Universitat Autònoma de Barcelona

Evolució cromosòmica en mamífers: cariotips ancestrals i punts de trencament evolutius

Memòria presentada per **Marta Farré Belmonte** per a optar al grau de Doctor en Biologia Cel·lular per la Universitat Autònoma de Barcelona.

Aquest treball ha estat realitzat a la Unitat de Biologia Cel·lular del Departament de Biologia Cel·lular, Fisiologia i Immunologia de la Universitat Autònoma de Barcelona, sota la co-direcció de la Dra. **Aurora Ruiz-Herrera Moreno** i la Dra. **Montserrat Bosch Gallego**.

Dra. Aurora Ruiz-Herrera Moreno Dra. Montserrat Bosch Gallego Marta Farré Belmonte

RESUM

Per a poder entendre la dinàmica evolutiva dels genomes és imprescindible conèixer com estan organitzats els cromosomes de les diferents espècies i determinar quins tipus de reorganitzacions cromosòmiques estan implicades en els processos d'especiació i en esdeveniments macroevolutius que afecten als grans grups taxonòmics. És per això que en aquesta tesi ens hem plantejat definir el cariotip ancestral de mamífers, amniotes i tetràpodes per a poder determinar les regions conservades (*Homologous Syntenic Blocks*, HSBs) i les regions de trencament evolutiu (*Evolutionary Breakpoint Regions*, EBRs) partint del genoma humà com a referència. Gràcies a la inclusió del genoma d'espècies *outgroup* (granota i gall), hem pogut millorar el cariotip ancestral de tetràpodes i amniotes, definint noves associacions sintèniques com a caràcters sinapomòrfics dels amniotes i dels mamífers. Igualment, hem analitzat la distribució de les EBRs en el genoma humà, veient que aquestes regions no estan distribuïdes a l'atzar i un 20% d'elles han estat re-utilitzades al llarg de l'evolució. Hem relacionat la distribució de les EBRs amb l'abundància de seqüències repetitives en el genoma humà, trobant un enriquiment de repeticions en tàndem en aquestes EBRs i una co-localització amb certs elements mòbils o transponibles (AAAT-*Alus*). A més a més, hem estudiat el paper del constrenyiment selectiu sobre el manteniment de les regions conservades i hem vist que certes reorganitzacions cromosòmiques no es troben en la natura ja que disrupcions en les regions afectades provocarien canvis d'expressió gènica possiblement letals per la progènie. Finalment, hem estudiat el paper de les reorganitzacions cromosòmiques en el procés d'especiació, on hem posat de manifest que regions genòmiques implicades en inversions són regions de baixa recombinació en relació a les regions no reorganitzades i per tant podrien donar lloc a un procés d'aïllament reproductiu per l'acúmul d'incompatibilitats genètiques en els híbrids. Per poder explicar les nostres observacions hem proposat un model on la presència d'heterocariotips flotants en el node d'especiació provocaria una supressió de recombinació en les regions invertides encara observable en les espècies actuals.

SUMMARY

The study of the genome organization as well as how chromosomal reorganizations are involved in speciation and adaptation processes are the key points to better understand the evolutionary dynamics of vertebrate genomes and their inter- and intra-specific phylogenetic relationships. In this thesis we described the ancestral karyotype for mammals, amniotes and tetrapods in order to determine the homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in the human genome. Using the chicken and frog genomes as outgroups, we were able to improved previously described ancestral karyotypes for tetrapods and amniotes and we defined new syntenic associations as an amniote or mammal synapomorphies. We also analysed the distribution of EBRs in the human genome, showing that EBRs are not randomly distributed and 20% are reused during the evolutionary period. The distribution of EBRs is related to the abundance of repetitive sequences, exhibiting an enrichment of specific tandem repeats in EBRs and co-localizing with mobile elements (AAAT-*Alu*). Furthermore, we studied the selective constrain on the maintenance of conserved regions. We observed that certain reorganizations are not found in natural populations because disruptions of the regions involved in reorganizations would lead to changes in gene expression probably lethal for the progeny. Finally, we studied the relation between chromosomal reorganizations and speciation. We showed that regions affected by reorganizations have lower recombination rates than regions not rearranged, thus, an increase of genic incompatibilities in these regions could lead to reproductive isolation by the existence of a barrier of gen flow. In order to explain our observations we proposed the floating heterokaryotypes model, where the presence of heterokaryotypes in the speciation node resulted on a suppression of recombination in the rearranged regions, which is still detected on the extant species.

BAC	Bacterial artificial chromosome
CO	Cross over
dHJ	Double Holliday Junction
DSB	Double Strand Break
EBR	Evolutionary breakpoint region
FISH	Fluorescent *in situ* hybridization
GRIMM	Genome rearrangements in man and mouse
H3K27ac	Acetilació de la lisina 27 de l'histona 3
H3K4me3	Trimetilació de la lisina 4 de l'histona 3
H3K9ac	Acetilació de la lisina 9 de l'histona 3
Hi-C	High-throughput chromosome capture conformation
HSB	Homologous synteny block
inferCAR	Contiguous ancestral region
Kpb	Kilo parells de bases
LCR	Low copy repeat
LINE	Long interspearsed repeat
LTR	Long terminal repeat
Ma	Milions d'anys
MGR	Multiple genome rearrangement
Mpb	Mega parells de bases
NAHR	Non-allelic homologous recombination
NGS	Next generation sequencing
PAK	Placental ancestral karyotype
pb	Parell de bases
RGC	Rare genomic change
SD	Segmental duplication
SINE	Short interspersed repeat
TFBM	Turnover fragile breakage model
VNTR	Variable number tandem repeat

NOMENCLATURA DE LES ESPÈCIES

CATALÀ	ANGLÈS	LLATÍ
armadillo	armadillo	*Dasypus novemcinctus*
ascidia solitaria	sea squirt	*Ciona intestinalis*
babuí	baboon	*Papio hamadryas*
cavall	horse	*Equus caballus*
cérvol muntjac	Indian muntjank	*Muntiacus muntjak*
diamant mandarí	zebrafinch	*Taeniopygia guttata*
elefant	elephant	*Loxodonta africana*
eriço lila	purple sea urchin	*Strongylocentrotus purpuratus*
esturió rus	russian sturgeon	*Acispenser guedelnstaedtii*
gall	chicken	*Gallus gallus*
gat	cat	*Felis catus*
gos	dog	*Canis familiaris*
granota	frog	*Xenopus tropicalis*
humà	human	*Homo sapiens*
macaco	macaque	*Macaca mulatta*
manatí	manatee	*Trichechus sp.*
medaka	medaka	*Oryzias latipes*
musaranya comuna	common shrew	*Sorex araneus*
opòssum/sariga	opossum	*Monodelphis domestica*
orangutan	orangutan	*Pongo pygmaeus*
ornitorinc	platypus	*Ornithorhynchus anatinus*
peix globus	pufferfish	*Tetraodon nigrivirides*
peix globus	fugu	*Takifugu rubripes*
peix zebra	zebrafish	*Danio rerio*
porc	pig	*Sus scroffa*
rata	rat	*Rattus norvegicus*
ratolí	mouse	*Mus musculus*
ratolí Sitka	Sitka deer mouse	*Peromuscus sitkensis*
tenrec	tenrec	*Echinops telfain*
vaca	cattle	*Bos taurus*
ximpanzé	chimpanzee	*Pan troglodytes*

ÍNDEX

1 INTRODUCCIÓ

1.1 INTRODUCCIÓ A LA GENÒMICA COMPARATIVA

Per a poder entendre la dinàmica evolutiva dels genomes és imprescindible conèixer com estan organitzats els cromosomes de les diferents espècies i determinar quins tipus de reorganitzacions cromosòmiques estan implicades en l'especiació i en esdeveniments macroevolutius que afecten als grans grups taxonòmics. Els vertebrats són el grup D'ANIMALS més ampli dins dels cordats, inclouen unes 58.000 espècies (Baillie i col, 2004) i presenten una gran varietat de cariotips, des de 2n=6 del cérvol muntjac (*Muntiacus muntjak*) fins a 2n=250 de l'esturió rus (*Acispenser gueldenstaedtii*) (Gregory, 2007). En els últims anys, gràcies a dades citogenètiques i a la creixent disponibilitat de la seqüència de genomes de moltes espècies de vertebrats, l'estudi de les similituds i diferències estructurals entre els genomes (altrament anomenada genòmica comparativa) ha avançat molt. Això ha permès perfilar les relacions filogenètiques de les espècies (filogenòmica), proposar cariotips ancestrals, identificar regions cromosòmiques conservades a diferents nivells taxonòmics (altrament anomenades blocs de sintènia) i estudiar les regions no conservades (o disrupcions de sintènia) per desxifrar els mecanismes que han donat lloc a la organització genòmica de les espècies actuals.

1.1.1 Els mamífers

Els mamífers són un grup d'espècies dins de la classe de vertebrats que ha estat àmpliament analitzat i del que l'estudi de les seves relacions filogenètiques ha creat una enorme controvèrsia en els últims anys (Hällstrom i col, 2007, 2010; Asher i Helgen, 2010; Springer i col, 2011; Meredith i col, 2011). Es caracteritzen pel seu pelatge, els tres ossos de l'orella mitjana (martell, enclusa i estrep) i per l'alimentació de les cries amb llet gràcies a les glàndules mamàries. Actualment se n'han descrit unes 5500 espècies que s'agrupen en dos subclasses: **Prototheria** i **Theria**, que van divergir fa 218 milions d'anys (Ma) aproximadament (Meredith i col, 2011).

Els **Prototheria** estan representats pels *Monotremata*, és a dir l'ornitorinc i els equídnids (Augee, 2006). Aquests animals es caracteritzen perquè ponen ous i tant el sistema reproductiu, de defecació com l'urinari desemboquen en un únic conducte anomenat cloaca. Pel que fa als **Theria**, es divideixen en dos infraclasses: els *Metatheria* i els *Eutheria*, que van divergir fa 190 Ma (Meredith i col, 2011). Dels primers només en sobreviuen els marsupials, caracteritzats perquè tenen un període de gestació a l'úter matern molt curt i completen la major part del creixement a l'interior del marsupi (Dickman i col, 2007). En canvi, els *Eutheria* es caracteritzen per tenir un llarg període de creixement uterí, on l'embrió s'alimenta a través d'una placenta. Estudis basats en la comparació de seqüències de DNA (Springer i col, 1997; Madsen i col, 2001; Murphy i col, 2001; Nikolaev i col, 2007; Waters i col, 2007) agrupen als mamífers placentaris en 4 grans superordres que van divergir fa aproximadament 102 Ma (Meredith i col, 2011): *Xenarthra* (armadillos), *Afrotheria* (elefant, manatí i damans, entre d'altres), *Laurasiatheria* (rat-penats, musaranyes, talps, ungulats i carnívors) i *Euarchontoglires* (primats, rosegadors i conills). *Xenarthra* i *Afrotheria* són clades de l'hemisferi sud i, en canvi, *Laurasiatheria* i *Euarchontoglires* de l'hemisferi nord. Aquests últims s'agrupen com a *Boreoeutheria* (seguint la nomenclatura proposada per Asher i Helgen, 2010), i van divergir fa 92 milions d'anys (Meredith i col, 2011). La relació entre els tres grans grups de placentaris (*Afrotheria*, *Xenarthra* i *Boreoeutheria*) encara és motiu de debat ja que trobar senyals filogenètiques informatives en nodes molt ancestrals (més de 100 Ma) és problemàtic, per això s'han proposat tres grans hipòtesis (Fig.1). Actualment, la hipòtesi més acceptada és la que relaciona *Afrotheria* i *Xenarthra* com els mamífers més basals (*Atlantogenata*) basant-se en la reconstrucció d'arbres filogenètics a partir de *indels* en regions codificants (Murphy i col, 2007), alineant 60 Mpb en 41 espècies de mamífers (Prasad i col, 2008), utilitzant elements mòbils

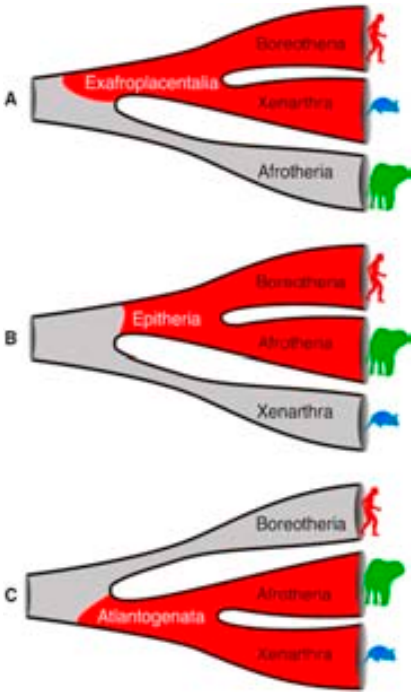
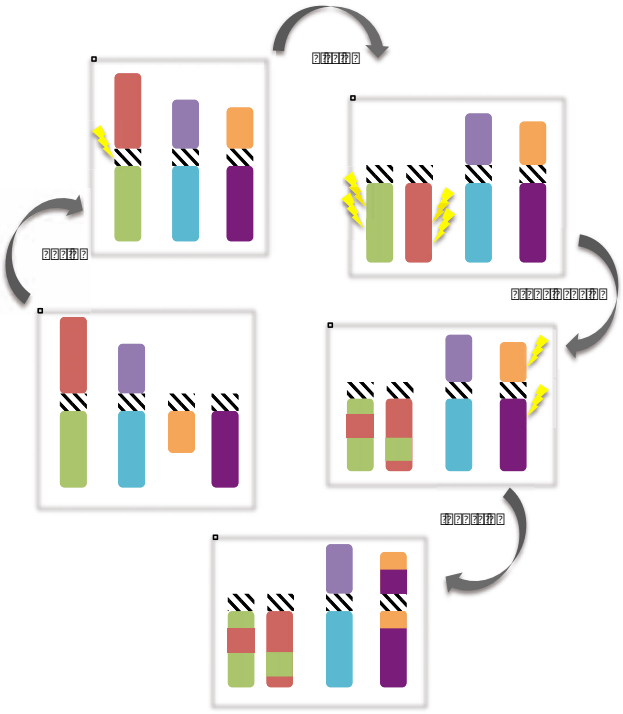


Figura 1. Diferents hipòtesis de l'origen dels mamífers placentaris. a) Exafroplacentalia. b) Ephiteria. c) Atlantogenata. Extret de Churakov i col, 2009.

[illegible] *uM* *Mi* *an* *id* *n* *ca* [illegible] úyy1 [illegible] úyy7 [illegible]

xFxFc [illegible] OURL1N [illegible] RTC N ÓR RIs 1D. [illegible] Ig CÓ 1 [illegible] I C U RU T1Ó [illegible] RTC [illegible] URN RCs N 1D. [illegible]

[illegible] úyyí [illegible]



1. U [illegible]

[illegible] *ci* *M* [illegible] *us* [illegible]

cromosoma per donar-ne dos; (iii) les translocacions, on es trenca un fragment d'un cromosoma i s'uneix a un cromosoma diferent i (iv) les inversions, on s'inverteix un fragment del cromosoma, que si involucra el centròmer serà una inversió pericèntrica i si no l'involucra serà paracèntrica (Fig. 2). En canvi, en les reorganitzacions cromosòmiques desequilibrades hi ha un canvi en el contingut de DNA, com ara les duplicacions o les delecions.

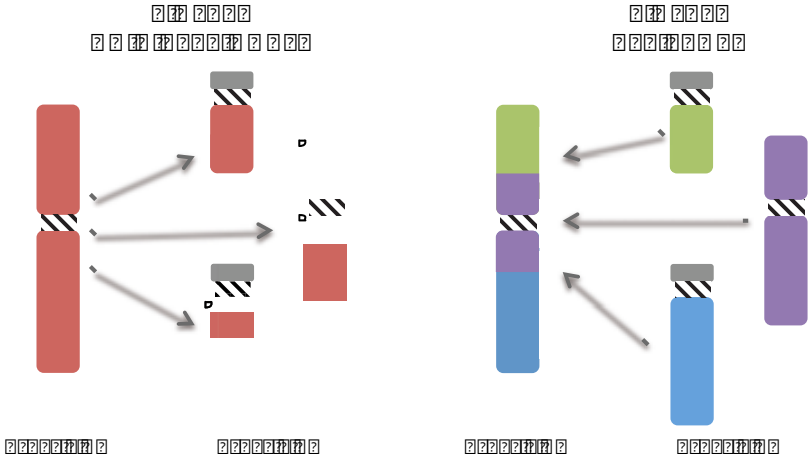
Per tal d'estudiar la contribució de les reorganitzacions cromosòmiques a la formació de noves espècies, en els últims anys s'han desenvolupat diferents metodologies per a poder determinar quina regió cromosòmica és homòloga entre els cariotips de diferents espècies, és a dir, establir els blocs sintènics.

1.1.2.1 Citogenètica comparativa

Tradicionalment, la citogenètica comparativa ha sigut l'estratègia utilitzada per a conèixer les homologies entre cariotips de diferents espècies i determinar els tipus de reorganitzacions cromosòmiques produïdes al llarg del procés evolutiu. A principis dels anys 70 es van començar a desenvolupar les tècniques de bandeig cromosòmic que van permetre identificar i diferenciar els cromosomes (bandes G, Q i R). Ràpidament van iniciar-se els estudis de comparació del patró de bandes R i G dels cariotips de diferents espècies, sobretot dins del grup dels primats (de Grouchy i col., 1972; Egozcue i col., 1973a i 1973b) que van permetre definir les homologies i reorganitzacions cromosòmiques necessàries per homologar els cariotips de les espècies estudiades.

A partir dels anys 90 es va començar a emprar la tècnica de la hibridació *in situ* fluorescent entre espècies (Zoo-FISH) (Wienber i col. 1990, Jauch i col. 1992), donant lloc a una nova disciplina, la citogenètica molecular comparativa. Aquesta tècnica es basa en la utilització de sondes de DNA corresponents a cromosomes sencers d'una espècie (sondes de pintat cromosòmic) que s'hibriden sobre preparacions cromosòmiques d'una altre espècie. En estudis de genòmica comparativa, aquesta tècnica s'aplica seguint dues estratègies diferents: i) el pintat cromosòmic unidireccional o ii) el pintat cromosòmic recíproc (Wienberg i Stanyon, 1998) (Fig. 3). En el pintat cromosòmic unidireccional, les sondes de l'espècie A s'hibriden sobre cromosomes de l'espècie B, per tant la informació que s'obté només es refereix a una

dr3□□TGRNNGq8□□e□N□elCSRe□R3aCern□Nor P r df P CN□RNaor N□BRdnt TGRd□Nr P a3RP Ren□□□P l□ 3□□UCl oCT□NGq□TR□dr eTRd□TR3□BRdaéNRR□□□dr l oR□Nor P r dr P Rd□TR3□BRdaéNRR□□□SzrlmReCern□□Cae CeOdoP□NGq□aRoa□Rdn□l3Go3Rd□oRDCedaNor P r df P Cpt Rd□Ur P f3r Dt Rd□Nr edRd□TRd□RemoR□Tt Rd□ RdaéNRd□□Cae□Nr P 3□□TGRNNGq□Rl r3t.mC□□TR3Rd□oRr oD□eCcc□NCed8□□


□□□ □□□□□
□□□□□□□

amb la freqüència amb la que s'observen junts en els gàmetes: una freqüència elevada indica que els loci estan lligats i per tant físicament propers. Aquesta tècnica es va començar a emprar comparant el genoma humà i el de ratolí (Lalley i col, 1978), però ràpidament es van estudiar altres espècies de mamífers, com el gat, el gos i la vaca (O'Brien i col, 1982, 1993, 1997a, 1997b), el cavall (Caetano i col, 1999) i altres espècies de primats, com el ximpanzé (Crouau-Roy i col, 1996) i el babuí (Rogers i col, 1995).

Pel que fa als mapes híbrids de radiació, s'obtenen irradiant amb raig X el genoma d'una espècie abans de fusionar-lo amb cèl·lules d'una altre espècie. La radiació trenca els cromosomes en fragments petits que es mantenen en les cèl·lules híbrides o es perden amb el temps. La distància entre dos loci es pot estimar a partir de la freqüència a la que aquests dos loci es troben en la mateixa cèl·lula. Una freqüència elevada indica que els loci són propers ja que la irradiació no ha provocat trencaments en la regió que els separa (Cox i col, 1990). Utilitzant aquesta estratègia s'han construït mapes híbrids de radiació, entre d'altres espècies, per a rata (Watanabe i col, 1999), gos (Mellersh i col, 2000), vaca (Band i col, 2000), porc (Hawken i col, 1999), gat (Murphy i col, 2000), macaco rhesus (Murphy i col, 2005a) i l'ocell diamant mandarí (zebrafinch) (Geilser i col, 1999) que s'han comparat amb el mapa híbrid d'humà. L'ús de mapes híbrids de radiació segueix sent una aproximació molt utilitzada per tal d'establir els blocs sintènics entre humà i les espècies de les quals no se n'ha seqüenciat el genoma. Un bon exemple és el treball publicat per Murphy i col·laboradors l'any 2005b. Mitjançant els mapes híbrids de radiació de gat, vaca, gos, porc i cavall i els genomes seqüenciats de rata i ratolí van establir els blocs sintènics conservats entre aquestes espècies i l'humà. Aquest treball va fundar les bases per a estudis posteriors de genòmica comparativa fent servir genomes complerts.

1.1.2.3 Estudis bioinformàtics o "in silico"

Des de l'any 2001 amb la publicació de la seqüència del genoma humà (Lander i col, 2001) hi ha hagut un creixent interès en la seqüenciació de genomes d'altres espècies de mamífers. Aquesta "revolució molecular" ha desembocat a la disposició de la seqüència genòmica de 39 espècies de mamífers (Ensembl release 66 – www.ensembl.org) en diferents graus de seqüenciació que permeten comparar els

genomes i anotar els gens. Gràcies al desenvolupament de noves metodologies, com ara els seqüenciadors Illumina, 454 i iTorrent, s'han abaratit els costos de seqüenciació, cosa que ha permès plantejar-se l'obtenció de genomes complerts d'altres espècies amb interès biomèdic i agrogenòmic. De fet, entre d'altres hi ha en marxa dos grans consorcis a nivell internacional que tenen com a objectiu la seqüenciació massiva de genomes de varies espècies i individus: el G10K project (G10K Community Scientists, 2009) i el 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010). El primer pretén seqüenciar el genoma de 10,000 espècies de vertebrats a una baixa resolució (2x *coverage*), comprenent unes 200 espècies de mamífers. Això ajudarà als estudis filogenòmics i aportarà claus per a elucidar els mecanismes evolutius. En canvi el segon té com a objectiu la seqüenciació del genoma de 2500 individus humans, de 29 poblacions diferents repartides per tots els continents, a una resolució de 4x, per tal d'investigar les diferències genòmiques en les poblacions humanes.

La disponibilitat d'aquesta quantitat de genomes s'ha acompanyat del desenvolupament de nous algoritmes matemàtics per a poder identificar regions conservades entre espècies mitjançant l'alineament dels genomes. D'entre els més utilitzats trobem el GRIMM (*Genome Rearrangements in Man and Mouse*) (Pevzner i Tesler, 2003a), el CHAINNET (Kent i col, 2003), el inferCARs (*Contiguous Ancestral Regions*) (Ma i col, 2006) i, més recentment, el SyntenyTracker (Donthu i col, 2009). Tots ells parteixen de les posicions de marcadors ortòlegs en les dues espècies que es volen comparar (veure CAIXA 1 per a la codificació de les dades); normalment els marcadors són regions codificants però en algun dels casos es poden utilitzar seqüències conservades no-codificants. Tot i que es fonamenten en tres grans passos comuns (ancoratge, filtratge i extensió), l'aproximació a cadascun dels passos varia en funció del mètode (Fig. 4). En qualsevol cas, els resultats no són sempre concordants; per exemple, Lemaitre i Sagot (2008a) van estudiar els blocs sintènics conservats entre humà i ratolí comparant el GRIMM i el CHAINNET i van observar que tant el número com la llargada dels blocs sintènics variava en funció del mètode emprat. La falta de convergència entre els mètodes actuals provoca que encara es segueixin desenvolupant mètodes més precisos i rigorosos per tal d'estudiar les regions conservades entre els genomes de diferents espècies.

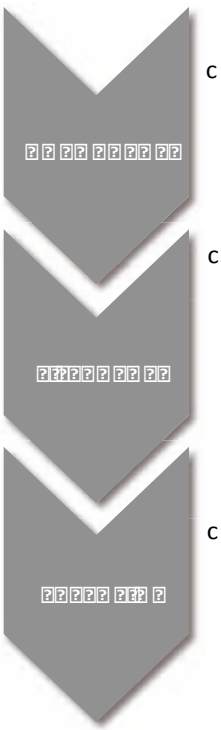


Fig. 1. Caption

Table 1
Table caption

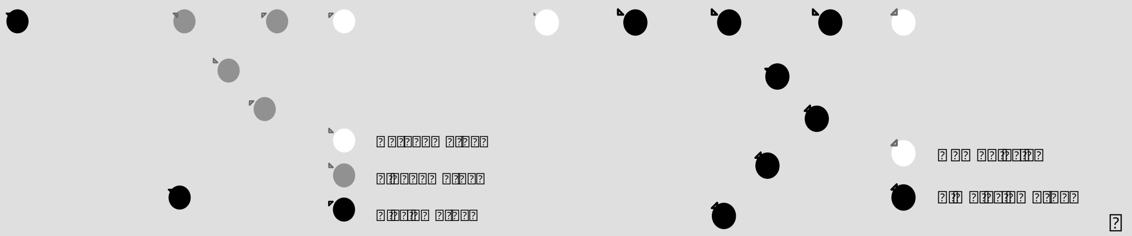
Header 1	Header 2	Header 3
Row 1	Data	Data
Row 2	Data	Data
Row 3	Data	Data
Row 4	Data	Data

1.2 CARIOTIPS ANCESTRALS

Els genomes de les espècies actuals poden comparar-se a un trencaclosques de blocs sintènics o regions ortòlogues, on l'ordre dels blocs en el genoma és el resultat d'una sèrie de reorganitzacions específiques de cada llinatge. Aquest ordre es pot fet servir per estudiar les relacions filogenètiques entre diferents grups taxonòmics (filogenòmica), ja que permet tenir una visió holística de l'evolució del genoma d'un conjunt d'espècies. L'establiment de sintenies conservades entre espècies permet predir la localització de gens en regions ortòlogues, amb aplicacions directes sobre l'estudi d'animals model per a malalties humanes. A més a més, la reconstrucció de cariotips ancestrals proporciona les bases per a l'estimació de taxes de canvi cromosòmic i la direccionalitat filogenètica d'aquests canvis. Per últim, la caracterització de marcadors genètics que no presentin alts graus d'homoplasia (veure CAIXA 2), com ara els canvis genòmics rars (*Rare Genomic Changes*, RGCs, Rokas i Holland, 2000) on s'hi inclouen les reorganitzacions cromosòmiques, els *indels* o els elements mòbils, ajuda a establir les relacions filogenètiques entre els diferents grups d'espècies, com ara els mamífers (Robinson i Ruiz-Herrera, 2008).

