The amphioxus hairy family: differential fate after duplication.

Carolina Minguillón, Georgia D. Panopoulou and Jordi Garcia-Fernàndez

Abstract

Several representatives of the hairy family have been so far isolated from a range of protostome and vertebrate species. In protostomes such as *Drosophila* they have been implicated in segmentation, peripheral nervous system development, and in somatic sex determination. In *Tribolium*, only a role during segmentation has been suggested for its single *hairy* gene. On the contrary, hairy genes are expressed in many places in vertebrate embryos and have been implicated in numerous functions, such as somitogenesis, neurogenesis, and endocrine tissue development. In order to gain insight into the timing of acquisition of these roles by the hairy family we have cloned and study the expression pattern of the hairy family in amphioxus. The cephalochordate amphioxus is widely believed to be the living invertebrate more closely related to vertebrates whose genome escaped the massive gene duplications that took place early during vertebrate evolution. Surprisingly, we have isolated six hairy genes from the "pre-duplicative" amphioxus genome, plus two genomic sequences that most probably represent pseudogenes. *In situ* hibridizations on amphioxus embryos show that hairy genes have undergone a process of subfunctionalization and non-functionalization predicted in the DDC model (for Duplication-Degeneration-Complementation) that has been described for the manteinance of duplicate genes in the genome by subfunctionalization. Only the summation of all Amphihairy genes expression resembles the expression pattern of vertebrate hairy genes, i. e. in the central nervous system, presomitic mesoderm (PSM), somites, notochord and gut. Noticeably, none amphioxus hairy gene seems to cycle within the amphioxus PSM.

Keywords: ***Amphioxus, hairy, bHLH, subfunctionalization, DDC model, duplication, Evo-Devo***

INTRODUCTION

Genes in the hairy family encode class E of basic helix-loop-helix (bHLH) transcription factors (Ledent and Vervoort, 2001). These are defined by (i) a proline in the sixth residue within the basic domain that allows them to bind preferentially to sequences refered as "N-boxes" in their target genes (Van Doren *et al*., 1994), (ii) an orange domain, that mediates the specificity of their biological action in vivo (Dawson *et al*., 1995) and, (iii) a 4-amino acid motif, WRPW, located at the C-terminus of the protein that interacts with the general co-repressor protein Groucho to act as active repressors (Jimenez *et al*., 1997).

In protostomes, hairy family members have been isolated from *Drosophila melanogaster*, *Caenorhabditis elegans*, and the beetle *Tribolium*. There are three members of the hairy family in the *D. melanogaster* genome called *hairy* (*h*), *deadpan* (*dpn*), and *similar to deadpan* (*side*), that seem to have arosen by independent duplications in the fly lineage (Moore *et al*., 2000). In contrast, there is only one highly diverged hairy-like gene in the *C. elegans* genome called *lin-22* (Wrischnik and Kenyon, 1997). A single hairy homolog has been isolated from the beetle *Tribolium* (Sommer and Tautz, 1993). In *Drosophila*, *h* acts during segmentation as a primary pair-rule gene (Carroll *et al*., 1998), and also during the development of the fly peripheral nervous system as a pre-pattern gene (Fisher and Caudy, 1998). *dpn* function is required during somatic sex determination as an autosomal factor, and also during the development of the peripheral nervous system where it acts as a precursor gene (Fisher and Caudy, 1998). In contrast to *hairy* and *dpn*, little is known about *side* function during *Drosophila* development. It was identified after the fly genome project. It is expressed in specific subsets of cells in the central nervous system (Moore *et al*., 2000). In *Tribolium*, the *hairy* gene also functions during segmentation, but is not involved in nervous system development (Sommer and Tautz, 1993). The nematode gene *lin-22* is more closely related to the hairy family than to other families of bHLH proteins, although Ledent and Vervoort (2001) argue that it is not a proper hairy gene. Mutant *C. elegans* for *lin-22* function exhibit neuronal defects, most probably due to a premature differentiation of specific neurons (Wrischnik and Kenyon, 1997).

Hairy genes have been also isolated from vertebrates (i. e. mouse, chicken, *Xenopus* and zebrafish). There are two hairy genes in chicken (Palmeirim *et al*., 1997; Jouve *et al*., 2000), in the zebrafish (Pasini *et al*., 2000; Leve *et al*., 2001), and in *Xenopus* (Davis *et al*., 2001). In the mouse genome, only a single hairy gene has

been so far identified (Sasai *et al*., 1992). They have been named in a way that can lead to confussion. They have been called HES (in mammals) or HER (in the fish) for hairy-enhancer of split related plus a number that most probably reflect the temporal order of cloning. Thus, it is not easy to realize that the mouse *HES1* gene is the pro-ortholog of the zebrafish *her6* and *her9* genes, unless a phylogenetic tree is made. On the other hand, zebrafish *her1* is not a hairy gene, but an enhancer of split gene. For mouse *HES1* and both chicken hairy genes a striking dynamic expression pattern in the presomitic mesoderm (PSM), where they cycle with a temporal periodicity that corresponds to the formation of one somite, has been observed (Palmeirim *et al*., 1997; Jouve *et al*., 2000)]. Besides the cycling pattern in PSM, their expression is also detected in several other tissues as endoderm-derived tissues (Jensen *et al*., 2000), the notochord, and the central nervous system (CNS) (Sasai *et al*., 1992). Accordingly, mice mutant for the *HES1* gene exhibit severe defects in neural and endocrine development. Briefly, these defects are thought to be due to the premature differentiation of postmitotic neurons or endocrine cells, respectively (Ishibashi *et al*., 1995; Jensen *et al*., 2000). In neither the zebrafish nor *Xenopus* is there a dynamic expression pattern of hairy genes in the PSM comparable to that in amniotes. During somitogenesis, the zebrafish hairy gene *her6* is expressed in the posterior part of each segmented somite and in stripes in the anterior PSM. Within the CNS, *her6* is expressed first in the prospective forebrain, and later in hindbrain segmentation with a very dynamic, segmentally restricted pattern. Low levels of *her6* expression are also present in the notochord (Pasini *et al*., 2000). The zebrafish hairy gene *her9* is also expressed during CNS development but in contrast to its paralog, neither in segmented somites nor in the PSM. Within the CNS, *her9* is predominantly expressed in the fore- and midbrain, and transiently in the hindbrain, leaving a non-expressing gap at the midbrain-hindbrain boundary (MHB), and it is also expressed in the midline mesoderm. Later in development, expression in the head is detectable in the anterior pituitary, in the optic stalk, in the eye, as well as in the anterior hypothalamus (Leve *et al*., 2001). In the frog, the two hairy homologs, *X-hairy2* and *X-hairy1*, are expressed in the CNS, somites, and PSM, Both genes have identical expression patterns as a band prefiguring a new somite formation in the anterior PSM and are also transcribed weakly in segmented somites. Within the neuroectoderm, they exhibit a non-ovelapping expression pattern (Davis *et al*., 2001).

It is thus to be very pleiotropic a characteristic of vertebrate hairy genes, in contrast to the protostome hairy family members. In order to gain insights into where in evolution the multiplicity of functions was acquired, we studied the hairy family in the cephalochordate amphioxus. Amphioxus is the closest living invertebrate relative to vertebrates and has not undergone the massive gene duplications (up to polyploidization) that took place early during vertebrate evolution (Holland and Garcia-Fernàndez, 1996). Hence, amphioxus has been widely used as a model system to study the function of a gene family in vertebrates, represented by a single gene in the chordate ancestor that may be very similar to that of moderm amphioxus. Surprisingly, we have isolated six canonical hairy family members from the "pre-duplicative" amphioxus genome that we have called *AmphihairyA* to *F*, and two genomic fragments of non-canonical hairy genes, that probably represent pseudogenes, that we have called *Amphihairy-like1*

Table I. Degenerate and specific oligonucleotides used in this report

primer pair	sequence 5' to 3'	amino acid sequence	PCR product length	cDNA region	kind*
h1/h2	AARCCNATHATGGARAA (h1) TCNGCYTTYTCNARYTT (h2)	KPIMEK KLEKAD	122 bp	bHLH	Degenerate for hairy genes
h1/h3	AARCCNATHATGGARAA (h1) TTYTCNAGYTTNSWRTG (h3)	KPIMEK HSKLEK	116 bp	bHLH	Degenerate for hairy genes
h5/h3rep	GTACGGCGGGATCCCCG (h5) TCACCACGGCCTCCACA (h3rep)	YGGIPV MWRPWX	262 to 441 bp[1]	3' coding	Degenerate for Amphihairy genes
RT-CF/RT-CR	CTACGCACCCAACTCTCC (F) AAGCAGTGAAGTCGTGCAC (R)	YAPNSP trailer	311 bp	3' coding/trailer	Specific for AmphihairyC
RT-EF/RT-ER	TACCACCTGACTAGCAGCG (F) GCTATCCGGATGGTACCC (R)	YHLTSS trailer	235 bp	3' coding/trailer	Specific for AmphihairyE
RT-FF/RT-FR	CGCTCCTAGCCCAGCCT (F) GATCGTACTACTACCAGGG (R)	LLAQP PWXtrailer	207 bp	3' coding/trailer	Specific for AmphihairyF
h1/SPLIT-R	AARCCNATHATGGARAA (h1) ARDATRTCNGCYTTYTC (R)	KPIMEK EKADIL	128 bp[2]	bHLH	Degenerate for hairy genes
SPLIT-F/h2	MGNMGNCGNMGNATHAA (F) TCNGCYTTYTCNARYTT (h2)	RRARIN KLEKAD	97bp[2]	bHLH	Degenerate for hairy genes

[1] The product length is variable depending on which Amphihairy gene is amplified.
[2] The length refers to the coding region amplified.

and *2*. We have analyzed their expression pattern during amphioxus development by whole mount *in situ* hybridization, and by RT-PCR for those genes for which no expression was detected by the former technique, namely *AmphihairyE* and *F*. Strikingly, Amphihairy genes seem to have undergone a process that was first described for duplicated genomes such as the zebrafish or the maize genome. This is the DDC model (for Duplication-Degeneration-Complementation), by which duplicate genes subdivide the complex role of the pre-duplicative gene had, ensuring the maintenance of duplicates in the genome by subfunctionalization (Force *et al*., 1999).

