched *cis*–regulatory motifs discovered upstream of co–expressed genes involved in those responses.

1 Introduction

Barley (*Hordeum vulgare* L.) is the fourth cereal crop in relevance worldwide. Like most crops, its production is affected by environmental stresses, drought being the most important among them (Cattivelli et al., 2008). Drought is already prominent at several major agricultural areas throughout the world (Luck et al., 2015), and its effects are predicted to worsen due to growing water demand, shrinking water supply and increased seasonal variability (Barnabas et al., 2008; Luck et al., 2015). An increment of overall temperature is also expected (Barnabas et al., 2008; IPCC, 2014). Actually, many stresses often occur in combination, as is the case of drought and heat, thus being more harmful (Challinor et al., 2014; Mickelbart et al., 2015). However, modern breeding has been directed mainly towards increasing yield, without considering yield stability as a major goal (Mittler, 2006). Therefore, attention is growing towards minimizing the gap between yields under optimal and stress conditions (Cattivelli et al., 2008), to cope with current yield variability (Keating et al., 2010), and to contribute to adaptation to global change (Challinor et al., 2014).

An appropriate strategy to achieve this goal is the exploitation of genetic diversity not yet incorporated into elite cultivars (Dwivedi et al., 2016). As in other crops, current barley cultivars exhibit a narrower genetic basis than wild progenitors (*Hordeum vulgare* ssp. *spontaneum*) and landraces, which are the primary source of useful genes for breeding programs (Fischbeck, 2003; Dawson et al., 2015). Furthermore, in environments with low productivity, landraces and old cultivars often outperform modern genotypes (Ceccarelli et al., 1998; Pswarayi et al., 2008; Yahiaoui et al., 2014). In comparison with wheat, barley has been grown in a wider range of environmental conditions, and is the predominant crop in marginal areas with little precipitation. Accordingly, it is sown in large expanses of the Mediterranean–climatic regions (Ceccarelli, 1994; Ryan et al., 2009), where drought can occur at any moment during the life cycle of crops, being particularly frequent during the terminal stages (Turner, 2004), when different components of grain yield can be largely influenced (Fischer and Turner, 1978; Saini and Westgate, 1999; Araus et al., 2002). Therefore, barley landraces adapted to such conditions could bear genes useful for breeding programs aiming to obtain better yields under drought.

Technical advances in the last decade have potential to improve crop breeding processes (Rivers et al., 2015). High throughput sequencing technologies are providing new powerful tools to study the association between plant genotypic and phenotypic variation (Varshney et al., 2014; Dawson et al., 2015). One of these, RNAseq (Mortazavi et al., 2008), is currently employed with different aims in crop genetics, like polymorphism detection and transcript profiling (Varshney et al., 2009). The latter can be used to analyze gene expression networks involved in different processes; for example, those related with resistance to abiotic stresses. However, analyses of *cis*–regulatory elements of transcription factors (TFs) and of promoters of genes involved in a given response have been rare in barley, likely due to the absence of adequate genomics resources.

In this work, two contrasting barley genotypes were subjected to prolonged water deficit, either alone or combined with heat. Spanish barley landrace SBCC073 was the best yielding genotype, among 159 landraces and 25 old and modern cultivars, in field trials in Spain in which average yield was below 3 t ha^{-1} (Yahiaoui et al., 2014). Here, it was compared to a modern cultivar, Scarlett, sensitive to water stress (Sayed et al., 2012). *De novo* assemblies of transcriptomes of both genotypes were obtained and gene expression changes evaluated both in developing inflorescences and leaves. Metabolic pathways, biological processes, molecular functions, co–expression clusters and *cis*–regulatory elements of drought–modulated genes are reported.

2 Materials and Methods

2.1 Plant material and drought experiments

Seeds of Spanish barley landrace SBCC073 (http://www.eead.csic.es/EEAD/barley/core.php?var=73) and of cultivar Scarlett were sown. Seedlings were allowed to grow for one week and then were vernalized for 24 days, in order to synchronize flowering. At the end of the vernalization period, plants at the 3-leaf stage were transferred to 28.0 x 20.8 cm (height x diameter) black plastic pots (one seedling per pot) with standard substrate made of peat, fine sand and perlite Europerl B-10 (Europerlita Española SA, Barcelona, Spain), from a mix with 46 kg, 150 kg and 50 L, respectively. Two series of pots were placed in a greenhouse (natural photoperiod, controlled maximum temperature 28ºC, average daily temperature 25±2°C during the day and 21±3°C at night) and in a growth chamber (16h light / 8h dark, 21 ºC daytime / 18 ºC night temperature). Additional pots filled only with substrate were used to estimate dry weight and field capacity (FC). Soluble fertilizer was provided with irrigation. Plants were treated with fungicide (Triadimenol 25%) to prevent powdery mildew build-up.

Drought treatments started 30 days after transplant at the end of the vernalization period. Water application was not interrupted abruptly. Instead, it was gradually reduced to resemble a slow drying soil, based on weight of each pot relative to the estimated FC. Pots were weighted, watered, rotated and their positions swapped every two days. Once the target fraction of FC was reached, the pots were watered to keep such weight constant. Treatment levels in the growth chamber were 70% and 20% FC, whereas an intermediate level of 50% FC was applied in the greenhouse. At the sampling date, all plants in the water-stress treatments had been at the target fraction of FC for at least 14 days. Temperature and relative humidity in the greenhouse were automatically recorded.

2.2 Measurement of phenotypic traits

Several traits were recorded 60 days after transplant. Leaf water potential (LWP) in leaves was measured at noon using a Scholander chamber (SF-PRES-70, Solfranc Tecnologías SL, Vila-Seca, Spain). Stomatal conductance (SCo) was measured, starting at 9 am, using a leaf porometer (Decagon Devices, Pullman, WA, USA). Relative water content (RWC) was also estimated, as described in Talame et al. (2007). For each plant, three independent measurements were taken for LWP, SCo and RWC. In addition, tiller number (TN) and number of tillers reaching at least Zadoks stage 49 (Zadoks et al., 1974), i.e., visibly emerging spikes (VSN) were counted. All measures were taken at two biological replicates.

2.3 RNA extraction and transcriptome sequencing

Two tissues, young inflorescences and leaves (including last expanded leaves and flag leaves), were sampled at 60 days after transplant. Fresh material was harvested and frozen in liquid N_2 before RNA extraction with the NucleoSpin® RNA Plant kit (Macherey-Nagel, Düren, Germany). RNA quality was assessed with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and with Bioanalyzer 2100 hardware (Agilent, Santa Clara, CA, USA; average RIN: 6.7 for leaves, 8.1 for flowers). Barcoded cDNA libraries were prepared at CNAG (Barcelona, Spain) following Illumina TruSeq standard procedures, and eventually sequenced in an Illumina HiSeq2000 sequencer, using a full flow-cell, 4 samples per lane, to produce 2x101 bp paired-end reads. The whole dataset consisted of 2 biological replicates from greenhouse plants (2 tissues x 2 replicates x 2 genotypes), 2 biological replicates of developing inflorescences and 3 biological replicates of leaves from plants subjected to drought and well irrigated plants in the growth chamber (5 x 2 genotypes x 2 treatments).

2.4 RNAseq data preprocessing and transcriptome assembly

Raw reads were sequentially processed with FASTQC v0.10.0 (Andrews, 2010) and Trimmomatic v0.22 (Bolger et al., 2014), discarding stretches of mean Phred score <28 and cropping the first nucleotides to ensure a per–**position A, C, G, T frequency near 0.25. Only reads of length ≥ 80** nucleotides were kept for further analysis. Surviving reads were error–corrected with Musket v1.0.6 (Liu et al., 2013) and default parameters. Then, reads were assembled following two different procedures, *de novo* and reference–guided.

De novo assemblies were obtained using Trinity r2013–02–25 recommended procedures (Haas et al., 2013). First, reads from sample replicates were pooled together and *in silico* normalized, to a maximum coverage of 30. This procedure was repeated with the resulting read sets to obtain, for each genotype, a final set of normalized reads. These were used for *de novo* assembly of SBCC073 and Scarlett transcriptomes.

A reference–guided assembly (RGA) was generated with the Tuxedo pipeline (Trapnell et al., 2012). First, clean reads were mapped to the IBGSC cv. Morex assembly (Mayer et al., 2012) with Tophat2 (v2.0.9; ––b2–very–sensitive, ––b2–scor–min C,–28,0 –read–mismatches 4 –read–gap–length 12 –read–edit–dist 12 –G 21Aug12_Transcript_and_CDS_structure.gff). This mapping procedure was performed in two steps, a first one to exclude reads with multiple mappings to the whole reference assembly (–M, –g 1, ––no–discordant) and a second one to identify reads mapping unambiguously to gene coding loci (–g 2, ––no–discordant, ––no–mixed). Mappings were used as input for Cufflinks (v2.2.1). Individual assemblies were merged with the reference Morex assembly with Cuffmerge.

2.5 Correction, validation and annotation of de novo transcriptomes

Clean reads were mapped back to the *de novo* transcriptomes using Trinity script *alignReads.pl* with Bowtie (Langmead et al., 2009). In addition, the newly assembled isoforms were mapped to Morex, Bowman, Barke WGS (Whole Genome Shotgun) assemblies (Mayer et al., 2012) and Haruna Nijo flcDNAs (Matsumoto et al., 2011) with the script *bmaux_align_fasta* from the Barleymap package (Cantalapiedra et al., 2015) (hierarchical=yes query–mode=cdna thres–id=98 thres–cov=10), keeping together sequences matching the same reference sequence. Sequences in each of these groups were clustered with WCD–express v0.6.3 (Hazelhurst and Liptak, 2011) using threshold=24, which is equivalent to a 98% identity cut–off.

Presence of these isoforms in existing references was further confirmed by aligning them iteratively to additional sequence repositories. These were the Haruna Nijo genome assembly (Sato et al., 2016), genome contigs of Chinese Spring wheat (Mayer et al., 2014), barley ESTs from HarvEST assembly 36 (Close et al., 2007), the MIPS repeat database (Nussbaumer et al., 2013), and sequences from *Hordeum*, *Brachypodium*, *Triticum*, *Oryza* or *Aegilops* in the nt NCBI database (ftp.ncbi.nlm.nih.gov/blast/db). Alignment to Morex, Bowman and Barke WGS assemblies, and to Haruna Nijo genome and flcDNAs was repeated with a more stringent coverage threshold (thres–cov=80). Finally, transcripts were scanned for the presence of sequencing vectors by comparison with the EMVec database (ftp://ftp.ebi.ac.uk/pub/databases/emvec/) and as a result 64 sequences were removed.

Gene annotation of assembled contigs was performed with the script *transcripts2cdsCPP.pl* (–n 50) from GET_HOMOLOGUES–EST (v 04052016, https://github.com/eead–csic–compbio/get_homologues), which uses Transdecoder (https://transdecoder.github.io/) and blastx alignments to SwissProt proteins to define CDS sequences. Clusters obtained with GET_HOMOLOGUES–EST (get_homologues–est.pl –t 0 –M –S 96 –A –L), requiring percentage sequence identity > 96, were used to obtain reciprocal correspondences between transcripts from SBCC073 and Scarlett assemblies. PFAM domains in translated CDS sequences were also annotated (*get_homologues-est.pl* –D).

2.6 Analysis of gene expression

Differential expression contrasts were performed for each genotype, tissue and treatment; both for isoforms and genes. For this purpose, we compared three different pipelines.

For the first one, estimation of expression levels of isoforms and genes was done with RSEM v.1.2.11 (Li and Dewey, 2011), using Bowtie2 (Langmead and Salzberg, 2012) and otherwise default **parameters. RSEM 'expected counts' were used as input for different**ial expression analyses with the **'glm'** functions of the R (R Development Core Team, 2008) Bioconductor package edgeR v3.8.6 (Robinson et al., 2010) **(false discovery rate function "BH" set to 0.001). A minimum CPM (counts per million) of 0.4, equivalent to around 10 RSEM 'expected counts' based on a linear regression (R–** square = 1, intercept ~ 0, slope = 25), was required in at least half of the samples to include an isoform or a gene in the analysis.

A second method relied on kallisto v0.42.5 (Bray et al., 2016) **to obtain 'expected counts' and to** generate 100 bootstrap samples for each replicate, followed by test for differential expression with sleuth v.0.28.0 Wald test (Pimentel et al., 2016), using the previously generated bootstrap samples.

For the third method, Cuffquant and Cuffdiff v.2.2.1 (Trapnell et al., 2013) were used to test differential expression, with FDR 0.05, on the RGA transcripts.

Principal component analyses (PCA) of the resulting expression estimates from kallisto were done with the function PCA from R package FactoMineR 1.29 (Lê et al., 2008). Correlation analysis was performed using the R package corrplot 0.73 (Wei and Simko, 2014).

2.7 RT–qPCR validation

Reference genes for calculating relative expression were either searched in the literature or selected from our RNAseq data. The latter were those with the smallest coefficient of variation of expression values across samples, among isoforms not reported as differentially expressed (DE) by edgeR. DE isoforms to be checked with RT–qPCR were chosen randomly from bins covering the range of edgeR logFC. All the selected DE isoforms had TPM (transcripts per million) greater than 1. Primers for both reference genes and DE isoforms were designed with Primer Express 3.0 (Applied Biosystems). Conservation of the target sequences was checked in both SBCC73 and Scarlett isoforms. Whenever possible, one of the primers of the pair was set over an exon–exon junction and t**owards the 3' end.**

The same DNase I–treated RNA samples used for RNAseq were utilized for the RT–qPCR assays. First strand cDNA synthesis was made from 2 µg of total RNA to a final volume of 40 µl containing oligo(dT)20 for priming and SuperScript III Reverse Transcriptase (Invitrogen, Cat.No. 18080–044). All the RT–qPCR reactions were performed in an ABI7500 (Applied Biosystems, Foster City, CA, USA) with the following PCR profile: 95ºC 10 min pre–denaturation step; 95ºC 15 sec denaturation and 60ºC 50 sec annealing (40 cycles), followed by a melting curve 60ºC–95ºC default ramp rate. The efficiency of primers was obtained from calibration curves with 1:5 dilution series and at least 4 points fitted in a linear regression with R–square over 0.99. We used NormFinder (Andersen et al., 2004) to analyze the stability value of the reference genes. Relative change of expression was calculated according to Pfaffl (2001), but using the geometric mean of three reference genes as normalization factor (Vandesompele et al., 2002).

2.8 Functional annotation of differentially expressed isoforms

Software CPC (Kong et al., 2007) was used to tag DE isoforms as coding or non–coding, and to obtain Uniref90 best hits. In addition, contained CDS sequences were deduced and PFAM protein domains annotated, as explained earlier for all the isoforms of each transcriptome. GO terms for each DE isoform were obtained with in–house script barleyGO (http://www.eead.csic.es/compbio/soft/barleyGO.tgz). Enrichment tests for PFAM domains and GO terms were performed in R using the Fisher exact test (p–value < 0.01). For the GO terms, we used the R package topGO (Alexa and Rahnenfuhrer, 2016).

DE isoforms were searched in metabolic pathways databases, including KEGG (Kanehisa et al., 2016), PlantReactome (Tello–Ruiz et al., 2016) and PlantCyc (Plant Metabolic Network, 2016). For KEGG, we obtained the list of genes of *Oryza sativa* ("osa"), from which we retrieved Orthology identifiers and pathways. DE isoforms were aligned to those genes with blastn (–perc_identity 75 –num_alignments 1), discarding hits with low query coverage in the alignment ('qcovs' < 70). PlantReactome (file "gene_ids_by_pathway_and_species.tab") was explored with Morex gene identifiers to obtain the pathways involved in differential expression. The gene identifiers were derived from mappings of *de novo* assemblies to the Morex reference genome from the validation step using the Barleymap package, as explained above. In the case of PlantCyc, we obtained the blast set "plantcyc.fasta" and enzymes annotation ("PMN11_June2016/plantcyc_pathways.20160601"), and used a custom script to match annotated enzymes with blastx (–evalue 0.00001 –num_alignments 1), filtering hits with percentage identity ≥ 75. Enzymes and pathways were grouped in broader categories manually, by merging their textual descriptions in KEGG and PlantCyc.

2.9 Comparison with related studies

The literature was surveyed to obtain protein and transcript sequences which had been previously associated with response to water deprivation in barley. These drought–related sequences were aligned with Blast[p|x] to genes from the Haruna Nijo genome assembly, which allowed mapping them to their corresponding DE isoforms from this study.