1.2.1 Establiment de cariotips ancestrals

Per tal de determinar el cariotip ancestral (número i estructura cromosòmica) d'un conjunt de taxons s'han de conèixer els blocs sintènics conservats entre els genomes d'aquestes espècies. Un cop s'han establert aquestes sintènies s'aplica el principi de parsimònia segons el qual si dues espècies comparteixen la mateixa forma cromosòmica, el més probable és que aquesta forma sigui ancestral i no que les dues hagin patit el mateix canvi. Per tant, es considera que la forma cromosòmica més estesa entre les espècies analitzades és, probablement, la forma ancestral. Quan es detecten dues formes cromosòmiques diferents en dos o més taxons cal la utilització d'una espècie fora del grup taxonòmic estudiat (*outgroup*) per tal de clarificar quina és la forma plesiomòrfica i quina l'apomòrfica (veure CAIXA 2). Un cop es coneixen les homologies cromosòmiques entre varies espècies i les reorganitzacions necessàries per a poder homologar els cariotips es pot reconstruir un cariotip ancestral per a les espècies estudiades i establir la filogènia de cada cromosoma.

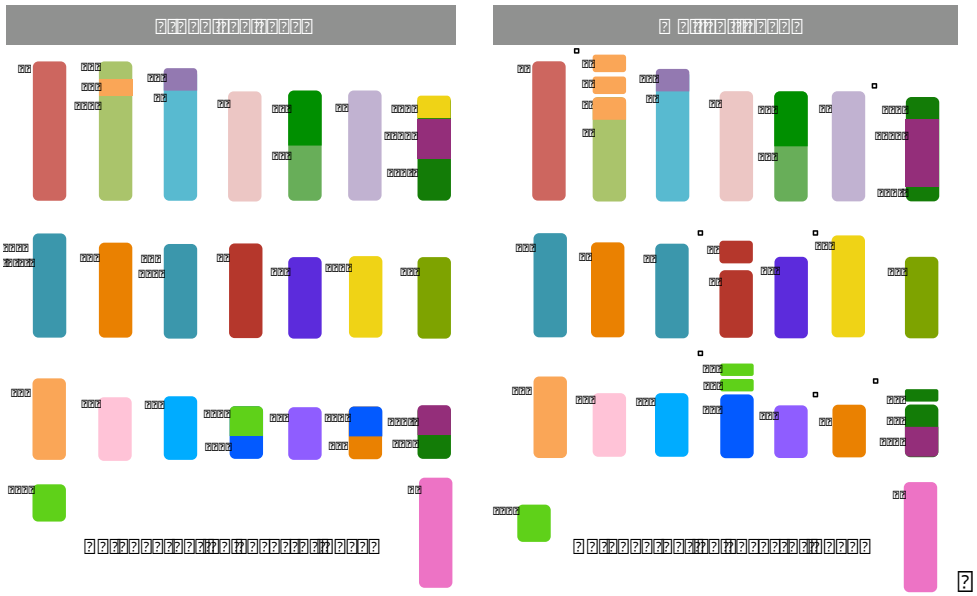


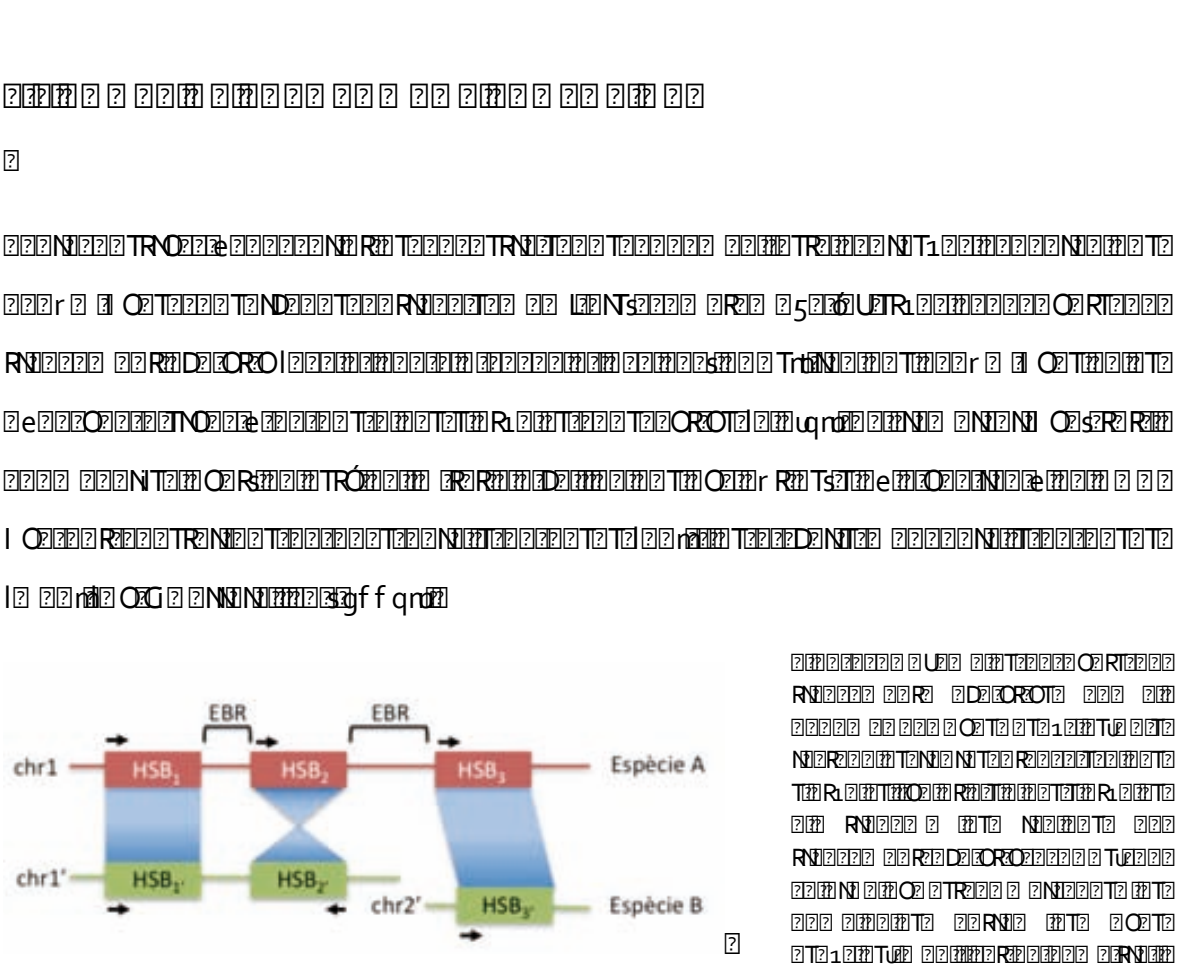
Graphodatsky i col, 2002; Nie i col, 2002), per a *Perissodactyla* (Trifonov i col, 2003), *Cetartiodactyla* (Bielec i col, 1998; Fronicke i col, 2001; Biltueva i col, 2004), *Chiroptera* (Volleth i col, 2002) i *Eulipotyphla* (Dixkens i col, 1998). Dins del superordre de *Euarchontoglires*, s'han estudiat els grups de *Rodentia* (Stanyon i col, 2003; Li i col, 2004), *Lagomorpha* (Korstanje i col, 1999; Robinson i col, 2002), *Scandentia* (Müller i col, 1999) i *Primates* (Wienberg, 2005; Ruiz-Herrera i col, 2005; Stanyon i col, 2008; Misceo i col, 2008). Això ha portat a proposar una hipòtesi robusta sobre quina seria la composició del cariotip ancestral dels *Boreoeutheria* (Chowdhary i col, 1998; Froenicke i col, 2003; Richard i col, 2003; Yang i col, 2003; Svartman i col, 2004, 2006; Ferguson-Smith i Trifonov, 2007; Graphodatsky i col, 2011), que està format per n=23 cromosomes amb 16 cromosomes sencers o fragments cromosòmics homòlegs als cromosomes humans: 1, 5, 6, 2q13-qter, 7b, 2p-q13, 9, 11, 10q, 13, 8q, 17, 18, 20, 19p i X i per associacions sintèniques corresponents als cromosomes humans: 4q/8p/4pq, 3/21, 14/15, 10p/12pq/22qter, 19q/16q, 16p/7a i 12q/22q (Fig. 5).

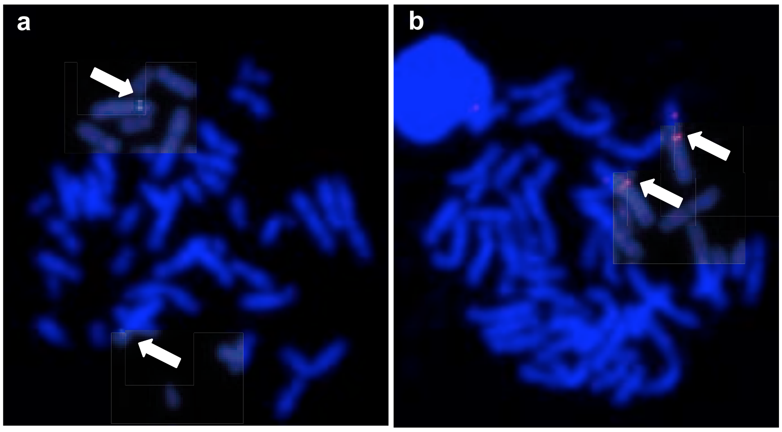
Paral·lelament als estudis de citogenètica molecular, el desenvolupament de nous algoritmes matemàtics basats en la comparació de genomes complerts ha permès avançar en l'estudi de l'evolució cromosòmica i la determinació de cariotips ancestrals (Froenicke i col, 2006). De fet, s'han proposat diferents aproximacions matemàtiques, que, partint de criteris de parsimònia i un cop obtinguts els blocs sintènics a nivell bioinformàtic (veure secció 1.1), reconstrueixen l'orientació d'aquests blocs en un possible cariotip ancestral. La primera aproximació *in silico* proposada per a estudiar genomes sencers es basava en la comparació de l'ordre de gens ortòlegs d'humà, de ratolí i de gat per on inferien un cariotip ancestral de les tres espècies (Bourque i Pevzner, 2002). Posteriorment, i fent servir el mateix algoritme (*Multiple Genome Rearrangement*, MGR) es va ampliar l'estudi analitzant dades provinents de mapes d'híbrids de radiació (RH) de vaca, gos, porc i cavall, proposant un cariotip de *Boreoeutheria* de n=24 (Murphy i col, 2005b). Malgrat que el número haploide obtingut per tècniques *in silico* estava dins del rang proposat per tècniques citogenètiques, els blocs sintènics conservats eren bastant diferents. Això va despertar un aferrissat debat entre els defensors de les dues aproximacions (Froenicke i col, 2006; Bourque i col, 2006) però que es va calmar l'any 2006 quan Ma i col·laboradors van desenvolupar un

nou algoritme (inferCARs) per reconstruir cariotips ancestrals. Partint dels genomes complerts d'humà, ratolí, rata i gos, Ma i col·laboradors van establir les regions ancestrals conservades i, ajudant-se amb dades citogenètiques prèvies, van proposar un cariotip ancestral de *Boreoeutheria* que reconstruïa pràcticament els mateixos blocs sintènics que l'aproximació citogenètica (Fig. 5).

La disposició de la seqüència completa de cada vegada més genomes i el fet que la metodologia de Zoo-FISH no es pugui aplicar entre eutheria i metatheria degut a la alta divergència de seqüències ha desembocat en la convergència de les dues metodologies, donant embranzida a l'estudi dels cariotips ancestrals. Per una banda, Robinson i Ruiz-Herrera (2008), tenint en compte dades citogenètiques prèvies i incloent el genoma de dos espècies *outgroups* (l'opòssum i el gall), van definir noves simplesiomorfies pels mamífers placentaris. Per altra banda, l'any 2006, Kohn i col·laboradors van crear la tècnica de *E-painting* (o pintat cromosòmic electrònic) on partint de les coordenades genòmiques de gens ortòlegs entre varies espècies, s'ordenen els gens en els cromosomes d'una de les espècies i així es veuen clarament les regions sintèniques en les altres. Un cop es coneixen les regions conservades es pot inferir el cariotip ancestral com si fossin dades obtingudes per Zoo-FISH. Aquesta metodologia ha permès proposar cariotips per a nodes més ancestrals, com ara els amniotes (Nakatani i col 2007) o els vertebrats (Kohn i col 2006). Tot i així no s'ha arribat a un consens pel que fa al cariotip ancestral dels vertebrats ni dels amniotes degut sobretot a que el mostreig ha sigut incomplert fins al moment i que les metodologies aplicades han estat molt diferents. És per això, que sota aquest escenari, és necessari la introducció d'*outgroups* adients per a poder definir amb un elevat grau de confiança quins caràcters es consideren sinapomorfies i quins simplesiomorfies, com també definir la composició de cariotips de nodes més ancestrals.




EBR
EBR
chr1
HSB1
HSB2
HSB3
Espècie A
chr1'
HSB1'
HSB2'
chr2'
HSB3'
Espècie B


a
b

Pel que fa a nivell bioinformàtic, els punts de trencament s'han estudiat com el subproducte d'algoritmes que es centraven en la detecció de blocs sintènics (GRIMM i inferCARs, entre d'altres) (veure apartat 1.1.3). Només recentment, Baudet i col·laboradors (2010) han desenvolupat un algoritme on l'objectiu principal és acotar els punts de trencament, el CASSIS; partint de la detecció de blocs sintènics intenten alinear fraccions de la seqüència entre els blocs sintènics i refinar al màxim els punts de trencament. Això permetrà l'estudi en profunditat d'aquestes regions i l'associació amb certes característiques del DNA sense que s'emmascarin pel soroll de fons de les regions flanquejants.

1.3.1 Models de distribució de les regions evolutives

Des dels inicis de la genòmica comparativa, l'estudi de les regions evolutives ha anat lligat a la formalització de models que tracten d'explicar la seva presència i distribució en els genomes. L'any 1984 Nadeau i Taylor van proposar el ***Random Breakage Model*** (model de trencaments aleatoris) d'evolució cromosòmica. Aquest model sosté que les reorganitzacions cromosòmiques estan distribuïdes de manera uniforme i independent al llarg del genoma. Fent servir els mapes de lligament d'humà i de ratolí i comparant la distribució de 83 gens ortòlegs entre aquestes dues espècies van demostrar que, tal i com havia postulat anteriorment Ohno l'any 1973, hi havia grans regions dels genomes que estaven conservades entre aquestes dues espècies. Nadeau i Taylor (1984) van definir 13 segments que contenien com a mínim dos gens ortòlegs consecutius i van demostrar que la distribució de la llargada d'aquestes regions conservades seguia una funció exponencial que descriu els resultats d'un procés aleatori (Fig. 8a). Aquests autors, per tant, van concloure que l'evolució cromosòmica patida entre humà i ratolí seguia el model de trencaments aleatoris. A més a més van estimar en 178 +/- 38 el número de reorganitzacions que van tenir lloc entre les dues espècies. Durant els anys següents, es van anar identificant més gens ortòlegs entre humà i ratolí i es van re-analitzar les dades, confirmant el model de trencaments aleatoris (DeBry i Seldin, 1996; Nadeau i Sankoff, 1998).

Però l'any 2003, després de la publicació de les seqüencies del genoma humà i del de ratolí juntament amb el desenvolupament de nous algoritmes per alinear els dos genomes, Pevzner i Tesler van proposar el ***Fragile Breakage Model*** (model de trencaments fràgils). Aquests autors van establir 281 regions conservades entre humà i ratolí de més d'una Mpb i 190 regions de menys d'una Mpb. La distribució de les llargades de les regions més grans seguien una funció exponencial però les regions més petites s'escapaven d'aquesta funció, invalidant per tant el model de trencaments aleatoris (Fig. 8b). A més a més, Pevzner i Tesler (2003a) van estimar que eren necessàries unes 245 reorganitzacions (transposicions, inversions, etc.) i 258 punts de trencament evolutiu per tal d'homologar els genomes d'humà i de ratolí. Aquest número de reorganitzacions els va portar a la conclusió que alguns punts de trencament s'havien utilitzat més d'una vegada per donar lloc a diferents reorganitzacions; com a mitjana cada punt de trencament s'havia fet servir 1,9 vegades. Per tant, el model de trencaments fràgils proposa que hi ha regions del genoma que són més propenses a patir trencaments, de manera que algunes d'aquestes regions han estat re-utilitzades al llarg de l'evolució.

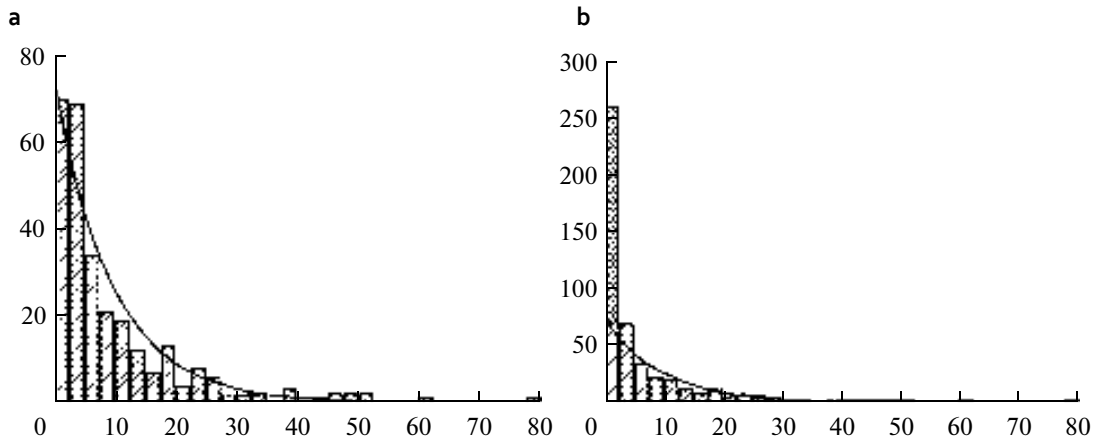


Figura 8. Distribució de les llargades dels blocs sintènics entre el genoma humà i el de ratolí. a) Distribució dels blocs sintènics > 1 Mpb i b) distribució de tots els blocs sintènics, incloent-hi els < 1 Mpb. La línia continua negra representa la curva de la funció exponencial. Es veu clarament com la distribució dels blocs sintènics petits escapen de la funció exponencial i per tant, del model de trencaments aleatoris. Modificat de Bourque i Tesler, 2008

Aquest nou model va desembocar en una nova polèmica entre els partidaris del model aleatori i els del model de fragilitat (Pevzner i Tesler, 2003b; Trinh i col, 2004; Sankoff i

Trinh, 2005; Peng i col, 2006; Sankoff, 2006; Alekseyev i Pevzner, 2007). El punt de discòrdia és, sobretot, la taxa de re-utilització dels punts de trencament evolutiu. Sankoff i col·laboradors sostenen que aquesta taxa de re-utilització no és una propietat biològica de l'evolució dels cromosomes, sinó que és un artefacte del mètode a l'hora d'estimar el nombre de blocs sintènics: el fet de no tenir en compte els blocs sintènics petits (< 1Mb) pot produir un increment de la taxa de re-utilització dels punts de trencament (Trinh i col, 2004; Sankoff i Trinh, 2005), que també es pot veure afectada pel fet que els punts de trencament són regions on no s'han pogut alinear els dos genomes (Sankoff, 2006). Cal remarcar que s'entén com a re-utilització dels punts de trencament evolutiu al fet que un mateix punt de trencament es localitzi en dues espècies diferents però no en l'ancestre comú (Murphy i col, 2005b).

Tot i així, el model de fragilitat té el suport d'altres arguments. S'ha demostrat la re-utilització dels punts de trencament evolutiu, no només a nivell teòric (Pevzner i Tesler, 2003), sinó a nivell empíric, fent servir tècniques citogenètiques (Froenicke, 2005; Ruiz-Herrera i col, 2005) i tècniques bioinformàtiques de comparació de múltiples espècies (Bourque i col, 2004; Murphy i col, 2005b; Ruiz-Herrera i col, 2006; Ma i col, 2006; Kemkemer i col, 2009; Larkin i col, 2009). A més a més, s'ha vist que alguns d'aquests punts de trencament també estan implicats en reorganitzacions descrites en alguns tipus de càncer (Murphy i col, 2005b; Ruiz-Herrera i Robinson, 2008) i en regions involucrades en síndromes humans (Antonell i col, 2005).

La demostració de l'existència de regions cromosòmiques que han sigut reutilitzades durant l'evolució dels mamífers ens porta a plantejar-nos les següents qüestions: aquestes regions són més fràgils degut a la seqüència de DNA i/o a l'organització del genoma o són regions on la selecció en contra dels trencaments és mínima?

1.3.2 Factors que determinen la distribució dels punts de trencament evolutius

Hem vist que una reorganització cromosòmica ve donada per una mala reparació d'un trencament de doble cadena del DNA (Fig. 2) i està delimitada pels punts de

trencament. Per tant, per tal de poder entendre la distribució genòmica dels punts de trencament evolutiu s'han de tenir en compte els factors que poden desembocar en la formació de DSBs i en la fixació d'aquestes reorganitzacions al llarg de l'evolució de les espècies.

1.3.2.1 Factors dependents de la seqüència del DNA: Les seqüències repetitives

Se sap que la major part del DNA d'una determinada espècie està format per seqüencies repetitives, per exemple en el genoma humà representa aproximadament un 50% de la seqüència (Lander i col, 2001). Estudis realitzats fins al moment apunten a una possible implicació dels elements repetitius en la formació de DSBs (Hedges i Batzer, 2005; Bacolla i col, 2008). La implicació d'aquest tipus de seqüències en desordres humans ha estat àmpliament estudiada en la literatura (Knight i col, 1993; Campuzano i col, 1996; Usdin i col, 2000), però la seva implicació com a detonants de canvis genòmics evolutius es poc coneguda.

1.3.2.1.1 Tipus de seqüències repetitives

Elements mòbils o transponibles

Els elements mòbils van ser descoberts inicialment en el blat de moro (McClintock, 1984) però s'ha vist que existeixen en la majoria d'organismes, des de procariotes fins a eucariotes (Capy, 1998). Constitueixen una gran fracció dels genomes d'eucariotes, per exemple representen aproximadament el 50% del genoma dels grans simis (humà: Lander i col, 2001; ximpanzé: Mikkelsen i col, 2005; orangutan: Locke i col, 2011). En funció del seu mecanisme de transposició, trobem dues grans classes d'elements mòbils (revisat a Richard i col, 2008):

Elements de Classe I: transposen a través d'un intermediari de RNA. El DNA és transcrit a RNA i després és convertit a DNA gràcies a una transcriptasa inversa, freqüentment codificada pel propi element mòbil. Aquesta còpia de DNA es re-insertarà en el genoma en una posició diferent a l'original. Aquest procés es coneix com a retrotranscripció i a aquest tipus d'elements com a retroposons. Es caracteritzen per l'absència d'introns, una cua rica en adenines i repeticions directes que flanquegen

l'element produïdes en el moment de la inserció. Dins d'aquesta classe podem subdividir els elements mòbils com a retroposons virals i no virals o com a autònoms o no autònoms, en funció del grau d'autosuficiència en els mecanismes de replicació (Fig. 9). Els retroposons virals es caracteritzen per la presència de repeticions d'entre 250 a 600 pb als extrems, anomenades *Long Terminal Repeats* (LTRs) i els elements autònoms d'aquest grup codifiquen per almenys 3 gens (*gag*, *pol* i *env*). En canvi, els retroposons no virals no presenten LTRs i s'agrupen majoritàriament en dos tipus: LINEs (*Long Interspersed Elements*), d'unes 6 kb de llargada i que codifiquen per una endonucleasa i una transcriptasa inversa (Jurka, 1997); SINEs (*Short Interspersed Elements*), d'unes 600 pb i no codifiquen per cap enzim implicat en la retrotransposició, per tant depenen d'altres elements per a poder mobilitzar-se. Dins d'aquest grup trobem els elements *Alu* i LINE1 que representen gairebé el 34% del genoma humà (Lander i col, 2001).

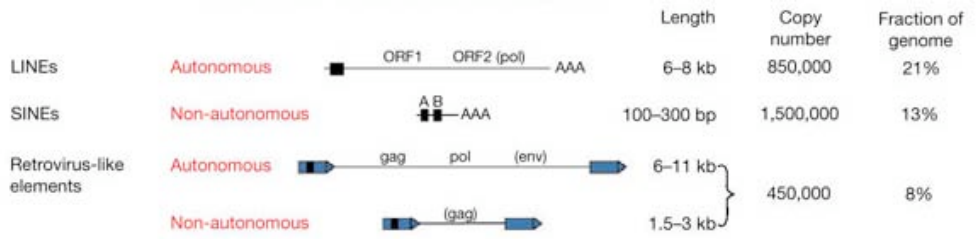


Figura 9. Classes de retroposons en el genoma humà. Extret de Lander i col. 2001

Elements de Classe II: són elements que codifiquen per una transposasa que els escindeix del genoma. Es mobilitzen emprant un intermediari de DNA de cadena única o doble. El punt d'escissió és reparat utilitzant la seqüència homòloga o la cromàtide germana com a motlle, per tant, es crea un duplicat de l'element mòbil en la regió d'inserció. La còpia escindida s'inserirà en un punt diferent del genoma. Aquests elements es coneixen també com a transposons.

Duplicacions segmentals (SD)

Les duplicacions segmentals, altrament conegudes com a *Low Copy Repeats* (LCRs), són fragments de DNA duplicats que mapen en dos o més llocs del genoma, tenen més d'una Kpb i un elevat grau d'homologia (Bailey i Eichler, 2006). Se n'han trobat en tots

els mamífers estudiats fins al moment: en ratolí representen un 6% del genoma, en rata un 2%, en vaca un 5,6%, en macaco un 1,55%, en orangutan un 1,18%, en ximpanzé un 4% i en humà un 5% (Marques-Bonet i col, 2009a). La distribució de les duplicacions segmentals en el genoma varia en funció de l'espècie, en la majoria de mamífers estan organitzades en tàndem, excepte en primats que predomina la distribució en llocs dispersos del genoma (Liu i col, 2009). A més a més s'ha observat que la distribució no és a l'atzar, sinó que hi ha una acumulació en regions pericentromèriques i subtelomèriques (Liu i col, 2009).