MATERIALS AND METHODS

Amphihairy cDNA clones isolation

Degenerated oligonucleotides were designed over the alignment of the bHLH domain of chicken c-hairy 1, mouse HES-1, zebra fish HER-6, *Xenopus X*-hairy 1, *Drosophila* hairy, and *Tribolium* hairy proteins in Palmeirim *et al* (1997). cDNA obtained from RNA of adult animals was used as a template for the PCR reaction (1min at 94ºC, and 35 cycles of 10 seconds at 94ºC, 20 seconds at 42ºC, 30 seconds at 72ºC) with the degenerate primers h1 and h2 (see table I). As no bands of the expected size were obtained, a semi-nested PCR with the degenerate primers h1 and h3 (table I) under the same conditions was performed. A band of the expected size (116 bp) was subcloned into pBLUESCRIPT. Sequencing showed its similarity to the hairy family genes.

The cloned fragment was used to screen a cDNA library constructed from mRNA isolated from 6- to 20-h postfertilization amphioxus embryos (Langeland *et al*., 1998). Approximately $4x10^5$ pfu were screened at moderate stringency (55ºC, Church´s buffer) with the PCR product. Nine cDNAs were identified, escissed, and completely sequenced on automated sequencers [ABI Prism (Perkin-Elmer)]. This led to the identification of four different cDNA clones coding for putative representatives of the hairy family in amphioxus, that we called *AmphihairyA*, *B*, *C* and *D* (table II). The h1 and h2 primers were used to amplify by PCR most of the bHLH domain of *AmphihairyA* to *D*, and all four products were used in conjunction to perform a further screening over the same cDNA library under less stringent condicions (50ºC, Church´s buffer). Twenty-seven positives were recovered (table II) and restriction mapped to assign them into groups. One representant of each group was then fully sequenced. This yielded the isolation of a further cDNA representing a hairy family gene in amphioxus that we named *AmphihairyE*.

Table II. Summary of clones isolated

	Amphihairy					
	A	***B***	***C***	***D***	***E***	***F***
1st screening	1	1	2	5	0	0
2nd screening	8	1	11	6	1	0
Random sequencing	2	2	1	1	0	1

A further cDNA clone, *AmphihairyF*, was identified in a random EST sequence project.

Phylogenetic analysis

The sequences used in the phylogenetic comparisons with the Amphihairy genes reported here (Accession numbers XXXXXX, XXXXXX, XXXXXX, XXXXXX, XXXXXX, XXXXXX, XXXXXX, XXXXXX), were obtained from the public databases, and aligned using the Clustal X method and refined by eye. The whole proteins were used to construct a Neighbour-joining tree (Clustal X). Topology robustness was assessed by 1000 bootstrap resampling of the data.

Intron amplification

In order to amplify the putative second intron of Amphihairy genes we performed PCR experiments on different extractions of genomic DNA from single individuals with the degenerate primers h1 (exon 2) and SPLIT R (exon 3) using about 10 ng of genomic DNA as a template (2 min at 94ºC, and 35 cycles of 20 seconds at 94ºC, 20 seconds at 48ºC and 30 seconds at 72ºC). As no bands were visible, after primer-cleaning, we performed two nested PCR for each individual with the degenerate primers SPLIT F (exon 2) and h2 (exon 3) using 5 μl of the former purified PCR product (2 min at 94ºC, and then 35 cycles of 20 seconds at 94ºC, 20 seconds at 50ºC and 30 seconds at 72ºC). See table I for primers nomenclature and amplifying regions.

The PCR products were cloned into pBluescript using two different approaches. Firstly, a nested PCR was resolved on an agarose gel and the bands escissed one by one. Secondly, another nested PCR reaction was primer-cleaned and directly used for ligation and subsequent cloning. This yielded the isolation of all the former Amphihairy genes plus two new products that, by sequence, represent non-canonical hairy family representatives (*Amphihairy-like1* and *2*).

Southern blot analysis

In order to ascertain whether different Amphihairy cDNA clones represented different Amphihairy genes or polymorphic alleles of the same gene(s), we performed genomic southern blots. The 3' coding region (that included neither the conserved bHLH domain nor the orange domain) of the different Amphihairy genes was obtained by PCR over the cDNA clones with the primers h5 and h3rep (table I), and used as a 32-P-labelled probe to hybridize a *Pst*I-, *EcoRI*-, *Hpa*II- or *Msp*I-digested genomic DNA from single *B. floridae* individuals.

Obtaining embryos and ***in situ*** hybridization

Ripe adults of the Florida lancelet, *Branchiostoma floridae*, were collected from Old Tampa Bay, Florida, USA, during the summer breeding season. The males and females were spawned electrically in the

laboratory, and selected developmental stages were raised by methods in Holland and Holland (1993).

In situ hibridizations were according to methods in Holland *et al*. (1996). The 3' coding region of each Amphihairy gene (the same used for the genomic Southern experiments) was used as a template for the DIG-labelled antisense probe.

After photographied as whole mounts, selected embryos were contrasted in 1% Poinceau S, 1% acetic acid, dehydrated through an ethanol series and embedded in Spurr's resin. Serial 3 m sections were obtained with a glass knife, mounted in DePeX and photographed under Nomarski optics.

RT-PCR experiments

We performed RT-PCR experiments with the Amphihairy genes for which we did not detect any expression in the whole mount hybridizations, *AmphihairyE* and *AmphihairyF*. Specific primers (RT-EF, RT-ER, RT-FF and RT-FR; see table I) were designed over the most divergent region of the cDNA clones (i.e. the most 3' coding region). Multiplex separate reactions were performed for each gene in conjunction with the use of *AmphihairyC* as an internal positive control. The primers used for *AmphihairyC* (RT-CF and RT-CR; table I) were designed in such a way that the region amplified was largest than the region amplified for *AmphihairyE* or *AmphihairyF*.

Total RNA was extracted from 12-, 15-, 18-, and 21-hour embryos by standard methods (Bayascas, 1997) and was retro-transcribed using the M-MLV-RT enzyme (Roche) and oligodT, following manufacturer's instructions. The reaction was then column-purified (QIAGEN) and 3 μl of a 1:100 dilution was used as a template for the PCR (2 min at 94ºC, and 30 cycles of 15 seconds at 94ºC, 10 seconds at 60ºC and 20 seconds at 72ºC). As positive controls the full-length cDNA clones of *AmphihairyC*, *E* and *F* as a template were used. After electrophoresis, gels were blotted and hybridized with *AmphihairyC* plus *AmphihairyE* or *AmphihairyC* plus *AmphihairyF* specific probes at very high stringency (70ºC, Church's buffer) to avoid cross-hybridization.

RESULTS

Isolation and characterization of the amphioxus hairy family

A PCR survey with degenerate primers able to amplify hairy family genes yielded the isolation of several clones with the highest sequence similarity to those members of the hairy family in other species. One of the PCR products was used to screen a cDNA library at moderate stringency. This yielded the identification of four different cDNA clones, that we called *AmphihairyA*, *AmphihairyB*, *AmphihairyC*, and *AmphihairyD*. To study whether more genes of the hairy family were present in amphioxus, we performed a further cDNA screening at lower stringency conditions, using nearly all the region coding for the bHLH domain of the four formerly isolated cDNA clones as probes. This experiment resulted in the identification of the already identified Amphihairy cDNA clones plus a novel one that was named *AmphihairyE*. The random EST sequencing of either a 5- to 6-hour embryos or a 26-hour embryos libraries (Panopoulou *et al*., 1998) yielded the isolation of all the Amphihairy clones already isolated but *AmphihairyE*, plus a novel clone, that we called *AmphihairyF*. Table II summarizes which and how many clones were obtained in each screening and the random sequencing.

An alignment of all six hairy clones isolated shows that the putative proteins are very similar (Fig. 1). An alignment with other hairy proteins shows that this similarity is mainly observed within three regions of the proteins, i. e. the bHLH DNA binding and dimerization domain, the orange domain that gives specificity among the hairy-related proteins, and the C-terminal WRPW domain of interaction with the co-repressor protein Groucho (data not shown). From the alignments we concluded that *AmphihairyA* to *F* are *bona fide* hairy genes.

Phylogenetic analysis

Table III summarizes the percentage identity of all six Amphihairy proteins bHLH domain, with those of selected hairy proteins from both vertebrates and invertebrates. Amphihairy proteins bHLH domain is

Table III. Percentage of similarities of Amphihairy proteins bHLH domains to other hairy proteins

		Amphihairy					
		A	**B**	**C**	**D**	**E**	**F**
GROUP I	**mHES1/hHES1**	97	88	95	93	97	86
	c-hairy2	95	90	97	91	95	86
	zfHER6	95	86	93	91	95	86
	***X*-hairy1**	97	88	95	93	97	86
GROUP II	**c-hairy1**	95	88	93	91	95	84
	***X*-hairy2/zfHER9**	97	88	95	93	97	86
protostomes	**Dm hairy**	81	75	81	78	81	76
	Dm Dpn	76	74	78	76	76	69
	Tc hairy*	85	79	83	84	85	75
	Ce-lin22	55	57	55	53	55	53
E(spl)	**mHES2**	64	62	64	64	64	64
	hHES2	66	66	66	64	66	62