2.10 Clustering and identification of cis–regulatory elements of co–expressed genes

DE isoforms were clustered based on their TPM values (from kallisto). Distance between each pair of isoforms was calculated with Pearson correlation. This metric was weighted with Euclidean distance, under the hypothesis that isoforms sharing their expression pattern, but differing in magnitude, might have promoters which could be overlooked when clustered together with Pearson correlation only. These distances were used to perform hierarchical clustering (R package hclust, method="complete"). To declare the final number of clusters, the dendrogram was pruned when 95% of clusters had an internal average distance below 0.001% of the initial average distance of all DE isoforms.

The following procedure was used to recover promoter sequences corresponding to the genes present in the expression clusters. DE isoforms from each cluster were mapped to transcripts from the Morex WGS assembly (Blastn –perc_identity 98). For each cluster containing 10 or more genes, repeat–masked promoter sequences (–1000, +200 nucleotides around TSS) were retrieved from the RSAT::Plants server (http://plants.rsat.eu, version Hordeum_vulgare.082214v1.29) (Medina–Rivera et al., 2015). As negative controls, promoter sequences were retrieved from randomly generated gene clusters of the same size. Enrichment in GO terms and motif discovery with oligo–analysis and dyad–analysis were performed following the protocol of (Contreras–Moreira et al., 2016). Motif scores within upstream regions of co–expressed genes and their orthologous genes in *Brachypodium* *distachyon* reference (v1.0.29) (International Brachypodium Initiative, 2010), were obtained with the program matrix-scan from RSAT::Plants. These scores were also calculated for motifs generated by permutation of the bases of each discovered motif. Therefore, two types of evidences were used to assess the reliability of discovered motifs: i) their statistical significance compared to the negative controls, and ii) their matrix-scan scores compared to the scores of permuted motifs. Discovered motifs were annotated by comparison to plant regulatory motifs in the footprintDB repository (Sebastian and Contreras-Moreira, 2014). The highest scoring motif, in terms of footprintDB 'Ncor' score, was selected as the best hit. The full report on the promoter analysis, including source code, is available at http://floresta.eead.csic.es/rsat/data/barley_drought_clusters.

Finally, deduced peptide sequences of DE isoforms annotated as transcription factors with iTAK (http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi), were used to predict their putative DNA-binding motifs with footprintDB.

3 Results

3.1 Growth of Scarlett and SBCC073 plants subjected to drought

Two different experiments were set up, in which plants were placed in a growth chamber or in a greenhouse. The growth chamber was kept at strictly controlled environmental conditions, whereas the greenhouse underwent a natural photoperiod (August – September, 2012, starting with 14 h 23 min and ending with 11 h 46 min daylight, http://www.fomento.gob.es/salidapuestasol/2012/Zaragoza-2012.txt) and controlled, but more variable, temperature and humidity. Both daytime and night temperatures in the greenhouse were higher than in the growth chamber, whereas relative humidity was similar on average (Supplementary Figure S1). In both settings, water stress was imposed after initiation of the stem elongation stage. Growth chamber plants were watered in order to conserve 70% field capacity (FC) (controls, C), or instead subjected to reduced irrigation, up to 20% FC (drought, D). Greenhouse plants were irrigated to an intermediate 50% FC (mild drought and heat, MDH). These experiments are outlined in Figure 1.

Daily loss of water, based on the weights of pots, was largest in C plants, intermediate under MDH and lowest under D (Supplementary Figure S2). The same trend was observed for leaf water potential (LWP), summarized in Figure 1. LWP was proportional to the three imposed water regimes, with plants subjected to drought (D and MDH) showing larger absolute LWP that those well-watered. The largest value corresponded to Scarlett plants under D, in which SBCC073 plants had values comparable to those of both SBCC073 and Scarlett plants under MDH. Likewise, minimum values for stomatal conductance (SCo) were recorded for plants under D (Table 1). However, the largest SCo was found under MDH. Relative water content (RWC) was lowest for plants under D, in both genotypes, whereas under MDH, it was closer to that of C plants in SBCC073, and closer to that of plants under D in Scarlett. Tiller number (TN) and visible spike number (VSN) were also affected by water deprivation, being larger in C than under D, both in SBCC073 and Scarlett. Under MDH, similarly to the RWC observations, TN was less affected in SBCC073 than in Scarlett.

3.2 Assembly and validation of Scarlett and SBCC073 transcriptomes

Sequencing of cDNA libraries, derived from leaf (LF) and young inflorescence (YI) transcripts, yielded 1.18 billion paired-end sequence reads. From this dataset, we assembled separate *de novo* transcriptomes for Scarlett and SBCC073, as well as a reference-guided assembly (RGA) (Supplementary Figure S3).

The *de novo* assemblies yielded similar numbers and lengths of isoforms for both genotypes (Table 2). These sets, with 103,623 genes in SBCC073 and 113,962 in Scarlett, were comparable but larger than the annotated gene sets for the Morex cultivar (Mayer et al., 2012), with 75,258 high and low confidence genes, and with the results from the RGA (75,204 genes). Validation and correction of the *de novo* isoforms was performed in three stages. First, the clean reads were mapped back to the assembled transcripts, to compute the fraction of well aligned pairs of reads (both reads mapped, correct orientation and insert size), which was near 83% for both cultivars. Second, *de novo* subcomponents were revised for re–clustering. This requires some explanation. Whereas RGA contigs are isoforms associated to known genes from the reference, *de novo* assembly generates contigs which are isoforms clustered in so called subcomponents. In some cases, these subcomponents accumulate closely related sequences, for instance from paralogous genes or expressed pseudogenes, which should be separated. Therefore, this second step consisted in re–clustering isoforms from subcomponents to genes, by alignment to annotated references (see Methods), and assigning them to different loci when appropriate. The final number of genes in the *de novo* assemblies was 112,923 in SBCC073 and 123,582 in Scarlett. Third, the isoforms were matched to a variety of genomic and transcriptomic sequence repositories of barley, wheat and other grasses. In total, 93% of SBCC073 and 87% of Scarlett genes could be confirmed. These sequence comparisons are further illustrated in Figure 2. Note that at least 10% alignment coverage was required in all cases. Further, the alignment against Morex, Barke, Bowman and Haruna Nijo was repeated, with a more strict minimum coverage of 80%. This test confirmed that 88,293 (78% of SBCC073) and 92,713 (75% of Scarlett) genes map with high confidence to previously reported barley sequences.

3.3 Analysis of gene expression

Clean paired–end reads were mapped back to SBCC073 and Scarlett assemblies, to estimate expression counts for each transcript. These estimates were subsequently used to identify DE tags (genes and isoforms) between stressed treatments and C, for each tissue and genotype. For this purpose, we compared three different pipelines, which rely on different software for each of the two steps: RSEM–edgeR, kallisto–sleuth and Cuffquant–Cuffdiff. In addition, a set of isoforms from YI were randomly chosen to test their expression by RT–qPCR, using genes selected from the literature and from our RNAseq expression data as references (Supplementary Table S1).

The results of differential expression computed with kallisto–sleuth had the best agreement with those of RT–qPCR (Figure 3). The outcome of the RSEM–edgeR pipeline was comparable to kallisto–sleuth after discarding a few outliers. Moreover, PCA and clustering of samples, using expression estimates from kallisto, showed good correlation between replicates (Supplementary Figures S4 and S5). When the expression estimates, obtained with the three methods, were directly compared, RSEM–edge and kallisto–sleuth showed the best agreement (Supplementary Figures S6–S8, Supplementary Table S2). In order to reduce false positives, final DE tags were obtained from the intersection between those two methods.

Overall, the response differed between genotypes in YI, and between treatments in LF (Figure 4). Under D, we found almost no response in SBCC073, either in YI or LF samples, whereas in Scarlett, YI samples had many DE tags. On the contrary, abundant changes in gene expression were observed under MDH, with the exception of YI from SBCC073, which remained mostly unaltered. Regarding the proportion of up–regulated tags over total DE tags, in LF under MDH it was close to 50%, in both genotypes, whereas in YI from Scarlett plants it increased under D (62.6% in isoforms, 61.4% in genes) and decreased dramatically under MDH. There was high agreement between DE genes and

DE isoforms in all contrasts, aalthough some DE genes were different to those found when analyzing isoforms (Supplementary Table S3). On the other hand, common DE tags between different contrasts were negligible, with the exception of LF under MDH, in which Scarlett and SBCC073 shared a low but sizable fraction (Supplementary Figure S9).

Finally, overall gene expression changes (number of DE tags and cumulative logFC from each contrast) were compared with the physiological measurements. Some large correlations were obtained (Supplementary Table S4), although these results must be considered with care due to the small sample size. For LWP, we found a positive correlation with YI overall logFC of isoforms (r 0.97, p–value 0.03) and number of DE tags (r 0.99, p–value 0.01). SCo exhibited strong positive correlation with gene expression changes in LF (ranges: r 0.95 – 0.98, p–values 0.05 – 0.02), to which VSN showed strong negative correlation (ranges: r –0.91 – –0.96, p–values 0.04 – 0.09).

DE isoforms were annotated combining different strategies, as described in Materials and Methods. The main annotation results are detailed in the following sections, whereas the complete results are provided in Supplementary File S1.

3.4 Differentially expressed isoforms in leaves under drought

As explained in the previous section, just a few isoforms were DE in LF under D. In both genotypes, we found an up–regulated isoform encoding a polyamine oxidase, involved in spermine and spermidine degradation. In addition, an isoform corresponding to a chlorophyll apoprotein from photosystem II was down–regulated in Scarlett. However, this change was not observed in SBCC073, which instead showed induction of transcripts of three proteins, namely ABA/WDS (abscisic acid / water deficit stress) induced protein, ribonuclease T2 and calcineurin–like phosphoesterase. Other DE isoforms were annotated as non–coding or of unknown function.

3.5 Differentially expressed isoforms in leaves under mild drought and heat

There were more DE tags in LF under MDH, and involved a more diverse array of gene functions than under D. The same polyamine oxidase induced in LF under D was also observed up–regulated in Scarlett under MDH. Intriguingly, in SBCC073 we found up–regulated a transcript encoding a spermidine synthase.

Some GO terms were enriched in both genotypes, including "phosphorelay signal transduction system", "pyrimidine–containing co**mpound biosynthesis process", "response to temperature stimulus", "response to water deprivation" and "thiamine biosynthetic process" (**Supplementary File S2). Other pathways and cellular processes involved in the responses of both genotypes were starch phosphorylation, chorismate biosynthesis, L–ascorbate biosynthesis and recycling, DMNT biosynthesis (a volatile homoterpene), and other proteins involved in protein folding, proteolysis and defense response (Figure 5). We also found in both genotypes up–regulation of isoforms annotated as CCA1/LHY MYB–related TF (Supplementary Table S5). Moreover, we found another DE gene annotated as MYB–related TF in both genotypes, which is similar to *Arabidopsis thaliana* TCL2, and an additional uncharacterized MYB–related TF in SBCC073 only. At the same time, down–regulation of other genes related with circadian rhythm was detected, like adagio–like protein 3 and a PRR1 (HvTOC1) transcription regulator. In SBCC073, we found also down–regulation of another circadian clock related gene, annotated as APRR3. Another gene up–regulated in both genotypes was annotated as protein kinase CIPK9. Regarding transporters, repressed transcripts encoding aquaporins were noticed in both genotypes. There were a few other protein domains regulated in both genotypes, most of them repressed.

Differences between genotypes were also seen among DE transcripts in LF under MDH. For instance, in SBCC073 there was enrichment of terms such as “actin filament–based movement”, “ammonium ion metabolic process” and “defense response by cell wall thickening”, while in Scarlett a greater variety of response–related terms were obtained, such as “response to abscisic acid”, “response to bacterium”, “response to ethylene”, “response to hydrogen peroxide” or “response to wounding” (Supplementary File S2). Also, DE isoforms related to glycine betaine biosynthesis and to abscisic acid (ABA) biosynthesis were seen in SBCC073, whereas trehalose biosynthesis was involved in the response of Scarlett LF to MDH (Figure 5). Moreover, isoforms involved in cell wall, epidermis (wax esters) and membrane lipids (glycerophospholipids, ceramide) metabolism were up–regulated in Scarlett but not present among SBCC073 DE isoforms. This was also the case of some defense response metabolic pathways (benzoxazinoids and dhurrin biosynthesis), xanthophylls metabolism, several antioxidation related proteins (like baicalein peroxidase or glutathione S–transferase) or sulphur metabolism related proteins. We also found differences among TFs and protein kinases (PKs) (Supplementary Table S5). For instance, CIPK17 and a C2C2–Dof TF, whose best SwissProt hit is Arabidopsis protein CDF2, were up–regulated, and an AP2/ERF–AP2 TF (related to *Brassica napus* BBM2) down–regulated, all in SBCC073. Instead, repression of a TUB TF, similar to *O. sativa* subsp. *japonica* TULP7, and induction of both a bZIP TF and a jasmonate ZIM TIFY TF, the latter related to *O. sativa* subsp. *japonica* TIFY6B, was noticed in Scarlett. Besides aquaporins, already mentioned, DE isoforms related to transport processes were different between genotypes, being more abundant in Scarlett. These included lipid transfer proteins, phosphate, potassium, triose–phosphate, adenine, vacuolar amino acid and ABC transporters, and a repressed NUCLEAR FUSION DEFECTIVE 4 (NFD4) protein.

3.6 Differentially expressed isoforms in young inflorescences in SBCC073

In YI, the transcriptional responses were markedly different between genotypes, with only minor responses in plants of genotype SBCC073 under both treatments. Indeed, a single down–regulated transcript was identified in SBCC073 under D, annotated as Pollen Ole e 1 allergen/extension. Under MDH, a repressed isoform was annotated as “non–coding”, whereas four up–regulated isoforms corresponded to CCA1/LHY.

3.7 Differentially expressed isoforms in young inflorescences in Scarlett

In contrast with what was seen in SBCC073, YI from Scarlett showed abundant gene expression changes. Enriched GO terms found both under D and under MDH were scarce (Table 3), including cell wall–related processes “beta–glucan biosynthetic process”, “lignin metabolic process”, “phenylpropanoid metabolic process”, and “cell wall organization or biogenesis”, and others like “response to carbon dioxide” and “sucrose metabolic process”. Other shared DE tags included isoforms involved in tetrahydrofolate biosynthesis and a subtilase serine protease (Figure 6). Among DE TFs in YI, we found B3–ARF isoforms (Auxin response factors with B3 and PB1 domains) induced under both treatments (Supplementary Table S6). However, reciprocal alignment revealed that they belong to different genes (blastn, alignment coverage 48% and percentage of identity 63%). The most similar protein of the isoform in the D treatment was ARF21, also known as OsARF7b, whereas the closest homologue of the isoform found under MDH was ARF11.

Besides B3–ARF TFs, only a few other isoforms were up–regulated in Scarlett YI under MDH, corresponding to an elongation factor EF–1, a DNA topoisomerase, a kinesin motor domain, CCA1/LHY, and a condensing complex subunit protein. All the others were down–regulated, whose enriched GO terms included “cellulose biosynthetic process”, “xylan biosynthetic process”, “flavonoid biosynthetic process”, “mitotic chromosome condensation”, “plasmodesmata–mediated intercellular transport” and “mucilage extrusion from seed coat” (Table 3). Other differences with respect to the D treatment were the involvement of enzymes from thiamine biosynthesis, triglyceride catabolism, epoxidation, berberine alkaloid biosynthesis or auxin biosynthesis (Figure 6). Among repressed isoforms related with transporters, we found sugar and lysine-histidine transporters, a PRA1 family protein B2 (a protein family related to regulation of vesicle trafficking, (Kamei et al., 2008), and several ABC transporters (Supplementary Table S6). Other proteins (and protein domains) which were found DE only under MDH included an expansin-B3, a putative cell wall protein, a PMR5/Cas1p, and several germin-like proteins.

Under D, Scarlett YI showed almost twice as many induced than repressed isoforms. The number of enriched GO terms was greater than for all the other contrasts (Supplementary File S2), including numerous enriched processes (Table 3) and metabolic pathways (Figure 6), related with responses to abiotic stress (cell wall thickening, biosynthesis of wax, triglyceride mobilization, expansin-A7), development (seed, embryo and root development), central metabolism (starch, glucose, pyruvate, many amino acids, fatty acids biosynthesis, activation and beta-oxidation), hormones (ethylene, jasmonate), energy (ATP and NADP metabolism related proteins, F and V-type H+-transporting ATPases), nucleic acids and proteins metabolism, antioxidation, proteolysis, protein folding, numerous proteins involved in transport and vesicle trafficking (Supplementary Table S6), tRNA synthetases, an up-regulated MADS-MIKC TF whose best hit in SwissProt is *O. sativa* subsp. *japonica* MADS6, several PKs (like CIPK30) and phosphatases (like phosphoinositide phosphatase SAC7), proteins involved in interactions and signal transduction (SNF2, ASPR1 topless-related protein 1, 14-3-3 protein epsilon, CypP450), cytoskeleton proteins (tubulin, myosin, fimbrin and villin domains), and even processes related with photosynthetic tissues, like biosynthesis of chlorophyll a or tetrapyrrole, or induction of a Rubisco activase.