Repeticions en tàndem de número variable (VNTRs)

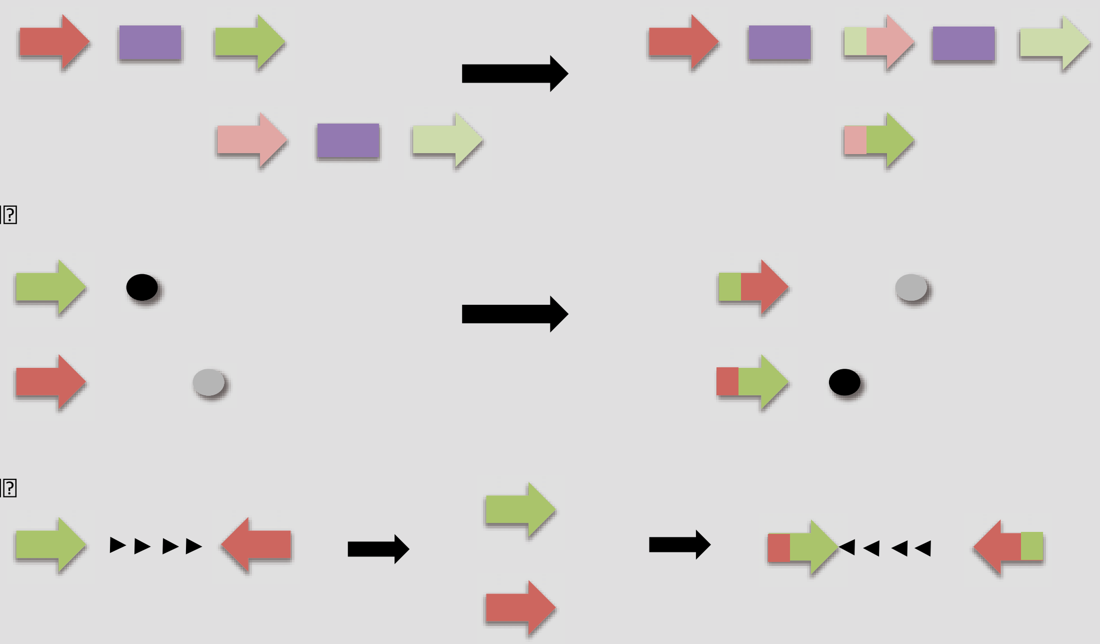
Les VNTRs són repeticions de seqüències curtes de DNA organitzades en clústers i que es troben disperses per tot el genoma. El nombre de repeticions d'aquestes seqüències pot ser variable entre individus, *loci* i al·lels del mateix *locus*. Tradicionalment el terme VNTR designava els *loci* minisatèl·lits (Nakamura i col, 1987) però ara ja s'hi inclouen també els microsatèl·lits.

Els microsatèl·lits són seqüències de DNA que consisteixen en la repetició en tàndem "n" vegades d'una unitat bàsica o motiu d'entre un i sis nucleòtids de longitud (per exemple, ATATAT) (Richard i col, 2008). Es poden classificar seguint dos criteris: segons el nombre de nucleòtids de la unitat bàsica de repetició o segons la composició nucleotídica de la seqüència repetitiva. Seguint el primer criteri trobem els microsatèl·lits **mononucleòtids**, **dinucleòtids**, **trinucleòtids**, **tetranucleòtids**, **pentanucleòtids** i **hexanucleòtids**. En canvi, seguint el segon criteri trobem microsatèl·lits **perfectes**, si estan formats per una única seqüència repetitiva sense cap interrupció en el patró de repeticions (per exemple, p.e. ACACACACACAC) o **imperfectes**, si tenen interrupcions per l'aparició d'un nucleòtid diferent en la seqüència (p.e. ACACACACATACACTCAG). També es parla de microsatèl·lits **compostos** quan es troben dos microsatèl·lits diferents consecutius (p.e. ATATATATATAGAGAGAG). La fracció de genoma que ocupen els microsatèl·lits varia en cada espècie, en l'espècie humana representen un 3% del genoma (Lander i col, 2001), en ximpanzé un 1,5% (Mikkelsen i col, 2005) i en ratolí ocupen un 2,7% del genoma (Waterson i col, 2002).

En canvi, els minisatèl·lits són repeticions en tàndem on la seva unitat de repetició té un rang d'entre 7 i 100 parells de bases (pb) (Näslund i col, 2005), i sovint té centenars de repeticions. Són molt utilitzats en l'anomenat *DNA fingerprinting*, que ha estat àmpliament utilitzat en genètica forense (Pena i Chakraborty, 1994).

1.3.2.1.2 Impacte de les seqüencies repetitives en l'evolució genòmica

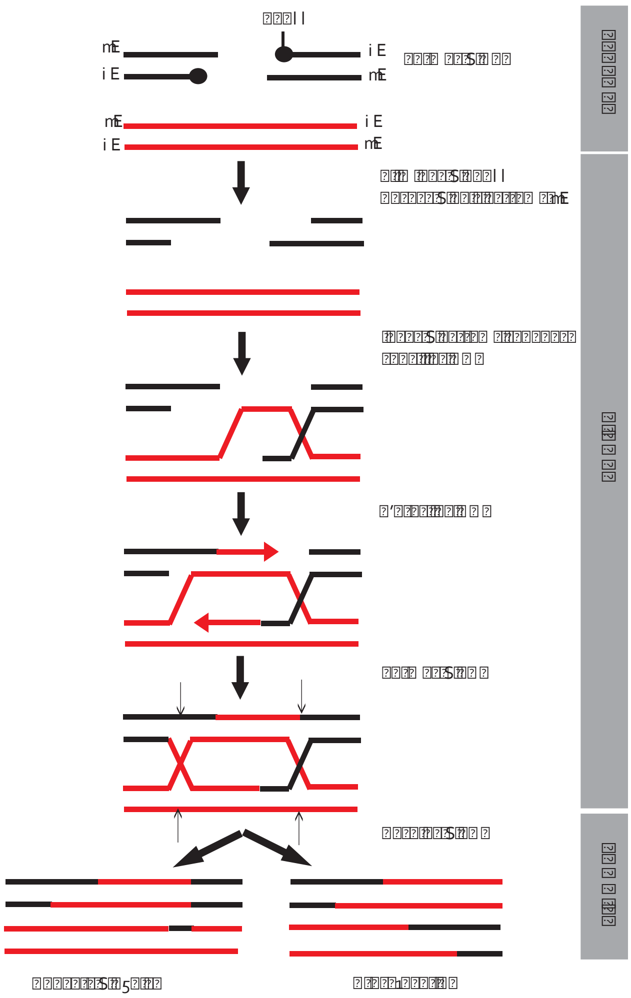
Tots tres tipus de seqüències repetitives (elements mòbils, duplicacions segmentals i repeticions en tàndem) estan relacionades entre elles i poden estar implicades en la formació de reorganitzacions cromosòmiques. S'han trobat duplicacions segmentals en sis dels 9 punts de trencament evolutiu de les inversions pericèntriques entre el genoma humà i el de ximpanzé (Kehrer-Sawatzki i col, 2008) com també entre el genoma humà i el de ratolí (Armengol i col, 2003). Tenint en compte que les duplicacions segmentals són regions amb una elevada homologia s'ha proposat que gràcies a la recombinació ectòpica o *Non-Allellic Homologous Recombination* (NAHR), les regions altament repetitives poden originar reorganitzacions cromosòmiques (veure CAIXA 3). El fet que moltes duplicacions segmentals tinguin elements mòbils en la seva regió terminal (Bailey i col, 2003) i que en els casos on no s'han trobat duplicacions segmentals s'hagi observat una associació dels punts de trencament evolutius d'inversions amb elements mòbils (Cáceres i col, 1999; Kerhrer-Sawatkzi i col, 2008; Lee i col, 2008) fa que també s'hagi proposat la recombinació ectòpica entre dos elements mòbils com a possible mecanisme per generar reorganitzacions cromosòmiques (Gray, 2000; Cordaux i Batzer, 2009). A més a més, s'ha descrit l'existència d'interrelacions entre els diferents tipus d'elements repetitius. Els elements mòbils, per exemple, poden ser una font de microsatèl·lits degut a la cua poli-A que presenten o a la regió intermitja rica en A/T (revisat a Cordaux i Batzer, 2009). Aquests microsatèl·lits poden ser causa d'inestabilitat genòmica ja que al formar estructures diferents de la conformació B del DNA (com ara *hairpins* o tetraplexs) poden induïr la formació de DSBs (Bacolla i col., 2008). En aquest sentit, Ruiz-Herrera i col·laboradors (2006) van demostrar que en els punts de trencament evolutiu del genoma humà al comparar-lo amb els genomes de ratolí, rata, vaca, gos, porc, gat, cavall i gall presentaven una acumulació del número de bases implicades en repeticions en tàndem i van proposar que aquestes repeticions



1.3.2.2 Factors independents de la seqüència del DNA: Recombinació meiòtica i selecció

Un altre procés que pot desembocar en la formació de reorganitzacions cromosòmiques degut als trencaments de doble cadena del DNA és la recombinació meiòtica. Aquest tipus de recombinació està altament regulada en els mamífers (Cole i col, 2010; 2012) i comença gràcies a l'acció de l'endonucleasa Spo11 (Keeney i col, 1997) durant la fase de leptotè en la profase I meiòtica (veure CAIXA 4) . L'aparició de forma programada d'una elevada quantitat de DSBs en aquest estadi posa en marxa el procés de reparació del DNA mitjançant la recombinació homòloga. Només una petita part d'aquests DSBs (10% en ratolí) acabarà donant lloc a punts de recombinació on es produeix intercanvi de material genètic entre cromosomes homòlegs i a la formació de quiames (*crossing overs*, COs). Estudis recents en llevats i humans (Petes, 2001; Meyers i col, 2005) han demostrat que els CO no es disctribueixen a l'atzar en el genoma, sinó que estan concentrats en determinades regions, anomenades *hotspots* de recombinació.

Les reorganitzacions cromosòmiques sorgeixen degut a l'aparició de DSBs en un cromosoma seguides per un procés de reparació il·legítim, on el fragment cromosòmic afectat s'uneix de manera incorrecte a altres regions cromosòmiques (Krimbas i Powell, 1992). En un marc evolutiu, si aquestes reorganitzacions cromosòmiques només es donen en pocs individus d'una població, al creuar-se amb individus no portadors dels canvis cromosòmics originaran descendència híbrida per a aquestes reorganitzacions. Aquesta nova generació pot tenir problemes de segregació en el procés de meiosis, donant lloc a la formació de gàmetes desequilibrats (Stebbins, 1958; King, 1993). Aquest procés pot desembocar en la formació de noves espècies ja que pot crear una barrera reproductiva i aïllament (White, 1969). Aquest model d'especiació cromosòmica és l'anomenat **model tradicional o de baixa fertilitat dels híbrids** (Fig. 12). En aquest model es classifiquen les reorganitzacions cromosòmiques en tres tipus, en funció de l'efecte que tindran en la població (King, 1993): i) les que són deletèries i s'eliminaran per selecció natural, ii) aquelles que podran originar polimorfismes en la població i iii) aquelles que redueixen l'eficàcia reproductiva dels heterozigots, provocant subdominància.



Partint d'una població petita i amb elevada consanguinitat, el model tradicional d'especiació cromosòmica prediu que els canvis cromosòmics estructurals (inversions, translocacions, duplicacions i delecions) desembocaran en una mala segregació meiòtica en els individus heterozigots, donant gàmetes desequilibrats i provocant una baixada de la fertilitat o fins i tot esterilitat (Fig. 12). Aquest model també es coneix com a model d'especiació estasipàtrica (White 1968). Hi ha varis exemples d'especiació que es poden explicar gràcies a aquest model: un exemple clàssic és el de

Figura 12. Models d'especiació cromosòmica. Partint d'una població on uns individus han patit una inversió cromosòmica es dóna un procés d'hibridació entre els individus els portadors de la inversió (a) i els ancestrals (b), resultant en individus híbrids portadors d'un heterocariotip. La teoria tradicional d'especiació postula que l'híbrid experimentarà sotadominància (c); en canvi, el model de supressió de recombinació sosté que els cromosomes reorganitzats patiran supressió de recombinació en la zona afectada per la reorganització, resultant en un increment de la divergència nucleotídica (e) o en la fixació d'al·lels favorables per a l'adaptació a un nou ambient (f). Extret de Brown i O'Neill. 2010.

Vandiemenella, un grup de llagostes australianes (White, 1968; King,1993), també s'ha estudiat en lemurs (Dutrillaux i Rumpler, 1977) i en èquids (Allen i Short, 1997). Però sobretot s'ha estudiat en rosegadors. Aquest grup es caracteritza per presentar cariotips molt reorganitzats, sobretot degut a translocacions Robertsonianes (Rb; fusions de dos cromosomes acrocèntrics per formar-ne un de metacèntric). S'ha vist que les translocacions Rb poden jugar un paper important en el procés d'especiació,

actuant com a mecanismes d'aïllament postzigòtic entre poblacions diferenciades cromosòmicament, ja sigui perquè impedeixen la formació d'híbrids o perquè produeixen híbrids amb baixa viabilitat o amb esterilitat parcial o completa (White i col, 1968; King, 1993). Però aquest model suposa una gran paradoxa. Si una reorganització és realment molt subdominant, serà ràpidament eliminada de la població perquè, en origen, es donarà en heterozigosis. I si una reorganització és poc subdominant, es mantindrà en la població però tindrà un efecte lleu en la fertilitat dels híbrids i no facilitarà l'aïllament i l'especiació. Per tant, només aquelles reorganitzacions que siguin subdominants podran contribuir a l'especiació però és poc probable que es fixin en les poblacions.

Per a poder solucionar la paradoxa generada pel model tradicional d'especiació cromosòmica s'ha descrit el **model de supressió de recombinació** (Fig. 12), que proposa que les reorganitzacions cromosòmiques no necessàriament redueixen l'eficàcia biològica sinó que redueixen el flux gènic degut a la supressió de recombinació en la zona afectada per la reorganització i que, per tant, podrien originar un aïllament reproductiu parcial (Noor, 2001; Rieseberg, 2001; Navarro i Barton, 2003). El fet que no hi hagi recombinació provocarà que les mutacions que es donin en les regions reorganitzades no passin d'una població a l'altre; en canvi, hi haurà flux gènic entre els cromosomes no reorganitzats. Això farà que en els cromosomes reorganitzats hi hagi més divergència gènica que en els no reorganitzats. Per tant, si l'especiació cromosòmica es dóna per una supressió de la recombinació, s'hauran d'acomplir tres condicions: i) les reorganitzacions cromosòmiques han de suprimir la recombinació, ii) la supressió del flux gènic en les regions reorganitzades ha de ser un factor important en l'aïllament reproductiu i iii) s'han de donar reorganitzacions cromosòmiques diferents en llinatges propers (Faria i Navarro, 2010). Entre d'altres exemples que recolzen aquest model, trobem els estudis de Rieseberg i col·laboradors (1999, 2001) dels gira-sols on van veure que hi ha més introgressió (flux gènic entre espècies degut a un procés d'hibridació i retrocreuament) en cromosomes no reorganitzats que en reorganitzats. Estudis en *Drosophila* també han detectat més diferenciació gènica en regions properes als punts de trencament de les inversions (Machado i col, 2002; Noor i col, 2007; Stevison i col, 2011). També s'ha vist que en *Anopheles* hi ha una taxa de divergència

més elevada en la regió invertida del cromosoma X (Besansky i col, 2003). Pel que fa a mamífers, la situació és més complexa i els resultats obtinguts fins al moment heterogenis. A nivell experimental s'ha detectat una supressió de la formació de quiasmes (veure CAIXA 4) en inversions pericèntriques de poblacions híbrides de ratolins Sitka (*Peromuscus sitkensis*) (Hale 1986) i una reducció en el nombre de *foci* de MLH1 (proteïna meiòtica marcadora de punts de recombinació) en cromosomes translocats de musaranya comuna (*Sorex araneus*) (Borodin i col, 2008). Per altra banda, emprant mètodes indirectes, com ara la taxa de divergència gènica o proteica entre regions reorganitzades i no reorganitzades, s'ha detectat una baixada de flux gènic en les regions reorganitzades en ratolí domèstic (Marques-Bonet i col, 2005; Franchini i col., 2010) i musaranyes (Yannic i col, 2009). Però, pel que respecte l'estudi de l'especiació entre humans i ximpanzés els resultats obtinguts fins al moment han sigut heterogenis (Navarro i Barton, 2003; Zhang i col, 2004; Mikkelsen i col, 2005; Marques-Bonet i col, 2007; Szamalek i col, 2007). Per tant, calen estudis amb més profunditat per tal de determinar si el model de supressió de recombinació és l'adequat per a explicar l'especiació en mamífers.

2 OBJECTIUS

2. OBJECTIUS

Aquesta tesi doctoral s'emmarca dins de l'activitat de recerca del grup d'Evolució Genòmica Animal del Departament de Biologia Cel·lular, Fisiologia i Immunologia de la UAB, que es focalitza en l'estudi de l'evolució cromosòmica en mamífers. A partir dels coneixements adquirits en el nostre grup de recerca ens hem plantejat aprofundir en l'estudi dels punts de trencament evolutiu (EBRs) emprant eines bioinformàtiques. Per tant, l'objectiu principal d'aquesta tesi doctoral és estudiar les EBRs (característiques, distribució i mecanismes de formació) en els genomes dels mamífers per tal d'avançar en l'estudi de l'evolució d'aquestes espècies. D'aquest objectiu principal se'n deriven els següents objectius concrets:

1. Definir els blocs sintènics conservats en els mamífers placentaris, així com els canvis que han patit al llarg de l'evolució. Per tant, caldrà establir un cariotip ancestral per al grup de mamífers utilitzant *outgroups* adequats com són *Xenopus tropicalis* (com a representant dels amfibis) i *Gallus gallus* (com a representant de les aus).
2. Analitzar la distribució de les EBRs en els genomes de mamífers, tenint en compte l'efecte de diversos factors, com són:
    a. Les repeticions en tàndem i els elements transponibles. Ens proposem estudiar la relació entre aquests elements a l'hora de determinar la distribució dels EBRs. Ens centrarem en l'estudi dels genomes d'humà i dels altres grans primats (ximpanzé i orangutan).
    b. Pressió o constrenyiment selectiu exercit sobre la fixació de les reorganitzacions cromosòmiques originades durant l'evolució. Com a model per a determinar l'efecte de la pressió selectiva estudiarem el cas del ratolí domèstic *Mus musculus domesticus.*
    c. Recombinació meiòtica. Estudiar l'efecte de la recombinació meiòtica sobre l'especiació cromosòmica. En aquest cas, el model d'estudi serà l'espècie humana i el ximpanzé.

3 RESULTATS

TREBALL 1

Molecular cytogenetic and genomic insights into chromosomal evolution

Ruiz-Herrera A, Farré M and Robinson TJ

Heredity (2012) 108, 28-36

Índex d'impacte (2010): 4,569

ÀREA: Evolutionary biology, Genetics & Heredity, QUARTIL 1

Heredity (2012) 108, 28–36

www.nature.com/hdy

Heredity

REVIEW

Molecular cytogenetic and genomic insights into chromosomal evolution

A Ruiz-Herrera[1,2], M Farré[1] and TJ Robinson[3]

This review summarizes aspects of the extensive literature on the patterns and processes underpinning chromosomal evolution in vertebrates and especially placental mammals. It highlights the growing synergy between molecular cytogenetics and comparative genomics, particularly with respect to fully or partially sequenced genomes, and provides novel insights into changes in chromosome number and structure across deep division of the vertebrate tree of life. The examination of basal numbers in the deeper branches of the vertebrate tree suggest a haploid (n) chromosome number of 10–13 in an ancestral vertebrate, with modest increases in tetrapods and amniotes most probably by chromosomal fissioning. Information drawn largely from cross-species chromosome painting in the data-dense Placentalia permits the confident reconstruction of an ancestral karyotype comprising $n=23$ chromosomes that is similarly retained in Boreoeutheria. Using *in silico* genome-wide scans that include the newly released frog genome we show that of the nine ancient syntenies detected in conserved karyotypes of extant placentals (thought likely to reflect the structure of ancestral chromosomes), the human syntenic segmental associations 3p/21, 4pq/8p, 7a/16p, 14/15, 12qt/22q and 12pq/22qt predate the divergence of tetrapods. These findings underscore the enhanced quality of ancestral reconstructions based on the integrative molecular cytogenetic and comparative genomic approaches that collectively highlight a pattern of conserved syntenic associations that extends back ~360 million years ago.
Heredity (2012) **108,** 28–36; doi:10.1038/hdy.2011.102; published online 23 November 2011

Keywords: ancestral karyotypes; comparative cytogenetics; conserved syntenies; syntenic segmental associations; FISH; phylogenomics

INTRODUCTION

How genomes are organized and which types of chromosomal rearrangements are implicated in speciation and macroevolutionary events are fundamental to understanding the dynamics of chromosomal evolution. Molecular cytogenetic data and the increasing availability of partially or fully sequenced genomes from a variety of vertebrate species have fueled advances in phylogenomics (phylogenetic reconstructions using genomic data). This has lead to hypothesized ancestral chromosome numbers, karyotypes and the identification of conserved chromosomal syntenies and segmental associations at different taxonomic levels.

Chromosome number variation has traditionally been considered a proxy for the structural modification of karyotypes, especially so in groups of organisms where detailed information such as the differential staining of chromosomes, the extent and location of heterochromatin, and the location and number of nucleolar organizers is lacking. A considerable body of early work on chromosome number variation was reviewed by White (1973), who expressed reservations on whether it would be possible to determine 'modal numbers' for groups of organisms the higher one progresses in the systematic hierarchy. More specifically, he was of the view that 'to speak of a type number for the Insecta, the Vertebrata or even the Mammalia would be absurd'. However, recently, various computational approaches have been used to estimate the extent of rearrangement events and to derive the putative genomic architecture of ancestral genomes by inferring evolutionary histories from entire genomes. This has led to suggestions of ancestral syntenies and chromosomal complements—each progressively more distant in divergence—for amniotes (~310 million years ago, mya), tetrapods (~360 mya) and even vertebrates (~450 mya) (Postlethwait *et al.*, 2000; Naruse *et al.*, 2004; Woods *et al.*, 2005; Kohn *et al.*, 2006; Nakatani *et al.*, 2007; Voss *et al.*, 2011).

Among vertebrates, phylogenomic investigations have focused principally on mammalian genome evolution, in large part reflecting the availability of chromosomal and genomic information for this clade. Extant mammals (represented by monotremes, marsupials and placental or eutherian mammals) last shared common ancestry nearly 162 mya (Hallström and Janke, 2010). Modern eutherian taxonomic schemes recognize four superordinal clades (Afrotheria, Xenarthra, Laurasiatheria and Euarchontoglires) largely on the basis of phylogenetic analysis of both nuclear and mitochondrial DNA (Hallström and Janke, 2010 and references therein) and insertion sites of retroelements (Nishihara *et al.*, 2005; Kriegs *et al.*, 2006; Waters *et al.*, 2007; Churakov *et al.*, 2009). Although it would appear that the terms 'Eutherian' and 'Boreoeutherian' have been used synonymously in comparative cytogenetic and phylogenomic studies, they do in fact represent different nodes. Eutheria refers to everything on the so-called 'eutherian' side of the 'metatherian'-'eutherian' dichotomy

[1]Departament de Biologia Cellular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Barcelona, Spain; [2]Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain and [3]Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Matieland, South Africa
Correspondence: Professor TJ Robinson, Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Stellenbosch, Matieland 7602, South Africa.
E-mail: tjr@sun.ac.za
Received 31 May 2011; revised 8 August 2011; accepted 12 August 2011; published online 23 November 2011

(that is, Afrotheria, Xenarthra, and Boreoeutheria, and all fossil relatives that are more closely related to this clade than to Marsupialia). Boreoeutheria on the other hand comprises Laurasiatheria (Waddell *et al*., 1999) and Euarchontoglires (Murphy *et al*., 2001a). It is also more accurate to refer to the 'eutherian' ancestral karyotype as that of Placentalia as the data we have to infer this from are solely from placentals (all extant members of the last common ancestor to Atlantogenata (Afrotheria+Xenarthra) and Boreoeutheria; Asher and Helgen, (2010))—a usage that we follow in this review.

Here we examine how molecular cytogenetics and the*in silico* analysis of genomic sequences have contributed to our understanding of mammalian chromosomal evolution and the identification of conserved genomic regions. Furthermore, we review and extend previous observations by providing new data on the presence of conserved syntenic segmental associations that track back to the origin of tetrapods.

CHROMOSOME NUMBER VARIATION IN VERTEBRATES

Chromosome number and the number of chromosomal arms are good summary statistics of karyotypic change and hence chromosomal evolution in groups of organisms. Although data on chromosome arm number variation (the nombre fondamental of Matthey (1945), usually abbreviated to NF) are sparse, information on chromosome numbers across high-level taxonomic groups (mammals, birds, reptiles and amphibians) is extensive (Figure 1). Early lists of animal haploid (n) or diploid (2n) numbers (reviewed by White, 1973) included those on insects, crustaceans, fishes and, with respect to mammals, those of Hayman and Martin (1969) for marsupials and Matthey (1958) for placentals. Although these early attempts often suffered from poor taxon representation, they nonetheless permitted several general conclusions one of which was that the haploid number of most animal species lies between 6 and 24.