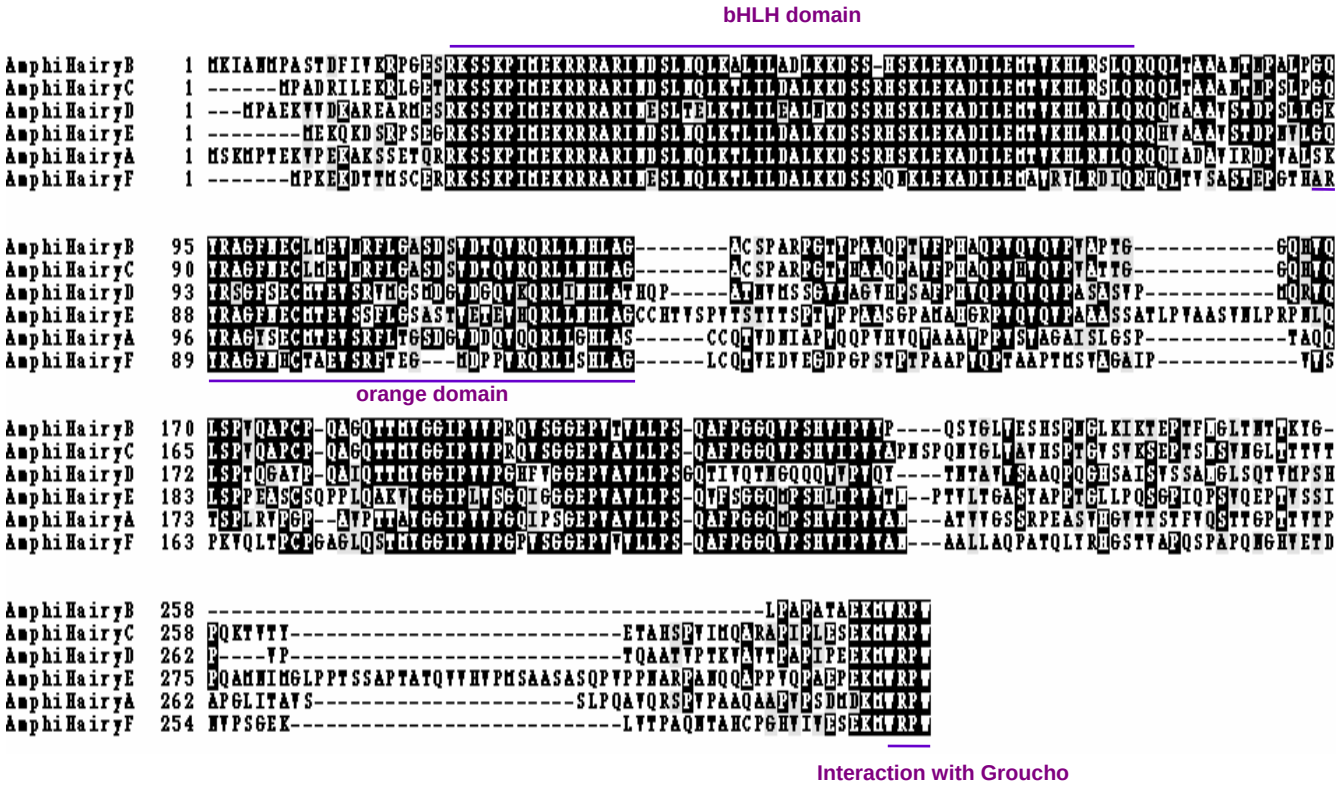

Figure 1. Alignment of the amphioxus hairy proteins conducted by ClustalW. Black boxes represent identical residues among the six Amphihairy proteins. The bHLH, the orange and the C-terminal WRPW domain are indicated.

more closely related to that of vertebrate hairy proteins (88-97%) than to those of other invertebrate hairy proteins (69-85%; note that the divergent lin-22 sequence is not considered).

To gain more insight into the relationships among hairy genes, we conducted a molecular phylogenetic analysis by the neighbor-joining method on the full hairy proteins, using mouse and human HES2 sequences as outgroups (Fig. 2). Vertebrate hairy proteins fell together (bootstrap value 100%) into two well supported groups (values 78 and 88%) that we denominate group I and group II. All the Amphihairy proteins form a monophylletic group that branches immediately outside these groups (as their sister group; 93%). The grouping of all Amphihairy proteins in a monophylletic group is also highly supported (96%), which mean that all they have originated by independent duplication in the cephalochordate lineage after its divergence from the phylogenetic tree main branch. Moreover, the positioning of Amphihairy proteins before the origin of the vertebrate groups I and II, is also supported by a high bootstrap value (100%), in agreement with the hypothesis that vertebrate genes have originated by duplication after the cephalochordate-vertebrate divergence.

"Non-canonical" hairy genes in the amphioxus genome

To study whether more hairy genes were present in the amphioxus genome, we performed PCR experiments on genomic DNA extracted from single individuals. We used degenerate primers (see materials and methods) able to amplify hairy family genes that were flanking a putative intronic position deduced by conserved intronic positions among vertebrate hairy genes (Nishimura *et al*., 1998). The PCR and nested PCR reactions were electrophoresed on an agarose gel, which made visible the existence of a major band of about 450 bp plus lots of bands ranging

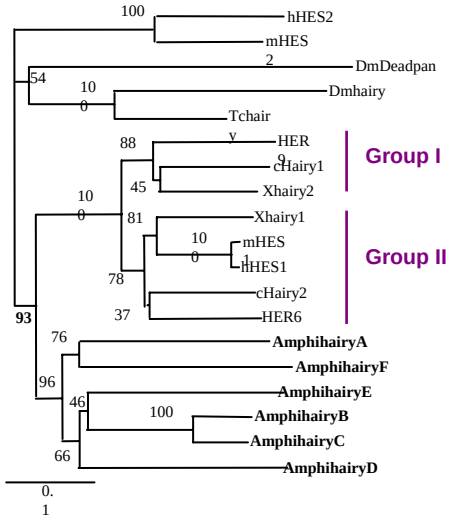

Figure 2. Neighbor-joining phylogenetic tree relating the amphioxus Hairy proteins with other selected Hairy proteins from vertebrate and protostome species. The trees are rooted using the human and the mouse HES2 sequences. The numbers refer to the bootstrap values. The abbreviations used for species and genes are: m, mouse; h, human; X, *Xenopus laevis*; c, chicken; Dm, *Drosophila melanogaster*; Tc, *Tribolium castaneum*; zf, zebrafish. Sequences were obtained from public databases.

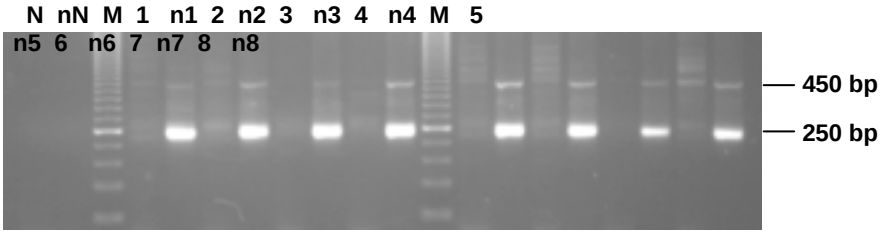

Figure 3. Degenerate genomic PCR. Numbers refer to DNA extracted from different single animals. The numbers on the right refer to the molecular weight of the bands obtained. Abbreviations used are: n, nested PCR reaction; N, negative control; M; molecular weight marker.

from 200 to 250 bp (Fig. 3). We purified the PCR reaction by either escissing the bands form the gel, or by primer-cleaning directly the nested PCR.

This approach yielded the isolation of the putative intron 2 plus flanking exonic regions of all the formerly isolated Amphihairy genes. The 450 bp band represented the sequence of *AmphihairyE* gene, whereas the smaller bands represented all the other Amphihairy genes. Apart from them, we also obtained two novel sequences that we called *Amphihairy-like1* and *Amphihairy-like2*. Figure 4 shows the alignment and percentages of similarity of these sequences (the putative exonic parts) with those of the formerly isolated Amphihairy genes and a selected vertebrate and invertebrate hairy genes. We argue that these sequences may represent pseudogenes in the amphioxus genome. First, although *Amphihairy-like1* and *2* sequences are more closely related to hairy than to Enhancer of split genes (E[spl]), (e. g. they are more similar to hairy genes than to the mouse HES2 gene which belongs to the E(spl) family), they are very divergent (46-62% whereas the proper amphioxus hairy genes are 75-92% similar to HES1). And second, we have been unable to isolate their cDNA clones although we isolated the cDNA clones that correspond to the *AmphihairyE* and *AmphihairyF* genes which are expressed at very low levels during embryogenesis, from embryonic libraries (see below). We think thus that these sequences may also represent ancient duplicates of a single hairy gene that have undergone a process of non-functionalization (see discussion).

AmphihairyA, ***B***, ***C*** and ***D*** genes expression in gastrula stages

The expression of four different Amphihairy genes was studied by whole mount *in situ* hibridizations: *Amphihairy A*, *B*, *C*, and *D*. In brief, all these four genes have a specific expression pattern whose summation resembles that of the single (for the mouse) or the two hairy genes that exist in other vertebrates.

No signal is detected at the blastula or early gastrula stages for any of the Amphihairy genes. During mid-gastrula stage, *AmphihairyA* is expressed in two domains: in the anterior endoderm and just outside the dorsal lip of the blastopore in the presumptive neural plate (arrow in fig. 5A).

Expression of *AmphihairyB*, *C*, and *D* is first detected at the very late gastrula stage in both the neural plate and the presumptive somitic mesoderm. In the presumptive somitic mesoderm all three genes are expressed in a striped pattern prefiguring the definitive somites. Specimens with a 2/3, 3 and 4 bands are shown for *AmphihairyB* (Fig. 5B, 5C and 5E, respectively). For the 2/3 stripe embryo only a weak signal is seen in the left side presumptive third somite (Fig. 5B, the pre-somitic bands are numbered). The expression pattern in the presumptive somitic mesoderm is very similar between *AmphihairyB*, *C* and *D*. For example, a 3-stripe pattern is shown for *AmphihairyC* (Fig. 5F) and a 4-stripe pattern is shown for *AmphihairyD* (Fig. 5G and H).

AmphihairyB, *C* and *D* are also expressed in the presumptive neural plate of the late gastrula. For *AmphihairyB*, expression is at about the level of the first and second somites (Fig. 5C, arrow) and is strongest laterally (Fig. 5D). For *AmphihairyC* the relatively weak signal in the neural plate is in two regions: an anterior one between the first and second somites (Fig. 5F, arrow) and a posterior one from third somitic stripe towards the posterior-most part of the gastrula (Fig. 5F, twin arrows). And last, *AmphihairyD* is expressed only in the posterior-most region of the neural plate between the third and the fourth somite (Fig. 5G, arrow).

In summary, *AmphihairyA* was the only hairy gene

		Amphihairy like1	Amphihairy like2	mHES1	product length
Amphihairy-like1	SSLNELKNLILGTVKDDINAPHHS	100	67	46	149bp
Amphihairy-like2	----------D-Y-N-STS..--	67	100	62	137bp
AmphihairyA	D---Q--T---DAL-K-SSR..--	50	71	92	260bp
AmphihairyB	D---Q--A---ADL-K-SS...--	50	62	75	236bp
AmphihairyC	D---Q--T---DAL-K-SSR..--	50	71	92	240bp
AmphihairyD	E--TE--T---EALNK-SSR..--	42	54	83	237bp
AmphihairyE	D---Q--T---DAL-K-SSR..--	50	71	92	475bp
AmphihairyF	E---Q--T---DAL-K-SSR..QN	42	58	88	212bp
mHES1	E--SQ--T---DAL-K-SSR..--	46	62	100	
Tc-hairy	N---E--T---DAM-K-PAR..--	50	67	75	
mHES2	E--SQ--G-V-PLLGAETSR...S-	29	37	54	

Figure 4. Alignment of Amphihairy-like1 and Amphihairy-like2 and mouse HES1 deduced amino acid sequence from the intronic PCR amplification to Amphihairy proteins and Tribolium Hairy protein and mouse HES1 and HES2 proteins. Lines represent identities to Amphihairy-like1 and dots represent gaps. Species abbreviations are the same as in figure 2. The triangle marks the intronic position, and the in the right column the intron length is found.