All these evidences indicate that responses to D and MDH of Scarlett YI were different, and that reproductive tissues were undergoing large gene expression changes, especially under D.

3.8 Comparison with related studies

We surveyed the literature reporting genes and proteins expressed in barley in response to water deprivation. The goal was to compare those sequences to the DE transcripts identified in this work. The studies listed in Table 4 include 5 microarray experiments, 7 based on proteomics, 1 RNAseq study, 1 QTL work, 1 surveying expression QTL and 1 meta-analysis. Most of them focused on barley plants under drought, with a few exceptions. The work “matsumoto2014” surveyed responses to desiccation, salt stress and ABA. In addition, both “ashoub2015” and “rollins2013” combined drought and heat stress. The meta-analysis “shaar-moshe2015” compared drought related genes from different plant species. Although many of these works (9) sampled leaves, other tissues were also analyzed in some of them (mainly shoots, roots, spikes and grain).

Out of 4389 DE tags (proteins, genes and transcripts) reported overall in the studies above, more than half (2730) were barley genes included in the meta-analysis “shaar-moshe2015” and, indeed, that study matches the largest number of DE tags of the current work. However, in relative terms, the most similar were those of “ashoub2013”, “ashoub2015”, “vitamvas2015”, “wang2015”, “kausar2013” and “rollins2013”, in decreasing order, whose DE tags were also found in the present study in proportions ranging from 52% to 32% (see white bars in Figure 7). Interestingly, these are all proteomics studies. DE transcripts from Scarlett YI under D matched the largest percentage of DE tags from the surveyed studies.

We also recorded the number of DE tags found in individual contrasts of our study, which had already been identified in previous studies. These figures for the four main contrasts of our study, Scarlett YI under D, Scarlett YI under MDH, SBCC073 LF under MDH and Scarlett LF under MDH, were 44%, 30%, 56% and 52%, respectively. The largest figures found for the leaf contrasts likely reflect the prevalence of studies which sampled LF tissues.

A total 470 DE isoforms were not found in previous studies, whereas 160 were in just one study and 47 in two. Only 19 DE isoforms were in common in three or more studies. These DE isoforms included several 70kDa and 90kDa heat shock proteins, a S–methyltransferase from S–adenosyl–L–methionine cycle and an N–methyltransferase involved in choline biosynthesis, transcripts related with photosynthesis and carbon fixation, a sucrose synthase, a phosphoglycerate mutase and a triose–phosphate isomerase, a glutathione peroxidase, an ATP synthase and a V–type H+–transporting ATPase subunit, an aspartate kinase, a protein with Potato inhibitor I family domain and a spermidine synthase (Table 5).

3.9 Analysis of co–expressed genes

DE isoforms were clustered based on their expression patterns across samples (Supplementary Figure S10), with the aim of identifying shared regulatory motifs in their upstream genomic regions. We obtained 23 clusters, 14 of them with more than 10 isoforms (Supplementary Table S7). Several clusters contained mostly isoforms from a given contrast while others had mixed DE tags from different treatments (Supplementary Figure S11–S12).

In order to validate the expression–based gene clusters, they were tested for Gene Ontology (GO) enrichment. Moreover, to test the hypothesis that co–expressed genes might share *cis*–regulatory sequences, their upstream sequences were subjected to motif discovery algorithms and the DNA motifs found were annotated. Finally, the resulting regulatory motifs were compared to the binding predictions of DE expressed TFs identified in this work, trying to link these TFs to clusters of DE tags.

The results are summarized in Figure 8. Upstream sequences of genes from cluster 1, with functional annotations related to the metabolism of carbohydrates, contain a wtATAAAAGw site, which is similar to motifs of TATA–binding proteins and Dof TFs (Yanagisawa, 2002). We observed a C2C2–Dof TF up–regulated in SBCC073 LF under MDH (see previous sections), although we were not able to identify DNA–binding domains associated to it. Therefore, we cannot confirm whether or not C2C2–Dof protein binds to this motif to regulate genes in cluster 1, but the possibility deserves further investigation. Promoter sequences of genes in clusters 9 and 10, which group mostly transcripts down–regulated in LF under MDH, contain sites identical to the consensus of CCA1/LHY, which belongs to the MYB/SANT family (Green and Tobin, 1999). These sites were independently predicted by oligo–analysis (AAAATATCTy) and dyad–analysis (aAAAkaTCTw), indicating that they are high–confidence predictions. Genes of this cluster are annotated as components of thiamine biosynthesis in the chloroplast. Accordingly, CCA1/LHY, which was up–regulated in SBCC073 and Scarlett samples under MDH, binds to the same motif (aAAATATCTkY). Cluster 12 had predicted yaCGTACGtr *cis*–elements. Genes in this cluster were induced in LF under MDH, and are annotated as heat shock proteins. Finally, genes in cluster 14 are annotated as components of salinity response, and share *cis*–elements of consensus smACACTbm.

Out of 11 DE TFs, 7 were associated with DNA–binding domains (Table 6), including CCA1/LHY (see above), the MYB–related TF of unknown function DE in SBCC073 LF under MDH, the MADS–MIKC up–regulated in Scarlett YI under D (AwRGaAAaww), the B3–ARF TFs induced in Scarlett YI either under D or MDH (yTTGTCtC), the bZIP up–regulated in Scarlett LF under MDH (cayrACACGTgkt) and the AP2/ERF–AP2 down–regulated in SBCC073 LF under MDH

(CACrrwTCCCrAkG). It is possible that these genes were in part regulating the changes in gene expression in response to the treatments. However, these could not be linked to the motifs identified in promoters.

4 Discussion

In this work, *de novo* assemblies of Spanish landrace SBCC073 and elite cultivar Scarlett were generated. These assemblies had a larger number of isoforms and genes than current barley references. This could be an effect of sequencing errors and non-coding sequences being expressed, but also of absence of actual transcripts from the references. Nonetheless, the use of all available reference sequences (Morex, Barke, Bowman, Haruna Nijo) led to the confirmation of a substantial percentage of those isoforms, allowing the identification of more assembled isoforms than using any of them separately. This highlights the variability in gene content between genome references, which poses a problem when working with non-reference genotypes as in the present study. In light of this, an advantage of *de novo* assemblies resides in recovering genotype-specific transcripts and in reducing mapping errors produced by polymorphisms. Therefore, using them as reference, as we have done in this study, allows diminishing the proportion of unmapped reads and increasing mapping accuracy, which is essential for gene expression assays. Moreover, we tested three different pipelines for differential expression, and those based on *de novo* assemblies had a better agreement with RT-qPCR results.

Plants from Scarlett and SBCC073 were subjected to severe drought and mild drought combined with heat, during the reproductive stage, and physiological responses were measured. Water-stressed plants showed reduced daily loss of water, increased absolute leaf water potential, changes in stomatal conductance, reduced tiller number and reduced spike number, at the end of the experiment. However, there were also differences between the genotypes, indicating different strategies of adaptation to stress. Absolute leaf water potential under severe drought was higher in Scarlett than in SBCC073. Moreover, under combined mild drought and heat, Scarlett exhibited the lowest tiller number, with relative water content comparable to plants under severe drought. In comparison, both measurements were close to that of well-watered plants in SBCC073, under the combined stress. Taken together, these results indicate that Scarlett was more susceptible to mild drought and heat than SBCC073. Experiments carried out in pots, like this, have the disadvantage of not mimicking natural conditions perfectly. On the other hand, experiments in controlled settings actually help to limit variation due to interaction with environment. For instance, rooting depth is kept out of the equation as, although the pots were large, the roots readily explored all soil volume. Hence, potential genotypic differences in soil exploring capacity cannot be held responsible for the genotypic disparities in physiological measurements. Given that soil conditions and water availability were similar for the two genotypes, it can be concluded that SBCC073 was more drought tolerant than Scarlett.

Regarding gene expression, the responses to the stresses were specific of each tissue and genotype. Drought almost did not impact SBCC073, whereas the combination of mild drought and heat only affected its leaves. In contrast, gene expression in both Scarlett tissues was strongly altered in the greenhouse, whereas severe drought alone impacted young inflorescences only.

Overall, we found few changes in leaves under severe drought stress. Although related studies found more differences in gene expression in leaves, most of them studied early responses and only a few addressed prolonged stresses, as in the present study. Processes involved in plant responses to water deficit are different depending on the temporal scale, being those related with drought resistance and grain production, like phenology adjustment, acclimation, fertility and harvest index, affected by medium– to long–term water scarcity (Passioura, 2004). Severe brief stresses, which are rare in the field, are more related with plant survival (Passioura, 2002). Nonetheless, another study focused on long–lasting water and heat stress (Ashoub et al., 2015) reported many gene expression changes. However, that study involved wild barley seedlings starting at the stage of two leaves. Leaves from adult plants, like the ones in our study, are expected to show different responses to drought than those of seedlings (Blum, 2005). Mature flowering plants could have a more limited transcriptional response to prolonged drought stress due to acclimation or enhanced tolerance, which could be achieved, for example, through selective senescence of older leaves or the development of a deep root system (Blum, 2005; 2009). Studies similar to ours, in which the stress conditions were maintained for a long period, and samples were taken from adult plants, have provided contrasting results. The closest result to ours was found by Rollins et al. (2013), who reported no changes in leaf proteome of mature barley plants under drought stress, but apparent changes due to heat. Others, however, did find differentially expressed genes in flag leaves of adult barley plants (Guo et al., 2009) or changes in protein expression in mature leaves of wheat drought tolerant genotypes (Ford et al., 2011).

In contrast with the drought treatment, we found numerous differentially expressed transcripts in leaves under combined drought and heat stress. There is scarce information about the optimum temperature for barley growth. We can assume that it is close to the one reported for wheat, whose optimum range is between 18 and 23 ºC (Slafer and Rawson, 1995; Porter and Gawith, 1999). A previous study showed that high temperature (25ºC) resulted in rapid progression through reproductive development in long days (Hemming et al., 2012). The temperatures in the greenhouse clearly exceeded that range for most of the experimental period and, therefore, experienced a combination of heat and drought stress, together with a wider range of variation for other environmental factors than control plants, such as a mild powdery mildew infection, presence of phytophagous insects, and variable natural photoperiod.

In such conditions, there were several DE isoforms in common in both genotypes. For example, transcription of CCA1/LHY was induced in Scarlett and SBCC073, in both leaves and young inflorescences. The observed changes in expression of CCA1/LHY might be related to photoperiod rather than to tolerance to stress, given that CCA1/LHY is a component of the circadian clock (Campoli et al., 2012; Deng et al., 2015), and other genes related with circadian clock were also differentially expressed in leaves under mild drought and heat, like HvPRR1/TOC1 (Campoli et al., 2012) and an homolog of Arabidopsis adagio–like protein 3. Even so, CCA1/LHY has been shown to be controlled by heat (Karayekov et al., 2013) and reported to play a key role in abiotic stress (Grundy et al., 2015) in other species. Also, among differentially expressed transcripts in leaves, the most recurrent were those related with polyamines (like spermine and spermidine), which were identified in leaves from both genotypes, under severe drought alone and under drought combined with heat. These are small aliphatic amines which have been associated to numerous stresses in plants, including osmotic stress and heat (Bouchereau et al., 1999), and their knock–out mutants in Arabidopsis show increased susceptibility to drought stress (Yamaguchi et al., 2007). However, their specific roles in drought stressed plants remain obscure (Capell et al., 2004; Do et al., 2013).

Besides that, Scarlett leaves displayed more numerous and functionally diverse differentially expressed transcripts than SBCC073, under mild drought and heat. Despite presenting comparable stomatal conductance to SBCC073, Scarlett showed increased responses in genes related to photosynthesis and carbon fixation metabolism, as well as antioxidant enzymes. Also, this genotype seems to react more actively to pathogen attack under MDH, as seen by the increased biosynthesis of molecules related to defense responses. Another interesting genotypic difference was that glycine betaine biosynthesis was induced in SBCC073, whereas in Scarlett trehalose biosynthesis was induced instead. These two compounds have an alleged osmoprotectant function in organisms. While glycine–betaine is well known in plants, including cereals (Ashraf and Foolad, 2007), trehalose is not common in plants (Majumder et al., 2009). These results point towards the presence of effects on different pathways, and different genotypic strategies to cope with the combination of stresses encountered in the greenhouse treatment.

In young inflorescences, there were noticeable changes in gene expression in Scarlett, but almost none in SBCC073, in both stress treatments. As in leaves, this could indicate that Scarlett inflorescences were suffering more from stress than those of SBCC073. A similar interpretation was made by (Hübner et al., 2015), who found a larger proportion of differentially expressed genes for this plant organ in response to stress in sensitive genotypes of wild barley. It is intriguing that inflorescences from Scarlett in the greenhouse showed primarily repressed transcripts, most of them related with metabolism of carbohydrates, reorganization of cell wall and biosynthesis of secondary metabolites. Also, two transcripts involved in indole–3–acetic acid (IAA) biosynthesis were repressed: an L–tryptophan transaminase, which catalyzes the conversion of tryptophan to indole–3–pyruvate, and an indole–3–pyruvate monooxygenase, which yields IAA. This is a key auxin, a phytohormone which regulates many critical developmental processes (Woodward and Bartel, 2005). Barley developing inflorescences are a source of IAA (Wolbang et al., 2004), involved in modulation of stem growth and of floret primordia development (Leopold and Thimann, 1949). We could speculate that this could be an attempt to delay spike development in the face of severe stress.

Differentially expressed transcripts were compared with those from related studies. Disparities with other studies partly reflect differences in experimental set up and vegetal material assessed, but other causes are also possible. Interestingly, agreement was better with works based on proteomics than on transcriptomics. This may reflect a statistical bias, due to the choice of strict significance thresholds in our case and in proteomics studies. In fact, the number of differentially expressed proteins reported from proteomics studies was low, which could explain in part the large percentage of coincidences. On the other hand, RNAseq sampling and expression range is different from that of microarrays (Ozturk et al., 2002), which predominated in the gene expression datasets used for comparison, which could favor obtaining results closer to those of proteomics. Actually, there was only one study using RNAseq in the comparison dataset (Hübner et al., 2015), but similarities with it were also scarce. These authors sequenced transcripts from barley immature spikelets subjected to prolonged water stress, which is rather similar to our experiment. However, they worked with wild barley, whereas this study employed a landrace and an elite cultivar. Wild barley holds much more diversity than cultivated types, with considerable variation in physiological and phenotypic characteristics, and presents specific environmental adaptations to stress like temperature and rainfall (Ellis et al., 2000; Hübner et al., 2013). Therefore, it is feasible that the responses to abiotic stresses of wild barley are different to those of cultivated genotypes. In addition, the methodology in that study, an approach based on RGA, was also different from the one adopted here. As mentioned above, we show that such method produced different outcomes than *de novo* assemblies.

Overall agreement between studies was limited, as seen by the few DE isoforms found in common in three or more studies. A previous meta–analysis of gene expression in response to drought (Shaar–Moshe et al., 2015) also detected few common differentially expressed transcripts between studies, although in this case the comparison involved different plant families. This notwithstanding, some processes are recurrently found in drought studies in barley, including ours, independently of the diversity of genotypes and environmental conditions employed. Hence, these could play central roles in the response of barley to abiotic stress. Many of these have already been discussed and reviewed, like the role of polyamines (see above) (Guo et al., 2009; Abebe et al., 2010; Ashoub et al., 2013), proteases (Ford et al., 2011; Ashoub et al., 2013), glycine betaine and other osmoprotectants (Abebe et al., 2010; Ashoub et al., 2013; Ashoub et al., 2015), ascorbic acid (Guo et al., 2009; Wendelboe-Nelson and Morris, 2012; Wang et al., 2015), lipoxygenases (Wendelboe-Nelson and Morris, 2012; Ashoub et al., 2015), aldehyde dehydrogenase (Guo et al., 2009), and also components of photosystem II, carbohydrates metabolism, heat shock proteins, methionine metabolism, or antioxidant enzymes like catalases, which are well known to be involved in stress responses in plants (Krasensky and Jonak, 2012; Marco et al., 2015).