Since these early investigations, the biggest advances in determining chromosome numbers in deep branches of the vertebrate tree of life have, not surprisingly, resulted from the *in silico* scans of sequenced genomes. Genomic comparisons between human and teleost fish species (medaka, zebrafish and tetraodon) permitted hypothesized ancestral vertebrate genome configurations with n=10–13 (Postlethwait *et al*., 2000; Naruse *et al*., 2004; Woods *et al*., 2005). Detailed analyses of the likely amniote and tetrapod compositions followed (Khon*et al*., 2006; Nakatani *et al*., 2007). These studies, based on data from human, chicken, zebrafish, medaka and pufferfish genomes (Khon*et al*., 2006), were subsequently expanded (Nakatani *et al*., 2007) to include the

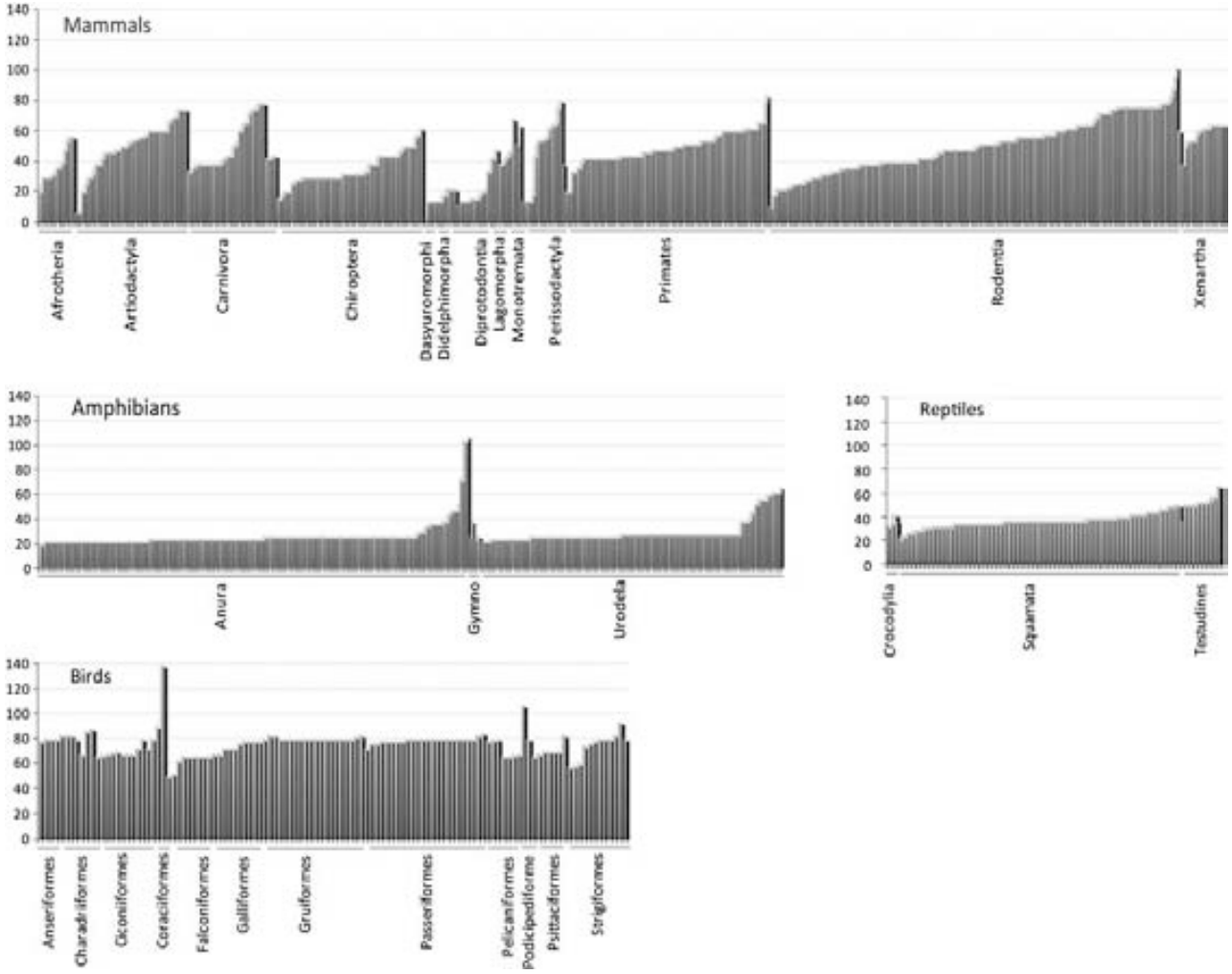


Figure 1 Chromosomal number variation among vertebrates. The *x* axis indicates the diploid chromosomal number, whereas the *y* axis groups species in different orders. The data for each taxonomic group are based on 515 species of mammals, 117 species of birds, 170 reptiles and 328 amphibians. Chromosomal data extracted from O'Brien *et al*. (2006) and Gregory (2011). A full color version of this figure is available at the *Heredity* journal online.

mouse, dog, tunicate (*Ciona intestinalis*) and sea urchin (*Strongylocentrotus purpuratus*). Khon *et al.* (2006) relied on 'E-painting'—the *in silico* identification of orthologous gene pairs to identify conserved genomic regions—whereas Nakatani *et al.* (2007) developed their own computational methodology to detect 'Ohnologs' (paralogs produced by two rounds of whole genome duplication) and thus conserved vertebrate linkage blocks. These studies, respectively, posit $n=18$ for the tetrapod ancestor (using the teleost pufferfish as outgroup), and $n=26$ for the amniote ancestor (using the pufferfish and medaka as outgroups).

The basal numbers retrieved by the various studies outlined above collectively permit inferences on the broader patterns of chromosome number changes across these groups. First, the low chromosome number suggested for the tetrapod ancestor increased to 26 in the amniote ancestor, most probably by multiple fissions. Previous attempts to reconstruct the ancestral tetrapod genome configuration (Kohn *et al.*, 2006; Voss *et al.*, 2011) have resulted in contradictory outcomes. Kohn *et al.*, 2006 proposed an ancestral tetrapod karyotype with $n=18$. On the other hand Voss *et al.* (2011), who studied the *Xenopus* ($n=10$) and *Ambystoma* ($n=14$) as part of an investigation into ancestral tetrapod chromosomes, proposed a high but unspecified chromosome number mirroring those usually found in birds. Their hypothesis is based on the observation that phylogenetically derived lineages (such as *Xenopus* and *Ambystoma*) have fewer chromosomes, indicating a tendency to have reduced chromosome numbers in these lineages. In contrast, birds (represented here by chicken) and mammals (by platypus, opossum and 11 placental species) are characterized by markedly different modes of chromosome number evolution.

Aves

The predominant mode of genome reorganization in Aves is chromosomal fission. Avian karyotypes are composed of microchromosomes and macrochromosomes but contrary to non-avian reptiles, birds are characterized by high chromosomal numbers that range from $n=20$ (or 21; see Nie *et al.*, 2009) to $n=69$ (De Smet, 1981; Figure 1). Descriptions of the ancestral avian karyotype are conventionally based only on macrochromosomes (Griffin *et al.*, 2007; Nanda *et al.*, 2011) and suggest that many of these have remained conserved within the group without disruption by interchromosomal rearrangements (reviewed in Ellegren, 2010). In fact, Griffin *et al.* (2007) have argued that the ancestral avian karyotype was similar to that of chicken, with macrochromosomes 1, 2, 3, 4q, 5, 6, 7, 8, 9, 4p and Z representing the ancestral state for chromosomes 1–10+Z; chromosome 4 was regarded as the most ancient linkage group within this karyotype.

Mammalia

A different situation holds for Mammalia where significant variation in chromosomal number is observed among Monotremata, Marsupialia and the eutherian placental mammals (Placentalia; Figure 1).

The three extant species belonging to Monotremata all have high diploid chromosome numbers with platypus characterized by $n=26$, and both the short-beaked and long-beaked echidnas having $n=32$ (O'Brien *et al.*, 2006). Although only one of these species was included in our analysis (the platypus, whose genome has been sequenced and is partially assembled), it is nonetheless clear that, as with Aves, fission events predominate in the karyotypic evolution of Monotrema.

Comprehensive cytogenetic studies on marsupials show that chromosomal numbers within the group range from $n=5$ to $n=16$ (Hayman, 1990). Whereas the majority of the families have conserved karyotypes (mainly $n=7$), the Macropodidae (kangaroos, wallabies and rat-kangaroos) shows evidence of more extreme chromosome reshuffling including fusion/fissions, inversions and centromere repositioning (O'Neill *et al.*, 2004 and references therein). Among marsupials, the South American opossum (*Monodelphis domestica*) is the only marsupial for which pair-wise alignments with the human genome are possible. Recently Westerman *et al.* (2010), using a combination of cytogenetics and sequence-based phylogenetics, have argued that the karyotype of the opossum ($n=9$) is highly conserved in relation to those of Australian marsupials confirming previous hypotheses (Rens *et al.*, 2001). *Monodelphis domestica* groups within the basal Didelphimorphia (Nilsson *et al.*, 2010; Westerman *et al.*, 2010) and is thought to have undergone two fissions from the hypothesized marsupial ancestral karyotype of $n=7$ (Rens *et al.*, 2001). If the marsupial ancestral estimate is correct (our small sample size precludes an estimate for Marsupialia given that only one fully sequenced genome is available), a dramatic decrease in chromosome number appears to have occurred in the marsupial lineage (presumably by serial fusion events) since its divergence from the mammalian common ancestor (with $n=23$—see mammalian ancestral configuration discussed below) $\sim$138 mya (Hallström and Janke, 2010).

The extremes in mammalian chromosome number occur in the species-rich Placentalia where these range from $n=3$ in the female Indian muntjac to a high of $n=51$ in the Red viscacha rat (O'Brien *et al.*, 2006). There is also substantial variation among Orders (Figure 1) reflecting the complex dynamics of mammalian chromosomal evolution. Recent studies based on cross-species chromosome painting analyses have estimated an ancestral haploid chromosome number that ranges from 22 to 25 for Placentalia (Chowdhary *et al.*, 1998; Froenicke *et al.*, 2003; Richard *et al.*, 2003; Yang *et al.*, 2003; Svartman *et al.*, 2004, 2006; Murphy *et al.*, 2005; Ferguson-Smith and Trifonov, 2007), with a consensus opinion settling on $n=23$ (see Ferguson-Smith and Trifonov, 2007). The rationale underpinning this, and the likely composition and uniqueness of the ancestral karyotype, as well as its correspondence with *in silico*-based studies of genome sequences, are discussed below.

ANCESTRAL PLACENTAL KARYOTYPES AND THE DETECTION OF SYNTENIES BASED ON FISH

Reconstructions of ancestral karyotypes across the placental mammalian tree rely heavily on molecular cytogenetic approaches that entail cross-species fluorescence *in situ* hybridization (Zoo-FISH; methodology reviewed by Rens *et al.*, 2006) using human and chromosome-specific DNA sequences from other species as probes. This has allowed the identification of orthologous regions defined by their correspondence with human chromosomes, and the delimitation of chromosomal rearrangements among species. These conserved regions span entire chromosomes, chromosomal arms, or chromosomal segments in closely and distantly related placental species permitting the generation of large-scale comparative maps among taxa. In the present context, it is important to make the distinction between segmental associations (the adjacent syntenies of some terminologies) and syntenic blocks that are retained *in toto* among lineages. The detection of segmental associations such as 4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q (each of which involve segments of human chromosomes that in combination correspond to complete chromosomes in the ancestral eutherian karyotype) in placentals, chicken and opossum was based on the evidence of the entire adjacent segment having been retained in representative genomes (Robinson and Ruiz-Herrera, 2008). However, the incomplete nature of the genome assemblies of platypus and frog does not permit the same level of resolution. We consequently used the junction as the defining character of a particular conserved segmental association based on the

premise that the independent assembly of a precisely shared association in different lineages was unlikely. Gene order within the abutting syntenic blocks may be altered by intrachromosomal rearrangement, and the size of these segments affected by subsequent translocation of parts to other regions of the genome (Robinson and Seiffert, 2004).

In most high-level reconstructions the identification of conserved syntenic blocks in multiple extant species (that is, commonality) was taken to reflect the retention of a shared ancestral evolutionary state leading to hypothesized ancestral karyotypes for Placentalia (Chowdhary *et al.*, 1998; Richard *et al.*, 2003; Yang *et al.*, 2003; Svartman *et al.*, 2004, 2006; Ferguson-Smith and Trifonov, 2007) and various orders of mammals, principally within Boreoeutheria. Reconstructions of the placental ancestral karyotype (PAK) have diploid numbers that vary from $n=22$ to $n=25$ (see Table 1 in Svartman *et al.*, 2004). The differences in interpretation are primarily related to the recognition of a single large chromosome (corresponding to HSA 1) in the placental ancestor (Murphy *et al.*, 2003), the detection of the 10q/12p/22q conserved syntenic segmental association (Froenicke *et al.*, 2003), and fusion of HSA1/19p (Yang *et al.*, 2003, $n=22$) based on its presence in Afrotheria (aardvark, elephant, golden mole and elephant shrew), at the time regarded as the most basal split in the eutherian tree (Murphy *et al.*, 2001a, b). The more recent studies appear, however, to have converged on $n=23$ for Placentalia (that is, the eutherian ancestral karyotypes of Froenicke *et al.*, 2003; Wienberg, 2004; Ferguson-Smith and Trifonov, 2007), and an identical $n=23$ in the boreoeutherian ancestral karyotype (BAK; Froenicke, 2005; Froenicke *et al.*, 2006; Robinson *et al.*, 2006).

The most definitive of the PAK constructs (Figure 2a) benefited from the availability of genome sequence information from two important outgroup species, the opossum and chicken. This permitted the distinction between shared ancestral characters (symplesiomorphies) and those that are unique to the ingroup Placentalia (that is, showing shared derived similarity and referred to as synapomorphies) allowing firm conclusions on the evolutionary history of each (Robinson and Ruiz-Herrera, 2008). The PAK is considered to comprise two chromosome pairs (corresponding to human chromosomes 13 and 18) and three conserved chromosome segments (10q, 19p and 8q in the human karyotype) that are probable symplesiomorphies as they are also present as unaltered orthologues in one or both outgroup species. Seven additional syntenic segmental associations (4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q), each involving human chromosomal segments from two or more human chromosomes, are also present in one or both outgroup taxa and are probable symplesiomorphies. Importantly, however, there are eight intact pairs (corresponding to human chromosomes 1, 5, 6, 9, 11, 17, 20 and the X) and three chromosomal segments (7b, 2p-q13 and 2q13-qter) that are derived characters, potentially consistent with placental monophyly. In summary therefore, the karyotype of the putative ancestor of Placentalia comprised 32 conserved segments (including the X) and nine syntenic segmental associations, several of which trace back to a common amniote ancestor (discussed below; Figures 2a and b).

Table 1 Number of orthologous genes and homologous synteny blocks in species established by pairwise comparisons to human

Species	*No. of orthologs genes*	*No. of HSBs*	*Median length (bp)*[a]	*Genome representation (%)*[b]
Chimpanzee	115 835	39	66 808 733	87.43
Orangutan	113 962	60	33 425 697	83.32
Macaque	117 410	72	29 853 043	91.37
Mouse	120 957	275	4 542 283	89.92
Rat	118 266	278	4 912 642	93.03
Cow	120 340	228	5 792 994	85.97
Dog	116 105	189	7 290 339	89.51
Armadillo	93 193	268	75 290	0.67
Elephant	113 657	141	6 285 988	54.66
Tenrec	100 672	580	78 347	1.64
Opossum	115 284	472	2 452 061	89.11
Platypus	102 547	957	116 845	37.15
Chicken	96 838	468	887 305	94.89
Frog	99 597	1128	224 864	37.95

Abbreviations: bp, base pair; HSBs, homologous synteny blocks.
[a]Median length of HSBs.
[b]Percentage of each genome covered by our scans.

There is, at this point, no evidence to suggest that the boreoeutherian ancestral karyotype (Froenicke *et al.*, 2006; Robinson *et al.*, 2006) underwent further modification from the hypothesized PAK (see above). The subsequent radiation of Boreoeutheria, however, showed extensive karyotypic modification in most lineages permitting hypothesized ancestral karyotypes for several orders of mammals, as well as the identification of syntenic segmental associations that underpin the monophyly of various supraordinal and ordinal groups (Robinson *et al.*, 2004; Wienberg, 2004; Froenicke, 2005; Ferguson-Smith and Trifonov, 2007; Ruiz-Herrera and Robinson, 2007, among others).

IN SILICO DETERMINATION OF THE ANCESTRAL BOREOEUTHERIAN KARYOTYPE AND EXTENT OF CONCORDANCE WITH THE CYTOGENETIC DATA

Advances from large-scale genome sequencing projects and the availability of new mathematical algorithms have revolutionized the study of chromosome evolution. The genomes of 35 mammalian species have been sequenced to differing degrees of completion (Ensembl database, version 59): 16 species of the Euarchontoglires (guinea pig, rat, mouse, rabbit, kangaroo rat, squirrel, tree shrew, tarsier, mouse lemur, bushbaby, marmoset, macaque, chimpanzee, orangutan, gorilla and human), 11 laurasiatherian representative (megabat, microbat, shrew, dolphin, pig, cow, alpaca, horse, dog, cat and hedgehog), three Afrotherian species (elephant, hyrax and tenrec), two xenathrans (sloth and armadillo), two species of Metatheria (wallaby and opossum) and the platypus as a prototherian representative. Of these, only the genomes of chimpanzee, rhesus macaque, orangutan, mouse, rat, cow, dog, horse and pig are sufficiently complete to allow pair-wise alignments with the human genome and the delimitation of syntenic blocks with a high degree of confidence.

Several sequenced-based reconstructions of the boreoeutherian ancestral karyotype have been attempted, often resulting in disparate outcomes compared with the findings suggested by FISH. In general terms, two different approaches can be distinguished when defining ancestral genomes in this way; (i) those that rely on the minimal number of rearrangements required to obtain the syntenies that lead to modern genomes (that is, MGR, Bourque and Pevzner, 2002) or (ii) models that focus on identifying conserved synteny blocks (Ma *et al.*, 2006). The former methodology was used in an early attempt at the reconstruction of a mammalian (but more correctly boreoeutherian) ancestral karyotype (human–rat–mouse) using the chicken as an outgroup (Bourque *et al.*, 2005). Although there is reasonable correspondence in the chromosome numbers suggested by MGR and cytogenetic data ($n=21$, cf. the $n=23$ posited by most chromosome painting strategies), the numbers of conserved segments and the numbers of syntenic associations were vastly different (Froenicke

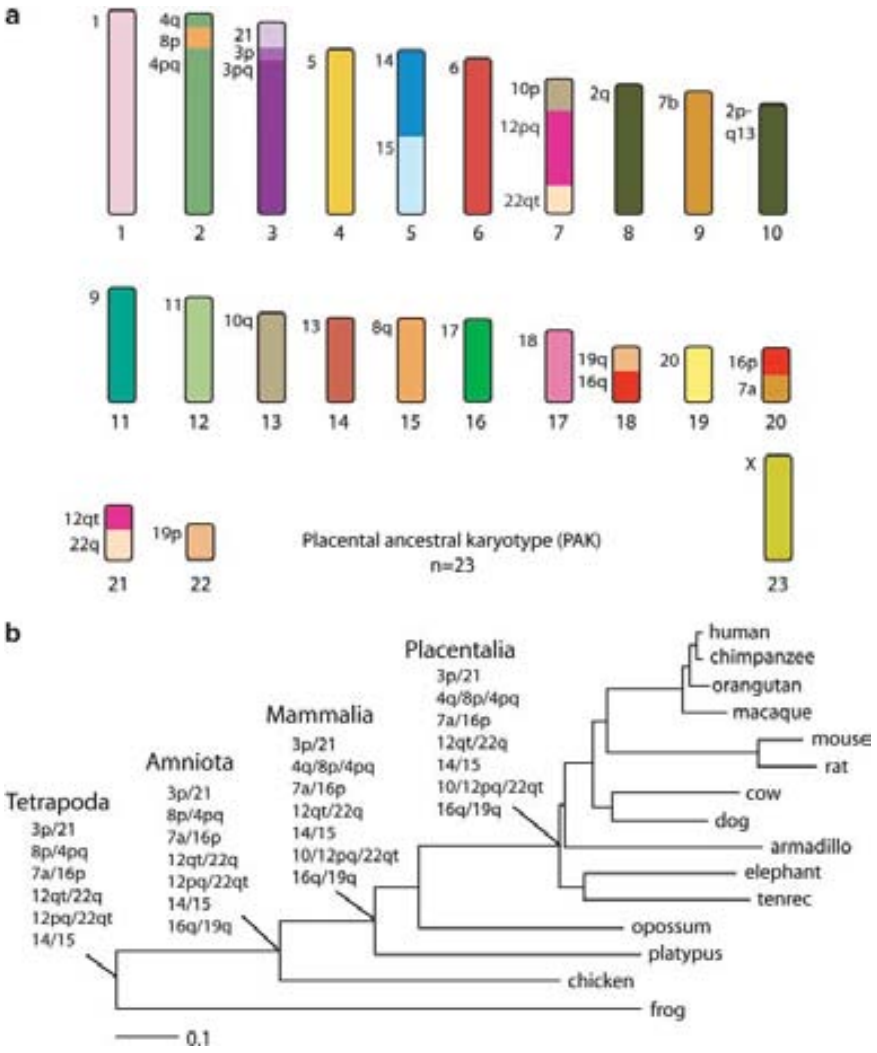


Figure 2 (**a**) Ancestral karyotype of Placentalia (PAK) defined by chromosomal correspondence to human chromosomes. Note that the HSA3/21 junction corresponds to human chromosomal segment 3p (depicted in violet), a region close to the centromere (from position 76.0 to 87.0 Mbp; Ruiz-Herrera and Robinson, 2007; Robinson and Ruiz-Herrera, 2008), and the conserved segmental association should more correctly be referred to as HSA3p/21. (**b**) Phylogenetic tree showing syntenic segmental associations detected at each ancestral node: p, short arm; pq, segment comprises parts of both the short and long arms; q, long arm; qt, terminal portion of the q arm.

et al., 2006). The MGR approach resulted in only four syntenic segmental associations (3/21, 4/8, 12a/22a and 12b/22b) being in common with those suggested by molecular cytogenetics. The degree of concordance was improved by Murphy *et al.* (2005), who used both genomic sequence data and information from radiation hybrid maps of eight species to obtain a more comprehensive view of the dynamics of genome organization in mammals. Their computational approach proposed an ancestral chromosome number of $n=24$ and showed that 80% of the conserved segments are in common with those detected by molecular cytogenetic approaches. However, only half of the syntenic segmental associations (specifically 3/21, 4/8×2, 7/16, 14/15, 12/22 and 16/19) were shared by both approaches (Robinson *et al*., 2006). Although it could be argued that the difference in the numbers of conserved segments is a reflection of the increased discrimination of the DNA sequence comparisons, several of the *in silico* syntenies fall within the limits detectable by FISH leading Froenicke *et al.* (2006) to question the effectiveness of the computational methodology. Using a different approach, in this case inferring contiguous ancestral regions within the completed genomes of human, dog, rat and mouse with chicken and opossum as outgroups, Ma *et al.* (2006) posit an ancestral boreoeutherian karyotype with $n=29$ but, importantly, with strong support for five of the ancestral syntenic segmental associations proposed by cytogenetic methods (4/8, 3/21, 14/15, 12/22×2).

Although there is consensus among the cytogenetic and computational approaches with respect to those conserved syntenies with strong probabilistic support (3/21, 4/8, 14/15, 12a/22a and 12b/22b), there are a meaningful number of ambiguous adjacent syntenies in conflict with the cytogenetic model (specifically 1/22, 5/19, 2/18, 1/10 and 2/20). This has led to the integration of available algorithms (Alekseyev and Pevzner, 2009) and to new methods of genome sequences analysis (Peng *et al*., 2009; Lin *et al*., 2010; Pham and Pevzner, 2010). It is anticipated that these efforts may, in future, provide more consistency to ancestral reconstructions based on *in silico* analysis and the degree of correspondence to the boreoeutherian construct suggested by the molecular cytogenetic analysis of more than 100 taxonomically diverse mammalian species.

IN SILICO IDENTIFICATION OF SYNTENIC SEGMENTAL ASSOCIATIONS AT DEEPER NODES OF THE VERTEBRATE TREE

As Zoo-FISH across the eutherian/metatherian boundary has been unsuccessful (with the exception of a small portion of the X that is conserved between the two lineages, Glas *et al*., 1999), there is a reliance on *in silico* methodologies to define the vertebrate protokaryotype, and to detect ancestral chromosomal syntenies that have been retained over deep diversification nodes. The recent publication (Hellsten *et al*., 2010) of the first amphibian genome to be sequenced—that of *Xenopus tropicalis*, a lineage that is thought to have diverged from amniotes ~360 mya—offers an opportunity to revisit putative ancestral karyotypes and conserved syntenies (which indicate the likely structure of ancestral chromosomes), deep within the vertebrate tree of life.

We used the SyntenyTracker (Donthu *et al*., 2009) to establish homologous synteny blocks (HSBs) between human and the genomes of 12 mammalian species (chimpanzee, orangutan, rhesus macaque, mouse, rat, cow, dog, armadillo, elephant, tenrec, opossum and platypus) plus the chicken and the frog (see online appendix for details). Table 1 provides the number of genes analyzed for each species and the number of HSBs detected, whereas the composition of the HSBs in the three progressively distant taxa to Placentalia—opossum, platypus and the chicken—is presented in Figure 3. Unfortunately, the draft frog genome is not assembled into chromosomes at this stage thus precluding the analysis of the whole karyotype and limiting our ability to unambiguously distinguish homologous and homoplasious syntenic associations between very distantly related species with potentially highly rearranged genomes (of which only portions can be traced in HSBs). The same shortcoming applies to platypus where several chromosomes remain unassembled (see below). The *in silico* chromosomal homologies identified by this approach permitted testing for PAK ancestral syntenic segmental associations at different phylogenetic levels (Figure 2b). This also allowed us to revisit the ancestral amniote and tetrapods genome compositions suggested by Kohn *et al*. (2006) and Nakatani *et al*. (2007) using syntenies identified in the frog, chicken, platypus and opossum.

Mammalian ancestral configuration

Our scans of the opossum and chicken genomes, analyzed as part of attempts to define placental chromosomal characters that define the monophyly of the group (Robinson and Ruiz-Herrera, 2008), revealed syntenic segmental associations (4q/8p/4pq, 3p/21, 14/15, 10p/12pq/22qt, 16q/19q, 7a/16p and 12qt/22q) that are shared with either (or, in some instances, both) opossum and chicken. No conserved human chromosomal segments were observed in the assembled platypus chromosomes (most probably because of low coverage of the annotated sequences). Nevertheless, there were several contigs (orthologous regions of small size, not yet assembled) that contained some of the syntenic segmental associations considered to be present in the PAK (Figure 2b). These were 4q/8p (Ultracontig173), 12qter/22q (Ultracontig252), 7a/16p (Ultracontig371), 3p/21 (Ultracontig388), 12q/22q (Ultracontig443), 16q/19q (Ultracontig517) and 22q12/12q24.3 (Ultracontig57; Table 2). On the basis of these data (constrained as they are by the partially complete platypus genome), and the chromosome number estimates presented above, our data suggest that the mammalian ancestral karyotype likely resembled the PAK in terms of chromosome number (n=23), and in the majority of the conserved syntenic segmental associations (only 10p/12pq/22qt and 14/15 were not detected in our scans of the platypus and the former has been regarded as a comparatively weakly supported ancestral chromosome form, see Froenicke *et al*., 2006). More detailed correspondence between the PAK and the ancestral karyotype for Mammalia is clearly dependent on progress in assembling the platypus genome.

Amniote ancestral configuration

The frog is an appropriate outgroup for defining syntenic segmental associations present in the ancestral amniote karyotype. The*X. tropicalis* genome is estimated at ~1.7 Gbp, distributed over 10 chromosomes or linkage groups (Hellsten *et al*., 2010). Of this, 769 Mb has been placed onto 691 scaffolds using genetic markers. This paucity of information is further underscored by 200 Mbp being assigned to linkage groups based on inference but without genetic markers (Hellsten *et al*., 2010), clearly necessitating further experimental studies. Despite this, our scans reveal that most of the ancestral placental syntenic segments are conserved in the frog genome (Figure 2b). In particular, the syntenies 3p/21, 4pq/8p, 7a/16p, 14/15, 12qt/22q and 12pq/22qt are present in some of the *Xenopus* scaffolds; in contrast, there was no evidence of 4q/8p/4pq, 10p/12pq/22qt and 16q/19q (Table 2).

Previous reports have attempted the description of the ancestral amniote genome (Nakatani *et al*., 2007; Ouangraoua *et al*., 2009) based on different taxon representation and methodological approaches. Nakatani *et al*., (2007) defined an ancestral amniote karyotype (AAK) comprising n=26. According to the authors, the AAK would present the following ancestral syntenic segmental associations in its chromosomes (see Figure 4 in Nakatani *et al*., 2007): 12/7b/12/22/7/16/17/22 (AAK 1), Xq/18/8pq/6/5/3p/7/10q (AAK 2), 5 (AAK 3), 2/6/13/3/2 (AAK 4), 6/20 (AAK 5), 10pq (AAK 6), 14/15 (AAK 7), 4 (AAK 8); 1 (AAK 9), 15 (AAK 10), 12/22 (AAK 11), 19/16 (AAK 12), 3/11 (AAK 13), 3/11 (AAK 14), 17 (AAK 15), 17 (AAK 16), 1/16 (AAK 17), 20 (AAK 18), X/5 (AAK 19), 6/19 (AAK 20), 1 (AAK 21), 1 (AAK 22), 8/7/2 (AAK 23), 19p (AAK 24), 11 (AAK 25) and 18/9/5 (AAK 26). Interestingly, the ancestral syntenic segments 10/12/22, 4/8 and 3/21 were not reported in the Nakatani *et al*. (2007) construct. It seems probable that the 4/8 and 3/21 are included in one or more of the unassigned blocks in Nakatani *et al*. (2007) given that they are present in chicken (Robinson and Ruiz-Herrera, 2008) and also frog (present study). More puzzling, however, is 10/12/22, which is not detected in chicken, nor in the frog genome, but is present in opossum and several placental mammals. We therefore view 10/12/22 as a chromosomal signature for Mammalia. Its absence in the platypus genome is due to the low coverage of the assembled sequences or, alternatively, to disruption in the lineage leading to Prototheria.