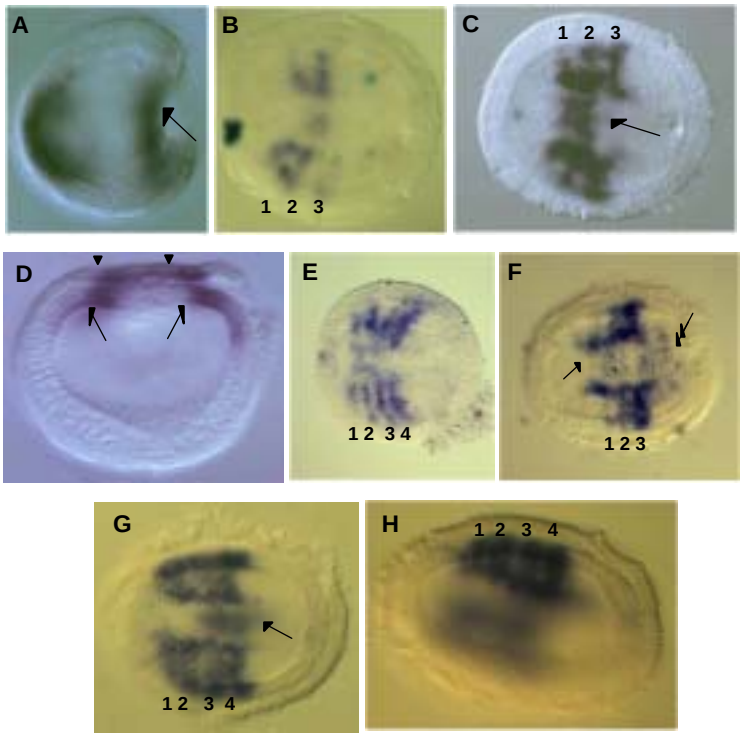

Figure 5. *AmphihairyA*, *B*, *C*, and *D* genes expression in amphioxus gastrulae by whole mount *in situ* hybridization. All embryos in dorsal views (A, B, C, E, F and G) have anterior toward the left. The lateral view (H) has anterior to the left and dorsal up. In the transverse optical disection (D), dorsal is up looking from posterior. The presumptive somitic bands are numbered and the arrows indicate expression domains in the neural plate (with the exception of D). A) *AmphihairyA* expression within the anterior inner endodermal cell layer and the posterior neural plate. B) Dorsal view of an *AmphihairyB* 2 to 3 stripe-pattern gastrula. C) Dorsal view of an *AmphihairyB* 3-stripe pattern embryo. D) Optical transverse dissection of the specimen in (C) showing the highest *AmphihairyB* expression levels at the edges of the neural plate (arrowheads). The expression in the dorso-lateral walls of the archenteron is shown by arrows. E) Dorsal view of an *AmphihairyB* 4-stripe pattern embryo. F) Dorsal view of a 3-stripe pattern embryo stained for the *AmphihairyC* probe. G) Dorsal view of a 4-stripe embryo stained for the *AmphihairyD* labelled probe. H) Lateral view of the gastrula in (G) showing the *AmphihairyD* gene expression as pre-somitic stripes.

expressed in the endoderm and not in the presumptive somitic mesoderm, region where the rest of Amphihairy genes seem to be co-expressed. In the medial neural plate all four genes have a striking pattern, being all them expressed there but in complementary patterns. *AmphihairyA* is expressed in the posterior-most region, then is expressed *AmphihairyC* but leaving a gap between somitic stripes 2 and 3 that is filled by *AmphihairyD* expression, and anteriorly are *AmphihairyB* and again *AmphihairyC* expressed.

AmphihairyA, ***B***, ***C*** and ***D*** genes expression in neurula stages

During neurula stage, the expression of the four Amphihairy genes becomes more gene-specific. *AmphihairyA* gene behaves accordingly to the pattern observed for earliest stages; namely it is expressed in the endoderm and the posterior-most part of the neural tube. In early neurula stages the signal is conspicuous along most of the gut and also in the posterior-most third of the dorsal nerve corde (Fig. 6A, arrow). The nascent Hatschek's left diverticulum also expresses the gene (arrowhead in Fig. 6A). In later neurulae, the signal remains in the posterior third of the neural tube (arrow in fig. 6B), whereas the signal in the gut is now restricted to specific regions. It is confined to its ventral posterior-most part, a middle part, and in the anterior-most part, the signal is restricted dorsally and in the Hatschek's left anterior diverticulum (arrowhead in fig. 6B) and the anterior wall of the gut.

AmphihairyB also continues to be expressed in the same tissues as it was in gastrula stages. It is strongly expressed in the neural plate and also in the posterior part of the formed somites and in the most posterior paraxial mesoderm. . In an oblique dorso-lateral view of a very young neurula, the signal is detected in stripes at the posterior-most part of the segmented somites (the arrow in fig. 6C marks the posterior border of the last formed somite) and very highly in the anterior-most PSM. It is also detected all along the neural plate (Fig. 6C). A similar pattern was observed for older neurulae within the somites and anterior PSM, but the signal in the neural plate is no longer detected over its entire length and appears more region-specific (Fig. 6D). In transverse sections (Fig. 6E to G from anterior to posterior) the signal in the neural plate (arrowhead in Fig. 6E), in the neural plate and the somites (arrowhead and arrows in Fig. 6F) or in the somites (arrows in Fig. 6G) is better observed. Later on the signal in the somites becomes extinguished, and is only

detected in the most posterior paraxial mesoderm (arrowheads in fig. 6H and 6I). In the anterior part of the neural tube, there is a gap between the anterior-most expression domain and the next domain (Fig. 6Hand 6I, lines).

During neurula stage, *AmphihairyC* (Fig. 6J) and *AmphihairyD* (Fig. 6R) have slightly different expression patterns. Although both where expressed in the neural plate (arrowheads in fig. 6K and 6S, respectively), they were complementarily expressed in the dorsal portion of the anterior gut. *AmphihairyC* was conspicuously expressed as two patches at the inmediate lateral to the midline anterior endoderm (arrows in fig. 6K), whereas *AmphihairyD* mRNA was present in the dorsal-most part (arrow in fig. 6S) and in two lateral domains of the anterior endoderm (double arrow in fig. 6S). In a medial section, both genes were also similarly transcribed in the neural plate (arrowhead in fig. 6L). *AmphihairyC* was highly expressed in the segmented somites (arrows in fig. 6L) and low expressed in the notochord. In a posterior transverse section, another difference came to light. Although both genes where highly expressed in the forming somites, *AmphihairyC* was the only one expressed in the neural plate (arrowhead in fig. 6M and asterisk in Fig. 6T). Moreover, whereas *AmphihairyC* is weakly expressed in the forming notochord (asterisk in fig. 6M),