In order to understand the role of differentially expressed genes, it is important to analyze how these genes are orchestrated. Here, this was accomplished by discovering potential *cis*-elements within upstream promoter sequences. Indeed, this study shows that RNAseq can be exploited to obtain biologically relevant conclusions from co-expressed genes using currently available barley genomic resources. As a proof of concept, the CCA1/LHY TF, up-regulated in leaves under mild drought and heat, was associated to two clusters of repressed transcripts, which harbor high-confidence CCA1 binding sites in their promoter sequences. Genes in those clusters were related to thiamine biosynthesis in the chloroplast, an early response to stress known to be linked to the circadian clock (Bocobza et al., 2013; Wang et al., 2016). Transcripts from thiamine biosynthesis were repressed in another study assessing barley under drought (Talame et al., 2007), indicating that thiamine could play an important role in drought response, maybe regulating function of enzymes for which it is a cofactor, enhancing tolerance to oxidative damage, or as a signaling molecule in adaptation mechanisms to abiotic stress (Tunc-Ozdemir et al., 2009; Goyer, 2010). Therefore, we were able to associate gene regulation apparently elicited by CCA1/LHY with a previously known stress response linked to regulation of thiamine biosynthesis, through analysis of DNA-binding motifs.

Besides CCA1/LHY, we were able to identify other promoters and DNA-binding affinities of TFs. A motif involved in the regulation of heat shock proteins matches a SBP zinc-finger protein SPL7, which has been described as a TF related to heat stress in rice (Yamanouchi et al., 2002). Genes from another cluster shared a motif whose best hits were Arabidopsis ZAT6, belonging to a family of zinc-finger repressors involved in responses to salt stress (Ciftci-Yilmaz et al., 2007), and AZF2, a C2/H2 zinc-finger which negatively regulates abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions (Kodaira et al., 2011). Moreover, among hundreds of differentially expressed transcripts, only 11 TFs were found in this study (including CCA1/LHY). As an example, we found differential expression of transcripts of a MYB-related protein, whose closest SwissProt homologues are single-repeat R3 MYB TFs from Arabidopsis. These are involved in epidermal cell fate specification, more specifically in regulation of trichome development (Gan et al., 2011). Therefore, this MYB-related protein could have a similar role of that of GT factors in wheat, which ahev been related to drought tolerance and trichome development (Zheng et al., 2016). Some of the TFs identified here have already been associated with abiotic stress in rice or Arabidopsis. In example, we found a bZIP TF whose DNA-binding motif corresponds to that of ABRE (ABA-responsive element) *cis*-element, and thus could be regulating ABA-responsive genes (Nakashima et al., 2014). We also found an AP2/ERF-AP2 TF differentially expressed in SBCC073 leaves. The AP2/ERF is a large family of plant-specific TFs, which includes dehydration-responsive element-binding (DREB) proteins, involved in the activation of drought responsive genes (Mizoi et al., 2012). However, the TF reported here was similar to BABY BOOM genes from *Brassica napus*, in which they promote embryo development (Boutilier et al., 2002). We also found differentially expressed transcripts related to a MADS-MIKC homologue of OsMADS6, related with floral organ and meristem identities in rice (Li et al., 2010), up-regulated in Scarlett developing inflorescences under drought; an uncharacterized MYB-related TF, in SBCC073 leaves only; a C2C2-Dof, similar to Arabidopsis CDF2, which regulates miRNAs involved in control of flowering time (Sun et al., 2015); a TF of the TIFY family, whose members are responsive jasmonic acid and to abiotic stresses (Ye et al., 2009); a TUBBY-like protein (TULP), which have been associated to sensitivity to ABA in Arabidopsis (Lai et al., 2004); and two transcripts matching different B3-ARF (auxin responding factor with B3 domains) from Arabidopsis. Therefore, the responses observed here seem to have only partial overlap with those already described in other plants. For example, NAC TFs (Nakashima et al., 2012) have not been found in this study. Taking advantage of DNA-binding motifs allows linking TFs and groups of co-expressed genes through their common interface, and provides an additional layer of insight on the dynamics of stress responses in plants. Signaling pathways in response to drought in barley, especially depending on type of stress, development stage, tissue and genotype, remain to be deciphered (Gürel et al., 2016), although it is expected that different responses and strategies will be favored in different agronomic contexts.

Well-adapted accession SBCC073 is currently being tested under water stress field conditions in populations derived from crosses, to search for QTL that control agronomic traits. The catalog of sequence transcripts and expression profiles from the current study will complement this population-based approach to unravel the genetic control of drought responses which impact grain yield.

5 Data accessibility

Raw reads of barley landrace SBCC073 and cultivar Scarlett have been deposited at ENA (study PRJEB12540).

Assembled transcripts of SBCC073 and Scarlett are available upon request.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

EI, PG, AC, conceived the experiment and designed the greenhouse and growth chamber experiment. CC, AC, and BC designed the sequencing experiments. CC and AC grew and dissected the plants, made physiological measurements and extracted RNA. CC, MG and BC analyzed RNAseq data. CC and AC performed RT-qPCR experiments. All authors read and approved the final manuscript.

8 Funding

This work was funded by DGA – Obra Social La Caixa [grant number GA-LC-059-2011] and by the Spanish Ministry of Economy and Competitiveness [projects AGL2010-21929, RFP-2012-00015-00-00 and AGL2013-48756-R]. Carlos P. Cantalapiedra is funded by [grant BES-2011-045905 linked to project AGL2010-21929].

9 Acknowledgments

CC thanks Takeshi Itoh and Tsuyoshi Tanaka and the Bioinformatics Unit at NIAS (Tsukuba, Japan) for their help in RNAseq analysis during a visit to their institute in 2013. BC thanks Jacques Van

Helden and Jaime Castro-Mondragón for their support and expertise in the analysis of regulatory sequences. CC contributes this work in the framework of the Plant Biology and Biotechnology PhD Program (Universitat Autònoma de Barcelona, Spain).

10 References

Abebe, T., Melmaiee, K., Berg, V., and Wise, R.P. (2010). Drought response in the spikes of barley: gene expression in the lemma, palea, awn, and seed. *Funct. Integr. Genomics* 10, 191–205. doi: 10.1007/s10142-009-0149-4

Alexa, A., and Rahnenfuhrer, J. (2016). "topGO: enrichment analysis for Gene Ontology. R package version 2.24.0". Available online at: https://bioconductor.org/packages/release/bioc/html/topGO.html (Accessed June 1, 2016)

Andersen, C.L., Ledet-Jensen, J., and Ørntoft, T.F. (2004). Normalization of real-time quantitative RT-PCR data: a model based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res.* 64, 5245–5250. doi: 10.1158/0008-5472.can-04-0496

Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data". Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed April 24, 2014)

Araus, J.L., Slafer, G.A., Reynolds, M.P., and Royo, C. (2002). Plant breeding and drought in C3 cereals: what should we breed for? *Ann. Bot.* 89, 925–940. doi: 10.1093/aob/mcf049

Ashoub, A., Baeumlisberger, M., Neupaertl, M., Karas, M., and Bruggemann, W. (2015). Characterization of common and distinctive adjustments of wild barley leaf proteome under drought acclimation, heat stress and their combination. *Plant Mol. Biol.* 87, 459–71. doi: 10.1007/s11103-015-0291-4

Ashoub, A., Beckhaus, T., Berberich, T., Karas, M., and Bruggemann, W. (2013). Comparative analysis of barley leaf proteome as affected by drought stress. *Planta* 237, 771–81. doi: 10.1007/s00425-012-1798-4

Ashraf, M., and Foolad, M.R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environ. Exper. Bot.* 59, 206–216. doi: 10.1016/j.envexpbot.2005.12.006

Barnabas, B., Jager, K., and Feher, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 31, 11–38. doi: 10.1111/j.1365-3040.2007.01727.x

Blum, A. (2005). Drought resistance, water-use efficiency, and yield potential – are they compatible, dissonant, or mutually exclusive? *Aust. J. Agric. Res.* 56, 1159–118. doi: 10.1071/AR05069

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res.* 112, 119–123. doi: 10.1016/j.fcr.2009.03.009

Bocobza, S.E., Malitsky, S., Araújo, W.L., Nunes-Nesi, A., Meir, S., Shapira, M., et al. (2013). Orchestration of thiamin biosynthesis and central metabolism by combined action of the thiamin pyrophosphate riboswitch and the circadian clock in Arabidopsis. *Plant Cell* 25, 288–307. doi: 10.1105/tpc.112.106385

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–20. doi: 10.1093/bioinformatics/btu170

Bouchereau, A., Aziz, A., Larher, F., and Martin–Tanguy, J. (1999). Polyamines and environmental challenges: recent development. *Plant Sci.* 140, 103–125. doi: 10.1016/S0168–9452(98)00218–0

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14, 1737–1749. doi: 10.1105/tpc.001941

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near–optimal probabilistic RNA–seq quantification. *Nature Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Campoli, C., Shtaya, M., Davis, S.J., and Von Korff, M. (2012). Expression conservation within the circadian clock of a monocot: natural variation at barley *Ppd-H1* affects circadian expression of flowering time genes, but not clock orthologs. *BMC Plant Biol.* 12. doi: 10.1186/1471–2229–12–97

Cantalapiedra, C.P., Boudiar, R., Casas, A.M., Igartua, E., and Contreras–Moreira, B. (2015). BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol. Breeding* 15. doi: 10.1007/s11032–015–0253–1

Capell, T., Bassie, L., and Christou, P. (2004). Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9909–9914. doi: 10.1073/pnas.0306974101

Cattivelli, L., Rizza, F., Badeck, F.–W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., et al. (2008). Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Res.* 105, 1–14. doi: 10.1016/j.fcr.2007.07.004

Ceccarelli, S. (1994). Specific adaptation and breeding for marginal conditions. *Euphytica* 77, 205–219. doi: 10.1007/BF02262633

Ceccarelli, S., Grando, S., and Impiglia, A. (1998). Choice of selection strategy in breeding barley for stress environments. *Euphytica* 103, 307–318. doi: 10.1023/A:1018647001429

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., and Chhetri, N. (2014). A meta–analysis of crop yield under climate change and adaptation. *Nat. Clim. Chang.* 4, 287–291. doi: 10.1038/NCLIMATE2153

Ciftci–Yilmaz, S., Morsy, M.R., Song, L., Coutu, A., Krizek, B.A., Lewis, M.W., et al. (2007). The EAR–motif of the Cys2/His2–type zinc finger protein Zat7 plays a key role in the defense response of Arabidopsis to salinity stress. *J. Biol. Chem.* 282, 9260–9268. doi: 10.1074/jbc.M611093200

Close, T.J., Wanamaker, S., Roose, M.L., and Lyon, M. (2007). "HarvEST: an EST database and viewing software", in *Plant bioinformatics: methods and protocols,* ed. D. Edwards, (Totowa, New Jersey: Humana Press), 161–77. doi: 10.1007/978–1–59745–535–0_7

Contreras–Moreira, B., Castro–Mondragon, J.A., Rioualen, C., Cantalapiedra, C.P., and Van Helden, J. (2016). "RSAT::Plants: Motif discovery within clusters of upstream sequences in plant genomes", in *Plant synthetic promoters: methods and protocols,* ed. R. Hehl, accepted.

Dawson, I.K., Russell, J., Powell, W., Steffenson, B., Thomas, W.T.B., and Waugh, R. (2015). Barley: a translational model for adaptation to climate change. *New Phytol.* 206, 913–31. doi: 10.1111/nph.13266

Deng, W., Clausen, J., Boden, S., Oliver, S.N., Casao, M.C., Ford, B., et al. (2015). Dawn and dusk set states of the circadian oscillator in sprouting barley (Hordeum vulgare) seedlings. *PLoS ONE* 10, e0129781. doi: 10.1371/journal.pone.0129781

Do, P.T., Degenkolbe, T., Erban, A., Heyer, A.G., Kopka, J., Köhl, K.I., et al. (2013). Dissecting rice polyamine metabolism under controlled long–term drought stress. *PLoS ONE* 8, e60325. doi: 10.1371/journal.pone.0060325

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K., and Ortiz, R. (2016). Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci.* 21, 31–42. doi: 10.1016/j.tplants.2015.10.012

Ellis, R.P., Forster, B.P., Robinson, D., Handley, L.L., Gordon, D.C., Russell, J.R., et al. (2000). Wild barley: a source of genes for crop improvement in the 21st century? *J. Exp. Bot.* 51, 9–17. doi: 10.1093/jexbot/51.342.9

Fischbeck, G. (2003). "Diversification through breeding", in *Diversity in barley (Hordeum vulgare),* eds. R. Von Bothmer, T. Van Hintum, H. Knüpffer & K. Sato, (Amsterdam: Elsevier Science B.V.), 29–52.

Fischer, R.A., and Turner, N.C. (1978). Plant productivity in the arid and semiarid zones. *Annu. Rev. Plant Physiol.* 29, 277–317. doi: 10.1146/annurev.pp.29.060178.001425

Ford, K.L., Cassin, A., and Bacic, A. (2011). Quantitative proteomic analysis of wheat cultivars with differing drought stress tolerance. *Front. Plant Sci.* 2. doi: 10.3389/fpls.2011.00044

Gan, L., Xia, K., chen, J., and Shucai, W. (2011). Functional characterization of TRICHOMELESS2, a new single–repeat R3 MYB transcription factor in the regulation of trichome patterning in Arabidopsis. *BMC Plant Biol.* 11, 176. doi: 10.1186/1471–2229–11–176

Goyer, A. (2010). Thiamine in plants: aspects of its metabolism and functions. *Phytochemistry* 71, 1615–24. doi: 10.1016/j.phytochem.2010.06.022

Green, R.M., and Tobin, E.M. (1999). Loss of the circadian clock–associated protein 1 in Arabidopsis results in altered clock–regulated gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4176–4179. doi: 10.1073/pnas.96.7.4176

Grundy, J., Stoker, C., and Carre, I.A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00648

Guo, P., Baum, M., Grando, S., Ceccarelli, S., Bai, G., Li, R., et al. (2009). Differentially expressed genes between drought–tolerant and drought–sensitive barley genotypes in response to drought stress during the reproductive stage. *J. Exp. Bot.* 60, 3531–44. doi: 10.1093/jxb/erp194

Gürel, F., Özturk, Z.N., Uçarli, C., and Rosellini, D. (2016). Barley genes as tools to confer abiotic stress tolerance in crops. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01137

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA–seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–512. doi: 10.1038/nprot.2013.084

Hazelhurst, S., and Liptak, Z. (2011). KABOOM! A new suffix array based algorithm for clustering expression data. *Bioinformatics* 27, 3348–55. doi: 10.1093/bioinformatics/btr560

Hemming, M.N., Walford, S.A., Fieg, S., Dennis, E.S., and Trevaskis, B. (2012). Identification of high–temperature–responsive genes in cereals. *Plant Physiol.* 158, 1439–1450. doi: 10.1104/pp.111.192013

Hübner, S., Bdolach, E., Ein–Gedy, S., Schmid, K.J., Korol, A., and Fridman, E. (2013). Phenotypic landscapes: phenological patters in wild and cultivated barley. *J. Evol. Biol.* 26, 163–174. doi: 10.1111/jeb.12043

Hübner, S., Korol, A.B., and Schmid, K.J. (2015). RNA–Seq analysis identifies genes associated with differential reproductive success under drought–stress in accessions of wild barley *Hordeum spontaneum*. *BMC Plant Biol.* 15, 134. doi: 10.1186/s12870–015–0528–z

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–8. doi: 10.1038/nature08747

Intergovernmental Panel on Climate Change (IPCC), (2014). "Climate change 2014 synthesis report". Available online at: http://www.ipcc.ch/pdf/assessment–report/ar5/syr/AR5_SYR_FINAL_All_Topics.pdf (Accessed July 12, 2016)

Kamei, C.L.A., Boruc, J., Vandepoele, K., Van den Daele, H., Maes, S., Russinova, E., et al. (2008). The PRA1 gene family in Arabidopsis. *Plant Physiol.* 147, 1735–1749. doi: 10.1104/pp.108.122226

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Karayekov, E., Sellaro, R., Legris, M., Yanovsky, M.J., and Casal, J.J. (2013). Heat shock–induced fluctuations in clock and light signaling enhance phytochrome B–mediated Arabidopsis deetiolation. *Plant Cell* 25, 2892–2906. doi: 10.1105/tpc.113.114306

Kausar, R., Arshad, M., Shahzad, A., and Komatsu, S. (2013). Proteomics analysis of sensitive and tolerant barley genotypes under drought stress. *Amino Acids* 44, 345–59. doi: 10.1007/s00726–012–1338–3

Keating, B.A., Carberry, P.S., Bindraban, P.S., Asseng, S., Meinke, H., and Dixon, J. (2010). Eco–efficient agriculture: concepts, challenges and opportunities. *Crop Sci.* 50, S–109–S–119. doi: 10.2135/cropsci2009.10.0594

Kodaira, K.S., Qin, F., Tran, L.S., Maruyama, K., Kidokoro, S., Fujita, Y., et al. (2011). Arabidopsis Cys2/His2 zinc–finger proteins AZF1 and AZF2 negatively regulate abscisic acid–repressive and auxin–inducible genes under abiotic stress conditions. *Plant Physiol.* 157, 742–756. doi: 10.1104/pp.111.182683

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., et al. (2007). CPC: assess the protein–coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–9. doi: 10.1093/nar/gkm391

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress–induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63, 1593–608. doi: 10.1093/jxb/err460

Lai, C., Lee, C., Chen, P., Wu, S., Yang, C., and Shaw, J. (2004). Molecular analyses of the Arabidopsis TUBBY–like protein gene family. *Plant Physiol.* 134, 1586–1597. doi: 10.1104/pp.103. 037820

Langmead, B., and Salzberg, S.L. (2012). Fast gapped–read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory–efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb–2009–10–3–r25

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* 25, 1–18. doi: 10.18637/jss.v025.i01

Leopold, A.C., and Thimann, K.V. (1949). The effect of auxin on flower initiation. *Am. J. Bot.* 36, 342–347.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA–Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi: 10.1186/1471–2105–12–323

Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H., et al. (2010). The AGL6–like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Res.* 20, 299–313. doi: 10.1038/cr.2009.143

Liu, Y., Schroder, J., and Schmidt, B. (2013). Musket: a multistage k–mer spectrum–based error corrector for Illumina sequence data. *Bioinformatics* 29, 308–15. doi: 10.1093/bioinformatics/bts690

Luck, M., Landis, M., and Gassert, F. 2015. "Aqueduct water stress projections: decadal projections of water supply and demand using CMIP5 GCMs". Technical Note. Washington, D. C.: World Resources Institute. Available: wri.org/publication/aqueduct–water–stress–projections (Accessed July 4, 2016)

Majumder, A.L., Sengupta, S., and Goswami, L. (2009). "Osmolyte regulation in abiotic stress.", in *Abiotic stress adaptation in plants: physiological, molecular and genomic foundation.,* eds. A. Pareek, S.K. Sopory & H.J. Bohnert, (The Netherlands: Springer), 349–370.