Tetrapod ancestral configuration

Our comparative genome analyses directed at establishing the likely composition of the tetrapod common ancestor are consistent with those of Kohn *et al*. (2006) with respect to the chromosomal number (n=18) and six conserved syntenic segmental associations that it likely contained (that is, 3p/21, 4pq/8p, 7a/16p, 12q/22q, 12pq/22qter and 14/15). We differ with respect to the involvement of 1/19p and 16q/19q suggested in Kohn *et al*. (2006). The inclusion of the former was based on its presence in Afrotheria (Yang *et al*., 2003). Interestingly this synteny is not found in chicken, platypus, nor in opossum, but 1p manifests as 1p/19p and 1p/19q in different scaffolds of the frog genome. This suggests the existence of 1p, 19p and 19q as separate syntenies in the tetrapod ancestral complement, and their independent assembly in the lineage leading to the frog. The presence of 1p/19q in opossum would therefore represent a convergent change (homoplasy).

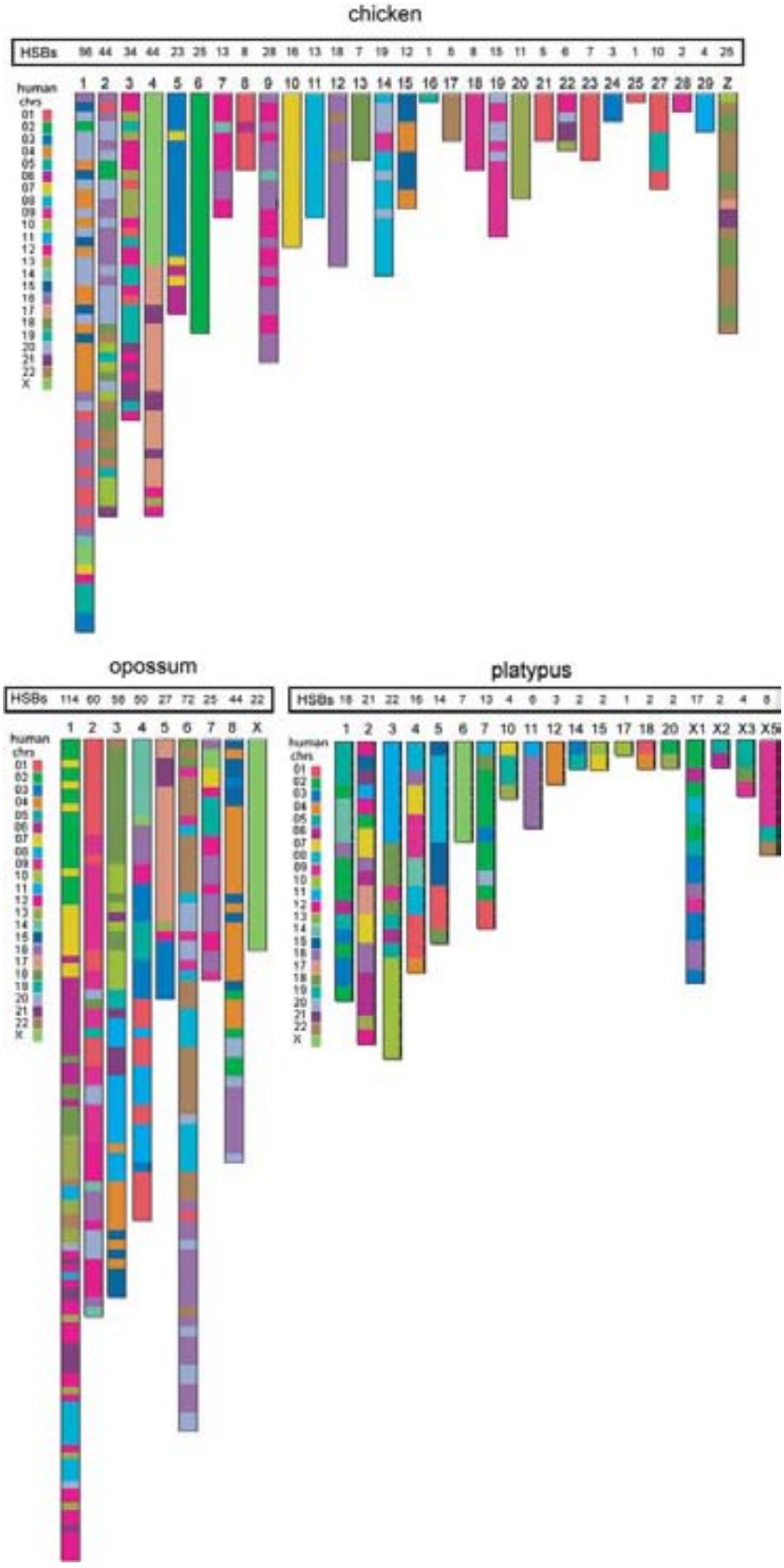


Figure 3 Conserved human chromosomal segments in the genomic assemblies of chicken, opossum and platypus. The human orthologous regions are color-coded and indicated as homologous syntenic blocks (HSBs) in the chromosomes of the respective species. The lengths of the chromosomes are based on homology coverage with the human genome and are not proportional to the chromosomal length.

Table 2 Presence (+) and absence (−) of ancestral syntenies detected in various outgroup species

Synteny	*Frog*[a]	*Chicken*	*Platypus*[a]	*Opossum*
3p/21	+	+	+	−
4q/8p/4pq	−	+	−	+
8p/4pq	+	−	−	−
8p/4q	+	−	+	−
7a/16p	+	+	+	+
10/12q/22q	−	−	−	+
12q/22q	+	+	+	+
14/15	+	+	−	+
19q/16q	−	−	+	+
1p/19p	+	−	−	−
1p/19q	+	−	−	+

Abbreviations: p, short arm; pq, segment; q, long arm.
[a]Low coverage, not completely assembled.

We found no evidence of the synteny 16q/19q in any of the *Xenopus* scaffolds (Table 2). On the basis of these conclusions, and the published data, we hypothesize that of all the ancient syntenic segments identified, at least 3p/21, 4pq/8p, 7a/16p, 14/15, 12qt/22q and 12pq/22qt predate the divergence of tetrapods (Figure 2b).

CLOSING COMMENTS AND FUTURE PROSPECTS

In this review we have examined how comparative molecular cytogenetic and computational approaches have contributed to the understanding of genome organization across deep divisions of the vertebrate tree of life. At first glance the diversity of karyotypes among extant species appears staggering. Placental mammals show a more pronounced and rapid rate of genomic reshuffling compared with birds and amphibians. It is clear from both Zoo-FISH and computational models of genome organization that the overwhelming pattern is, however, one of constrained change, most graphically illustrated by the high number of the conserved syntenies identified, and their retention in genomes of species from Boreoeutheria to Amphibia.

Superimposed on this conservative pattern are silos of rapid change where rearrangements have significantly altered the configuration and chromosome numbers of species, and this is most pronounced in Placentalia. Although reasons for these differences in tempo are still unclear, making this one of the most puzzling aspects of comparative cytogenetics, a burgeoning literature has identified regions at the junctions of synteny blocks that are rich in segmental duplications (Bailey and Eichler, 2006; Carbone *et al*., 2006; Kehrer-Sawatzki and Cooper, 2008), repeat content (Kehrer-Sawatzki *et al*., 2005; Ruiz-Herrera *et al*., 2006) and transposable elements (Bourque, 2009; Carbone *et al*., 2009; Delprat *et al*., 2009; Longo *et al*., 2009), predisposing these regions to rearrangement. In addition, transposable element activity and changes in DNA methylation patterns have been suggested as having a causative role in the structural modification of genomes in species as diverse as marsupials, rodents and primates (O'Neill *et al*., 1998; Brown *et al*., 2002; Carbone *et al*., 2009).

Although Robertsonian fusions and fissions appear frequently in studies of chromosomal rearrangement (as measured by changes in chromosome number), one of the most striking findings of comparative genomics is the high incidence of micro-inversions in the different genomes (Feuk *et al*., 2005; Lee *et al*., 2008; Zhao and Bourque, 2009). It may be that this largely undetected class of variation (inversions cannot be distinguished using whole chromosome painting, a data set that provides much of the basis for the recognition of ancestral constructs in Placentalia) functions as genomically localized barriers to recombination. In other words, the micro-inversions confer an adaptive advantage much in the same way as has been argued for speciation in the presence of gene flow (Rieseberg, 2001; Kirkpatrick and Barton, 2006; Butlin, 2010; Kirkpatrick, 2010 among others).

What is clear, however, is that the increasing availability of fully sequenced genomes (Haussler *et al*., 2009) will radically alter the field. These data, and anticipated improvements in methods of analysis, will result in comprehensive data sets that address current imbalances (large number of species but poor resolution provided by Zoo-FISH analysis, and the small number of species but high resolution provided by computational approaches), and provide fundamental insights to the mode and tempo of structural change in genomes that are presently intractable in terms of FISH analysis.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.7j0b8468.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

AR-H's laboratory is funded by the Spanish Ministry of Science and Innovation (CGL2010-20170) and the Barcelona Zoological Gardens (BSM, Zoo Barcelona). MF is a predoctoral student supported by the Universitat Autònoma de Barcelona (Spain). TJR's research is supported by grants from the South African National Research Foundation. We thank Lutz Froenicke and Adam Wilkins for their insightful comments and suggestions.

Alekseyev MA, Pevzner PA (2009). Breakpoint graphs and ancestral genome reconstructions. *Genome Res* **19**: 943–957.
Asher RJ, Helgen KM (2010). Nomenclature and placental mammal phylogeny. *BMC Evol Biol* **10**: 102.
Bailey JA, Eichler EE (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* **7**: 552–564.
Bourque G (2009). Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* **19**: 607–612.
Bourque G, Pevzner PA (2002). Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res* **12**: 26–36.
Bourque G, Zdobnov EM, Bork P, Pevzner PA, Tesler G (2005). Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res* **15**: 98–110.
Brown JD, Strbuncelj M, Giardina C, O'Neill RJ (2002). Interspecific hybridization induced amplification of Mdm2 on double minutes in a Mus hybrid. *Cytogenet Genome Res* **98**: 184–188.
Butlin RK (2010). Population genomics and speciation. *Genetica* **138**: 409–418.
Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J *et al.* (2009). Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet* **5**: e1000538.
Carbone L, Vessere GM, ten Hallers BF, Zhu B, Osoegawa K, Mootnick AR *et al.* (2006). A high-resolution map of synteny disruptions in gibbon and human genomes. *PLoS Genet* **2**: 223.
Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H (1998). Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Res* **8**: 577–589.
Churakov G, Kriegs JO, Baertsch R, Zemann A, Brosius J, Schmitz J (2009). Mosaic retroposon insertion patterns in placental mammals. *Genome Res* **19**: 868–875.
De Smet WHO (1981). The nuclear Feulgen-DNA content of the vertebrates (especially reptiles), as measured by fluorescence cytophotometry, with notes on the cell and chromosome size. *Acta Zool Pathol Antverpiensia* **76**: 119–167.
Delprat A, Negre B, Puig M, Ruiz A (2009). The transposon Galileo generates natural chromosomal inversions in Drosophila by ectopic recombination. *PLoS One* **4**: e7883.
Donthu R, Lewin HA, Larkin DM (2009). SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC Res Notes* **2**: 148.
Ellegren H (2010). Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol Evol* **25**: 283–291.
Ferguson-Smith MA, Trifonov V (2007). Mammalian karyotype evolution. *Nat Rev Genet* **8**: 950–962.

Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G *et al.* (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet* **1**: e56.

Froenicke L (2005). Origins of primate chromosomes - as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. *Cytogenet Genome Res* **108**: 122–138.

Froenicke L, Caldes MG, Graphodatsky A, Muller S, Lyons LA, Robinson TJ *et al.* (2006). Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome Res* **16**: 306–310.

Froenicke L, Wienberg J, Stone G, Adams L, Stanyon R (2003). Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of the human and the African elephant (*Loxodonta africana*) generated by chromosome painting. *Proc R Soc Lond B* **270**: 1331–1340.

Glas R, Marshall Graves JA, Toder R, Ferguson-Smith M, O'Brien PC (1999). Cross-species chromosome painting between human and marsupial directly demonstrates the ancient region of the mammalian X. *Mamm Genome* **10**: 1115–1116.

Gregory TR (2011). Animal Genome Size Database. http://www.genomesize.com.

Griffin DK, Robertson LB, Tempest HG, Skinner BM (2007). The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117**: 64–77.

Hallström BM, Janke A (2010). Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol* **27**: 2804–2816.

Haussler D, O'Brien SJ, Ryder OA, Barker FK, Clamp M, Crawford AJ, Hanner R, Hanotte O, Johnson WE, McGuire JA *et al.* (2009). Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J Hered* **100**: 659–674.

Hayman DL (1990). Marsupial cytogenetics. *Aust J Zool* **37**: 331–349.

Hayman DL, Martin PG (1969). Cytogenetics of marsupials. In: Benirschke K (ed.). *Comparative Mammalian Cytogenetics*. Springer-Verlag: New York.

Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V *et al.* (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**: 633–636.

Kehrer-Sawatzki H, Cooper DN (2008). Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res* **16**: 41–56.

Kehrer-Sawatzki H, Sandig CA, Goidts V, Hameister H (2005). Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* **108**: 91–97.

Kirkpatrick M (2010). How and why chromosome inversions evolve. *PLoS Biol* **8**: e1000501.

Kirkpatrick M, Barton N (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Kohn M, Hogel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JA *et al.* (2006). Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet* **22**: 203–210.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J (2006). Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol* **4**: e91.

Lee J, Han K, Meyer TJ, Kim HS, Batzer MA (2008). Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS One* **3**: e4047.

Lin CH, Zhao H, Lowcay SH, Shahab A, Bourque G (2010). webMGR: an online tool for the multiple genome rearrangement problem. *Bioinformatics* **26**: 408–410.

Longo MS, Carone DM, , Green ED, O'Neill MJ, O'Neill RJ (2009). Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. *BMC Genomics* **10**: 334.

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ *et al.* (2006). Reconstructing contiguous regions of an ancestral genome. *Genome Res* **16**: 1557–1565.

Matthey R (1945). L'evoltion de la formule chromosomiale chez les Vertebres. *Experentia* **1**: 78–86.

Matthey R (1958). Chromosomes & systematic position of various African Murinae (Mammalia, Rodentia). *Acta Trop* **15**: 97–117.

Murphy W, Froenicke L, O'Brien SJ, Stanyon R (2003). The origin of human chromosome 1 and its homologues in placental mammals. *Genome Res* **13**: 1880–1888.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001b). Molecular phylogenetics and the origins of placental mammals. *Nature* **409**: 614–618.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ *et al.* (2001a). Resolution of the early placental mammalian radiation using Bayesian phylogenetics. *Science* **294**: 2348–2351.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L *et al.* (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613–617.

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007). Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17**: 1254–1265.

Nanda I, Benisch P, Fetting D, Haaf T, Schmid M (2011). Synteny conservation of chicken macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet Genome Res* **132**: 165–181.

Naruse K, Hori H, Shimizu N, Kohara Y, Takeda H (2004). Medaka genomics: a bridge between mutant phenotype and gene function. *Mech Dev* **121**: 619–628.

Nie W, O'Brien PC, Ng BL, Fu B, Volobouev V, Carter NP *et al.* (2009). Avian comparative genomics: reciprocal chromosome painting between domestic chicken (*Gallus gallus*) and the stone curlew (*Burhinus oedicnemus*, Charadriiformes an atypical species with low diploid number. *Chromosome Res* **17**: 99–113.

Nilsson MA, Churakov G, Sommer M, Tran NV, Zemann A, Brosius J *et al.* (2010). Tracking marsupial evolution using archaic genomic retroposon insertions. *PLoS Biol* **8**: e1000436.

Nishihara H, Satta Y, Nikaido M, Thewissen JG, Stanhope MJ, Okada N (2005). A retroposon analysis of Afrotherian phylogeny. *Mol Biol Evol* **22**: 1823–1833.

O'Brien SJ, Menninger JC, Nash WG (2006). *An Atlas of Mammalian Chromosomes*. John Wiley & Sons, Inc: Hoboken, NJ.

O'Neill RJ, Eldridge MD, Metcalfe CJ (2004). Centromere dynamics and chromosome evolution in marsupials. *J Heredity* **95**: 375–381.

O'Neill RJW, O'Neill MJ, Graves JAM (1998). Undermethylation associated with retroelement activation and chromosome remodeling in an interspecific mammalian hybrid. *Nature* **393**: 68–72.

Ouangraoua A, Boyer F, McPherson A, Tannier E, Chauve C (2009). Prediction of contiguous regions in the amniote ancestral genome. *Lect Notes Comput Sci* **5542**: 173–185.

Peng Z, Elango N, Wildman DE, Yi SV (2009). Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. *BMC Genomics* **10**: 247.

Pham SK, Pevzner PA (2010). DRIMM-Synteny: decomposing genomes into evolutionary conserved segments. *Bioinformatics* **26**: 2509–2516.

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F *et al.* (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10**: 1890–1902.

Rens W, Fu B, O'Brien PC, Ferguson-Smith M (2006). Cross-species chromosome painting. *Nat Protoc* **1**: 783–790.

Rens W, O'Brien PC, Yang F, Solanky N, Perelman P, Graphodatsky AS *et al.* (2001). Karyotype relationships between distantly related marsupials from South America and Australia. *Chromosome Res* **9**: 301–308.

Richard F, Lombard M, Dutrillaux B (2003). Reconstruction of the ancestral karyotype of eutherian mammals. *Chromosome Res* **11**: 605–618.

Rieseberg LH (2001). Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**: 351–358.

Robinson TJ, Fu B, Ferguson-Smith MA, Yang F (2004). Cross-species chromosome painting in the golden mole and elephant-shrew: support for the mammalian clades Afrotheria and Afroinsectiphillia but not Afroinsectivora. *Proc Biol Sci* **271**: 1477–1484.

Robinson TJ, Ruiz-Herrera A (2008). Defining the ancestral eutherian karyotype: a cladistic interpretation of chromosome painting and genome sequence assembly data. *Chromosome Res* **16**: 1133–1141.

Robinson TJ, Ruiz-Herrera A, Froenicke L (2006). Dissecting the mammalian genome - new insights into chromosomal evolution. *Trends Genet* **22**: 297–301.

Robinson TJ, Seiffert E (2004). Afrotherian origins and interrelationships: new views and future prospects. *Curr Top Dev Biol* **63**: 37–60.

Ruiz-Herrera A, Castresana J, Robinson TJ (2006). Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol* **7**: R115.

Ruiz-Herrera A, Robinson TJ (2007). Chromosomal instability in Afrotheria: fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies. *BMC Evol Biol* **7**: 199.

Svartman M, Stone G, Page JE, Stanyon R (2004). A chromosome painting test of the basal eutherian karyotype. *Chromosome Res* **12**: 45–53.

Svartman M, Stone G, Stanyon R (2006). The ancestral eutherian karyotype is present in Xenarthra. *PLoS Genet* **2**: e109.

Voss SR, Kump DK, Putta S, Pauly N, Reynolds A, Henry R *et al.* (2011). Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res* **21**: 1306–1312.

Waddell PJ, Okada N, Hasegawa M (1999). Towards resolving the interordinal relationships of placental mammals. *Syst Biol* **48**: 1–5.

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007). Evolutionary history of LINE-1 in the major clades of placental mammals. *PLoS One* **2**: e158.

Westerman M, Meredith RW, Springer MS (2010). Cytogenetics meets phylogenetics: a review of karyotype evolution in diprotodontian marsupials. *J Hered* **101**: 690–702.

White MJD (1973). *Animal Cytology and Evolution*, 3rd edn. Cambridge University Press: London.

Wienberg J (2004). The evolution of eutherian chromosomes. *Curr Opin Genet Dev* **14**: 657–666.

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R *et al.* (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* **15**: 1307–1314.

Yang F, Alkalaeva EZ, Perelman PL, Pardini AT, Harrison WR, O'Brien PC *et al.* (2003). Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. *Proc Natl Acad Sci USA* **100**: 1062–1066.

Zhao H, Bourque G (2009). Recovering genome rearrangements in the mammalian phylogeny. *Genome Res* **19**: 934–942.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)

TREBALL 2

Assessing the role of tandem repeats in shaping the genomic architecture of great apes

Farré M, Bosch M, López-Giráldez F, Ponsà M, Ruiz-Herrera A

PLoS One (2011); 6 (11)

Índex d'impacte (2010): 4,411

ÀREA: Biology, QUARTIL 1

OPEN ACCESS Freely available online

PLoS one

Assessing the Role of Tandem Repeats in Shaping the Genomic Architecture of Great Apes

Marta Farré[1], Montserrat Bosch[2], Francesc López-Giráldez[3], Montserrat Ponsà[1], Aurora Ruiz-Herrera[1,4]*

1 Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, **2** Genètica de la Conservació Animal, IRTA, Cabrils, Spain, **3** Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, United States of America, **4** Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Abstract

Background: Ancestral reconstructions of mammalian genomes have revealed that evolutionary breakpoint regions are clustered in regions that are more prone to break and reorganize. What is still unclear to evolutionary biologists is whether these regions are physically unstable due solely to sequence composition and/or genome organization, or do they represent genomic areas where the selection against breakpoints is minimal.

Methodology and Principal Findings: Here we present a comprehensive study of the distribution of tandem repeats in great apes. We analyzed the distribution of tandem repeats in relation to the localization of evolutionary breakpoint regions in the human, chimpanzee, orangutan and macaque genomes. We observed an accumulation of tandem repeats in the genomic regions implicated in chromosomal reorganizations. In the case of the human genome our analyses revealed that evolutionary breakpoint regions contained more base pairs implicated in tandem repeats compared to synteny blocks, being the AAAT motif the most frequently involved in evolutionary regions. We found that those AAAT repeats located in evolutionary regions were preferentially associated with *Alu* elements.

Significance: Our observations provide evidence for the role of tandem repeats in shaping mammalian genome architecture. We hypothesize that an accumulation of specific tandem repeats in evolutionary regions can promote genome instability by altering the state of the chromatin conformation or by promoting the insertion of transposable elements.

Citation: Farré M, Bosch M, López-Giráldez F, Ponsà M, Ruiz-Herrera A (2011) Assessing the Role of Tandem Repeats in Shaping the Genomic Architecture of Great Apes. PLoS ONE 6(11): e27239. doi:10.1371/journal.pone.0027239

Editor: David Liberles, University of Wyoming, United States of America

Received April 6, 2011; **Accepted** October 12, 2011; **Published** November 4, 2011

Funding: Financial support from Ministerio de Ciencia y Tecnologia (BFU2004-03422 and CGL-2010- 20170) and the Universitat Autònoma de Barcelona (PhD fellowship to Marta Farré) are gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aurora.ruizherrera@uab.cat

Introduction

Since the earliest cytogenetic studies, evolutionary biologists have sought to understand how mammalian genomes are organized. The characterization of orthologous chromosomal segments among several mammalian species was initially performed by means of G-banding comparisons [1,2]. Advances in molecular cytogenetic techniques, such as cross-species in situ hybridization, increased the level of resolution for defining orthologous regions as well as the number of species studied [3]. As a result, the integration of cross-species chromosome painting studies performed in more than 100 mammalian species [4,5] has revealed that evolutionary breakpoints (i.e., the disruption of two orthologous chromosomal segments) are not homogeneously distributed but rather concentrated in certain regions across the human genome.

The multiple ongoing genome sequencing projects are producing an extraordinary amount of data to further refine genome comparisons at a deeper level of resolution: the DNA sequence level. The public availability of these data makes it possible to establish reliable comparisons among genomes, thus providing new insights into the driving forces that generate gene variation, adaptation and evolution. Different approaches have been developed in order to define homologous synteny blocks (HSBs; i.e. regions where the gene order has been conserved among species) and evolutionary breakpoint regions (EBRs; i.e. regions where the synteny has been disrupted by chromosomal reorganizations) among mammalian genomes. Early studies were based on pair-wise comparisons between human and mouse or human and rat genomes [6,7], using the human genome as a reference whereas recent approaches have gone even further by establishing pair-wise comparisons among several vertebrate species [8–11].

Confirming previous cytogenetic studies, in silico analysis lead to the fragile-breakage model, founded initially on mathematical algorithms [6,12]. According to this model, EBRs are located in specific regions and have been used repeatedly during evolution (i.e., "reused"). In a phylogenetic context, the term "breakpoint reuse" accounts for the recurrence of the same breakpoint in two different species, but not in the common ancestor, based on comparison with an outgroup lineage [8,11,13]. The assumption that some chromosome regions have been reused during mammalian chromosomal evolution leads evolutionary biologists to investigate whether there is any particular DNA configuration or composition driving genome instability. Are these evolutionary

regions physically unstable due to sequence composition and/or genome organization, or do they merely represent genomic areas where the selection against breakpoints is minimal?

An interesting aspect that has emerged from comparative genomic studies is the finding that breakpoint regions are rich in repetitive elements, for example tandem repeats [14], segmental duplications [15–17], and transposable elements [18–20]. Repetitive elements represent nearly 50% of the human genome [21]. Among them are tandem repeats, which consist of perfect (or slightly imperfect) copies of a motif in a head to tail fashion, and comprise about 3% of the human genome [21]. They can be classified into two groups, microsatellites and minisatellites. Microsatellites are short tandem repeats with 1-6 bp as a repeat unit, whereas minisatellites contain repeat units $\geq$7 bp [22]. Tandem repeats have been regarded as an important source of DNA variation and mutation [23]. Tandem repeats can form non-B DNA structures (i.e., DNA structures different from the Watson-Crick conformation), such as hairpins, cruciform or triplex conformations [24], promoting DNA instability and giving rise to chromosomal reorganizations [25].

While it is clear that tandem repeats are involved in the etiology of several human diseases [26–28], the evolutionary implications of these sequences remain elusive. Given that tandem repeats have been shown to be concentrated in evolutionary chromosomal bands in the human genome [10] our aim was to test this hypothesis in other primate species presenting a comprehensive study of the distribution of tandem repeats in great apes. Taking advantage of the sequenced genomes of 10 vertebrate species (chimpanzee, orangutan, rhesus macaque, mouse, rat, horse, dog, cow, opossum and chicken) available in the public databases, we analyzed the distribution of tandem repeats in relation to the distribution of evolutionary breakpoint regions in the human, chimpanzee, orangutan and macaque genomes, from which the ancestral chromosomal state is known. A comparative study among species is presented and its implications for mammalian chromosome evolution are discussed.