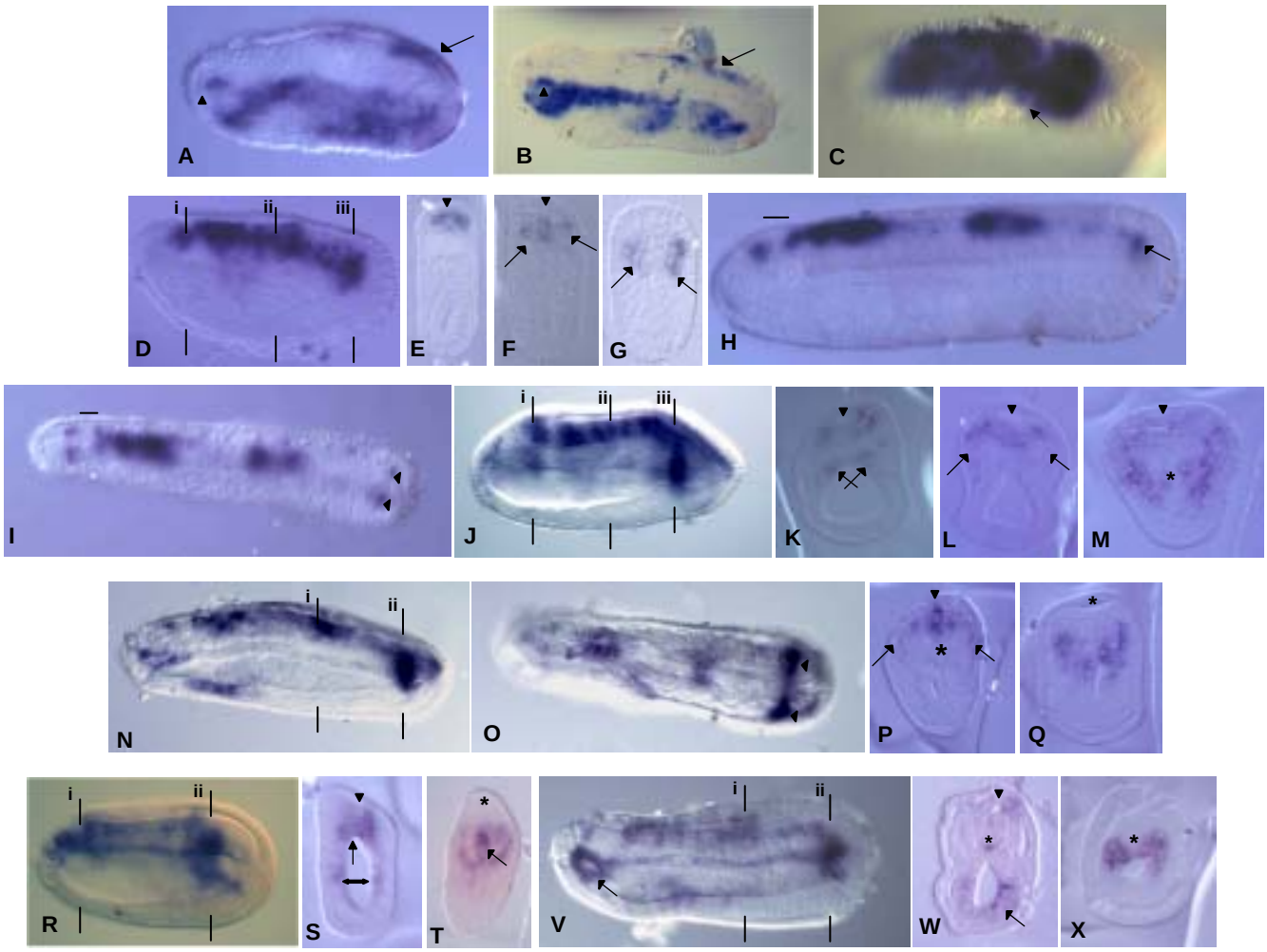

Figure 6. *AmphihairyA*, *B*, *C*, and *D* genes expression in amphioxus neurulae by whole mount *in situ* hybridization. Lateral views (A, B, C, D, H, J, N, R and V) have anterior to the left and dorsal up. In dorsal views (I and O), anterior is to the left. In transverse sections (E, F, G, K, L, M, P, Q, S, T, W and X), dorsal is up looking from posterior. A) *AmphihairyA* expression in the endoderm and the posterior dorsal nerve cord (arrow) of an early amphioxus neurula. B) *AmphihairyA* expression in the endoderm and the posterior dorsal nerve cord (arrow) of a late neurula. The expression in the Hatschek's left diverticulum is shown by an arrowhead in A and B. C) Oblique view of an early neurula that exhibits *AmphihairyB* expression in the whole neural plate and the posterior compartment of the segmented somites (the last somitic border is marked by an arrow). D) Lateral view of an early neurula showing *AmphihairyB* expression in the neural tube, the posterior PSM, and in the segmented somites. E) Cross-section through level i in F showing the expression of *AmphihairyB* in the neural tube (arrowhead). F) Cross-section through level ii in F showing the expression of *AmphihairyB* in the neural tube (arrowhead) and the segmented somites (arrows). G) Cross-section through level iii in F showing the expression of *AmphihairyB* in the posterior PSM (arrows). H) Lateral view of late neurula showing *AmphihairyB* expression in the neural tube and the posterior PSM (arrow). I) Dorsal view of the specimen in (H); the expression in the PSM is shown by arrowheads. The gap in the anterior neural tube domain is shown by a line in H and I. J) Lateral view of an early neurula stained for the *AmphihairyC* mRNA with the expression in the neural plate, the somites and the posterior PSM. K) Cross-section through level i in J showing the expression of *AmphihairyC* in the neural plate (arrowhead) and the anterior gut (arrows). L) Cross-section through level ii in J showing the expression of *AmphihairyC* in the neural tube (arrowhead) and the somites (arrows). M) Cross-section through level iii in J showing the expression of *AmphihairyC* in the posterior neural plate (asterisk) and mesoderm. The low expression levels in the forming notochord is shown by an asterisk. N) Lateral view of a late neurula stained for the *AmphihairyC* mRNA showing the expression in the neural plate, the anterior gut, the somites and the posterior PSM. O) Dorsal view of the especimen shown in N. The arrowheads show the expression in the PSM. P) Cross-section through level i in N showing the expression of *AmphihairyC* in the neural tube (arrowhead) and the segmented somites (arrows). The low expression in the dorsal notochord is shown by an asterisk. Q) Cross-section through level ii in N showing the expression of *AmphihairyC* in the posterior mesoderm and no longer in the neural plate (asterisk). R) Lateral view of an early neurula showing *AmphihairyD* expression in the anterior endoderm, the neural tube and the notochord. S) Cross-section through level i in R showing the expression of *AmphihairyD* in the neural plate (arrowhead) and the gut (arrow and double arrow). T) Cross-section through level ii in R showing the expression of *AmphihairyD* in the posterior mesoderm. The lack of *AmphihairyD* expression in the posterior neural plate is shown by an asterisk. V) Lateral view of a late neurula showing *AmphihairyD* expression in left gut diverticulum (arrow), the neural tube and the notochord. W) Cross-section through level i in V showing the expression of *AmphihairyD* in the neural plate (arrowhead), the ventral notochord (asterisk) and the gut (arrow). X) Cross-section through level ii in V showing the expression of *AmphihairyD* in the posterior mesoderm. The asterisk marks the expression in the forming notochord.

AmphihairyD is expressed at high levels in this territory (arrow in fig. 6T). During the late neurula stage, both genes are expressed in dorsal structures (Fig. 6N for *AmphihairyC* and 6V for *AmphihairyD*) with several differences. *AmphihairyC* signal is mainly seen in the neural tube (arrowhead in fig. 6P), the gut, the segmented somites (arrows in fig. 6P), and only at low levels within the dorsal notochord (asterisk in Fig. 6P). *AmphihairyD* signal is also detected in the neural tube (arrowhead in Fig. 6W) and in the gut (Fig. 6W, arrow). However, *AmphihairyD* signal is not detected within the segmented somites, but is conspicuous in the ventral notochord (asterisk in fig. 6W). In a posterior section through level ii in 6N and 6V, the signal is detected in the forming somites for both genes. However, *AmphihairyD* signal is stronger detected in the forming notochord (asterisk in fig. 6X) than *AmphihairyC* signal (Fig. 6Q).

Interestingly, all four genes are similarly expressed within the neural tube of late neurulae as they where expressed in the neural plate of gastrula stages. That is, *AmphihairyA* being the hairy gene expressed in a posterior-most domain of the neural tube (Fig.6B), *AmphihairyB* being highly expressed in the anterior-most domain (Fig. 6H), and *AmphihairyC* and *AmphihairyD* similarly expressed in between (Fig. 6N).

AmphihairyA, ***B***, ***C*** and ***D*** genes expression in larvae

During larval stages, *AmphihairyA* expression is similar to that observed in late neurulae. Its mRNA continues restricted to the posterior-most part of the dorsal nerve corde, and within the gut it is strongly expressed in three regions (Fig. 7A): in the most posterior gut it has a two-domain pattern; in a middle region there are scattered positive cells; and in the anterior gut it is expressed dorsally in the left anterior gut diverticulum that will become the Hatschek's pit (asterisk in fig. 6A), in the anterior wall of the gut, and in the pharynx endoderm more ventrally. Cross sections through level i in fig. 7A show *AmphihairyA* expression in the left anterior gut diverticulum (arrow in fig. 7B). A cross section through a more posterior level (ii in fig. 7A) shows the signal within the posterior neural tube (arrowhead in fig. 7C) and the gut (arrow in fig. 7C).

AmphihairyB expression is still restricted to the anterior part of the neural tube during larval stages (Fig. 7D), which is better observed in a transverse section through level i in fig. 7D (arrowhead in fig. 7E). There is still a gap between the anterior patch of *AmphihairyB* expression and the rest along the mid-anterior neural tube (line in Fig. 7D).

In contrast to earlier stages *AmphihairyC* is no longer

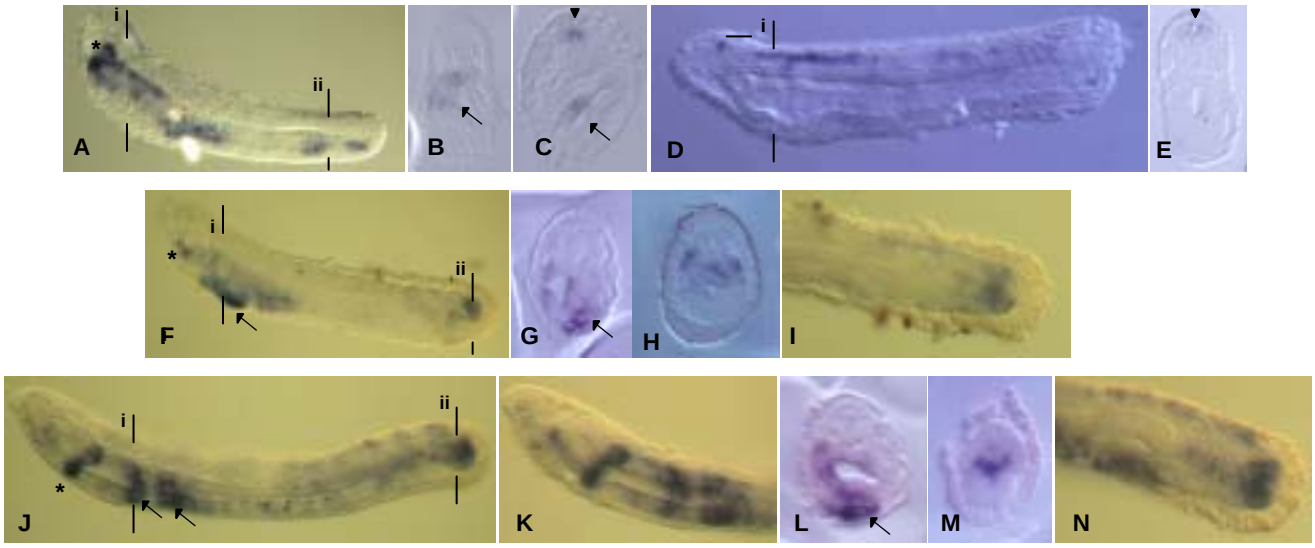

Figure 7. *AmphihairyA*, *B*, *C*, and *D* genes expression in amphioxus larvi by whole mount *in situ* hybridization. Lateral views (A, D, F, J) have anterior to the right and dorsal up. In dorsal and ventral views (I, K, N), anterior is to the right. In transverse sections (B, C, E, G, H, L, M), dorsal is up looking from posterior. A) *AmphihairyA* expression in the endoderm and the posterior dorsal nerve cord of an amphioxus larva. B) Cross-section through level i in A showing the expression of *AmphihairyA* in the gut (arrow). C) Cross-section through level ii in A showing the expression of *AmphihairyA* in the gut (arrow) and the posterior neural tube (arrowhead). D) *AmphihairyB* expression in the anterior half of the dorsal nerve cord. E Cross-section through level i in D showing the expression of *AmphihairyB* in the neural tube (arrowhead). F) Lateral view of an amphioxus larva showing *AmphihairyC* expression in the posterior tail bud, and in the anterior endoderm (arrow for the pharyngeal endoderm, and asterisk for the preoral pit). G) Cross-section through level i in F showing the expression of *AmphihairyC* in pharyngeal endoderm (arrow). H) Cross-section through level ii in F showing the expression of *AmphihairyC* in the posterior mesoderm. I) Dorsal view of the specimen in F showing the *AmphihairyC* expression within the tail bud. J) Lateral view of a larva showing *AmphihairyD* expression in the anterior club-shaped gland (asterisk) and the pressumptive first and second gill slits (arrows), and in the posterior mesoderm. K) Ventral view of the specimen in (J). L) Cross-section through level i in J showing the expression of *AmphihairyD* in the pharyngeal endoderm (arrow). M) Cross-section through level ii in J showing the expression of *AmphihairyD* in the posterior mesoderm. N) Dorsal view of the specimen in J showing the *AmphihairyD* expression within the tail bud.

detected in neural tissues. Its expression is mainly

observed in the posterior paraxial mesoderm associated with the tail bud, and in the anterior endoderm, where it is conspicuously expressed in a region ventral to the mouth and the branchial anlage (arrow in fig. 7F), and in the ventral part of the left anterior gut diverticulum or preoral pit (asterisk in fig. 7F). A cross section through level i in fig. 7F shows the expression in the ventral pharyngeal endoderm (arrow in fig. 7G), and in the posterior mesoderm through a more posterior level (Fig. 7H). In a dorsal view, the signal in the posterior tail bud is better observed (Fig. 7I).