Marco, F., Bitrián, M., Carrasco, P., Rajam, M.V., Alcázar, R., and Tiburcio, A.F. (2015). "Genetic Engineering Strategies for Abiotic Stress Tolerance in Plants", in *Plant biology and biotechnology,* eds. B. Bahadur, L. Sahijram, M.V. Rajam & K.V. Krishnamurthy, (India: Springer), 579–609. doi: 10.1007/978–81–322–2283–5_29

Matsumoto, T., Morishige, H., Tanaka, T., Kanamori, H., Komatsuda, T., Sato, K., et al. (2014). Transcriptome analysis of barley identifies heat shock and HD–Zip I transcription factors up–regulated in response to multiple abiotic stresses. *Mol. Breeding* 34, 761–768. doi: 10.1007/s11032–014–0048–9

Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., et al. (2011). Comprehensive sequence analysis of 24,783 barley full–length cDNAs derived from 12 clone libraries. *Plant Physiol.* 156, 20–8. doi: 10.1104/pp.110.171579

Mayer, K.F.X., Rogers, J., Doležel, J., Pozniak, C., Eversole, K., Feuillet, C., et al. (2014). A chromosome–based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* 345. doi: 10.1126/science.1251788

Mayer, K.F.X., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711–6. doi: 10.1038/nature11543

Medina–Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro–Mondragon, J.A., Delerce, J., et al. (2015). RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res.* 43, W50–W56. doi: 10.1093/nar/gkv362

Mickelbart, M.V., Hasegawa, P.M., and Bailey–Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Rev. Genet.* 16, 237–251. doi: 10.1038/nrg3901

Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. doi: 10.1016/j.tplants.2005.11.002

Mizoi, J., Shinozaki, K., and Yamaguchi–Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819, 86–96. doi: 10.1016/j.bbagrm.2011.08.004

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta* 1819, 97–103. doi: 10.1016/j.bbagrm.2011.10.005

Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00170

Nussbaumer, T., Martis, M.M., Roessner, S.K., Pfeifer, M., Bader, K.C., Sharma, S., et al. (2013). MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* 41, D1144–51. doi: 10.1093/nar/gks1153

Ozturk, Z.N., Talamé, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., et al. (2002). Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol Biol* 48, 551–573. doi: 10.1023/A:1014875215580

Passioura, J.B. (2002). Environmental biology and crop improvement. *Funct. Plant Biol.* 29, 537–546. doi: 10.1071/FP02020

Passioura, J.B. (2004). "Increasing crop productivity when water is scarce – from breeding to field management.", in *New directions for a diverse planet,* eds. R.A. Fischer, N. Turner, J. Angus, L. Mcintyre, M. Robertson, A. Borrell & D. Lloyd, (Brisbane, Australia: Proc. 4th Int. Crop Science Congress).

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45. doi: 10.1093/nar/29.9.e45

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2016). Differential analysis of RNA-Seq incorporating quantification uncertainty. *BioRxiv.* doi: 10.1101/058164

Plant Metabolic Network (2016). "PlantCyc database". Available online at: http://www.plantcyc.org (Accessed June 20, 2016)

Porter, J.R., and Gawith, M. (1999). Temperatures and the growth and development of wheat: a review. *Eur. J. Agron.* 10, 23–36. doi: 10.1016/S1161-0301(98)00047-1

Pswarayi, A., van Eeuwijk, F.A., Ceccarelli, S., Grando, S., Comadran, J., Russell, J., et al. (2008). Barley adaptation and improvement in the Mediterranean basin. *Plant Breeding* 127, 554–560. doi: 10.1111/j.1439-0523.2008.01522.x

R Development Core Team (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rivers, J., Warthmann, N., Pogson, B.J., and Borevitz, J.O. (2015). Genomic breeding for food, environment and livelihoods. *Food Secur.* 7, 375–382. doi: 10.1007/s12571-015-0431-3

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–40. doi: 10.1093/bioinformatics/btp616

Rollins, J.A., Habte, E., Templer, S.E., Colby, T., Schmidt, J., and von Korff, M. (2013). Leaf proteome alterations in the context of physiological and morphological responses to drought and heat stress in barley (Hordeum vulgare L.). *J. Exp. Bot.* 64, 3201–12. doi: 10.1093/jxb/ert158

Ryan, J., Abdel Monem, M., and Amri, A. (2009). Nitrogen fertilizer response of some barley varieties in semi–arid conditions in Morocco. *J. Agricult. Sci. Technol.* 11, 227–236.

Saini, H.S., and Westgate, M. (1999). Reproductive development in grain crops during drought. *Adv. Agron.* 68, 59–96. doi: 10.1016/S0065–2113(08)60843–3

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J., and Itoh, T. (2016). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Res.* 23, 21–28. doi: 10.1093/dnares/dsv033

Sayed, M.A., Schumann, H., Pillen, K., Naz, A.A., and Léon, J. (2012). AB–QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (*Hordeum vulgare L.*). *BMC Genet.* 13. doi: 10.1186/1471–2156–13–61

Sebastian, A., and Contreras–Moreira, B. (2014). footprintDB: a database of transcription factors with annotated cis elements and binding interfaces. *Bioinformatics* 30, 258–65. doi: 10.1093/bioinformatics/btt663

Shaar–Moshe, L., Hubner, S., and Peleg, Z. (2015). Identification of conserved drought–adaptive genes using a cross–species meta–analysis approach. *BMC Plant Biol.* 15, 111. doi: 10.1186/s12870–015–0493–6

Slafer, G.A., and Rawson, H.M. (1995). Base and optimum temperatures vary with genotype and stage of development in wheat. *Plant Cell Environ.* 18, 671–679. doi: 10.1111/j.1365–3040.1995.tb00568.x

Sun, Z., Guo, T., Liu, Y., Liu, Q., and Fang, Y. (2015). The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS Genet.* 11, e1005598. doi: 10.1371/journal.pgen.1005598

Talame, V., Ozturk, N.Z., Bohnert, H.J., and Tuberosa, R. (2007). Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis. *J. Exp. Bot.* 58, 229–40. doi: 10.1093/jxb/erl163

Tello–Ruiz, M., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., et al. (2016). Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 44, D1133–D1140. doi: 10.1093/nar/gkv1179

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA–seq. *Nat. Biotechnol.* 31, 46–53. doi: 10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., et al. (2012). Differential gene and transcript expression analysis of RNA–seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Tunc–Ozdemir, M., Miller, G., Song, L., Kim, J., Sodek, A., Koussevitzky, S., et al. (2009). Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. *Plant Physiol.* 151, 421–32. doi: 10.1104/pp.109.140046

Turner, N.C. (2004). Sustainable production of crops and pastures under drought in a Mediterranean environment. *Ann. Appl. Biol.* 144, 139–147.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real–time quantitative RT–PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, research0034.1–research0034.11.

Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

Varshney, R.K., Terauchi, R., and McCough, S.R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12, e1001883. doi: 10.1371/journal.pbio.1001883

Vitamvas, P., Urban, M.O., Skodacek, Z., Kosova, K., Pitelkova, I., Vitamvas, J., et al. (2015). Quantitative analysis of proteome extracted from barley crowns grown under different drought conditions. *Front. Plant. Sci.* 6, 479. doi: 10.3389/fpls.2015.00479

Wang, L., Ye, X., Liu, H., Liu, X., Wei, C., Huang, Y., et al. (2016). Both overexpression and suppression of an *Oryza sativa* NB-LRR-like gene OsLSR result in autoactivation of immune response and thiamine accumulaion. *Sci. Rep.* 6. doi: 10.1038/srep24079

Wang, N., Zhao, J., He, X., Sun, H., Zhang, G., and Wu, F. (2015). Comparative proteomic analysis of drought tolerance in the two contrasting Tibetan wild genotypes and cultivated genotype. *BMC Genomics* 16, 432. doi: 10.1186/s12864-015-1657-3

Wehner, G., Balko, C., Enders, M., Humbeck, K., and Ordon, F. (2015). Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. *BMC Plant Biol.* 15, 125. doi: 10.1186/s12870-015-0524-3

Wehner, G., Balko, C., Humbeck, K., Zyprian, E., and Ordon, F. (2016). Expression profiling of genes involved in drought stress and leaf senescence in juvenile barley. *BMC Plant Biol.* 16, 3. doi: 10.1186/s12870-015-0701-4

Wei, T., and Simko, V. (2014). "Corrplot: visualization of a correlation matrix". Available online at: http://CRAN.R-project.org/package=corrplot (Accessed September 24, 2014)

Wendelboe-Nelson, C., and Morris, P.C. (2012). Proteins linked to drought tolerance revealed by DIGE analysis of drought resistant and susceptible barley varieties. *Proteomics* 12, 3374–85. doi: 10.1002/pmic.201200154

Wolbang, C.M., Chandler, P.M., Smith, J.J., and Ross, J.J. (2004). Auxin from the developing inflorescence is required for the biosynthesis of active gibberellins in barley stems. *Plant Physiol.* 134, 769–776. doi: 10.1104/pp.103.030460

Woodward, A.W., and Bartel, B. (2005). Auxin: regulation, action, and interaction. *Ann. Bot.* 95, 707–735. doi: 10.1093/aob/mci083

Worch, S., Rajesh, K., Harshavardhan, V.T., Pietsch, C., Korzun, V., Kuntze, L., et al. (2011). Haplotyping, linkage mapping and expression analysis of barley genes regulated by terminal drought stress influencing seed quality. *BMC Plant Biol.* 11, 1. doi: 10.1186/1471-2229-11-1

Yahiaoui, S., Cuesta-Marcos, A., Gracia, M.P., Medina, B., Lasa, J.M., Casas, A.M., et al. (2014). Spanish barley landraces outperform modern cultivars at low-productivity sites. *Plant Breeding* 133, 218–226. doi: 10.1111/pbr.12148

Yamaguchi, K., Takahashi, Y., Berberich, T., Imai, A., Takahashi, T., Michael, A.J., et al. (2007). A protective role for the polyamine spermine against drought stress in Arabidopsis. *Biochem. Biophys. Res. Commun.* 352, 486–490. doi: 10.1016/j.bbrc.2006.11.041

Yamanouchi, U., Yano, M., Lin, H., Ashikari, M., and Yamada, K. (2002). A rice spotted leaf gene, Spl7, encodes a heat stress transcription factor protein. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7530–7535. doi: 10.1073/pnas.112209199

Yanagisawa, S. (2002). The Dof family of plant transcription factors. *Trends Plant Sci.* 7, 555–560. doi: 10.1016/S1360–1385(02)02362–2

Ye, H., Du, H., Tang, N., Li, X., and Xiong, L. (2009). Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. *Plant Mol. Biol.* 71, 291–305. doi: 10.1007/s11103–009–9524–8

Zadoks, J.C., Chang, T.T., and Konzak, C.F. (1974). A decimal code for the growth stages of cereals. *Weed Res* 14, 415–421. doi: 10.1111/j.1365–3180.1974.tb01084.x

Zheng, X., Liu, H., Ji, H., Wang, Y., Dong, B., Qiao, Y., et al. (2016). The wheat GT factor TaGT2L1D negatively regulates drought tolerance and plant development. *Sci. Rep.* 6. doi: 10.1038/srep27042

11 Figure legends

Figure 1. Design of stress treatments, and leaf water potential patterns. SBCC073 (73) and Scarlett (SC) plants were placed in a growth chamber and in a greenhouse. Growth chamber plants were either watered to 70% FC (control, C) or instead 20% FC (drought, D). Greenhouse plants were subjected to combined mild drought (50% FC) and heat stress (MDH). Drought treatments lasted 30 days (30d), after 24d of vernalization and 30d of normal irrigation. The bar plot shows average ± SEM absolute leaf water potential (LWP).

Figure 2. *De novo* assembled genes confirmed in existing barley references. Bars indicate the number of assembled genes of landrace SBCC073 (left) and cultivar Scarlett (right) which were confirmed by alignment to each other, and to several sequence repositories of barley and wheat (for list, see text). The total number of genes confirmed for each of the two assemblies is also shown (bottom black/grey bars). The alignments required 98% identity and a minimum alignment query coverage of either 10% (whole bars) or 80% (fraction of bars filled with a darker color).

Figure 3. Comparison of RT-qPCR and RNAseq gene expression results. Scatterplots show the logFC of isoforms obtained with RT-qPCR (horizontal axis) and with RNAseq (vertical axis). LogFC values from RNAseq were obtained with three different analysis methods: edgeR (left), sleuth (center) and Cuffdiff (right). Plots on the top show all available data, whereas plots on the bottom show data after removing the two most scattered data points (black arrows). Black lines correspond to a linear regression. N: number of data points; β: slope of regression; R^2: coefficient of determination; r: Pearson correlation coefficient.

Figure 4. Number of differentially expressed isoforms and genes. Number of up-regulated (up arrows) and down-regulated (down arrows) differentially expressed tags (isoforms, left; genes, right), for each contrast. Bars show the sum of both induced and repressed tags. LF: leaves. YI: young inflorescences. D: drought treatment. MDH: mild drought and heat treatment.

Figure 5. Metabolic pathways and cellular processes with differentially expressed isoforms from leaves under mild drought and heat. Metabolic pathways, cellular processes and proteins with differentially expressed isoforms are grouped into more general processes, within boxes. Bold categories include several differentially expressed isoforms from a given pathway or process, whereas non-bold names are from specific proteins. Green squares represent processes affected only in SBCC073 (73) plants, whereas red diamonds indicate those altered only in Scarlett (SC). Processes and proteins with changes in gene expression in both genotypes are marked with a black circle. A triangle links the metabolism of aromatic amino acids with downstream pathways of secondary metabolites obtained from them.

Figure 6. Metabolic pathways and cellular processes with differentially expressed isoforms from Scarlett young inflorescences. Metabolic pathways, cellular processes and proteins with differentially expressed isoforms are grouped into more general processes, within boxes. Bold categories include several differentially expressed isoforms from a given pathway or process, whereas non-bold names are from specific proteins. Green squares point out processes altered only under drought (D), whereas red diamonds indicate processes affected only in the mild drought and heat experiment (MDH). Processes and proteins with changes in gene expression in both treatments are marked with a black circle. A triangle links the metabolism of aromatic amino acids with downstream pathways of secondary metabolites obtained from them.

Figure 7. Percentage of differentially expressed tags from other studies which were identified in the present work. Bars indicate the percentage of differentially expressed tags (proteins, genes or isoforms) from other studies which were identified in this work. Each color represents the contribution of each contrast. The list of studies used for comparison is given in Table 4.

Figure 8. Enriched DNA motifs in promoters of differentially co-expressed isoforms. Gene Ontology enrichment and regulatory motifs discovered in 5 clusters of co-expressed isoforms. For each cluster, a plot is shown on the left with the expression profile, where LF and YI correspond to leaf and young inflorescence tissues, and G, D and C to greenhouse, chamber and control replicates, respectively. Regulatory motifs are shown on the right side of each cluster box, with the discovered consensus sequence on top and the most similar motif in footprintDB aligned below. Cluster 10 was found to be very similar to cluster 9, and thus is not shown. The evidence supporting the motifs of clusters 1, 9 and 10 is their significance (black bars) when compared to negative controls (grey bars). Motifs of clusters 12 and 14 (dark boxplots) have higher scores than their shuffled motifs (grey boxplots) when scanned along the cluster upstream sequences and their *Brachypodium distachyon* orthologues.