Results

Whole-genome comparisons and delimitation of homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in great apes

Definition of HSBs and EBRs. In order to establish the evolutionary genomic landscape in great apes, we initially delimitated HSBs and EBRs in the human, chimpanzee and orangutan genomes by means of pair-wise comparisons (see Material and Methods). The gorilla genome was not available at the moment of the initiation of the study and the rhesus macaque was included as an outgroup for the Hominoidea superfamily.

First, we determined the HSBs and EBRs in the human genome establishing pair-wise whole-genome comparisons with ten vertebrate species (chimpanzee, orangutan, rhesus macaque, mouse, rat, horse, dog, cow, opossum and chicken). The number of HSBs differed depending on the species compared, ranging from 81 HSBs between human and macaque to 470 HSBs between human and opossum (Table 1). HSBs represented more than 70% of the human genome, reaching 91.88% for the human/orangutan comparison (Table 1), reflecting the high conservation of mammalian genomes. The mean length of the HSBs ranged from 30.61 Mbp for human/macaque to 5.08 Mbp for the human/opossum pair-wise comparison. Likewise, the number of EBRs also differed among species, being low in the non-human primate species (35, 61 and 88 between human and macaque, chimpanzee and orangutan, respectively) and high in the human/opossum comparison (Table 1). Moreover, and in order to avoid possible artifacts derived from the low-coverage annotation, intervals longer than 4 Mbp between two HSBs were considered as gaps. Gap regions ranged from 3.79 to 17.82% of the human genome, depending on the genome analyzed (Table 1). The larger percentages of gap regions were found in the human/macaque, human/dog, human/opossum and human/chicken pair-wise comparisons. These differences were probably due to the low coverage of some of the genomes available in the databases (e.g., 5.2X coverage for the macaque genome) or to the large evolutionary distances between species (300 My between human and chicken and 180 My between human and opossum).

Given these results, we merged the coordinates of all pair-wise comparisons abovementioned in the human genome (see material and methods for detailed explanation and, Fig. 1) in order to have a broad view of the distribution of evolutionary breakpoint regions. As a result, we obtained a total of 1,353 HSBs and 898 EBRs, representing altogether 67.38% of the whole genome sequence (Table 1). The EBRs detected varied in size, from 3 bp to 3.5 Mbp, with a median length of 304 kbp. Regions of non-coverage (gaps) represented 23.95% of the whole genome whereas telomeric and centromeric regions accounted for the remaining 8.67% (Table 1).

We observed that EBRs were unevenly distributed among human chromosomes, given that some human chromosomes accumulated more EBRs than others, independently of their genomic length. We calculated the frequency of EBRs per megabase for each chromosome (Fig. 2), and estimated an average frequency of 0.3 EBR/Mbp in the human genome assuming a homogeneous distribution of the 898 EBRs across the genome (telomeres, centromeres and gap regions were excluded from the analysis). Comparing the observed frequencies with the estimated global frequency of EBRs (0.3 EBRs/Mbp), we observed a deviation ($\chi^2 = 7.7$, p-value = 0.005) from the homogeneous distribution of EBRs among chromosomes (Fig. 2). Chromosome 19 accumulated more EBRs (0.53 EBRs/Mbp), while chromosome 13 (0.18 EBRs/Mbp) and chromosome 14 (0.19 EBRs/Mbp) had less EBRs. Although these differences were found to be not significant after Bonferroni correction, the tendency was still observed in all species.

Once the evolutionary regions were defined in the human genome, we determined the HSBs and EBRs in the chimpanzee, orangutan and macaque genomes. In this case, we established pair-wise whole-genome comparisons with the human genome using the primate genomes as references performing chimpanzee/human, orangutan/human and macaque/human pair-wise alignments (Table 1). We detected 32 EBRs in the chimpanzee genome, 46 in orangutan and 27 in macaque, with a median length of 235 kbp, 32 kbp and 10 kbp, respectively. The percentage of homologous syntenic regions was greater in chimpanzee (84.55%) than in orangutan (83.81%) and macaque (79.64%), consistent with their phylogenetic relation to human.

Phylogenetic interpretation of evolutionary breakpoint regions. To have an estimation of the EBR reuse during mammalian evolution we placed the EBRs detected in the human genome in an evolutionary context (Fig. 3). Given that we applied maximum parsimony criteria, these rates represent estimates of change. This approach could led us to ignore the variability due to focusing only on the mapping that requires the fewest genomic changes and to underestimate the true rate of change [29]. Out of the 898 EBRs detected, 436 were species-specific (48.6%), 280 clade-specific (31.2%) and 182 (20.3%) were found in two or more species but not in their common ancestor (reused). Based on the phylogenetic distances described by Murphy and co-workers [30],

Table 1. Homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in primate genomes.

	HSBs			EBRs			Gaps		
Species compared	**N° regions**	**Total length (Mbp)**	**% human genome**	**N° regions**	**Total length (Mbp)**	**% human genome**	**N° regions**	**Total length (Mbp)**	**% human genome**
HSA-PTR	97	2,785	90.86	61	37	1.23	59	138	4.51
HSA-PPY	122	2,817	91.88	88	32	1.06	55	116	3.79
HSA-MMU	81	2,479	80.86	35	22	0.72	69	460	15.02
HSA-RNO	287	2,543	82.95	245	128	4.20	65	289	9.45
HSA-MMUS	324	2,727	88.97	291	81	2.64	56	152	4.99
HSA-ECA	188	2,764	90.17	154	49	1.62	55	147	4.81
HSA-BTA	336	2,726	88.93	301	80	2.62	58	255	8.32
HSA-CFA	173	2,388	77.90	128	38	1.24	68	546	17.82
HSA-MDO	470	2,390	77.96	424	302	9.86	69	269	8.78
HSA-GGA	361	2,176	71.00	288	244	7.97	96	537	17.53
merged HSA	1,353	1,403	43.14	898	788	24.24	576	779	23.95
PTR-HSA	89	2,832	84.55	32	28	0.85	11	83	2.94
PPY-HSA	109	2,888	83.81	46	36	1.04	8	256	7.43
MMU-HSA	66	2,466	79.64	27	15	0.48	19	380	12.28

Pair-wise genome comparisons were established in two directions; using as a reference the human genome (HSA-PTR, HSA-PPY, HSA-MMU, HSA-RNO, HSA-MMUS, HSA-ECA, HSA-BTA, HSA-CFA, HSA-MDO, HSA-GGA) or the primate genomes (PTR-HSA, PPY-HSA and MMU-HSA). The total numbers of HSBs, EBRs and gaps in the human genome after merging all pair-wise comparisons are also indicated.
PTR *Pan troglodytes*, PPY *Pongo pygmaeus*, MMU *Macaca mulatta*, RNO *Rattus norvegicus*, MMUS *Mus musculus*, ECA *Equus caballus*, BTA *Bos taurus*, CFA *Canis familiaris*, MDO *Monodelphis domestica* and GGA *Gallus gallus*.
doi:10.1371/journal.pone.0027239.t001

we estimated an average rate of 0.35 EBRs per million year (myr) for all mammals and 0.27 EBRs/myr for eutherian mammals. Out of the 280 clade-specific EBRs, 180 were marsupialia-specific (1 EBR/myr), 48 were placentalia-specific (0.27 EBR/myr) and 109 were mammalian-specific (0.35 EBR/myr) (Fig. 3). Among the mammalian species studied, the mouse and the rat genome presented the highest estimated rate of genomic changes (1.85 EBRs/myr and 1.95 EBRs/myr, respectively) whereas the macaque was the species with the lowest rate of change (0.2 EBR/myr). Within Laurasiatheria, the cow was the species with the highest rate (1.44 EBR/myr), followed by the dog (0.71 EBR/myr) and the horse (0.28 EBR/myr). Primates showed the lowest estimated rate of change (0.21 EBR/myr), ranging from 0.2 EBR/myr in macaque to 1.83 EBR/myr in chimpanzee.

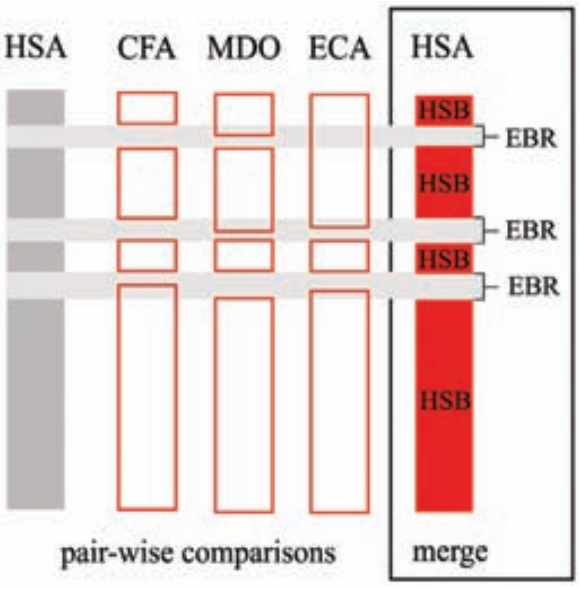


Figure 1. Representation of how homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) are defined in the human genome. Comparing two genomes at a time, we established pair-wise EBRs. Then, we merged those EBRs that overlap in the same human region, obtaining merged EBRs and HSBs. Abbreviations –PTR: *Pan troglodytes*, ECA: *Equus caballus*, MDO: *Monodelphis domestica*, HSA: *Homo sapiens*.
doi:10.1371/journal.pone.0027239.g001

Taking into account the putative ancestral hominoid karyotype [1,31] we interpreted the primate-specific EBRs found in each species of great apes. Chromosomes from orangutan, gorilla, chimpanzee and human are highly homologous and only few major reorganizations differentiate their karyotypes [1]. Since their divergence from a common ancestor 14 million years ago (mya) [32], some chromosomal forms have been maintained collinear (chromosomes 6, 13, 19, 20, 21, 22 and X) whereas others suffered inversions and/or lineage-specific fusions. Pair-wise whole-genome comparisons between great apes and human genomes allowed us to refine the number of rearrangements that occurred during hominoid evolution. New insertions were represented by one EBR whereas inversions were caused by two EBRs (Table 2). Regarding collinear chromosomes, we found reorganizations previously undetected in homologous chromosomes 13, 19 and X. In particular, we found a new insertion in orangutan chromosomes 13, X and 19 and a new inversion in chimpanzee chromosome 19. Even though chromosome 8 is collinear in chimpanzee, orangutan and human, we found one EBR due to an insertion in chimpanzee and orangutan but not in human homologous positions. Regarding the reorganized chromosomes, we corroborated the macro reorganizations found in chromosomes 1, 2, 3, 5, 7, 10, 12, 14, 15, 16, 17 and 18 [1,31]. In chromosome 11, which the orangutan represents the ancestral

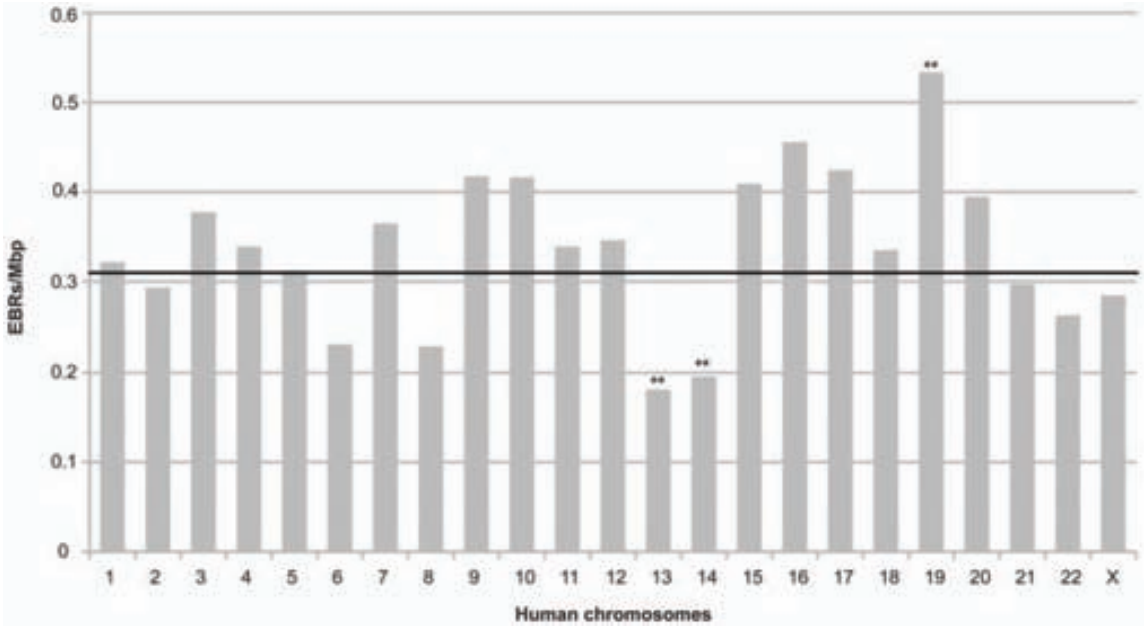


Figure 2. Distribution of EBRs across the human genome. Frequency of EBRs per megabase pair (Mbp) detected on each human chromosome. The dotted line represents the estimated frequency of EBRs per Mbp in the human genome.
doi:10.1371/journal.pone.0027239.g002

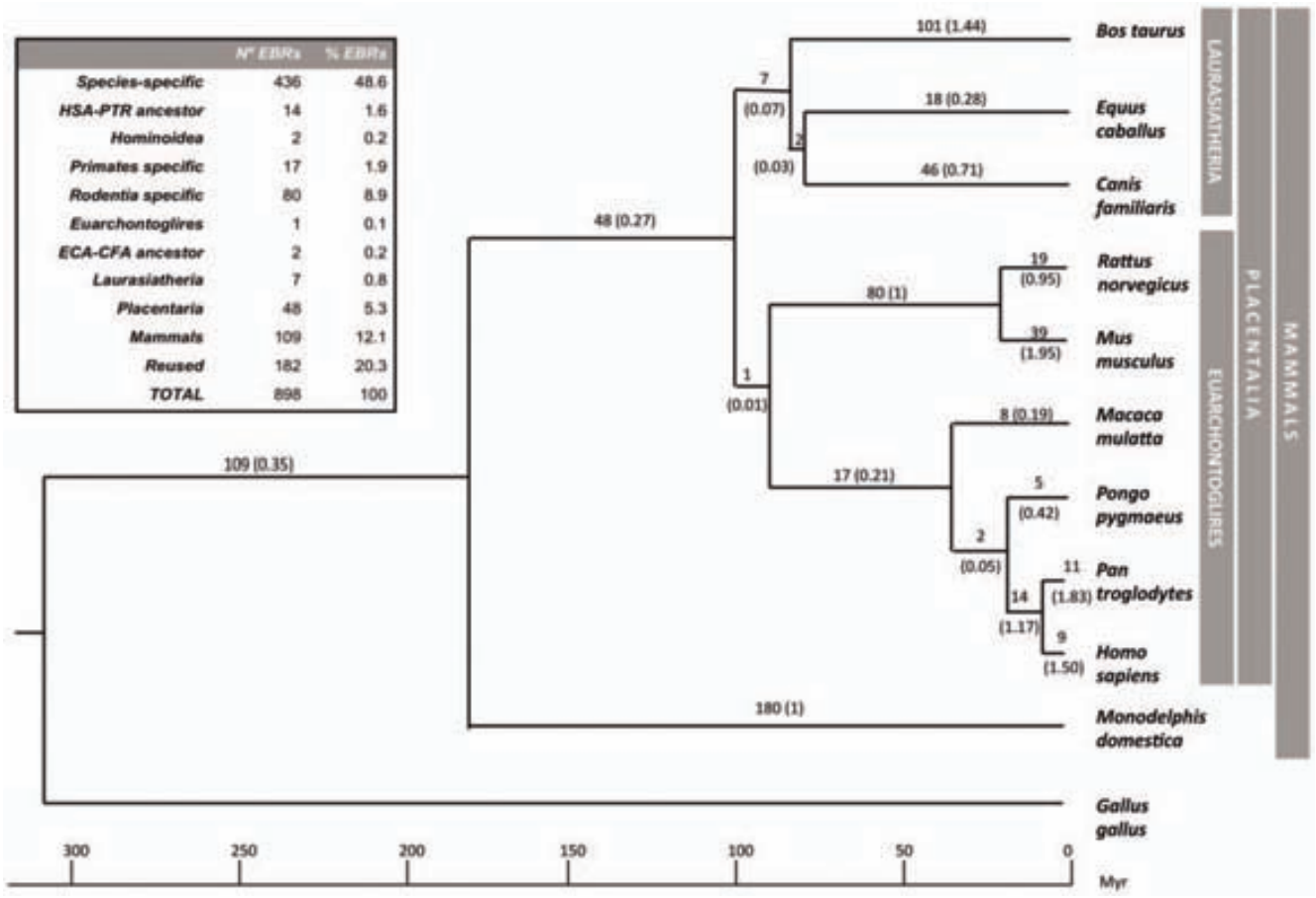


Figure 3. EBRs mapped in the phylogenetic tree of mammalian species included in our study. The phylogeny was based on previous studies [33,62]. The number of specific evolutionary breakpoint regions detected is plotted in each phylogenetic branch. The number of EBRs per million years detected for each lineage is displayed in brackets. Inset shows the number and percentage of EBRs found in our study.
doi:10.1371/journal.pone.0027239.g003

Table 2. Newly described reorganizations in human (HSA), chimpanzee (PTR) and orangutan (PPY) chromosomes.

Chromosome	HSA	PTR	PPY
4	Ancestral	Insertion (121,995,429-121,997,005)	Inversion previously found [a]
7	Inversion previously found [a]	Inversion (40,154,256-44,613,528)	Ancestral
8	Ancestral	Insertion (7,592,222-7,730,288)	Insertion (44,119,443-47,565,927)
9	Ancestral	Insertion (42,012,304-42,239,829)	Inversion previously found [a]
11	Inversion previously found [b]	Insertion (88,294,605-88,650,196)	Ancestral
13	Ancestral	Ancestral	Insertion (23,683,269-23,732,315)
19	Ancestral	Inversion (41,544,000-42,809,028)	Insertion (24,329,459-27,955,815)
X	Ancestral	Ancestral	Insertion (58,752,636-60,465,663)

The ancestral form and type of reorganization with the genomic location are shown. The genomic positions (start and end, NCBI build 36) of each insertion or inversion are also indicated.
[a][1],
[b][31].
doi:10.1371/journal.pone.0027239.t002

form, we verified the inversion found in human and chimpanzee, plus an additional EBR in chimpanzee resulted from an insertion of 355 kb. In chimpanzee chromosome 4 we found 3 EBRs, two as a result of the inversion previously described and one from an insertion of 1.5kb. Likewise, an insertion of 227 kb was found in chimpanzee chromosome 9 (Table 2).

Tandem repeats analysis

We elaborated a comprehensive study of the distribution of tandem repeats in great apes (macaque, orangutan, chimpanzee and human) with the aim to determine whether there is any correspondence between tandem repeats and the location of evolutionary breakpoint regions in these species.

Distribution of tandem repeats. Using the eTandem algorithm, we detected a total of 758,206 tandem repeats in the human genome, grouped into 242,539 different motif types with a repeat unit size ranging from 2bp to 100bp. Similar values were found in macaque, orangutan and chimpanzee: (i) 714,458 tandem repeats representing 229,023 motif types in chimpanzee, (ii) 697,824 tandem repeats grouped into 230,650 motif types in orangutan and (iii) 733,524 tandem repeats corresponding to 211,199 motif types in rhesus macaque. These data suggest that the overall content of tandem repeats in terms of number of tandem repeats is conserved during primate genome evolution. When studied more in detail, we found that the most representative and therefore more frequent motifs were the same in the genomes of all four primate species: CA, AT, AAAT, TC, CAAA and AAAG. These AT-rich tandem repeats accounted for approximately 30% of the whole tandem repeat content in these genomes.

Subsequently, we analyzed the density of tandem repeats in each primate chromosome in order to compare the distribution of tandem repeats among species (File S1). In the human genome, the overall density of tandem repeats varied from 11,682 bp/Mbp in chromosome 14 to 33,091 bp/Mbp in chromosome 19 (File S1). The same pattern was observed in each primate homologous chromosomes. In chimpanzee, the density ranged from 17,253 bp/Mbp (chromosome 14) to 36,446 bp/Mbp (chromosome 19). This pattern is also conserved in orangutan in the homologous chromosome 14, and even in rhesus macaque, where the chromosome 7, homologous to human chromosome 14, has the lowest density (20,460 bp/Mbp).

Since we observed different tandem repeat density and an uneven distribution of EBRs among primate chromosomes, we decided to analyze thoroughly the tandem repeats landscape of each primate chromosome considering their evolutionary history: which chromosomal form was maintained collinear or suffered any reorganization since their common hominoid ancestor according to previous reports [1,31]. We scrutinized each chromosome's complete sequence using moving non-overlapping windows of 0.1 Mb in order to analyze the distribution of tandem repeats in each of the primate genomes, using a Kolmogorov-Smirnov test (Fig. 4 and File S2). Those chromosomes that suffered the same evolutionary process seem to have the same tandem repeats distribution while those with different evolutionary history have a statistically different tandem repeats landscape. The tandem repeat distribution of five (PPY6, PTR10, PPY12, HSA18, and PTRX) out of 69 chromosomes analyzed did not correlate with their evolutionary history, suggesting that additional elements are influencing the dynamics of tandem repeats. Herein are the results of the comparison of tandem repeat distributions along each primate chromosome:

Chromosome 1. Human chromosome 1 is considered to be the derived form, showing a pericentric inversion when compared to chimpanzee and orangutan chromosome 1. The human tandem repeat landscape also differs from the other two great apes (HSA vs PTR: p-value = 0.006; HSA vs PPY: p-value = 0.000).

Chromosome 2. It is well known that human chromosome 2 derives from the ancestral form by a fusion of two hominoid homolog chromosomes [1]. The ancestral 2a form corresponds to HSA2pq and also has suffered a pericentric inversion in the human form, whereas the ancestral 2b form has not suffered further reorganizations. The tandem repeat contour is different between human and the other great apes regarding chromosome 2a form (HSA vs PTR: p-value = 0.000; HSA vs PPY: p-value = 0.000) but is maintained in the homologous chromosome 2b form (HSA vs PTR: p-value = 0.738; HSA vs PPY: p-value = 0.192).

Chromosome 3. Human and chimpanzee chromosomes are the derived forms, with an inverted region compared to orangutan chromosome. The tandem repeats distribution confirms this pattern (HSA vs PTR: p-value = 0.062; HSA vs PPY: p-value = 0.009).

Chromosome 4. All the great apes have a derivative chromosome 4 that evolved differently since their common ancestor. We found a different tandem repeats distribution between human and chimpanzee forms but the same

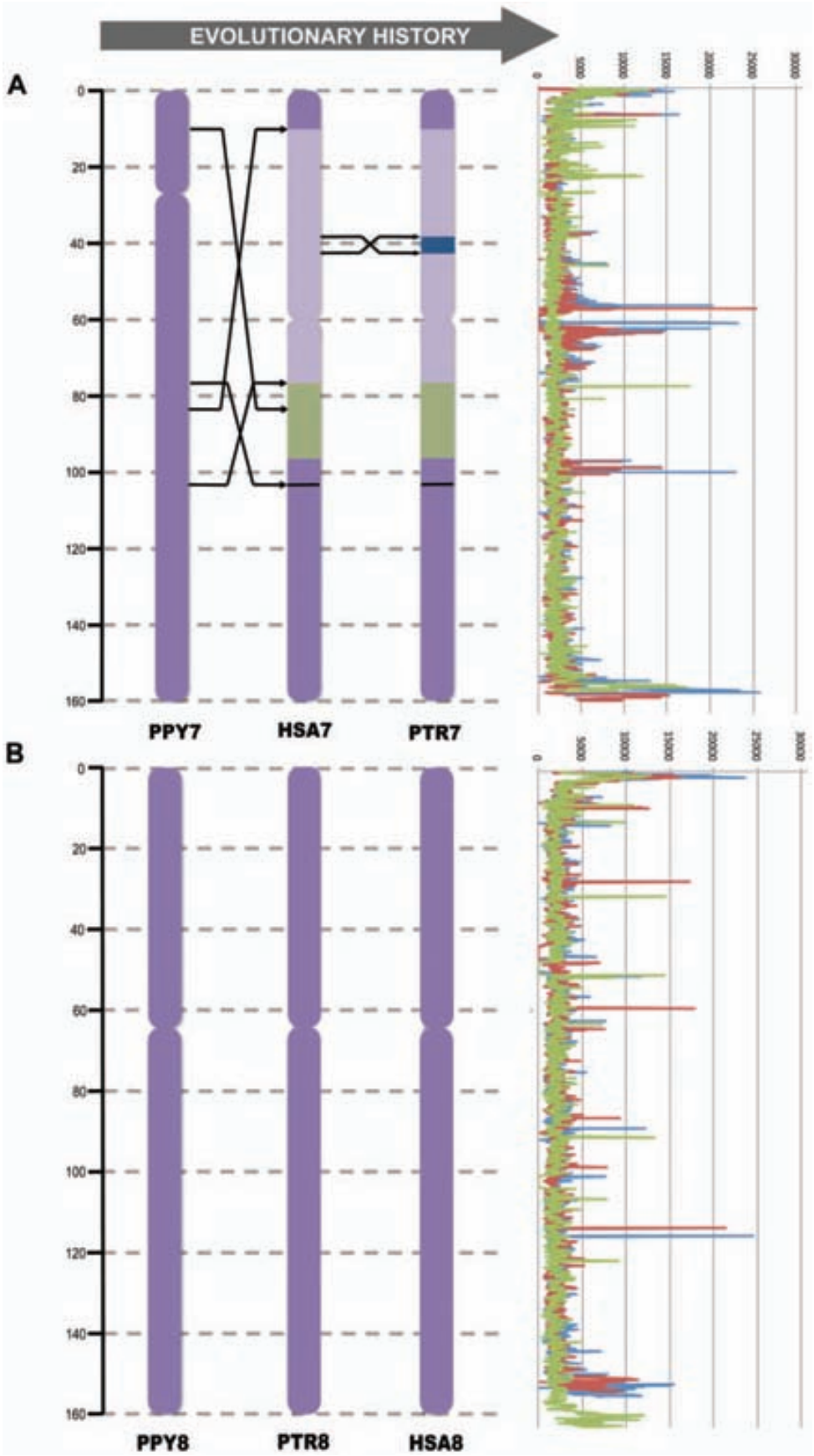

EVOLUTIONARY HISTORY
A
0
20
40
60
80
100
120
140
160
PPY7
HSA7
PTR7
B
0
20
40
60
80
100
120
140
160
PPY8
PTR8
HSA8

Figure 4. Tandem repeat content (bp) in human chromosomes 8 and 7 and its homologous in chimpanzee, orangutan and macaque. The image represents an example of a reorganized chromosome (a) and a collinear chromosome (b). In each case, the left panel shows the evolutionary history of each chromosome during hominoid evolution. The right panel shows the tandem repeat content in 100 kb windows in human (blue), chimpanzee (red), orangutan (green) and macaque (purple) genomes. Abbreviations –PPY: *Pongo pygmaeus*, PTR: *Pan troglodytes*, HSA: *Homo sapiens*.
doi:10.1371/journal.pone.0027239.g004

distribution between human and orangutan forms (HSA vs PTR: p-value = 0.022; HSA vs PPY: p-value = 0.272).