Last, *AmphihairyD* mRNA is not longer present in the neural tube during larval stages. It is only expressed in the anterior endoderm and in the posterior tail bud, similarly to *AmphihairyC*. In the anterior region of the larva, it is expressed in a 3-stripe pattern. The first one marks the region where the ventral duct of the club-shaped gland is developing (asterisk in fig. 7J), and the two posterior ones that are in the branchial anlage, may prefigure the first two gill slits (arrows in fig. 7J). This pattern is clearly seen from a ventral view (Fig. 7K). A section through level i in fig. 7I makes further visible *AmphihairyD* expression within the pharyngeal endoderm (arrow in fig. 7L), and in the posterior mesoderm through a more posterior level (Fig. 6M). It is also noticeable from a dorsal view, the signal within the entite tail bud, although a bit higher in its anterior part, the chordoneural hindge of the tail bud (Fig. 7N).

AmphihairyC and ***AmphihairyD*** genomic southern blots

As shown before, *AmphihairyC* and *AmphihairyD* exhibited a quite similar expression pattern during amphioxus embryogenesis. Although by sequence they are not the most similar ones (see figs. 1 and 2), we wanted to really ascertain whether they represented different hairy genes or very polymorphic alleles of a single hairy locus. With this aim, Southern blots of genomic DNA obtained from single individuals were hybridized with the most specific cDNA region obtained by PCR with the primers h5 and h3rep (table II) over the full length cDNA clones. As a different hibridizing pattern is observed for each gene and for each digestion, we concluded that they represent different hairy genes in the amphioxus genome (Fig. 8).

AmphihairyE and ***AmphihairyF*** genes expression

Although we tried several times by *in situ* whole mount hybridizations with *AmphihairyE* and *F* probes, we were unable to detect any expression on amphioxus embryos. We then thought in the possibility that these genes were expressed at too low levels to be readily detected by this technique. We thus planned RT-PCR experiments on embryonic and larval cDNA with specific primers for each of these genes plus and internal control, namely *AmphihairyC*. After blotting the gels and further hibridization with internal probes of *AmphihairyC* plus *AmphihairyE* or *AmphihairyC* plus *AmphihairyF* specific regions, we observed that the mRNA levels of *AmphihairyE* and *AmphihairyF* were not detected or detected much under the levels of AmphihairyC (Fig. 9). This made sense, as *AmphihairyC*, in contrast to the former genes, was already detected by whole mount experiments.

We thus concluded that neither *AmphihairyE* nor *F* are expressed during amphioxus embryogenesis, or they are expressed under detectable levels by in situ whole mount hybridizations. We favour the former, as even with the high-sensitive technique of PCR, we did not detect *AmphihairyE* or *F* expression. In the mouse, the hairy *HES1* gene is expressed during embryogenesis, but also in the adult (Sasai *et al*., 1992). It was then still possible that these two Amphihairy genes were expressed in the adult and no longer in the embryo. However, the same RT-PCR strategy was used on adult amphioxus cDNA, and we obtained the correspondent amplification band for AmphihairyC but not for *AmphihairyE* or *F* (data not shown).

DISCUSSION

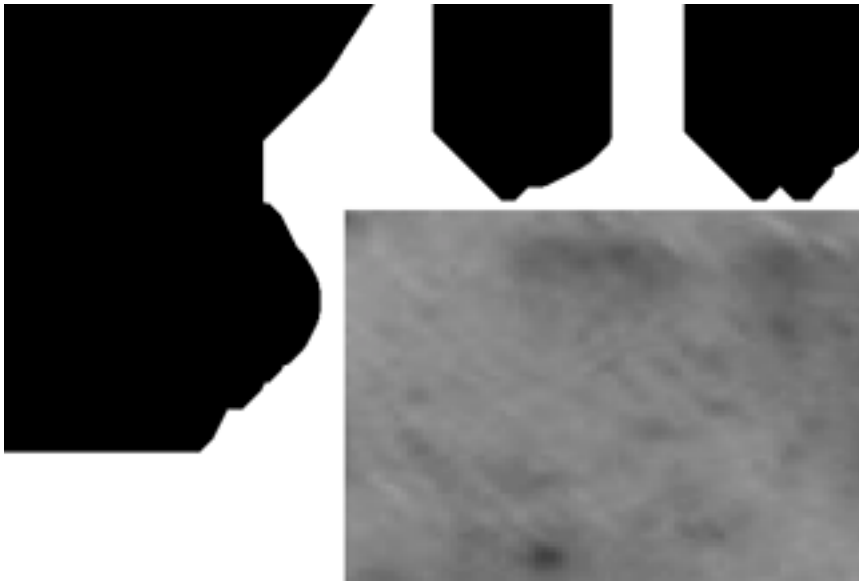

Figure 8. *B. floridae* genomic Southern Blots. A) Genomic DNA from single individuals digested with *Pst* I (P) or *Eco* RI (R) were probed for *AmphihairyC* (left panel) and *AmphihairyD* (right panel). B) Total genomic DNA from a single animal digested either with the methylation-sensitive enzyme *Hpa*II (lane H) or with its methylation-insensitive isoschizomer *Msp*I (lane M) probed for *AmphihairyC* (left panel) and *AmphihairyD* (right panel).

The amphioxus hairy family

Here we report the characterization of the big hairy family of amphioxus. This family is composed by a minimum of four genes that are expressed during *B. floridae* embryogenesis (*AmphihairyA* to *D*) and have undergone subfunctionalization, two hairy genes that are not longer expressed (*AmphihairyE* and *AmphihairyF*), and at least two very diverged hairy duplicate copies that most probably have undergone nonfunctionalization (*Amphihairy-like1* and *Amphihairy-like2*). Although other cases of gene duplication have been reported in amphioxus (Minguillón *et al*., 2002), to our knowledge, this is the first report of such an extreme case of duplication. Moreover, this is also the first case in which a differential fate after duplication can be asserted for the duplicate copies. In other cases of duplicate genes in amphioxus, only two copies were reported (one case with three), and either they were too similar to distinguish among their expression patterns (Holland *et al*., 1995), or only quantitative differences accounted for their expression patterns (Shimeld, 1997).

As all six "canonical" hairy genes are closely related to both vertebrate and protostome hairy genes (Fig. 2, table III) the nomenclature for their relationship with other hairy genes needs of extra concepts. For instance, each amphioxus hairy genes is semi-ortholog of the single beetle *hairy* gene, and the latter is pro-ortholog of all Amphihairy genes. On the other hand, each amphioxus hairy gene is trans-ortholog of each vertebrate or *Drosophila* hairy gene (for nomenclature see Sharman, 1999). Although phylogenetic trees cannot be constructed due to the short available sequence of Amphihairy-like1 and 2, we argue that they are hairy representatives and not members of other related families as the Enhancer of split family. Although very low, the highest similarity (of the Amphihairy-like sequences) is found with other hairy genes (Fig. 4). The existence of multiple hairy copies in the amphioxus genome does not enter in conflict with the "pre-duplicative" state of its genome. All the copies put together would constitute the pro-ortholog of vertebrate hairy genes (Fig. 2) and each Amphihairy gene is a trans-ortholog of each vertebrate gene, which mean that they have arose by independent duplication in the *B. floridae* genome. The fact that the cephalochordate lineage scaped the massive gene/whole genome duplications (see Wolfe, 2001 for discussion) do not imply that its genome is not evolving, meaning that it can suffer specific gene duplications, gene losses, etc. Hence, it has to be kept in mind that although amphioxus may resemble the ancestor of the vertebrates, it is not the ancestor, only its closest living relative, a priviledge position that did not include the freezing of its genome.

Why did the *B. floridae* genome undergo such an extreme case of duplication of the hairy genes is unknown. Has the duplication occurred recently in evolution and is specific to *B. floridae*, or in contrast, it is more ancient and the same representatives should be found in other *Branchiostoma* species? To really ascertain between both possibilities, it would be necessary the study of the hairy family in other amphioxus species, such *B. lanceolatum*. In fact, there are cases in the literature in which a gene has been duplicated only in one amphioxus species (Dalfó *et al*., 2001), and even duplicated different times in each amphioxus species (Karabinos and Bhattacharya, 2000). Therefore, it would be interesting to study this gene(s) in other species, and moreover, in the case that a single gene existed, to see whether its expression pattern is the summation of the patterns of the *B. floridae* hairy genes.

Differential fate after duplication of the amphioxus hairy genes

Several mathematical models have been developed to explain the future of paralog genes after duplication from a single ancestral gene. These models predict that duplicate genes initially have fully overlapping, redundant functions, such that one copy may shield the second from natural selection, if gene dosage is not critical. Because deleterious mutations occur more frequently than beneficial ones (Lynch and Walsh, 1998), the classical model predicts that one of the duplicate loci should most commonly deteriorate into a pseudogene (i. e. be fixed as a null allele) (Watterson, 1983). The classical model also considers a rarer alternative: maintenance of duplicate copies, owing to the fixation of a rare beneficial mutation in one copy that endows it with a novel function, while the other maintains the original role (Ohno, 1970). The DDC model (for Duplication-Degeneration-Complementation) was proposed by Force *et al*. (1999) as a possible explanation for the higher maintenance of duplicate genes in duplicated genomes observed than that expected under the classical mathematical models. According to this model, duplicate copies of an ancient single gene preserve their manteinance in the genome by subfunctionalization i. e. by differential degeneration of regulatory regions among the duplicate genes.

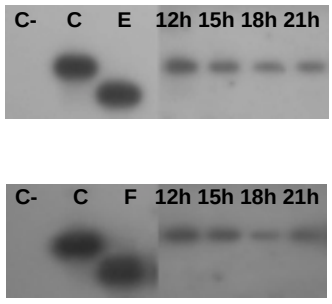

Figure 9. RT-PCR on RNA obtained from different amphioxus embryonic stages. 12h, 15h, 18h and 21h refer to hours after fertilization. Abbreviations used are: C-, negative control; C, *AmphihairyC* positive control; E, *AmphihairyE* positive control; F, *AmphihairyF* positive control.

Hence, paralog genes decouple a pleiotropic role that was carried out by their single ancestor gene and become thus all copies necessary for carrying out all the functions that the ancestral pre-duplicate gene did.