12 Tables

Table 1. Physiological measurements of plants in the drought experiments. Treatments corresponded to control (C) and drought (D) in the growth chamber, at 70% and 20% field capacity (FC), respectively; whereas greenhouse plants were kept at mild drought and heat (MDH, 50% FC). Physiological and morphological measurements were absolute leaf water potential (LWP), stomatal conductance (SCo), relative water content (RWC) of leaves, tiller number (TN) and visible spike number (VSN).

Treatment	LWP (bar)	SCo (mmol/m2s)	RWC	TN	VSN
		SBCC073			
C	8.09	33.57	0.94	13	4
MDH	14.10	40.93	0.97	11	1
D	14.95	23.02	0.82	8	3
		Scarlett			
C	6.00	12.45	0.92	16	2
MDH	13.47	39.00	0.85	5	0
D	18.15	0.25	0.87	11	2

Table 2. Statistics of *de novo* and reference-guided assemblies. Rows correspond to either *de novo* assemblies (SBCC073 and Scarlett) or reference-guided assembly (RGA). The upper part of the table shows the number of isoforms and genes, as obtained from the assembler, along with statistics on length of isoforms (N50 and mean length). The bottom half shows the number and percentage of annotated isoforms, and whether this annotation was obtained from alignment to SwissProt database or by CDS *de novo* prediction with Transdecoder.

Assembly	Isoforms	Genes	N50	Mean length	Annotated (%)	SwissProt	Transdecoder
SBCC073	303,872	112,923	2,589	1,603	195,184 (64%)	87,145	108,039
Scarlett	307,168	123,582	2,537	1,538	175,779 (57%)	84,310	91,469
RGA	146,427	75,204	4,085	2,512	96,107 (66%)	19,513	76,594

Table 3. Gene Ontology terms enriched in Scarlett young inflorescences. The upper left section shows the GO terms enriched in both experiments (MDH: mild drought and heat; D: drought). The upper right section shows the GO terms enriched only under MDH. The bottom section shows the GO terms enriched only among differentially expressed isoforms under D.

MDH and D	MDH only
Beta-glucan biosynthetic process	Cellulose biosynthetic process
Lignin metabolic process	Xylan biosynthetic process
Phenylpropanoid metabolic process	Plasmodesmata-mediated intercellular transport
Response to carbon dioxide	Mucilage extrusion from seed coat
Sucrose metabolic process	Flavonoid biosynthetic process
Cell wall organization or biogenesis	Mitotic chromosome condensation
D only	
ARF protein signal transduction	Growth
Aspartate family amino acid biosynthetic process	Hydrogen peroxide catabolic process
ATP generation from ADP	L-alanine catabolic process, by transamination
ATP hydrolysis coupled proton transport	L-phenylalanine catabolic process
Callose deposition in cell wall	Methionine biosynthetic process
Carbohydrate catabolic process	NADP metabolic process
Cell wall thickening	ncRNA transcription
Cellular response to starvation	Pentose-phosphate shunt
De-etiolation	Polycistronic mRNA processing
Embryo development ending in seed dormancy	Positive regulation of embryonic development
Ethylene biosynthetic process	Positive regulation of ribosome biogenesis
Glucose metabolic process	Primary root development
Glycerol catabolic process	Protein import into chloroplast stroma
Pyruvate metabolic process	Starch metabolic process
Response to metal iron	Sulfur amino acid biosynthetic process
Response to hormone	Translation elongation
Response to osmotic stress	Tricarboxylic acid metabolic process
S-adenosylmethionine biosynthetic process	Triglyceride mobilization
Seed development	Wax biosynthetic process

Table 4. Studies from the literature assessing protein or transcript expression changes in response to drought in barley. An alias was assigned to each study, to facilitate referring to them. There are different approaches in the comparison dataset, including microarrays (ma), proteomics (p), RNAseq (r), meta–analysis (me), a QTL study and one based on eQTLs. The genotypes used for the experiments involve barley cultivars (c), landraces (l) or wild barley (w). The type of stress applied was drought (d), heat (h), drought and heat combined (c), or dessication, salt and ABA in the case of **“matsumoto2014” (*).** Stresses were applied during different developmental stages, and the tissue sampled was varied also. Finally, the number of differentially expressed tags (transcripts, genes, proteins) included in the comparison dataset is shown (# DE tags).

Alias	Publication	Approach	Genotype	Stress	Develop. Stage	Tissue sampled	# DE tags
abebe2010	Abebe et al. (2010)	ma	c	d	Grain–filling	Lemma, palea, awn, seed	240
ashoub2013	Ashoub et al. (2013)	p	l	d	4–leaves	Leaf	25
ashoub2015	Ashoub et al. (2015)	p	w	d, h, c	2 leaves, 4 leaves	Leaf	40
guo2009	Guo et al. (2009)	ma	c, w, l	d	Flowering	Leaf (flag)	188
hubner2015	Hübner et al. (2015)	r	w	d	Flag leaf emerged	Spikelets	495
kausar2013	Kausar et al. (2013)	p	c	d	3–d old seedlings	Shoot	32
matsumoto2014	Matsumoto et al. (2014)	ma	c	*	4–d old seedlings	Root, shoot	66
rollins2013	Rollins et al. (2013)	p	c, l	d, h, c	Heading	Leaf	99
shaar–moshe2015	Shaar–Moshe et al. (2015)	me	–	d	–	–	2730
talame2007	Talame et al. (2007)	ma	c	d	4–leaves	Leaf	127
vitamvas2015	Vitamvas et al. (2015)	p	c	d	2–leaves	Crown	68
wang2015	Wang et al. (2015)	p	w, c	d	2–leaves	Leaf	26
wehner2015	Wehner et al. (2015)	QTL	c, l	d	7 days after sowing	–	33
wehner2016	Wehner et al. (2016)	eQTL	c, l	d	7 days after sowing	Leaf	14
weldelboe2012	Wendelboe–Nelson and Morris (2012)	p	c	d	7 days after sowing	Leaf, root	69
worch2011	Worch et al. (2011)	ma	c, w	d	Post–anthesis	Grain	137

Table 5. Differentially expressed isoforms found in three or more previous studies. Each row corresponds to a differentially expressed (DE) isoform which was observed in three or more previous studies. Fields include annotated gene name of each DE isoform, and the contrast in which it was declared as DE (73: SBCC073, SC: Scarlett, YI: young inflorescences, LF: leaves, D: severe drought treatment, MDH: mild drought and heat treatment). The presence of the DE isoform in a given study is highlighted with grey background.

DE isoform	Gene name	73–LF–MDH	SC–LF–MDH	SC–YI–D	abebe2010	ashoub2013	ashoub2015	guo2009	hubner2015	kausar2013	matsumoto2014	rollins2013	shaar–moshe2015	talame2007	vitamvas2015	wang2015	wehner2015	wehner2016	wendelboe2012	worch2011
00425	5–methyltetrahydropteroyltriglutamate–homocysteine S–methyltransferase			■		■	■			■		■		■	■				■	
01438	heat shock 70kDa protein 1/8		■		■	■	■	■					■	■						■
30291	photosystem II			■	■					■		■			■				■	
03771	Rubisco activase, chloroplastic		■	■			■			■		■		■		■				
23857	phosphoethanolamine N–methyltransferase	■						■					■	■	■					■
15018	heat shock 70kDa protein 1/8	■		■			■			■		■			■					
46536	sucrose synthase			■		■									■		■			■
49313	ribulose–bisphosphate carboxylase		■			■						■				■			■	
46824	2,3–bisphosphoglycerate–independent phosphoglycerate mutase			■			■					■			■					
22980	heat shock protein 90kDa beta			■								■	■			■				
03577	glutathione peroxidase			■					■						■		■			
43420	V–type H+–transporting ATPase subunit B			■		■						■							■	
19971	ATP synthase alpha/beta family, nucleotide–binding domain			■			■			■		■								
20214	triose–phosphate isomerase			■	■					■		■								
18227	spermidine synthase	■			■			■					■							
49597	Potato inhibitor I family	■	■		■								■	■						
33995	–unknown–		■		■			■					■							
01544	heat shock 70kDa protein 1/8		■						■	■					■					
15965	aspartate kinase		■		■								■	■						

Table 6. Predicted DNA motifs for differentially expressed transcription factors. DE isoforms which were annotated as TFs in all the contrasts (73: SBCC073, SC: Scarlett, YI: young inflorescences, LF: leaves, D: severe drought treatment, MDH: mild drought and heat treatment) are shown along with their iTAK–annotated Pfam domains, whether they were induced (up) or repressed (dn), the BLASTP E–value of homologous TFs, the sequence motif predicted by footprintDB and the best SwissProt hit, along with its gene name prefixed with acronym of the organism (At: *Arabidopsis thaliana*; Bn: *Brassica napus*; Os: *Oryza sativa* subsp. *japonica*).

Isoform	Pfam	Contrast	Up/Down– regulated	E–value	DNA motif	SwissProt
comp690102_c3	AP2/ERF–AP2	73–LF–MDH	dn	7.00E–79	CACrrwTCCCrAkG	Q8LSN2–BnBBM2
comp700847_c0	B3–ARF	SC–YI–D	up	7.00E–150	yTTGTCtC	Q6YZW0–OsARF21
comp61422_c0	B3–ARF	SC–YI–MDH	up	1.00E–98	yTTGTCtC	Q85983–OsARF11
comp59053_c0	bZIP	SC–LF–MDH	up	7.00E–42	cayrACACGTgkt	–
comp688195_c0	C2C2–Dof	73–LF–MDH	up	–	–	Q93ZL5–AtCDF2
comp67310_c0	CCA1/LHY	SC–YI/LF–MDH	up	0.00E+00	waGATAttt	Q6R0H1–AtLHY
comp53438_c1	CCA1/LHY	73–YI/LF–MDH	up	0.00E+00	waGATAttt	Q6R0H1–AtLHY
comp51250_c2	MYB–related	73–LF–MDH	up	5.00E–46	waGATwttww	–
comp61039_c0	MADS–MIKC	SC–YI–D	up	8.00E–61	AwRGaAAaww	Q6EU39–OsMADS6
comp689206_c7	MYB–related	73–LF–MDH	up	–	–	B3H4X8–AtTCL2
comp66417_c0	MYB–related	SC–LF–MDH	up	–	–	B3H4X8–AtTCL2
comp64196_c0	TIFY	SC–LF–MDH	up	–	–	Q6ES51–OsTIFY6B
comp702448_c0	TUB	SC–LF–MDH	dn	–	–	Q7XSV4–OsTULP7

Figure 1.TIF

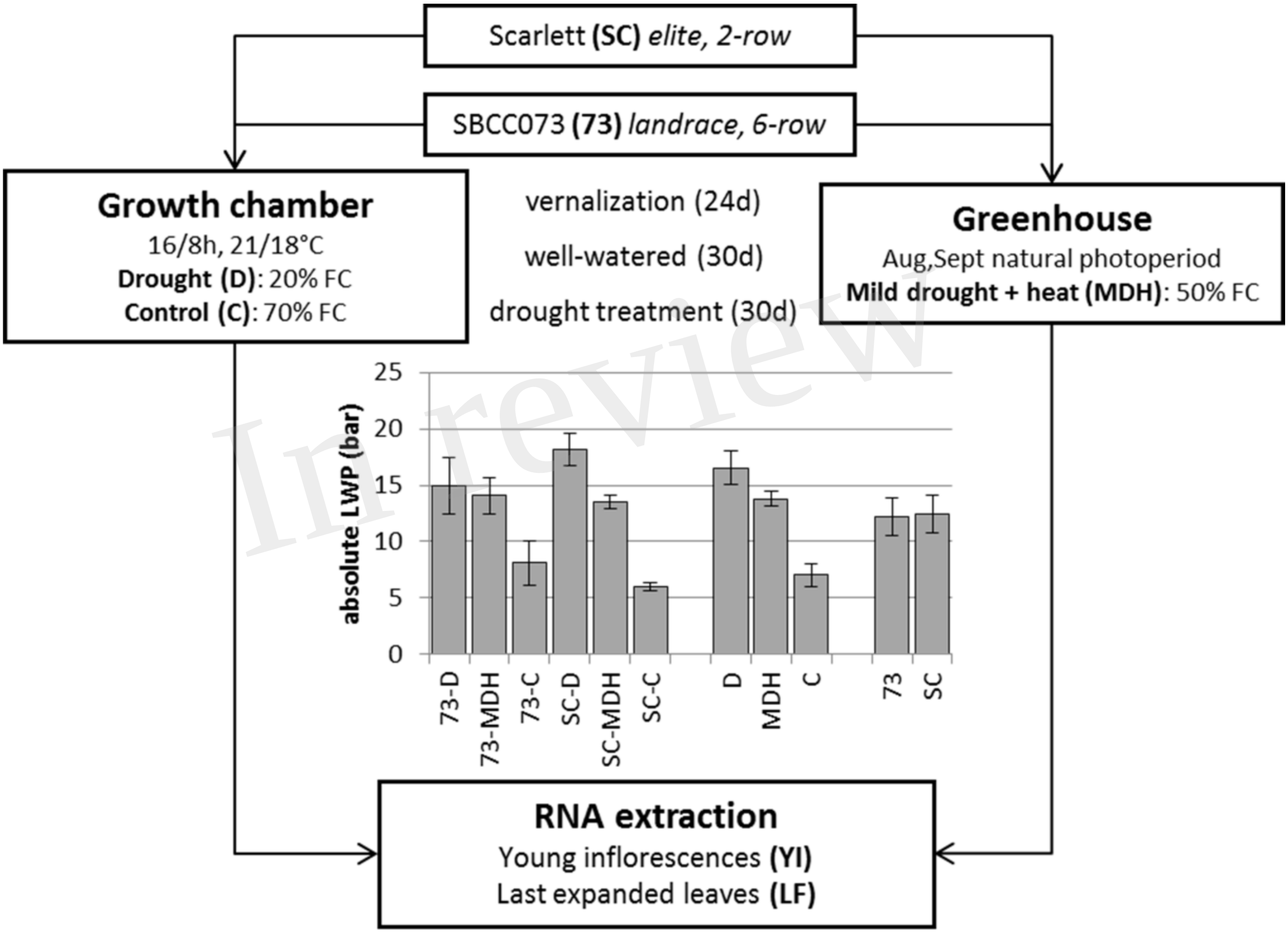

Scarlett (SC) elite, 2-row
SBCC073 (73) landrace, 6-row
Growth chamber
16/8h, 21/18°C
Drought (D): 20% FC
Control (C): 70% FC
vernalization (24d)
well-watered (30d)
drought treatment (30d)
Greenhouse
Aug,Sept natural photoperiod
Mild drought + heat (MDH): 50% FC
absolute LWP (bar)
0
5
10
15
20
25
73-D
73-MDH
73-C
SC-D
SC-MDH
SC-C
D
MDH
C
73
SC
RNA extraction
Young inflorescences (YI)
Last expanded leaves (LF)