Chromosome 5. Human chromosome is considered the ancestral form, whereas the chimpanzee and the orangutan have derived forms due to pericentric inversions. The tandem repeats landscape is consistent with this pattern (HSA vs PTR: p-value = 0.031; HSA vs PPY: p-value = 0.001).

Chromosome 6. The three species shared the same chromosome form, which is considered to be ancestral. We found the same tandem repeat profile between human and chimpanzee (HSA vs PTR: p-value = 0.069) but it differs between human and orangutan (HSA vs PPY: p-value = 0.003).

Chromosome 7. The orangutan chromosome represents the ancestral form, while human and chimpanzee share a pericentric inversion. We found the same tandem repeats pattern in human and chimpanzee (HSA vs PTR: p-value = 0.203) but this was different in orangutan (HSA vs PPY: p-value = 0.050) (Fig. 4a).

Chromosome 8. The three hominoid species share the same form but we detected an insertion of ~3Mb in the orangutan chromosome 8 (Table 2). This difference is reflected in the tandem repeats landscape, being equal between human and chimpanzee (p-value = 0.128) but different in orangutan (p-value = 0.009) (Fig. 4b).

Chromosome 9. All three species have different chromosomal forms, being the orangutan chromosome the ancestral one. Tandem repeats distribution is consistent with these differences (HSA vs PTR: p-value = 0.002; HSA vs PPY: p-value = 0.000).

Chromosome 10. Orangutan chromosome 10 is considered to be the ancestral form, which differs from human and chimpanzee forms by a paracentric inversion. We found that human and orangutan have a different tandem repeat pattern (p-value = 0.001) as well as human and chimpanzee (p-value = 0.010), although the same pattern between these two species was expected.

Chromosome 11. The ancestral chromosome form is conserved in orangutan, which differs from the human chromosome by a pericentric inversion and from chimpanzee by a pericentric inversion and an insertion of ~400 Kb (Table 2). These differences are also reflected in the tandem repeat distribution (HSA vs PTR: p-value = 0.016; HSA vs PPY: p-value = 0.000).

Chromosome 12. Human and orangutan share the same form, which is considered the ancestral. Chimpanzee differs from them by a pericentric inversion. In this case, the tandem repeats landscape is different between human and chimpanzee (p-value = 0.050) and between human and orangutan (p-value = 0.004).

Chromosome 13. Human and chimpanzee share the same form and have the same tandem repeats pattern (p-value = 0.072), while orangutan have a ~100Kb insertion (Table 2) and shows a different tandem repeats pattern (p-value = 0.003).

Chromosome 14. All great apes share the same chromosome form and also the same tandem repeats landscape (HSA vs PTR: p-value = 0.051; HSA vs PPY: p-value = 0.051).

Chromosome 15. All great apes have different chromosome forms and different tandem repeats profile (HSA vs PTR: p-value = 0.004; HSA vs PPY: p-value = 0.001).

Chromosome 16. All great apes have different chromosome forms and different tandem repeats profile (HSA vs PTR: p-value = 0.001; HSA vs PPY: p-value = 0.000).

Chromosome 17. Human and orangutan share the same ancestral form, while chimpanzee suffered a pericentric inversion. This pattern is in agreement with the tandem repeats distribution (HSA vs PTR: p-value = 0.030; HSA vs PPY: p-value = 0.106).

Chromosome 18. Chimpanzee and orangutan share a chromosome form ancestral to great apes, which differs from the human by a pericentric inversion. This is not observed in the tandem repeats profile, given that all the species share the same distribution (HSA vs PTR: p-value = 0.095; HSA vs PPY: p-value = 0.206).

Chromosome 19, 20, 21 and 22. All great apes share the same chromosome form and also the same tandem repeats landscape [HSA19 (PTR: p-value = 0.127; PPY: p-value = 0.161) HSA20 (PTR: p-value = 0.138; PPY: p-value = 0.051) HSA21 (PTR: p-value = 0.106; PPY: p-value = 0.111) HSA22 (PTR: p-value = 0.082; PPY: p-value = 0.051)].

Chromosome X. Human and chimpanzee share the same ancestral form while orangutan has a ~2Mb insertion (Table 2). Tandem repeat pattern is in agreement with human-orangutan evolution (p-value = 0.021) but not with human-chimpanzee history (p-value = 0.000).

Tandem repeats are accumulated in evolutionary breakpoint regions. Once we studied the distribution of tandem repeats across whole genomes, we analyzed whether tandem repeats were differentially accumulated in EBRs and/or HSBs and if this pattern was conserved among species. In all cases, we analyzed two parameters: (i) number of tandem repeat loci, and (ii) number of base pairs implicated in tandem repeats. By this way we took into account not only the number of repeats but also the density of tandem repeats in each genomic region.

We observed 189,330 tandem repeat loci in EBRs and 360,314 loci in HSBs in the human genome. Assuming a homogeneous distribution of tandem repeat loci in these genomic regions, we expected 183,213 and 366,431 tandem repeat loci in EBRs and HSBs, respectively, showing that the observed tandem repeat loci are significantly deviated (p-value < 0.001). Mirroring these results, we also detected that EBRs contained significantly more base pairs implicated in tandem repeats than HSBs in the human genome (contingency analysis, p-value < 0.001).

Therefore, and to have a general overview of the genomic landscape, we used the EBRs and HSBs defined in the human genome to analyze whether there was any specific repeat accumulated in each different genomic region by means of contingency analysis. Out of the 242,539 different motif types found in the human genome, no specific repeat motif was exclusively present in EBRs or HSBs. However, 17 different microsatellite motifs were significantly accumulated in EBRs (p-values ≤ 0.0016) (Table 3). Although we did not detect any pattern regarding the repeat motif and the GC content in the whole tandem repeat content, we found five microsatellites (AAAT, TTTG, TTTC, TATTT and ATTTTT) present in a extremely high frequency in the human genome (more than 1000 repeat units and AT content $\geq$ 80%) (Table 3). Of these overrepresented tandem repeats, the AAAT motif was by far the

Table 3. Microsatellite motifs significantly accumulated in EBRs.

Motif	EBRs		HSBs		p-value
	observed	expected	observed	expected	
aaat*	8186	7373	14336	15148	3.05 E-21
tttg*	3930	3739	7492	7682	0.0018
tttc*	3187	2996	5966	6156	0.0005
tattt*	2488	2328	4624	4783	0.0009
attttt*	1635	1513	2987	3108	0.0017
agg	1011	880	1678	1808	0.000011
agaggg	231	186	339	383	0.0012
ggggga	156	120	212	247	0.0012
tggggg	100	69	111	141	0.0002
cccagc	75	44	61	91	0.000005
gccggg	66	43	68	90	0.0008
ccggc	36	16	15	34	0.000002
ggcagg	36	20	27	42	0.0007
actg	34	19	25	39	0.0008
ggggat	24	11	11	23	0.0002
ctgacc	23	9	5	18	0.000005
ggctct	13	5	4	11	0.0016

Asterisks indicate the overrepresented motifs (more than 1000 repeat units detected, see text for details).

doi:10.1371/journal.pone.0027239.t003

most frequent among all EBRs. We, then, analyzed if the distribution of the AAAT motif was dependant on the type of EBRs and we observed an accumulation of this motif in the EBRs not related to primates (p-value < 0.001).

Given the similarity of these microsatellites rich in AT content to the standard L1 cleavage site for classical retrotransposition (5′-TTAAA-3′, [33]), we examined a possible association with any L1 and/or *Alu* sequences in the human genome. In doing so, we considered five possible scenarios: (i) the repeat is not contiguous to any transposable element, (ii) the repeat is upstream or (iii) downstream of the transposable element, (iv) the motif is in-between two transposable elements or (v) two repeat motifs surround one transposable element. Notably, we observed that the AAAT motif was the only repeat significantly associated with *Alu* elements but only when it is located upstream of the transposable element (TE) in EBRs ($\chi^2 = 9.33$, p-value = 0.002) but not in HSBs ($\chi^2 = 1.99$, p-value = 0.07). Moreover, we found more AAAT motifs associated with *Alu* repeats in primate-specific EBRs than in the other types of EBRs (p-value < 0.001). Regarding the other over-represented repeats (Table 3), none of them was significantly associated with TE elements when EBRs and HSBs were compared (data not-shown). In order to understand the observed association, we analyzed if the distribution of *Alu* sequences was dependant on the type of EBRs (i.e. EBRs primate-specific) given that it is well known that there was a burst of *Alu* transposition in the lineages leading to primates around ~40 mya [34]. Out of the 1,212,896 *Alu* repeats found in the human genome, 281,019 were located in EBRs. This value represents almost half of the expected number of *Alu* loci assuming a random distribution and shows a depletion of *Alu* sequences in these EBRs (p-value < 0.001). However, when we focused only on the primate-specific EBRs, we found a significant accumulation of *Alu* sequences in these regions (p-value < 0.001). Therefore, our observations indicate that primate-specific EBRs are enriched in *Alu* repeats, but depleted in AAAT motifs when compared to other types of EBRs, although the AAAT motifs found in primate-specific EBRs are significantly associated with *Alu* sequences.

Discussion

Homologous synteny and evolutionary breakpoint regions in mammalian genomes

Since the initial whole-genome analysis performed by Murphy and collaborators [8], several studies have described those evolutionary genomic regions involved in the reshuffling of mammalian genomes [6,7,10,11,35]. Although the focus of these studies was the precise delimitation of the evolutionary breakpoints, the results published to date are far from being consistent. Discrepancies are probably due to differences in the versions of the genomes and the source of the data analyzed (e.g., radiation hybrid maps or whole-genome DNA sequences), differences in the level of resolution of the technique applied, and because the sets of species examined were only partially overlapping. By analyzing the whole-genome sequences of 10 vertebrate species (chimpanzee, rhesus macaque, orangutan, mouse, rat, cow, dog, horse, opossum and chicken) we identified 1,353 vertebrate HSBs (Table 1). This number of homologous synteny blocks is very similar to the previous studies [9,11], reflecting the high degree of conservation among mammalian genomes. However, we identified substantially fewer EBRs in the human genome (n = 898; median size = 304 Kb), than previously published [11] probably due to the conservativeness of our approach. Since we excluded centromeric, telomeric and gap regions in our analysis in order to avoid low coverage regions and, therefore, false positives, EBRs and HSBs, represented 67.38% of the human genome. Importantly, when analyzing the distribution of EBRs along the human genome relative to the position in each chromosome we observed a non-homogenous distribution of EBRs among chromosomes (Fig. 2). Specifically, human chromosomes 13 and 14 accumulated fewer and chromosome 19 accumulated more EBR/Mbp than expected. The same pattern was observed in great apes and macaque. Using the non-human primate genomes as a reference we found 32 and 46 EBRs in chimpanzee and orangutan genomes, respectively, non-homogeneously distributed along chromosomes. These results confirm the existence of "hot spot" regions for chromosome evolution supporting the fragile breakage model of chromosome evolution [5–11].

Based on chromosomal painting studies, Froenicke [4] established the average rate of chromosomal exchange in eutherian mammals to be 0.19 rearrangements/myr and 0.39 EBR/myr. Combining the data derived from the comparison of 10 mammalian species we estimated a similar rate of evolution (0.35 EBRs/myr).We found that 20.3% of the 898 EBRs detected have been reused during the eutherian evolution. This proportion is higher than the 7-8% described in previous studies [9,11] but in agreement with initial studies [8]. What it is clear is that a fraction of the mammalian genomes (ranging from 20% to 7%) has suffered recurrent chromosome reorganizations during evolution. We also placed the EBRs detected in an evolutionary context; as an example, we detected 180 EBRs in the lineage leading to the opossum. Since its divergence from the common therian ancestor ~180 mya, the marsupial species has accumulated a rate of 1 EBR/myr. In placental mammals, the two rodent species studied (mouse and rat) accumulated more clade-specific EBRs (80) than other clades, with a rate of 1 EBR/myr, showing a high rate of EBRs, as previously described in the literature [36]. Primates, on

the other hand, show a wide range of rearrangement rates with chimpanzee showing the highest rate of genomic reorganization (1.83 EBRs/myr).

Moreover, and considering the evolutionary history of each hominoid chromosome [1,31] we were able to refine the rearrangements that occurred during genome evolution in great apes. Among chromosomes that have been conserved since their common ancestor, we found new insertions in orangutan chromosomes 13, 19 and X and an inversion in chimpanzee chromosome 19. In addition, we defined more rearrangements in the reorganized chromosomes. For instance, we found one insertion in chimpanzee chromosome 4, 9 and 11. Even though the great apes genomes are highly conserved, when their sequences are analyzed more in detail, these rearrangements show that they are organized as conserved blocks that had suffered additional reshuffling.

The distribution of EBRs across chromosomes, the high reuse degree of EBRs and the reconstruction of the likely chromosomal architecture of ancestral mammalian genomes have revealed that evolutionary breakpoints are clustered in regions that are prone to disruption, promoting the subsequent reorganization of chromosomes [37,38]. However, one question remains open: Is there any sequence composition and/or genome organization accounting for the distribution of evolutionary regions? To shed light on this pivotal issue, we have characterized the tandem repeats in the evolutionary regions detected.

Tandem repeats distribution and its evolutionary implications

We were able to elaborate a tandem repeat database distributed into five different regions (telomeres, centromeres, HSBs, EBRs and gaps) along the genomes of great apes and the macaque. We detected that the overall content of tandem repeats were similar in these closely related species (758,206 tandem repeats in human, 714,458 tandem repeats in chimpanzee, 697,824 in orangutan and 733,524 in the macaque). Moreover, out of the total content of tandem repeats, we observed that six tandem repeat motifs (CA, AT, AAAT, TC, CAAA and AAAG) were highly represented in the primate genomes. The presence of the same six microsatellites in the primate species is somehow surprising despite their common ancestor because microsatellites are highly mutable (in humans: 10^{-4} mutations per locus per generation, [39]). However, this conservation is coherent with the microsatellite turnover theory (i.e. cycles of expansions/deletions and stabilization/reactivation) and suggests that microsatellites fluctuate as a whole [40].

Once we studied the overall content of tandem repeats in the primate genomes, we focused on the distribution of tandem repeats in each chromosome. We observed that not all the chromosomes have the same tandem repeat density (bp implicated in repeats/Mbp of genome) (File S1). The human chromosome 14 and its homologous in the non-human primate species had the lowest density while the human chromosome 19 and its homologous had the highest tandem repeat density. These differences among chromosomes could be due to several factors, such as (i) random amplification and appearance of new repeats, (ii) some selective pressure that restricts the spread of the repeats or, (iii) artifacts of the sequencing procedure itself. Since we have analyzed the tandem repeats distribution in all great ape chromosomes and found the same overall content of tandem repeats, we discard both random amplification and biases in the sequencing procedure. To further analyze these differences, we used sliding windows of 100kb to compare the distribution of tandem repeats in each chromosome of the primate species (Fig. 4 and File S2). We found a non-homogeneous distribution of tandem repeats, with a high accumulation in the pericentromeric and telomeric regions, mirroring previous results [10]. But, more importantly, we found differences in tandem repeat distributions among species, suggesting that they might be correlated with the evolutionary history of each primate chromosome. Roughly, our qualitative comparisons of chromosome evolution suggest that the tandem repeats landscape might have been conserved in collinear chromosomes, but altered in those reorganized chromosomes (Fig. 4 and File S2). Further analysis will be necessary in order to corroborate this hypothesis.

The analysis of the human genome revealed specific features not found in the other primate species analyzed. Excluding regions of high complexity from our analyses (telomeres, centromeres and gaps) EBRs in the human genome accumulated more tandem repeat base-pairs than HSBs ($p \leq 0.05$ and $p \leq 0.001$). This result confirms previous observations [10] indicating that tandem repeats are elements that could promote genome reorganization during the evolutionary process. With the aim to investigate whether there was any particular DNA configuration or composition driving genome instability we analyzed more in detail the distribution of tandem repeats across the human genome. Although no specific repeat motif was exclusively present in EBRs or HSBs, 17 different microsatellites motifs were significantly accumulated in EBRs. Notably, out of these overrepresented tandem repeats, the AAAT was the most frequently detected. It has been described that this motif could form single-stranded coils [24], favoring chromatin instability and increasing the likelihood to break.

Additionally, the observed association of some tandem repeats with L1 and *Alu* elements provides indications for the possible role of transposable elements in shaping the distribution of mammalian large-scale chromosomal changes. Transposable elements, such as *Alu* and LINEs, are well known to induce genomic reorganizations and structural variation through multiple pathways, including unequal homologous recombination and alternative transposition, for instance [20,41–44,]. Although the association between microsatellites and transposable elements has been previously reported [45,46], the origin of this association remains unclear. The association of AAAT and transposable elements found in our study can be explained by (at least) two non-mutually exclusive hypotheses. One possibility is that the presence of the AAAT microsatellite in certain regions could derive from transposable elements already inserted in the genome. This interpretation is plausible, since both L1 and *Alu* are characterized by a 3′ poly(dA)-rich tail and an internal tandem repeat region [47]. Furthermore, Abrusan and Krambeck [48] have described that these transposable elements are enriched in AT-rich regions in the human genome. Alternatively, AAAT repeats could represent likely target regions for L1 and/or *Alu* insertions since this motif closely resembles the canonical cleavage sites for these elements (5′-TTAAA-3′) [33]. Cleavage on both strands is required, resulting in an intermediate equivalent to double-strand breaks (DSBs) at the early stage of the reverse transcription reaction. Gasior and collaborators [49] demonstrated an excess of DSBs in the L1 transposition process. The reasons for this high quantity of DSBs are unknown. Although the host cell would successfully repair most of the DSBs created, a fraction of these can be misrepaired, and eventually induce chromosomal alterations [49].

As a preliminary survey to favor one of these two hypotheses, we analyzed the distribution of *Alu* sequences and AAAT motifs in the different types of EBRs. We observed an accumulation of *Alu* repeats and depletion in number of AAAT motifs in primate-specific EBRs when compared with the other types of EBRs. Considering the massive transposition of *Alu* sequences that

occurred in the lineages leading to primates around ~40 mya [34] a very appealing scenario is to consider the AAAT motif as the target site of insertion of *Alu* sequences. Under such scenario, the enrichment of AAAT-*Alu* association observed in primate-specific EBRs could represent signatures of ancient insertions. Favoring this hypothesis, it has been previously shown that *Alu* density strongly correlates with L1 target site insertion motif and regions more prone to DSBs formation [50]. At this point, however, a detailed pair-wise comparison between closely related species in which the ancestral state of a novel insertion can be identified would allow us to distinguish if the accumulation of AAAT motifs is the cause or consequence of L1 and *Alu* insertions.

Summarizing, our results provide evidences for the role of both tandem repeats and transposable elements in evolution. A plausible hypothesis is to consider that an accumulation of tandem repeats in certain genomic regions might form secondary structures in the DNA and, therefore, promotes genome instability that could lead to evolutionary chromosomal changes. Moreover, certain tandem repeats (i.e. AAAT) could work as target sites, promoting the insertion of transposable elements and, eventually, leading to genomic reorganizations by non-allelic homologous recombination (NAHR) [43]. Previous studies have reported how breakpoint regions are rich in segmental duplications [15–17,51,52], high repeat content [10,14], transposable elements [19,20,53,54] or long regulatory regions. This heterogeneity in results is suggesting that additional elements, not only the DNA sequence per se, are affecting breakage susceptibility. Recent data indicate that the permissiveness of some regions of the genome to undergo chromosomal breakage could be determined by changes in chromatin conformation [19,32]. In this sense, transposable elements have been reported to be associated with the epigenetic status of the genome and regulation of gene expression [55,56], but also the length and type of tandem repeats can determine the conformation of the chromatin [57]. Although at this point a cause/effect between tandem repeats and genomic instability cannot be determined, we can anticipate, as a working hypothesis, that certain properties of local DNA sequences such as repetitive elements related to open chromatin configurations can be involved in the origin/resolution of chromosomal reorganizations.

Materials and Methods

Definition of evolutionary breakpoint regions

We included in our analysis the whole-genome sequences of 10 vertebrate species available in the public databases (Ensembl [58]). These species were chosen based on the availability of their completed whole-genome sequences and they included: *Pan troglodytes* (CHIMP2.1, assembly of March, 2006), *Macaca mulatta* (Mmul_1, assembly of February, 2006), *Pongo pygmaeus* (PPYG2, assembly of April, 2007), *Mus musculus* (NCBIm37, assembly of April, 2007), *Rattus norvegicus* (RGSC 3.4, assembly of December, 2004), *Bos taurus* (Btau_4.0, assembly of October, 2007), *Canis familiaris* (CanFam 2.0, assembly of May, 2005), *Equus caballus* (EquCab2, assembly of September, 2007), *Monodelphis domestica* (MonDom5, assembly of October, 2006) and *Gallus gallus* (WASHUC2, assembly of May, 2006). In addition, we used the human genome (NBCI build 36, assembly of March, 2006) as a reference.

We first defined the homologous synteny blocks (HSBs) and the evolutionary breakpoint regions (EBRs) in the human genome. To do so, we downloaded the pair-wise whole-genome comparisons detailed in the Ensembl genome browser (release 52) between the human reference genome and those of other vertebrate species (chimpanzee, orangutan, macaque, mouse, rat, horse, cattle, dog, opossum and chicken). These pair-wise comparisons were based on sequence homology. For each pair-wise comparison between the human genome and any of the vertebrate species, we established homologous syntenic regions, defining the start and end positions according the Ensembl database (in pb) (Fig. 1). Then, we manually grouped together those syntenic regions spaced less than 4 Mbp, with the same orientation and located in the same chromosome to form a single HSB. In order to avoid possible artifacts derived from the low-coverage annotation of whole-genome sequences, intervals between two contiguous HSBs larger than 4 Mbp in size were considered to be "gaps". Subsequently, we merged the coordinates of all pair-wise comparisons in the human genome by means of Perl scripts in order to define the total number, position and length of HSBs and EBRs in reference genome (Fig. 1). EBRs were considered as the interval between two contiguous HSBs as described in Ruiz-Herrera and collaborators [10] and were defined by sequence coordinates in any of the nine mammalian species compared with human plus the chicken (Fig. 1). We calculated the percentage of coverage of each pair-wise comparison using the human genome total length (data in Mbp) excluding the Y chromosome from the analysis. Furthermore, we labeled as telomeric/subtelomeric the 2 Mbp at the ends of each human chromosome and as centromeric/pericentromeric the 2 Mbp regions flanking the unknown nucleotides (Ns) as described elsewhere [10]. Thus, the human reference genome was classified into 5 types of genomic regions (telomeres, centromeres, HSBs, EBRs and gaps) in order to proceed with the subsequent analysis.

The definition of homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) for the primate genomes (chimpanzee, orangutan and macaque) was done following the same approach as with the human genome. In all cases, we wrote Perl scripts to parse the pair-wise comparisons data and to cross-reference the coordinates of all types of genomic regions.

Phylogenetic interpretation of evolutionary breakpoint regions

Extant mammals are classified into three major groups: monotremes, marsupials and placental mammals (eutherians) that split off from their last shared common ancestor nearly 240 mya [59]. Among placental mammals, four superordinal clades (Afrotheria, Xenarthra, Laurasiatheria and Euarchontoglires/Supraprimates) are recognized based on the phylogenetic analysis of both nuclear and mitochondrial DNA [60]. We followed the phylogeny proposed by Murphy et al. [30] for our phylogenetic interpretations. We classified all EBRs into different types depending on which species they have occurred in: i) species-specific, ii) clade-specific, when the EBR is found in species of the same order or superorder and iii) reused, if the EBR is found in two taxa but not in their common ancestor. We used the maximum parsimony criterion to place events in the tree and obtained the rate estimates in each branch.

Tandem repeat analysis

We analyzed the distribution of tandem repeats in the human genome using the eTandem algorithm (part of EMBOSS 6.0.1 package [61]). We run the eTandem algorithm with a minimum repeat unit of 2 bp and a maximum repeat unit of 100 bp. The resulting output files were computed for the detection of overlapping tandem repeats and the canonical motif was reported for each repeat (a canonical motif is intended as all possible rotations and reverse complementation; e.g., AC is the canonical form of AC, CA, GT, and TG). We merged the positions of all the

canonical motifs detected with those of the different types of genomic regions described in the previous section.

For the analysis of human retroelements, we obtained the genomic positions of all the human L1 and *Alu* sequences described in the UCSC database (http://genome.ucsc.edu) in order to analyze if tandem repeats were immediately contiguous to any transposable element sequence. First, we analyzed if a given motif was associated with L1 or *Alu* sequences within EBRs or HSBs designing a Perl script to compare the positions of the tandem repeats and the transposable elements. By means of a χ^2 test we evaluated this association, using the total number of tandem repeats contiguous to L1 and/or *Alu* sequences as the sample and the position in EBRs or HSBs as the factor. In order to analyze the distribution of *Alu* repeats in the different type of EBRs, we applied a χ^2 test and we calculated the expected *Alu* loci in each genomic region assuming a homogeneous distribution.

Using Perl scripts, we computed the overlapping degree of tandem repeats, searched the canonical motifs, and merged the positions of tandem repeats with the different types of genomic regions the human genome was classified and with the transposable elements.

Statistical analyses

We performed the statistical analyses using the JMP 7 package. Centromeric, telomeric and gap regions were excluded before any statistical analyses were performed given that they represent regions of high complexity overall.

To assess if EBRs were evenly distributed across human chromosomes we estimated an average frequency of 0.3 EBR/Mbp assuming a homogeneous distribution of EBRs across the human genome. We used a χ^2 test with a Bonferroni correction (p-value $= 0.0022$) to evaluate any possible deviation.

For the analysis of tandem repeat distribution, we first compared whether EBRs accumulate more base-pairs involved in tandem repeats than HSBs using a χ^2 test. We also analyzed the tandem repeat loci in EBRs and HSBs using the same test. We computed the expected number of tandem repeats in each region by assuming a homogeneous distribution of the total tandem repeat loci along the genome and then distributed them proportionally to the length of each genomic region (EBRs or HSBs). Then, we used a χ^2 test with the Bonferroni correction to assess whether a tandem repeat motif accumulates significantly in a certain type of genomic region (p-value = 0.0017).

Finally, to compare the tandem repeat distribution along primate chromosomes we counted the base-pairs of tandem repeats in 100 kb windows for each chromosome. In order to analyze whether the primate genomes had the same tandem repeat landscape, we performed Kolmogorov-Smirnov tests by pairs, comparing all hominoids' chromosomes. P-values smaller than 0.005 indicated that the distribution of base-pairs implicated in tandem repeats were significantly different among species.