AmphihairyA, *B*, *C* and *D* genes have expression patterns that generally do not overlap, although some of the differences are subtle. Briefly, *AmphihairyA* is mainly expressed within the endoderm and to a lesser extent at the posterior neural tube, *AmphihairyB* in the neural tube, posterior paraxial mesoderm, and somites, and *AmphihairyC* and *D* are more widely expressed in all three embryonic layers. There are several cases of combinatorial patterns among the Amphihairy genes all along amphioxus development. More noticeable are those within the neural plate first and the neural tube latter, and those in some anterior endoderm-derived structures in amphioxus larvae. The single mouse gene *HES1* is expressed all along the neural tube, but strikingly, only the summation of expression for all four hairy genes of amphioxus covers the entire amphioxus neural plate and neural tube. *AmphihairyB* is expressed at the very front of the animal, *AmphihairyA* at the posterior, and *AmphihairyC* and *D* genes in between. It is also interesting that a combination of two hairy genes (*AmphihairyA* plus *AmphihairyB*) is required to include all the cells of Hatschek's left diverticulum, a suspected homolog of the vertebrate adenohypophysis. In contrast, zebrafish *her9* is expressed in the whole pituitary. Hence, if we assume that this subfunctionalization has originated by differential lossess of *cis* regulatory sequences, one must be struck by the high degree of complexity in the regulatory regions of hairy genes. Only for these two examples, we have to think in at least three distinct regulatory elements driving the expression of an ancestral pre-duplicative hairy gene within the neural tube, and at least two driving the expression in the left anterior diverticulum, that have been differentially lost by the duplicate copies during or after the process of duplication.

The hairy family genes in *B. floridae* seem to have undergone different processes after duplication from a single hairy gene in its lineage. First, as shown above, *AmphihairyA*, *B*, *C*, and *D* genes seem to have undergone subfunctionalization, a process guaranteeing their maintenance in the genome as duplicate genes. Second, the other two hairy "canonical" genes, *AmphihairyE* and *F*, seem to be in the process of subfunctionalization, as they do not have any obvious function in developing or adult amphioxus, as shown by the RT-PCR experiments (Fig. 9), or are expressed in very particular instants of development that escaped the analyses. And finally, the two *Amphihairy-like* genomic sequences seem to be already nonfunctional (pesudogenes), as we were unable to recover them after exhaustive cDNA screening or random EST sequencing but also since they have a very divergent "hairy-like" sequence.

Insights into amphioxus early somitogenesis from Amphihairy genes expression

The amphioxus hairy gene expression data shed some light on specific aspects of cephalochordate development. First, they are relevant to the long-standing discussion on the asymmetry of amphioxus somitogenesis and somites. Classical morphologists had noticed that the amphioxus somite-rows were inherently asymmetric, with the left-side somites a bit advanced in their development, and thus offset a little anteriorly to their right-side counterparts. However, the extent of this asymmetry was controversial. Conklin (1932) and Hatschek (1893) agreed that somites are roughly symmetrical until the 7- or 8-somite stage, but Cerfontaine (1906) postulated that asymmetry is present from the first pair, with a slight delay in the development of the right side with respect to that of the left side. Our results support Cerfontaine observations, as the asymmetry is already molecularly detected in means of *AmphihairyB* expression in a late gastrulae even when the somites are not yet visible (Fig. 5B). In this embryo, three pre-somitic stripes are visible on the left side, whereas only two are seen on the right. Still during somitogenesis, all three amphioxus hairy genes that are expressed in the presumptive somitic mesoderm (*AmphihairyB*, *C*, and *D*) shed light on the formation of the first four pair of amphioxus muscular somites. These somites have the peculiarity that they bud off virtually simultaneously from the dorso-lateral walls of the archenteron. Hatschek (1893) claimed to have observed embryos with a single pair of somites, whereas Conklin (1932) never detected them and claimed that always more than one somite were present in all the embryos he analyzed. Regardless of their simultaneous or sequencial antero-posterior appearance, they are molecularly prefigured one by one, as we have detected gastrulae with a 2-stripe, 2- to 3-stripe, 3-stripe, and 4-stripe pattern (Fig. 5). Moreover, the maturation of those four first muscular somites also appears as a sequencial process, as the intensity of *AmphihairyB*, *C* and *D* expression is stronger in posterior (and thus younger) pre-somites than in the anterior (older) ones (Fig. 5).

Similarities among amphioxus and vertebrate hairy genes

Within the central nervous system, the mouse gene *HES1* is expressed in undifferentiated neuronal precursor cells in the ventricular zone, and its transcription decreases as neurogenesis proceeds until expression is no longer detected in mature neurons or glial cells (Sasai *et al*., 1992). Accordingly, mutant mice for the *HES1* gene exhibit severe neural defects. Moreover, in their brain there is an up-regulation of some neural bHLH factors and postmitotic neurons appear prematurely. It thus appears that *HES1*, like the *Drosophila hairy* gene, acts as a negative regulator of

neurogenesis, and that its down-regulation is required for precursor cells to enter the differentiation processes (Ishibashi *et al*., 1995). It seems reasonable to think that amphioxus hairy genes are carrying out a similar function within the amphioxus nerve cord. In the mouse, *HES1* is expressed all along the neural tube which corresponds to the summation of the expression patterns of all four Amphihairy genes. In larval stages, only the anterior- and the posterior-most parts of the neural tube are positive for a hairy gene (*AmphihairyB*, Fig. 7D, and *AmphihairyA*, Fig. 7A, respectively). It is tempting to speculate that these differences may account for differential maturation rates along the neural tube. Unfortunately, no work has been done about this differentiation process to really ascertain whether amphioxus hairy genes are regulated and function in a similar way than *HES1* in the mouse. In the zebrafish, the hairy ortholog *her9* is expressed in the mid- and hindbrain but in the MHB (Leve *et al*., 2001). It is striking thus to note a gap also in the anterior amphioxus neural tube (*AmphihairyB* expression in fig. 6 and 7). The existence of a tripartite brain in cephalochordates is still on debate (Ferrier *et al*., 2001 and references therein) and the gap of expression in this region suggestively weights to the presence of a MHB.

In endodermal derivatives, *HES1* is expressed in, and required for, the proper development of the endocrine islet cells of the mouse pancreas as well as other dispersed endocrine cells along the entire gut (Jensen *et al*., 2000). It was suggested that *HES1* functions as a general negative regulator of endodermal endocrine differentiation, in an analogous way that it does within neural precursors (Jensen *et al*., 2000). Similarly, the amphioxus hairy genes are expressed in the developing gut and certain derivatives (*AmphihairyA*, *C* and *D* only at larval stages). Amphioxus does not have a discrete pancreas but has several types of endocrine cells incorporated into the gut epithelium, some of which are possibly homologous of the pancreas-islet cells of mammals (Holland *et al*., 1997a). Interestingly, patches of *AmphihairyA* expression in the gut may represent regions with pressumptive endocrine cell types. All hairy genes but *AmphihairyB* are also conspicuosly expressed in combinatorial patterns in some endoderm-derived glands. Briefly, in the left gut diverticulum, *AmphihairyA* is predominantly expressed in the dorso-posterior region and *AmphihairyC* in its ventral region, whereas *AmphihairyD* is conspicuously expressed in the club-shaped gland. Hatschek's left gut diverticulum contributes to the Hatschek's pit in the adult, a structure thought to be homologous to the vertebrate adenohypophysis (Whittaker, 1997), an organ where the zebrafish hairy gene *her9* is expressed (Leve *et al*., 2001). The homology of the club-shaped gland has been difficult to assess (Goodrich, 1930), and based on specific gene expression data some authors support the hypothesis of being an amphioxus evolutionary novelty (Jackman, 2000). The expression of *AmphihairyD* in this misterious gland do not resolve any discussion about the existence of an homologous structure to the club-shaped gland in vertebrates, as it may merely reflect the future endocrine nature of this gland.

In summary, the addition of the expression patterns of *AmphihairyA* to *D* genes resembles *a grosso modo* the expression of the single mouse *HES1* gene or the multiple hairy genes in other vertebrate species. They are expressed in the CNS (all of them in complementary patterns), in the posterior paraxial mesoderm (*AmphihairyB*, *C*, and *D*), in the posterior compartment of the segmented somites (*AmphihairyB*), in the gut (*AmphihairyA* always and *AmphihairyC* and *D* in larval stages), and in the notochord (*AmphihairyC* and *D*). Hence, the common ancestor of cephalochordates and vertebrates already possessed a single hairy gene of a very pleitropic nature, in contrast to protostome hairy genes.

Do the Amphihairy genes cycle within the amphioxus posterior paraxial mesoderm?

There is not a single technique available to demonstrate whether one (or more) amphioxus hairy genes are cycling within the PSM in a similar way that they do in amniotes. Amphioxus somitogenesis is divided in two different phases. During an earlier one, the somites originate by the budding off of the dorso-lateral walls of the archenteron (which could be considered a sort of PSM) forming their coeloms by enterochoely. The first eight pair of muscular somites are formed during this early phase from paraxial mesoderm formed during gastrulation (Holland *et al*., 1997b). In contrast, during the second phase, somites arise directly from the proliferative tail bud by a schyzocoelic process (Schubert *et al*., 2001) without the intervention of any PSM between the tail bud and the nascent somites. Hence, if any hairy gene has a cycling behaviour, it will only be visible during the early phase of amphioxus somitogenesis. Then only *AmphihairyB*, *C*, and *D* could in principle cycle during the first phase of somitogenesis, as they are expressed at the right place at the right time. Nevertheless, the first eight muscular somites arise very quickly (within a few hours) and thus there is not much time for any cycling gene transcription.

In the chicken, the onset of the dynamic expression of hairy genes correlates with ingression of the paraxial mesoderm territory from the epiblast into the primitive streak. Hence, the number of oscillacions experienced by somitic cells is correlated with their position along the anterio-posterior axis (Jouve *et al*., 2002). If any hairy gene is being regulated accordingly to this cyclic behaviour, we should have been able to detect their "on and off" expression within or around the blastopore, the equivalent structure to the primitive streak, and this was not the case. Besides, we have detected

Amphihairy genes expression within the paraxial mesoderm (*AmphihairyB*, *C*, and *D*) not as pulses, but as stripes that prefigure the first muscular somites. Moreover, we saw gastrulae with 2, 3 or 4 stripes (Fig. 5), which may indicate that they are formed one by one, in the anterior PSM in a way much more similar to that of hairy genes in lower vertebrates such as the zebrafish *her6* gene or both *Xenopus* hairy genes. The expression of these genes is seen as one to three stripes at the anterior PSM that prefigure the regions where new somites will be added (Pasini *et al*., 2001; Davis *et al*., 2001). Then, we suggest that the cycling behaviour of the hairy family may be an amniote novelty, and not a secondary loss of this behaviour in the fish and frog lineages. None gene that exhibits a cyclic expression in higher vertebrates behaves so in *Xenopus*. On the contrary, there are some cycling genes in the zebrafish, as the Notch ligand *DeltaC* and the E(spl) family genes *her1* and *her7* (Jiang *et al*., 2000; Holley *et al*., 2000; cited in Leve *et al*., 2001). Also an E(spl) family gene, *Hes7*, cycles within the mouse PSM (Bessho *et al*., 2002). Therefore, it is possible that other genes related to the Notch signalling pathway or other genes from the E(spl) family cycle in the amphioxus PSM. Another possibility would be that somitogenesis in amphioxus would not need such an accurate control as happens with *Xenopus*. Interestingly, somitogenesis is also inherently asymmetrical in the frog, being the right side of the embryo temporally advanced in segmentation (Davis *et al*., 2001). It is thus plausible that there is no way to coordinate an asymmetrical cycling pattern, and that those organisms, namely amphioxus and the frog, use other strategies.