Figure 2.TIF

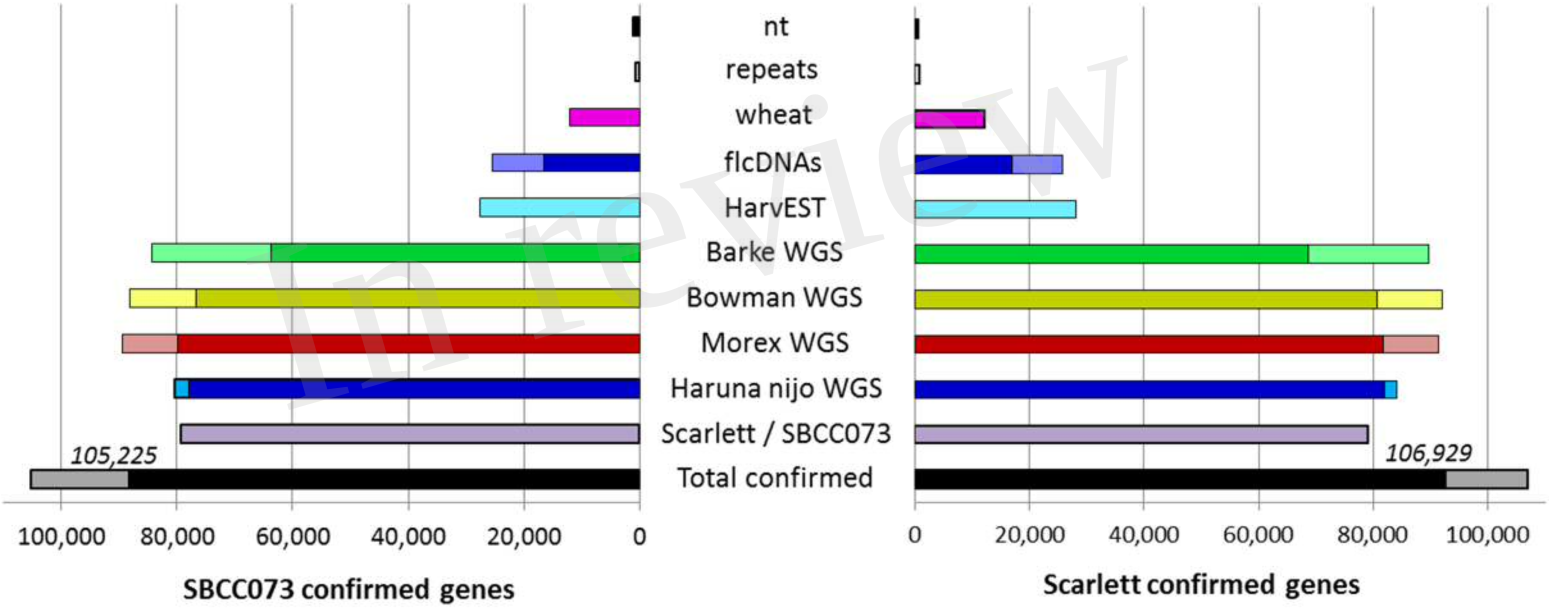

nt
repeats
wheat
flcDNAs
HarvEST
Barke WGS
Bowman WGS
Morex WGS
Haruna nijo WGS
Scarlett / SBCC073
Total confirmed
105,225
106,929
100,000
80,000
60,000
40,000
20,000
0
0
20,000
40,000
60,000
80,000
100,000
SBCC073 confirmed genes
Scarlett confirmed genes

Figure 3.TIF

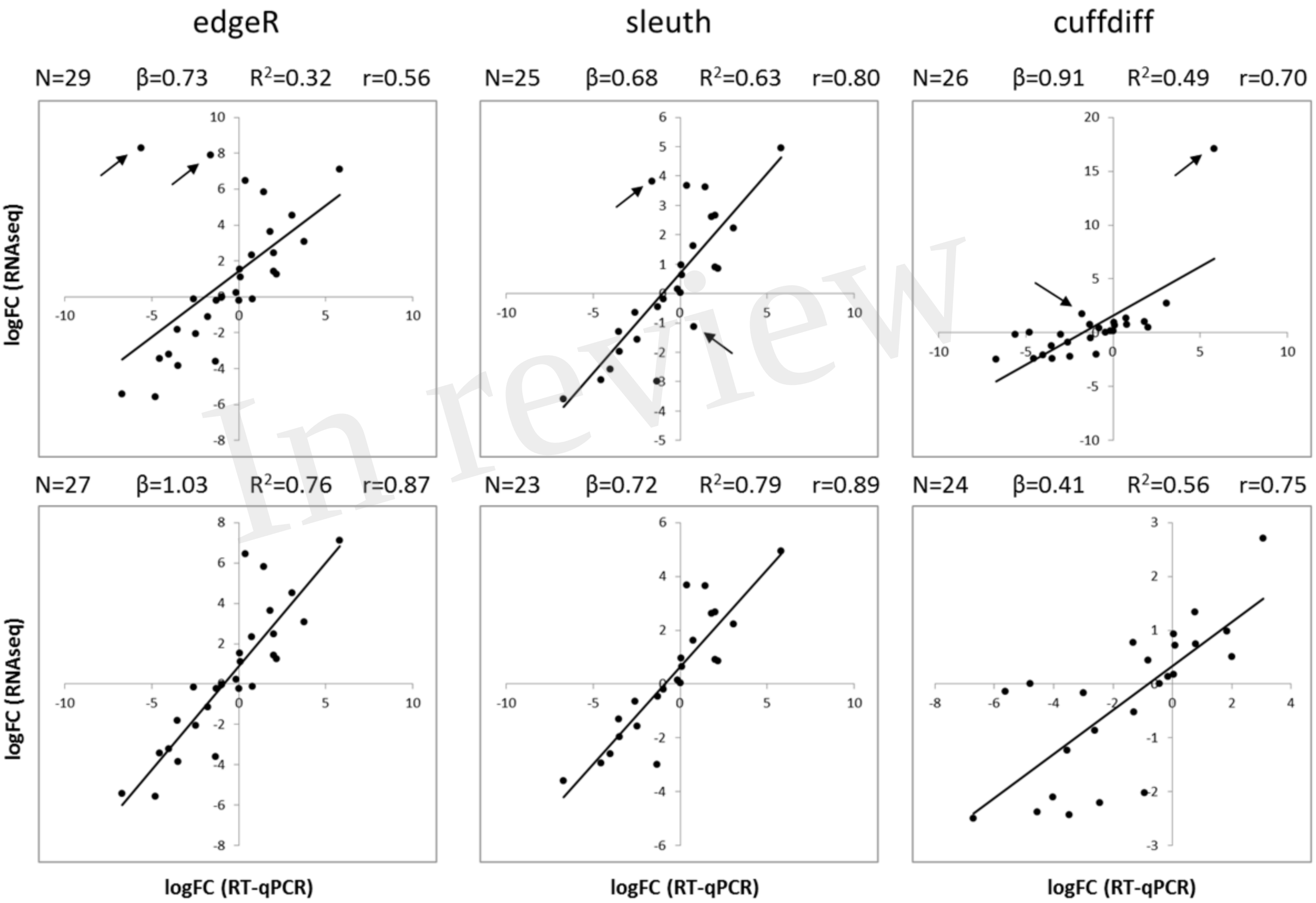
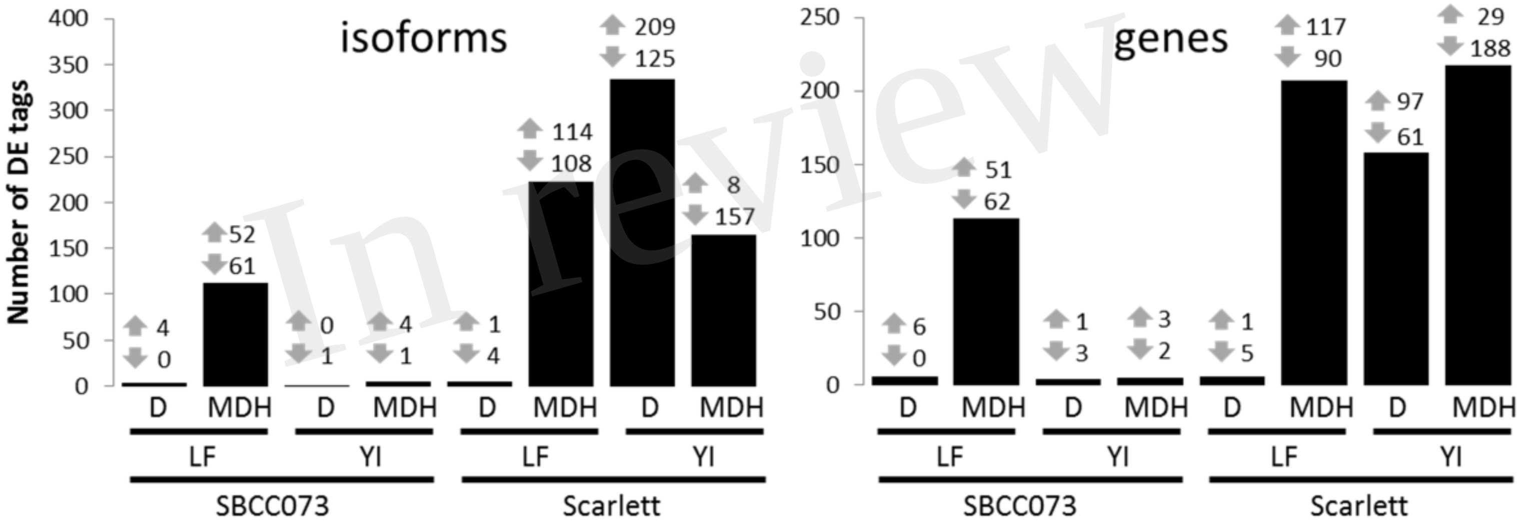
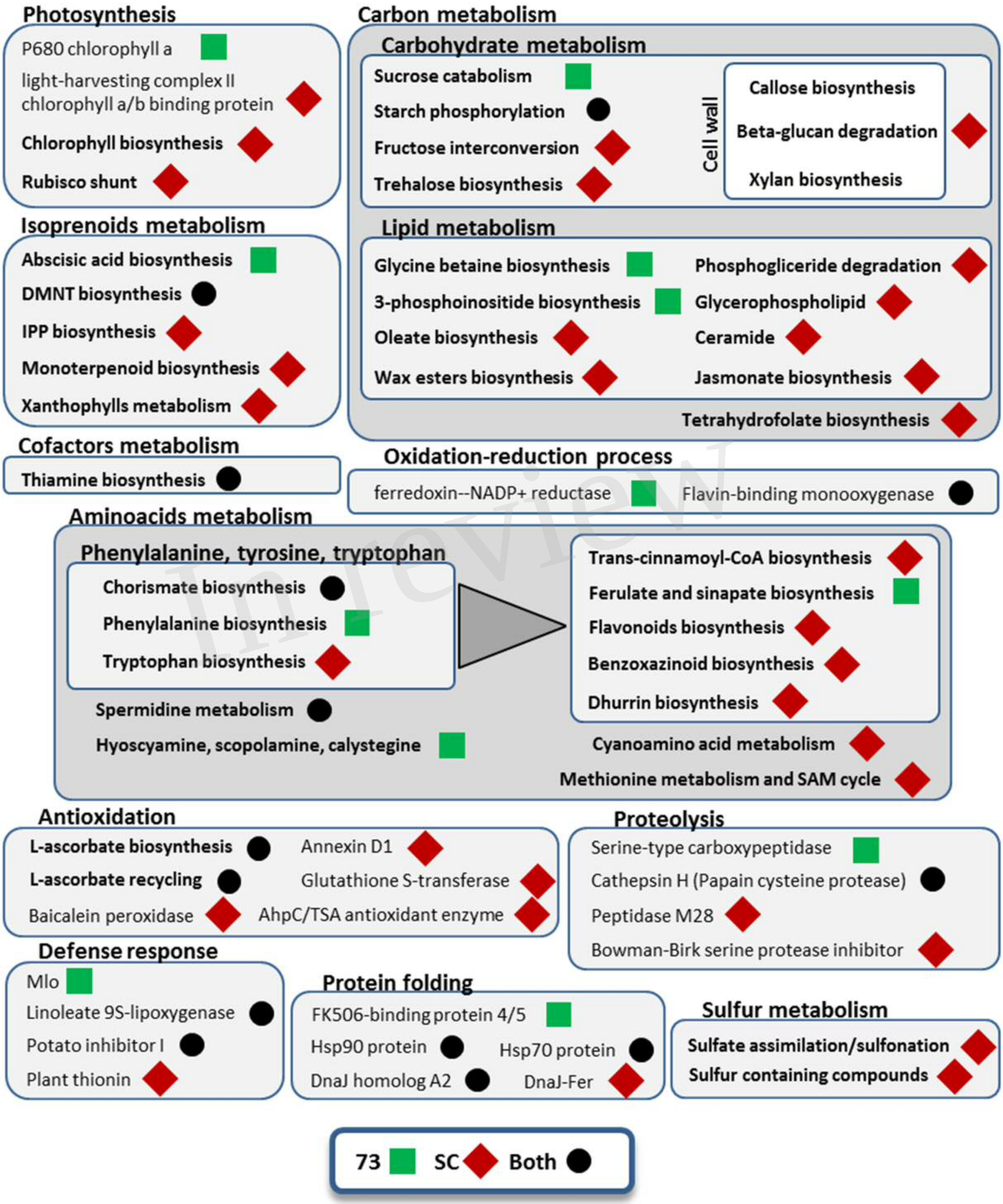

edgeR
sleuth
cuffdiff
N=29 β=0.73 R^2=0.32 r=0.56
N=25 β=0.68 R^2=0.63 r=0.80
N=26 β=0.91 R^2=0.49 r=0.70
N=27 β=1.03 R^2=0.76 r=0.87
N=23 β=0.72 R^2=0.79 r=0.89
N=24 β=0.41 R^2=0.56 r=0.75
logFC (RNAseq)
logFC (RNAseq)
logFC (RT-qPCR)
logFC (RT-qPCR)
logFC (RT-qPCR)
-10
-5
0
5
10
-8
-6
-4
-2
0
2
4
6
8
-5
-4
-3
-2
-1
1
2
3
-10
-5
5
10
15
20

Figure 4.TIF



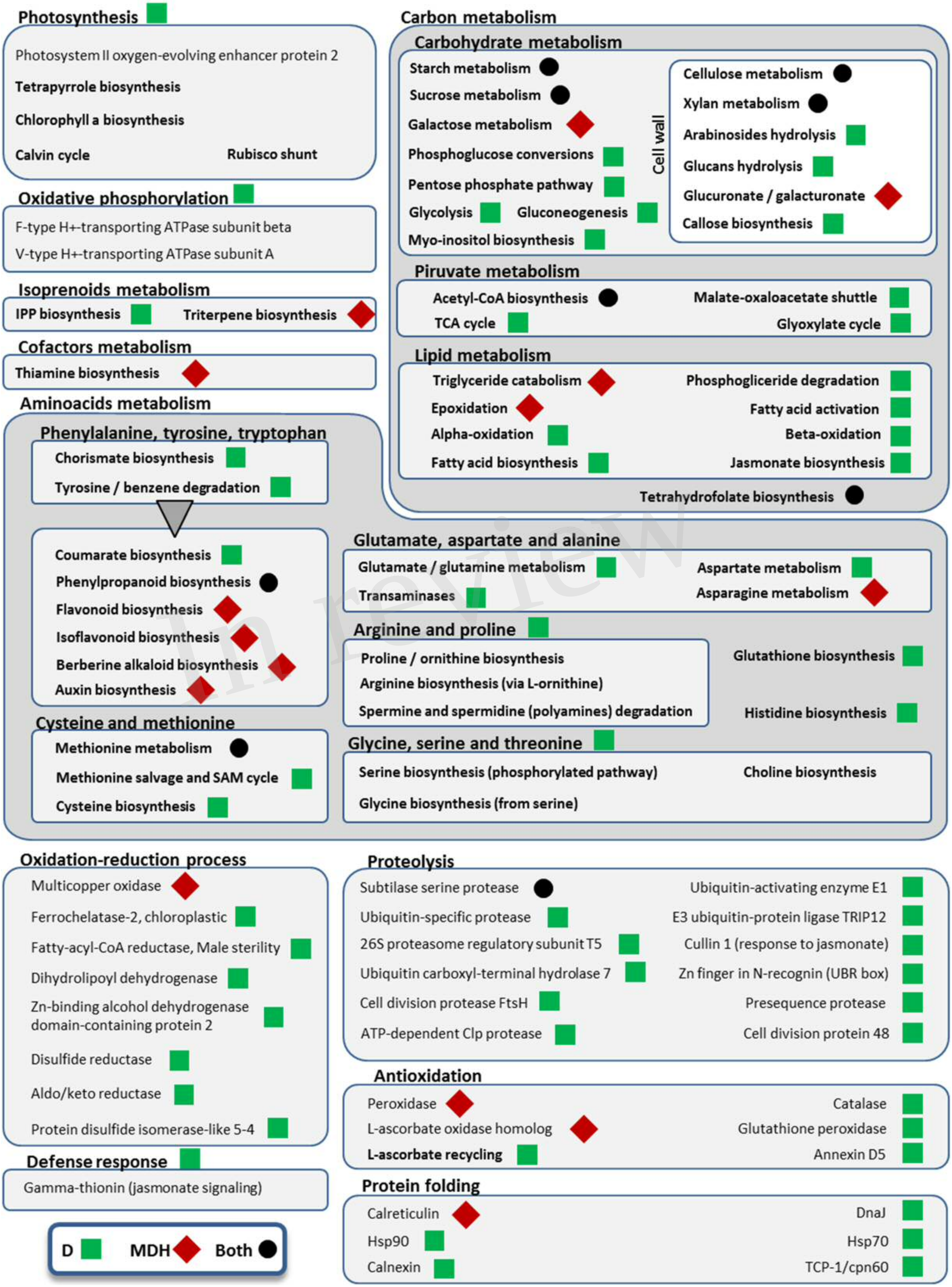
isoforms
Number of DE tags
0
50
100
150
200
250
300
350
400
4
0
52
61
0
1
4
1
1
4
114
108
209
125
8
157
D
MDH
D
MDH
D
MDH
D
MDH
LF
YI
LF
YI
SBCC073
Scarlett
genes
0
50
100
150
200
250
6
0
51
62
1
3
3
2
1
5
117
90
97
61
29
188
D
MDH
D
MDH
D
MDH
D
MDH
LF
YI
LF
YI
SBCC073
Scarlett

Figure 5.TIF



Photosynthesis
P680 chlorophyll a 🟩
light-harvesting complex II chlorophyll a/b binding protein ♦
Chlorophyll biosynthesis ♦
Rubisco shunt ♦
Isoprenoids metabolism
Abscisic acid biosynthesis 🟩
DMNT biosynthesis ●
IPP biosynthesis ♦
Monoterpenoid biosynthesis ♦
Xanthophylls metabolism ♦
Cofactors metabolism
Thiamine biosynthesis ●
Carbon metabolism
Carbohydrate metabolism
Sucrose catabolism 🟩
Starch phosphorylation ●
Fructose interconversion ♦
Trehalose biosynthesis ♦
Cell wall
Callose biosynthesis
Beta-glucan degradation
Xylan biosynthesis ♦
Lipid metabolism
Glycine betaine biosynthesis 🟩
3-phosphoinositide biosynthesis 🟩
Oleate biosynthesis ♦
Wax esters biosynthesis ♦
Phosphogliceride degradation ♦
Glycerophospholipid ♦
Ceramide ♦
Jasmonate biosynthesis ♦
Tetrahydrofolate biosynthesis ♦
Oxidation-reduction process
ferredoxin--NADP+ reductase 🟩
Flavin-binding monooxygenase ●
Aminoacids metabolism
Phenylalanine, tyrosine, tryptophan
Chorismate biosynthesis ●
Phenylalanine biosynthesis 🟩
Tryptophan biosynthesis ♦
Trans-cinnamoyl-CoA biosynthesis ♦
Ferulate and sinapate biosynthesis 🟩
Flavonoids biosynthesis ♦
Benzoxazinoid biosynthesis ♦
Dhurrin biosynthesis ♦
Spermidine metabolism ●
Hyoscyamine, scopolamine, calystegine 🟩
Cyanoamino acid metabolism ♦
Methionine metabolism and SAM cycle ♦
Antioxidation
L-ascorbate biosynthesis ●
L-ascorbate recycling ●
Baicalein peroxidase ♦
Annexin D1 ♦
Glutathione S-transferase ♦
AhpC/TSA antioxidant enzyme ♦
Proteolysis
Serine-type carboxypeptidase 🟩
Cathepsin H (Papain cysteine protease) ●
Peptidase M28 ♦
Bowman-Birk serine protease inhibitor ♦
Defense response
Mlo 🟩
Linoleate 9S-lipoxygenase ●
Potato inhibitor I ●
Plant thionin ♦
Protein folding
FK506-binding protein 4/5 🟩
Hsp90 protein ●
Hsp70 protein ●
DnaJ homolog A2 ●
DnaJ-Fer ♦
Sulfur metabolism
Sulfate assimilation/sulfonation ♦
Sulfur containing compounds ♦
73 🟩 SC ♦ Both ●

Figure 6.TIF



Photosynthesis 🟩
Photosystem II oxygen-evolving enhancer protein 2
Tetrapyrrole biosynthesis
Chlorophyll a biosynthesis
Calvin cycle
Rubisco shunt

Oxidative phosphorylation 🟩
F-type H+-transporting ATPase subunit beta
V-type H+-transporting ATPase subunit A

Isoprenoids metabolism
IPP biosynthesis 🟩
Triterpene biosynthesis ♦

Cofactors metabolism
Thiamine biosynthesis ♦

Aminoacids metabolism

Phenylalanine, tyrosine, tryptophan
Chorismate biosynthesis 🟩
Tyrosine / benzene degradation 🟩

Coumarate biosynthesis 🟩
Phenylpropanoid biosynthesis ●
Flavonoid biosynthesis ♦
Isoflavonoid biosynthesis ♦
Berberine alkaloid biosynthesis ♦
Auxin biosynthesis ♦

Cysteine and methionine
Methionine metabolism ●
Methionine salvage and SAM cycle 🟩
Cysteine biosynthesis 🟩

Carbon metabolism

Carbohydrate metabolism
Starch metabolism ●
Sucrose metabolism ●
Galactose metabolism ♦
Phosphoglucose conversions 🟩
Pentose phosphate pathway 🟩
Glycolysis 🟩
Gluconeogenesis 🟩
Myo-inositol biosynthesis 🟩

Cell wall
Cellulose metabolism ●
Xylan metabolism ●
Arabinosides hydrolysis 🟩
Glucans hydrolysis 🟩
Glucuronate / galacturonate ♦
Callose biosynthesis 🟩

Piruvate metabolism
Acetyl-CoA biosynthesis ●
TCA cycle 🟩
Malate-oxaloacetate shuttle 🟩
Glyoxylate cycle 🟩

Lipid metabolism
Triglyceride catabolism ♦
Epoxidation ♦
Alpha-oxidation 🟩
Fatty acid biosynthesis 🟩
Phosphogliceride degradation 🟩
Fatty acid activation 🟩
Beta-oxidation 🟩
Jasmonate biosynthesis 🟩

Tetrahydrofolate biosynthesis ●

Glutamate, aspartate and alanine
Glutamate / glutamine metabolism 🟩
Transaminases 🟩
Aspartate metabolism 🟩
Asparagine metabolism ♦

Arginine and proline 🟩
Proline / ornithine biosynthesis
Arginine biosynthesis (via L-ornithine)
Spermine and spermidine (polyamines) degradation

Glutathione biosynthesis 🟩
Histidine biosynthesis 🟩

Glycine, serine and threonine 🟩
Serine biosynthesis (phosphorylated pathway)
Glycine biosynthesis (from serine)
Choline biosynthesis

Oxidation-reduction process
Multicopper oxidase ♦
Ferrochelatase-2, chloroplastic 🟩
Fatty-acyl-CoA reductase, Male sterility 🟩
Dihydrolipoyl dehydrogenase 🟩
Zn-binding alcohol dehydrogenase domain-containing protein 2 🟩
Disulfide reductase 🟩
Aldo/keto reductase 🟩
Protein disulfide isomerase-like 5-4 🟩

Defense response 🟩
Gamma-thionin (jasmonate signaling)

Proteolysis
Subtilase serine protease ●
Ubiquitin-specific protease 🟩
26S proteasome regulatory subunit T5 🟩
Ubiquitin carboxyl-terminal hydrolase 7 🟩
Cell division protease FtsH 🟩
ATP-dependent Clp protease 🟩
Ubiquitin-activating enzyme E1 🟩
E3 ubiquitin-protein ligase TRIP12 🟩
Cullin 1 (response to jasmonate) 🟩
Zn finger in N-recognin (UBR box) 🟩
Presequence protease 🟩
Cell division protein 48 🟩

Antioxidation
Peroxidase ♦
L-ascorbate oxidase homolog ♦
L-ascorbate recycling 🟩
Catalase 🟩
Glutathione peroxidase 🟩
Annexin D5 🟩

Protein folding
Calreticulin ♦
Hsp90 🟩
Calnexin 🟩
DnaJ 🟩
Hsp70 🟩
TCP-1/cpn60 🟩

D 🟩 MDH ♦ Both ●

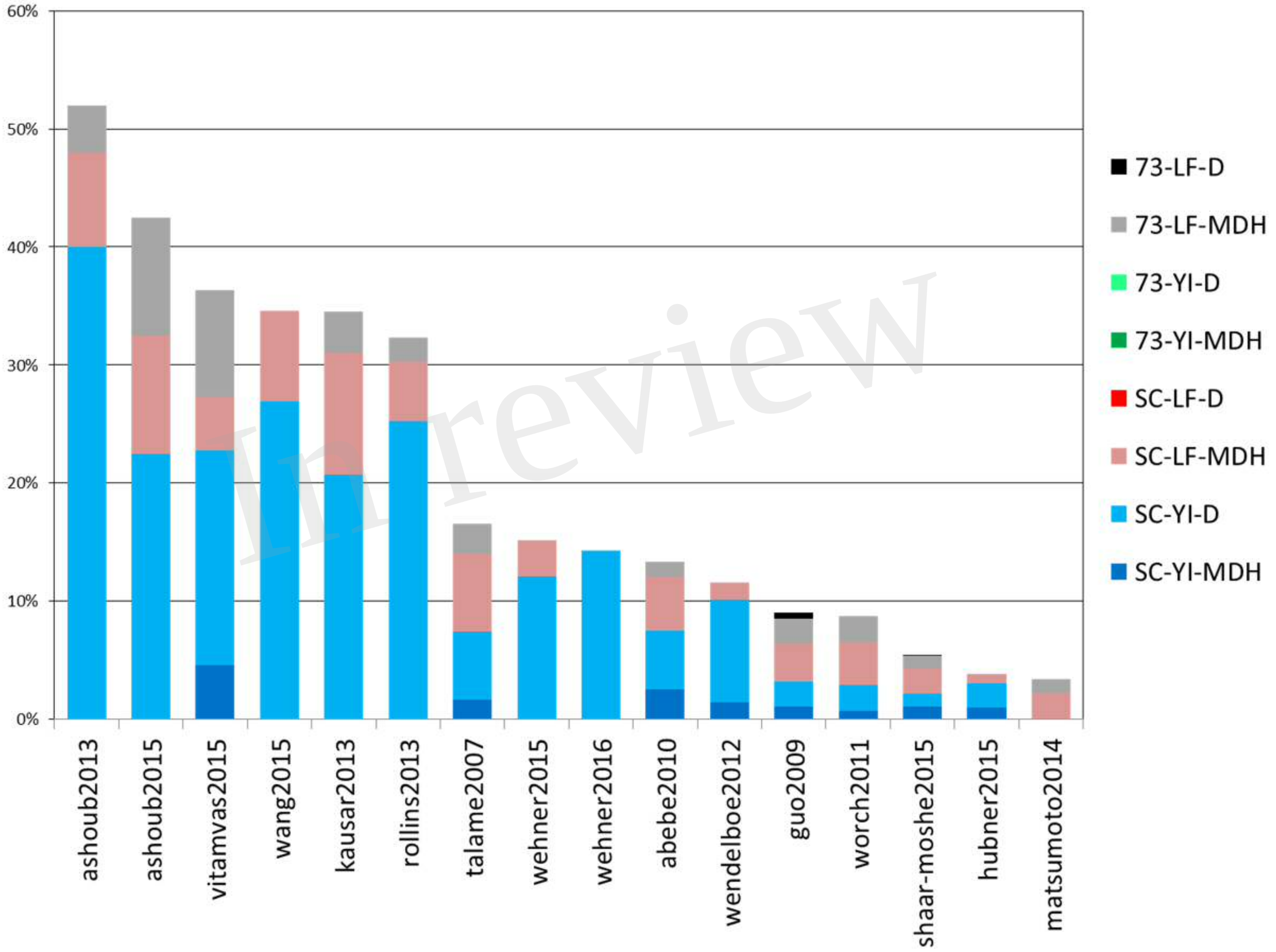
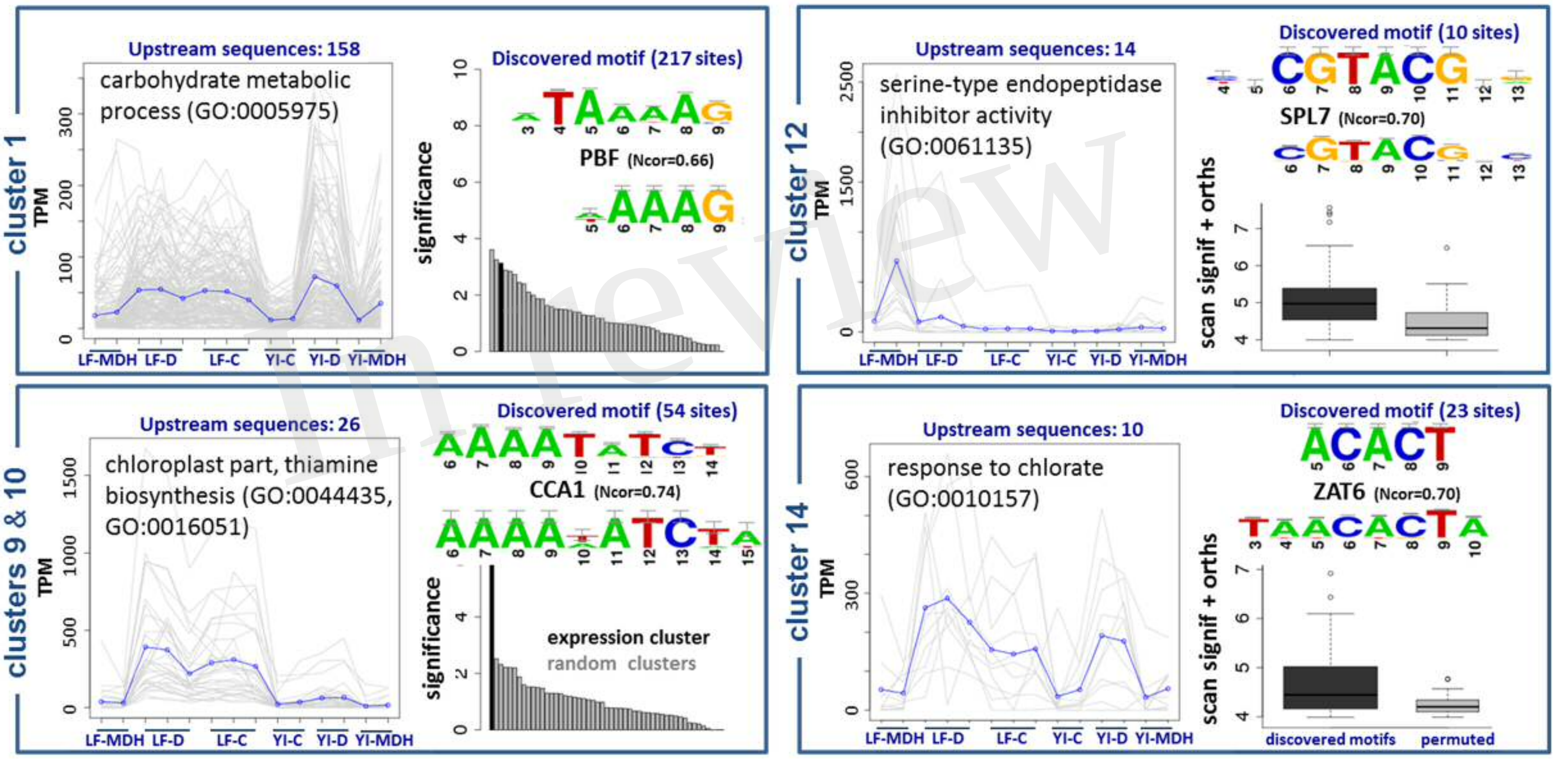

60%
50%
40%
30%
20%
10%
0%
ashoub2013
ashoub2015
vitamvas2015
wang2015
kausar2013
rollins2013
talame2007
wehner2015
wehner2016
abebe2010
wendelboe2012
guo2009
worch2011
shaar-moshe2015
hubner2015
matsumoto2014
73-LF-D
73-LF-MDH
73-YI-D
73-YI-MDH
SC-LF-D
SC-LF-MDH
SC-YI-D
SC-YI-MDH

Figure 8.TIF


cluster 1
Upstream sequences: 158
carbohydrate metabolic process (GO:0005975)
TPM
0
100
200
300
LF-MDH
LF-D
LF-C
YI-C
YI-D
YI-MDH
Discovered motif (217 sites)
PBF (Ncor=0.66)
significance
0
2
4
6
8
10
cluster 12
Upstream sequences: 14
serine-type endopeptidase inhibitor activity (GO:0061135)
TPM
0
1500
2500
LF-MDH
LF-D
LF-C
YI-C
YI-D
YI-MDH
Discovered motif (10 sites)
SPL7 (Ncor=0.70)
scan signif + orths
4
5
6
7
clusters 9 & 10
Upstream sequences: 26
chloroplast part, thiamine biosynthesis (GO:0044435, GO:0016051)
TPM
0
500
1000
1500
LF-MDH
LF-D
LF-C
YI-C
YI-D
YI-MDH
Discovered motif (54 sites)
CCA1 (Ncor=0.74)
significance
0
2
4
expression cluster
random clusters
cluster 14
Upstream sequences: 10
response to chlorate (GO:0010157)
TPM
0
300
600
LF-MDH
LF-D
LF-C
YI-C
YI-D
YI-MDH
Discovered motif (23 sites)
ZAT6 (Ncor=0.70)
scan signif + orths
4
5
6
7
discovered motifs
permuted