Acknowledgements

We are indebted to Jim Langeland for cDNA library, to the Serveis Científico-Tècnics of the Universitat de Barcelona for the sectioning and to Elisa Serra for unvaluable help. We thank colleagues from Departament de Genètica for discussions and Seb Shimeld, Ricard Albalat, and Nic Holland for discussion, support, and advice. This work was supported by grants from DGESIC to J.G.F (Ministerio de Educación y Cultura PB98-1261-C02-02). C.M. holded a CIRIT (Generalitat de Catalunya) fellowship.

References

Bayascas, J. R. (1997). Estudi comparatiu de gens reguladors del desenvolupament: caracterització de gens amb homeobox del complexe Hox i de la família Otx a planàries. PhD Thesis. University of Barcelona, Spain.

Bessho Y, Sakata R, Komatsu S, Shiota K, Yamada S, Kageyama R (2001). Dynamic expression and essential functions of Hes7 in somite segmentation. *Genes Dev.* 15, 2642-7.

Cerfontaine, P. (1906). Recherches sur le développement de l'Amphioxus. *Arch. de Biol.,* T. 22.

Conklin, E. G. (1932). The embryology of amphioxus. *J. Morphol.* 54, 69-151.

Carroll, S.B., Laughon, A., and Thalley, B.S. (1988). Expression, function, and regulation of the *hairy* segmentation protein in the *Drosophila* embryo. *Genes Dev.* 2, 883-890.

Dalfó, D., Cañestro, C., Albalat, R., Gonzàlez-Duarte, R. (2001). Characterization of a microsomal retinol dehydrogenase gene from amphioxus: retinoid metabolism before vertebrates. *Chem. Biol. Interact*. 130-132, 359-370.

Davis RL, Turner DL, Evans LM, Kirschner MW. (2001). Molecular targets of vertebrate segmentation: two mechanisms control segmental expression of Xenopus hairy2 during somite formation. *Dev Cell.* 1, 553-65.

Dawson, S.R., Turner, D.L., Weintraub, H, and Parkhurst, S.M. (1995). Specificity of the hairy/Enhancer of split Basic Helix-Loop-Helix (bHLH) Proteins Maps outside the bHLH domain and Suggests Two Separable Modes of Trnascriptional Repression. *Mol. Cel. Biol.* 15, 6923-6931.

Ferrier DE, Minguillon C, Cebrian C, Garcia-Fernandez J. (2001). Amphioxus Evx genes: implications for the evolution of the Midbrain-Hindbrain Boundary and the chordate tailbud. *Dev Biol.* 237, 270-81.

Fisher, A., and Caudy, M. (1998). The function of hairy-related bHLH repressor proteins in cell fate decissions. *BioEssays* 20, 298-306.

Force, A., Lynch, M., Pickett, F.B., Amores, A, Yan, Y., and Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* 151, 1531-1545.

Goorich, ES. (1930). The development of the club-shaped gland in amphioxus. *Q. J. Microsc. Sci*. 74, 155-164.

Hatschek, B. (1893). The amphioxus and its development. Swan Sonnenschein, London.

Holland, L. Z., Holland, P. W. H., and Holland, N. D. (1996). Revealing homologies of distantly related animals by *in situ* hybridization to developmental genes: amphioxus versus vertebrates. In "Molecular Zoology: Advances, Strategies and Protocols". (J.D. Ferraris and S.R. Palumbi, Eds.), pp. 267-282; 473-483. Wiley, New York.

Holland, L. Z., Kene, M., Williams, N. A., and Holland, N. D. (1997b). Sequence and embryonic expression of the amphioxus *engrailed* gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segment-polarity homolog in *Drosophila*. *Development* 124, 1723-1732.

Holland, N. D., and Holland, L. Z. (1993). Embryos and larvae of invertebrate deuterostomes. In "Essential Developmental Biology: a Practical Approach" (C. D. Stern and P. W. H. Holland, Eds.) pp. 21-32. IRL Press, Oxford.

Holland, P.W.H., Koschorz, B., Holland, L.Z., Herrmann, B.G. (1995). Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development* 121, 4283-4291.

Holland, P.W.H., and Garcia-Fernàndez, J. (1996). Hox genes and chordate evolution. *Dev Biol.* 173, 382-95.

Holland, P.W.H., Patton, S.J., Brooke, N.M., and Garcia-Fernàndez, J. (1997a). Genetic patterning of ectoderm and endoderm in amphioxus: from homeobox genes to hormones. In: "Advances in Comparative Endocrinology" (S. Kawashima and S. Kikuyama, Eds.). pp. 247-252. Moduzzi Editore, Bologna.

Holley SA, Geisler R, Nusslein-Volhard C. (2001) Control of her1 expression during zebrafish somitogenesis by a delta-dependent oscillator and an independent wave-front activity. *Genes Dev.* 14, 1678-90.

Ishibashi M, Ang SL, Shiota K, Nakanishi S, Kageyama R, Guillemot F. (1995). Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* 9, 3136-48.

Jackman WR (2000). Segmentation and molecular patterning of the amphioxus hindbrain, pharyng and somites. PhD Thesis, University of Oregon, USA.

Jensen J, Pedersen EE, Galante P, Hald J, Heller RS, Ishibashi M, Kageyama R, Guillemot F, Serup P, Madsen OD. (2000). Control of endodermal endocrine development by Hes-1. *Nat Genet.* 24, 36-44.

Jiang YJ, Aerne BL, Smithers L, Haddon C, Ish-Horowicz D, Lewis J. (2000). Notch signalling and the synchronization of the somite segmentation clock. *Nature* 408, 475-9.

Jiménez, G., Paroush, Z., and Ish-Horowicz, D. (1997). Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev.* 11, 3072-3082.

Jouve C, Palmeirim I, Henrique D, Beckers J, Gossler A, Ish-Horowicz D, Pourquie O. (2000). Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. *Development* 127, 1421-9.

Jouve C, Iimura T, Pourquie O. (2002). Onset of the segmentation clock in the chick embryo: evidence for oscillations in the somite precursors in the primitive streak. *Development* 129, 1107-17.

Karabinos, A., Bhattacharya, D. 2000. Molecular evolution of calmodulin and calmodulin-like genes in the cephalochordate *Branchiostoma*. *J. Mol. Evol.* 51, 141-148.

Langeland JA, Tomsa JM, Jackman WR Jr, Kimmel CB. (1998). An amphioxus snail gene: expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. *Dev Genes Evol.* 208, 569-77.

Ledent, V., and Vervoort, M. (2001). The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis. *Genome Res.* 11, 754-770.

Leve C, Gajewski M, Rohr KB, Tautz D. (2001). Homologues of c-hairy1 (her9) and lunatic fringe in zebrafish are expressed in the developing central nervous system, but not in the presomitic mesoderm. *Dev Genes Evol.* 211, 493-500.

Lynch, M., and Walsh, J.B. (1998). Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Minguillón, C., Ferrier, D. E. K., Cebrián, C., and Garcia-Fernàndez, J. (2002). Gene duplications in the prototypical cephalochordate amphioxus. *Gene* In press.

Moore, A.W., Barbel, S., Jan, L.Y., and Jan, Y.N. (2000). A genomewide survey of basic helix-loop-helix factors in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10436-10441.

Nishimura M, Isaka F, Ishibashi M, Tomita K, Tsuda H, Nakanishi S, Kageyama R. (1998). Structure, chromosomal locus, and promoter of mouse Hes2 gene, a homologue of Drosophila hairy and Enhancer of split. *Genomics* 49, 69-75.

Ohno, S. 1970. Evolution by gene duplication. George Allen and Unwin, London.

Palmeirim I, Henrique D, Ish-Horowicz D, Pourquié O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*. 91, 639-648.

Panopoulou GD, Clark MD, Holland LZ, Lehrach H, Holland ND. (1998). AmphiBMP2/4, an amphioxus bone morphogenetic protein closely related to Drosophila decapentaplegic and vertebrate BMP2 and BMP4: insights into evolution of dorsoventral axis specification. *Dev Dyn.* 213, 130-9.

Pasini A, Henrique D, Wilkinson DG. (2001). The zebrafish Hairy/Enhancer-of-split-related gene her6 is segmentally expressed during the early development of hindbrain and somites. *Mech Dev.* 100, 317-21.

Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. (1992). Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. *Genes Dev.* 12B, 2620-34.

Sharman AC. (1999). Some new terms for duplicated genes. *Semin Cell Dev Biol.* 10, 561-3.

Schubert, M., Holland, L. Z., Stokes, M. D., and Holland N. D. (2001). Three amphioxus *Wnt* genes (*AmphiWnt3, AmphiWnt5, and AmphiWnt6*) associated with the Tail Bud: the evolution of somitogenesis in chordates. *Dev. Biol.* 240, 262-273.

Shimeld, S.M. 1997. Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. *Dev. Biol.* 183, 74-85.

Sommer, R.J., and Tautz, D. (1993). Involvement of an orthologue of the *Drosophila* pair-rule gene *hairy* in segment formation of the short germ-band embryo of Tribolium. *Nature* 361, 448-450.

Van Doren M, Bailey AM, Esnayra J, Ede K, Posakony JW. (1994). Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of achaete. *Genes Dev.* 8, 2729-42.

Watterson, G.A. (1983). On the time for gene silencing at duplicate loci. *Genetics* 105, 745-766.

Whittaker, J. R. (1997). Cephalochordates, the Lancelets. In "Embryology: constructing the organism" (S. F. Gilbert and A. M. Raunio, Eds.) pp. 365-381. Sinauer Associates Inc. Publishers, Sunderland.

Wolfe KH. (2001). Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet.* 2, 333-41.

Wrischnik, L.A., and Kenyon, C.J. (1997). The role of *lin-22*, a *hairy/Enhancer of split* homolog, in patterning the peripheral nervous system of *C. elegans*. *Development* 124, 2875-2888.