Supporting Information

File S1 Density of tandem repeats in each primate chromosomes. The density is expressed in base-pairs (bp) of a tandem repeat sequence per megabase-pairs (Mbp) of a chromosome sequence.
(PDF)

File S2 Tandem repeat content (bp) in non-overlapping 100 kb windows. For each chromosome, the tandem repeat distribution for human (black), chimpanzee (dark green), orangutan (light green) and macaque (orange) is shown. In each case, the Spearman's p test comparing chimpanzee (PTR), orangutan (PPY) and macaque (MMU) with human (HSA) is indicated. C: centromere, N: distal telomere. * Statistically significant p-value <0.0001.
(PDF)

Acknowledgments

C. Gilbert and W. Sanseverino are acknowledged for comments on the draft of this manuscript. We are also thankful to the editor and the anonymous reviewers whose comments and suggestions improved the manuscript.

Author Contributions

Conceived and designed the experiments: ARH MB MF. Performed the experiments: MF. Analyzed the data: MF FLG MB ARH. Contributed reagents/materials/analysis tools: MP. Wrote the paper: ARH MF. Designed the analysis: MF ARH. Contributed to the writing of the initial draft of the manuscript: MB FLG MP.

References

1. Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215: 1525–1530.
2. Clemente IC, Ponsa M, Garcia M, Egozcue J (1990) Evolution of the Simiiformes and the phylogeny of human chromosomes. Hum Genet 84: 493–506.
3. Wienberg J, Stanyon R (1997) Comparative painting of mammalian chromosomes. Curr Opin Genet Dev 7: 784–791.
4. Froenicke L (2005) Origins of primate chromosomes - as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. Cytogenet. Genome Res 108: 122–138.
5. Ruiz-Herrera A, Garcia F, Mora L, Egozcue J, Ponsa M, et al. (2005) Evolutionary conserved chromosomal segments in the human karyotype are bounded by unstable chromosome bands. Cytogenet. Genome Res 108: 161–174.
6. Bourque G, Pevzner PA, Tesler G (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res 14: 507–516.
7. Zhao S, Shetty J, Hou L, Delcher A, Zhu B, et al. (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14: 1851–1860.
8. Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, et al. (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309: 613–617.
9. Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, et al. (2006) Reconstructing contiguous regions of an ancestral genome. Genome Res 16: 1557–1565.
10. Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol 7: R115.
11. Larkin DM, Pape G, Donthu R, Auvil L, Welge M, et al. (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19: 770–777.
12. Pevzner P, Tesler G (2003) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci U S A 100: 7672–7677.
13. Sankoff D (2009) The where and wherefore of evolutionary breakpoints. J Biol 8: 66.
14. Kehrer-Sawatzki H, Szamalek JM, Tanzer S, Platzer M, Hameister H (2005) Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. Genomics 85: 542–550.
15. Bailey JA, Eichler EE (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet 7: 552–564.
16. Kehrer-Sawatzki H, Cooper DN (2008) Molecular mechanisms of chromosomal rearrangement during primate evolution. Chromosome Res 16: 41–56.
17. Carbone L, Vessere GM, ten Hallers BF, Zhu B, Osoegawa K, et al. (2006) A high-resolution map of synteny disruptions in gibbon and human genomes. PLoS Genet 2: 223.
18. Caceres M, Ranz JM, Barbadilla A, Long M, Ruiz A (1999) Generation of a widespread Drosophila inversion by a transposable element. Science 285: 415–418.
19. Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, et al. (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. PLoS Genet 5: e1000538.
20. Longo MS, Carone DM, NISC Comparative Sequencing Program, Green ED, O'Neill MJ, et al. (2009) Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10: 334.

21. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.
22. Naslund K, Saetre P, von Salome J, Bergstrom TF, Jareborg N, et al. (2005) Genome-wide prediction of human VNTRs. Genomics 85: 24–35.
23. Armour JA (2006) Tandemly repeated DNA: why should anyone care? Mutat.Res 598: 6–14.
24. Bacolla A, Larson JE, Collins JR, Li J, Milosavljevic A, et al. (2008) Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome Res 18: 1545–1553.
25. Kolb J, Chuzhanova N, Hogel J, Vasquez KM, Cooper DN, et al. (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. Chromosome Res 14: 469–83.
26. Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, et al. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74: 127–134.
27. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271: 1423–1427.
28. Usdin K, Grabczyk E (2000) DNA repeat expansions and human disease. Cell Mol Life Sci 57: 914–931.
29. Nielsen R (2002) Mapping mutations on phylogenies. Syst.Biol 51: 729–739.
30. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, et al. (2001) Molecular phylogenetics and the origins of placental mammals. Nature 409: 614–618.
31. Muller S, Wienberg J (2001) "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. Hum Genet 109: 85–94.
32. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, et al. (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9: 585–598.
33. Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A 94: 1872–1877.
34. Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33: 311–320.
35. Lemaitre C, Zaghloul L, Sagot MF, Gautier C, Arneodo A, et al. (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 10: 335.
36. Graphodatsky AS, Yang F, Dobigny G, Romanenko SA, Biltueva LS, et al. (2008) Tracking genome organization in rodents by Zoo-FISH. Chromosome Res 16: 261–274.
37. Froenicke L, Caldes MG, Graphodatsky A, Muller S, Lyons LA, et al. (2006) Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res 16: 306–310.
38. Robinson TJ, Ruiz-Herrera A (2008) Defining the ancestral eutherian karyotype: a cladistic interpretation of chromosome painting and genome sequence assembly data. Chromosome Res 16: 1133–1141.
39. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5: 435–445.
40. Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. Bioessays 28: 1040–1050.
41. Gray YH (2000) It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet 16: 461–468.
42. Ostertag EM, Kazazian HH (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res 11: 2059–2065.
43. Lee J, Han K, Meyer TJ, Kim HS, Batzer MA (2008) Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PLoS One 3: Genome Res.
44. Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. Nat Rev Genet 10: 691–703.
45. Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Natl Acad Sci U S A 93: 6470–6475.
46. Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res 18: 30–38.
47. Babushok DV, Kazazian HH (2007) Progress in understanding the biology of the human mutagen LINE-1. Hum Mutat 28: 527–539.
48. Abrusan G, Krambeck HJ (2006) The distribution of L1 and Alu retroelements in relation to GC content on human sex chromosomes is consistent with the ectopic recombination model. J Mol Evol 63: 484–492.
49. Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357: 1383–1393.
50. Kvikstad EM, Makova KD (2010) The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. Genome Res 20: 600–613.
51. Goidts V, Szamalek JM, Hameister H, Kehrer-Sawatzki H (2004) Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. Hum Genet 115: 116–122.
52. Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. Genome Res 19: 934–942.
53. Bourque G (2009) Transposable elements in gene regulation and in the evolution of vertebrate genomes. Curr Opin Genet Dev 19: 607–612.
54. Delprat A, Negre B, Puig M, Ruiz A (2009) The transposon Galileo generates natural chromosomal inversions in Drosophila by ectopic recombination. PLoS One 4: e7883.
55. Hua-Van A, Le Rouzic A, Boutin TS, Filee J, Capy P (2011) The struggle for life of the genome's selfish architects. Biol Direct 6: 19.
56. Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8: 272–285.
57. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. Science 324: 1213–1216.
58. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, et al. (2007) Ensembl 2007. Nucleic Acids Res 35: D610–7.
59. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci U S A 100: 1056–1061.
60. Nikolaev S, Montoya-Burgos JI, Margulies EH, NISC Comparative Sequencing Program, Rougemont J, et al. (2007) Early history of mammals is elucidated with the ENCODE multiple species sequencing data. PLoS Genet 3: e2.
61. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
62. Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17: 1254–65.

TREBALL 3

Selection against Robertsonian fusions involving housekeeping genes in the House mouse: integrating data from gene expression arrays and chromosome evolution

Ruiz-Herrera A, Farré M, Ponsà M and Robinson TJ

Chromosome Research (2010); 18 :801-808

Índex d'impacte (2010): 3,130

ÀREA: Genetics & Heredity, QUARTIL 2

Chromosome Res (2010) 18:801–808
DOI 10.1007/s10577-010-9153-8

Selection against Robertsonian fusions involving housekeeping genes in the house mouse: integrating data from gene expression arrays and chromosome evolution

Aurora Ruiz-Herrera · Marta Farré · Montserrat Ponsà · Terence J. Robinson

Received: 11 June 2010 / Revised: 3 August 2010 / Accepted: 4 August 2010 / Published online: 2 September 2010

Abstract Monobrachial homology resulting from Robertsonian (Rb) fusions is thought to contribute to chromosomal speciation through underdominance. Given the karyotypic diversity characterizing wild house mouse populations [*Mus musculus domesticus*, (MMU)], variation that results almost exclusively from Rb fusions (diploid numbers range from 22 to 40) and possibly whole arm reciprocal translocations (WARTs), this organism represents an excellent model for testing hypotheses of chromosomal evolution. Previous studies of chromosome size and recombination rates have failed to explain the bias for certain chromosomes to be involved more frequently than others in these rearrangements. Here, we show that the pericentromeric region of one such chromosome, MMU19, which is infrequently encountered as a fusion partner in wild populations, is significantly enriched for housekeeping genes when compared to other chromosomes in the genome. These data suggest that there is selection against breakpoints in the pericentromeric region and provide new insights into factors that constrain chromosomal reorganizations in house mice. Given the anticipated increase in vertebrate whole genome sequences, the examination of gene content and expression profiles of the pericentromeric regions of other mammalian lineages characterized by Rb fusions (i.e., other rodents, bats, and bovids, among others) is both achievable and crucial to developing broadly applicable models of chromosome evolution.

Keywords mouse · Rb fusions · housekeeping genes · chromosome evolution · expression arrays

Abbreviations

MMU	Mus musculus domesticus
Rb	Robertsonian
HSKP	Housekeeping
DSBs	Double strand breaks
PEV	Position effect variegation

Responsible Editor: Herbert Macgregor.

A. Ruiz-Herrera (✉) · M. Farré · M. Ponsà
Departament de Biologia Cel.lular,
Fisiologia i Immunologia,
Universitat Autònoma de Barcelona,
Campus UAB,
08193 Cerdanyola del Vallès, Barcelona, Spain
e-mail: aurora.ruizherrera@uab.cat

A. Ruiz-Herrera
Institut de Biotecnologia i Biomedicina,
Campus UAB,
08193 Cerdanyola del Vallès, Barcelona, Spain

T. J. Robinson
Evolutionary Genomics Group,
Department of Botany and Zoology,
University of Stellenbosch,
Private Bag X1,
Matieland 7602, South Africa

Introduction

It has been argued that chromosomal reorganization may contribute to speciation due to underdominance associated with meiotic abnormalities in heterozygotes (White 1978). This is generally considered likely to occur in small, inbred populations, or when rearrangements are weakly underdominant individually but strongly underdominant in combination (King 1993). More recently, a number of related studies have proposed that chromosomal rearrangements can reduce gene flow and potentially contribute to speciation by the suppression of recombination (Rieseberg 2001). Under this hypothesis, it is thought that chromosomal rearrangements have a minimal influence on fitness, but by suppressing recombination, they contribute to a reduction of gene flow. Paradoxically, the spread and subsequent fixation of an underdominant rearrangement is problematic from a population genetic perspective, given that a chromosomal reorganization that disrupts gene flow is less likely to be fixed. However, there is general consensus that chromosomal speciation can result where multiple centric fusions [i.e., Robertsonian (Rb) fusions] show monobrachial homologies (i.e., one arm in common; see Baker and Bickham 1986). Hybrids resulting from crosses between individuals that show fixed monobrachial differences can result in complex chain or ring configurations at meiosis that impede normal segregation and may therefore lead to speciation.

The house mouse [*Mus musculus domesticus* (MMU)] represents an excellent model for testing hypotheses of chromosomal evolution. Although the standard karyotype of the mouse (*Mus musculus*, MMU) is characterized by $2n=40$ acrocentric chromosomes, a wide range of diploid numbers (from 22 to 40) have been detected in house mouse populations over the last 30 years—variation resulting almost exclusively from Rb fusions and/or WARTs (Gazave et al. 2003; Pialek et al. 2005). The most recent review of chromosomal variation in the house mouse recognizes a total of 97 different metacentric "populations" distributed across Europe and the Mediterranean basin (Pialek et al. 2005). These populations formed very recently (10,000 years, Britton-Davidian et al. 1989), making genetic and morphological differences among them small or almost inexistent. Of the 171 different possible metacentric combinations, only 106 have been described in wild populations (Gazave et al. 2003; Pialek et al. 2005), indicating that not all chromosomes contribute equally to the observed chromosomal variation.

One of the chromosomes less frequently involved in Rb fusions is MMU19 (Pialek et al. 2005). Only two different Rb fusions involving MMU19 have been described in wild populations (Rb2.19 and Rb11.19; Britton-Davidian et al. 2000). Although chromosome size and recombination rates have been mooted as possible causes of this bias (Nachman and Searle 1995; Qumsiyeh 1994; Gazave et al. 2003), support for these suggestions has not been forthcoming, suggesting that additional factors may be involved in the process. In an attempt to advance our understanding of this phenomenon, we have analyzed the gene content and expression profiles of mouse chromosomes (the autosomes and X chromosome), taking advantage of whole sequence genome and gene expression data available in the public domains. Our analyses show that the pericentromeric region of MMU19 is significantly enriched for housekeeping genes when compared to other chromosomes in the genome, suggesting the possibility that there is selection against breakpoint disruption in this region, hence, the chromosome's infrequent involvement in Rb fusions.

Material and methods

Positions of the reference sequence mouse genes (RefSeq) were obtained from the NCBIm37 assembly using the BioMart browser of Ensembl. Only genes with a known function were analyzed; novel genes with unknown function, pseudogenes and RNA genes were not included in the analysis. Gene expression data were obtained from the Gene Expression Atlas (Su et al. 2004) available through the BioGPS portal (Wu et al. 2009). This atlas represents a whole genome gene expression array targeting 36,182 mouse transcripts from 61 different tissues. Probe sets available on this platform were converted to official gene symbols, their positions calculated in the mouse reference sequence genome and in each mouse chromosome, and the median expression values of each calculated.

We used two different cutoffs in order to make the distinction between housekeeping (HSKP) and tissue-specific genes. The first consisted of defining the

minimal signal for each gene from probes on the chip (value ≥5); this value corresponds to half of the "median expression" (Vinogradov and Anatskaya 2007) in all 61 tissues analyzed. The second cutoff value applied was the median expression, as recommended by the authors of the Gene Expression Atlas (Su et al. 2004). Mouse genes were classified as HSKP if they showed significantly elevated expression levels (i.e., in excess of the cutoff above) in all tissues of the array. Mouse genes were classified as tissue-specific if the coefficient of variation (Cv) was >200 (Vinogradov and Anatskaya 2007). We grouped all genes with a known function in windows of 2 Mbp along each mouse chromosome in order to analyze the density of genes [number of genes per megabase (Mbp) of chromosome]. Our analyses showed that the distribution of genes within each 2 Mb window did not fit a normal distribution. Consequently, the Mann–Whitney test with the Bonferroni correction was applied to establish median comparisons using the chromosomes as a factor and the 2 Mb windows as the sample. We used a Chi-square (χ^2) test to compare the total chromosomal gene density and also the gene density in the pericentromeric regions of mouse chromosomes. We similarly compared the density of HSKP genes in the pericentromeric regions in the complete mouse complement by χ^2. A Fisher's test was implemented to compare the number of HSKP and tissue-specific genes in MMU19 and the remaining mouse chromosomes.

Results and discussion

The mouse genome [NCBIm37 mouse assembly (July 2007), Ensembl release 49] currently has 22,931 protein-coding genes; of these, 15,593 have a known function (Table 1). When analyzing the position of these genes, we noted an absence of annotated genes between 0 Mbp and 3 Mbp in each mouse chromosomes, reflecting the presence of highly repetitive centromeric sequences in these regions. In our analysis of the distribution of genes per megabase pairs (Mbp), we arbitrarily set the limits of the pericentromeric region as encompassing sequences spanning 0–16 Mbp for each mouse chromosome. This was done in order to ensure that regions close enough to the centromeric sequences were included for all the mouse chromosomes.

The distribution of total protein-coding genes reveals statistical differences among mouse chromosomes. Three chromosomes show a significant concentration (p=0.0001) of protein-coding genes in relation to their genomic size: MMU11 (10.4 genes/Mbp), MMU17 (8.2 genes/Mbp), and MMU19 (9.1 genes/Mbp). However, when the distribution of genes was analyzed for each mouse chromosome, the highest concentration of MMU19 genes fell in the pericentromeric region. We noted two distinct peaks, one within 3–5 Mbp and another one within 10–13 Mbp (Fig. 1a). This represents a pattern not detected in any of the other mouse chromosomes. Given the implication that this holds for the reshuffling of the mouse genome by Rb fusions, we extended our investigation to include the analysis of gene content and distribution to the remaining chromosomes. These data show unequivocally that MMU19 has accumulated significantly more genes in the pericentromeric region (defined here as the genomic region spanning 0–16 Mbp) than other chromosome in the genome (χ^2=32.89, DF=19, p=0.035; Table 1).

We similarly examined the distribution of HSKP and tissue-specific genes across the mouse genome in an attempt to explain the observed differences in gene accumulation among chromosomes. Genes can be considered as HSKP when they are constitutively expressed in all tissues/cells in order to maintain basic cellular functions (Butte et al. 2001). Clustering of HSKP genes has been reported in human and mouse genomes (Williams and Hurst 2002; Singer et al. 2005), but the implications that this holds for structural rearrangement of chromosomes have not been examined. We define an HSKP gene as one that exhibits elevated expression levels in all 61 tissues that are included in the Gene Expression Atlas (Su et al. 2004). Although large-scale transcriptome studies have attempted to approximate the number of HSKP genes in human and mouse, a general consensus is lacking (Eisenberg and Levanon 2003; Su et al. 2004; Freilich et al. 2005; Zhu et al. 2008; Wang and Rekaya 2009), and we consequently used two different cutoff criteria in our determinations (see "Material and methods"). The number of HSKP genes/chromosome varied from 19.14% to 27.89%, depending on the chromosome examined, when using the median expression cutoff. This proportion rose to 57%, on average, when the expression

Table 1 Distribution of genes in the mouse genome

Chromosome	Length (bp)	Total chromosomal length			Pericentromeric region		
		Number of genes	Number of HSKP (%)		Number of genes	Number of HSKP (%)	
			(a)	(b)		(a)	(b)
1	197,195,432	896	218 (24.33)	477 (53.24)	38	5 (13.15)	20 (52.63)
2	181,748,087	1,209	271 (22.41)	686 (56.74)	54	9 (16.66)	33 (61.11)
3	159,599,783	777	178 (22.91)	421 (54.18)	30	5 (16.66)	16 (53.33)
4	155,630,120	953	203 (21.31)	559 (58.66)	40	7 (17.50)	27 (67.50)
5	152,537,259	936	199 (21.26)	552 (58.97)	40	7 (17.50)	23 (57.50)
6	149,517,037	823	220 (26.73)	402 (48.79)	35	4 (11.43)	20 (57.14)
7	152,524,553	1,204	292 (24.25)	667 (55.39)	120	33 (27.5)	56 (46.66)
8	131,738,871	799	177 (22.15)	459 (57.45)	84	24 (28.57)	40 (47.62)
9	124,076,172	850	170 (20.00)	493 (58.00)	51	14 (27.45)	20 (39.21)
10	129,993,255	704	138 (19.61)	414 (58.81)	50	11 (22.00)	28 (56.00)
11	121,843,856	1,272	251 (19.73)	776 (61.01)	88	15 (17.04)	55 (62.50)
12	121,257,530	491	94 (19.14)	291 (59.26)	49	11 (22.45)	26 (53.06)
13	120,284,312	569	128 (22.49)	320 (56.24)	49	17 (34.69)	24 (48.97)
14	125,194,864	526	102 (19.39)	305 (57.98)	27	7 (25.92)	13 (48.14)
15	103,494,974	600	131 (21.83)	343 (57.16)	48	13 (27.08)	23 (47.91)
16	98,319,150	550	120 (21.81)	308 (56.00)	103	20 (19.42)	51 (49.51)
17	95,272,651	778	156 (20.05)	474 (61.01)	72	12 (16.66)	41 (56.94)
18	90,772,031	412	87 (21.12)	230 (55.82)	55	10 (18.18)	29 (52.72)
19	61,342,430	559	127 (22.7)	320 (57.24)	256*	45** (17.59)	159 (62.11)
X	166,650,296	685	189 (27.89)	298 (43.50)	90	21 (23.33)	44 (48.88)
Total	2,638,992,663	15,593	3,451	8,795	1,123	290	748

The total gene content (from telomere to telomere) and the pericentromeric regions (from 0 to 16 Mbp) are shown for each mouse chromosome. The number and percentage of housekeeping (HSKP) genes are shown for each of the cutoffs used in our study

[a] Cutoff median expression

[b] Cutoff 50% of median expression

[*] Significant differences using Chi-square test (χ^2 =32.89, DF=19, p=0.035). [**] Significant differences according to Chi-square test (p=0.05)

threshold was 50% of the median expression (Table 1). This approximate threefold increase was maintained when considering only HSKP genes located in the pericentromeric region (Table 1). Remarkably, MMU19 was the only chromosome to show a significant concentration of HSKP genes in the pericentromeric region (45 HSKP genes, p=0.05) compared to all other chromosomes in the complement at the cutoff median expression (Fig. 1b and c; Table 1).

The analysis of tissue-specific genes located in the pericentromeric regions was similarly informative. A total of 274 tissue-specific genes were detected representing 31 different tissue types. Although we did not find tissue-specific genes that were exclusively present in the pericentromeric region of MMU19, this chromosome nonetheless accumulates the highest number of different tissue-specific genes in the mouse complement (Fig. 2). This accumulation was, however, not statistically significant (χ^2=2.17, DF=19, p= 0.546). The tissue-specific genes concentrated in the pericentromeric region of MMU19 include 58 genes that are expressed in the immune system, testis, brain, skeletal and smooth muscle, liver, lung, osteoblasts, retina, olfactory bulb, kidney, epidermis, bladder, adrenal gland, and adipose tissues (Fig. 2). The

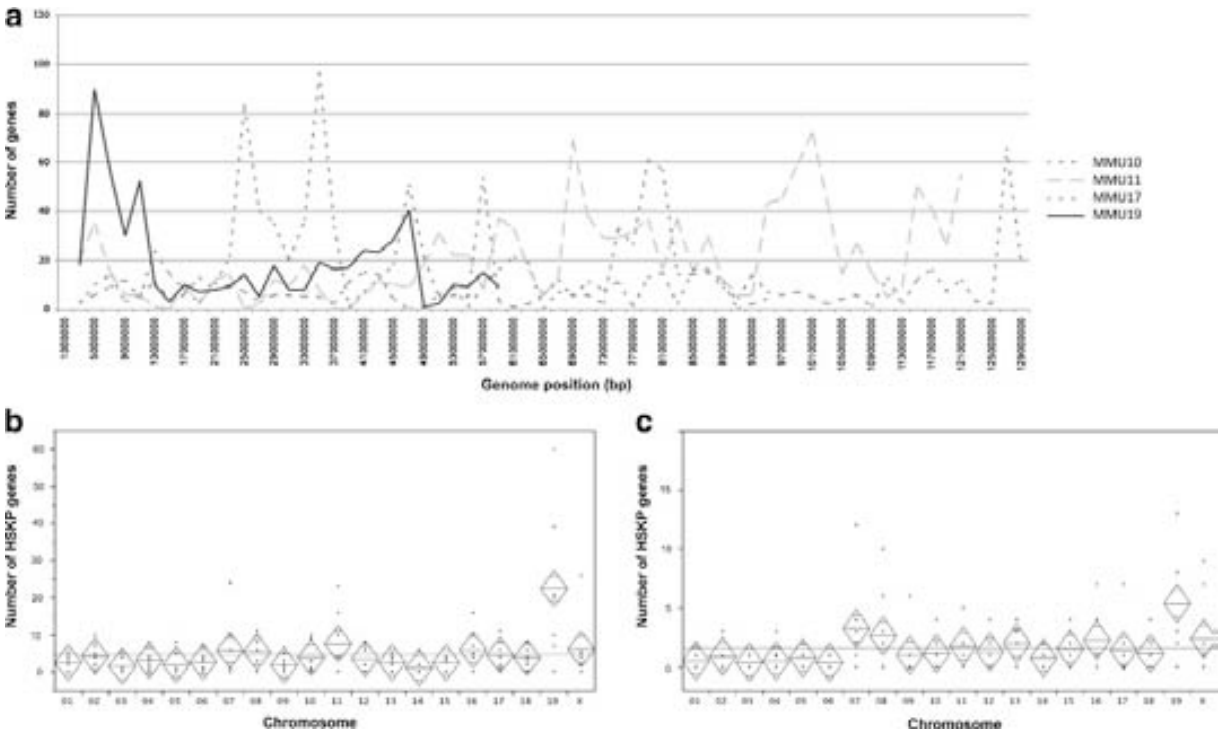


Fig. 1 Analysis of gene content in the mouse genome. **a** Distribution of genes in 2 Mb windows along mouse chromosomes 10 (MMU10), 11 (MMU11), 17 (MMU17), and 19 (MMU19). Note the high concentration of genes located in the first 16 Mbp of MMU19 compared to the other three chromosomes. **b** and **c** Number of HSPK genes identified in the pericentromeric region of each mouse chromosome applying different cutoffs: **b** median expression and **c** 50% of median expression (see "Material and methods" for more details). Each *dot* corresponds to the number of genes in each 2 Mb window. *Polygons* represent the mean values for each chromosome and the *grey line* the mean of all chromosomes. Note that mouse chromosome 19 differs significantly from other mouse chromosomes using the median cutoff

elevated occurrence of both HSKP and tissue-specific genes in pericentromeric MMU19 suggests a constraining role—either cells carrying deleterious breakpoints within this region do not survive, or there is an alteration of the gene expression pattern following chromosomal reorganization. This manifests in wild populations as selection against Rb fusions involving MMU19.

Studies dealing with the behavior of telomeres and centromeres, the two structures required for chromosome integrity and segregation, have attempted to clarify the molecular mechanisms underpinning the formation of Rb fusions (Slijepcevic 1998; Kalitsis et al. 2006). The centromeric regions of mouse telocentric chromosomes (which represent the partners in brachial combinations) are each characterized by a large block of major satellite deoxyribonucleic acid (DNA) flanked by a small block of minor satellite DNA adjacent to the telomeric repeats (Garagna et al. 1995, 2001). The distance between the telomeric and the minor satellite repeats is estimated to span 1.8–11 kb (Kalitsis et al. 2006). It has been argued that following Rb fusions, telomeric sequences are lost but a component of the minor satellite DNA is retained between two regions of major satellite DNA in the resulting metacentric chromosome (Garagna et al. 2001, 2002). This type of chromosomal reorganization necessitates double strand breaks (DSBs) that are repaired by recombination events between highly repetitive sequences. Breakpoints occurring within a genomic region that has a high concentration of HSKP and/or tissue-specific genes would impact on the cell's ability to maintain basic cellular functions, and persistence in meiosis is unlikely. Therefore, carriers of these rearrangements do not contribute offspring.

Another possible explanation for our observations is the stochastic epigenetic silencing of relocated

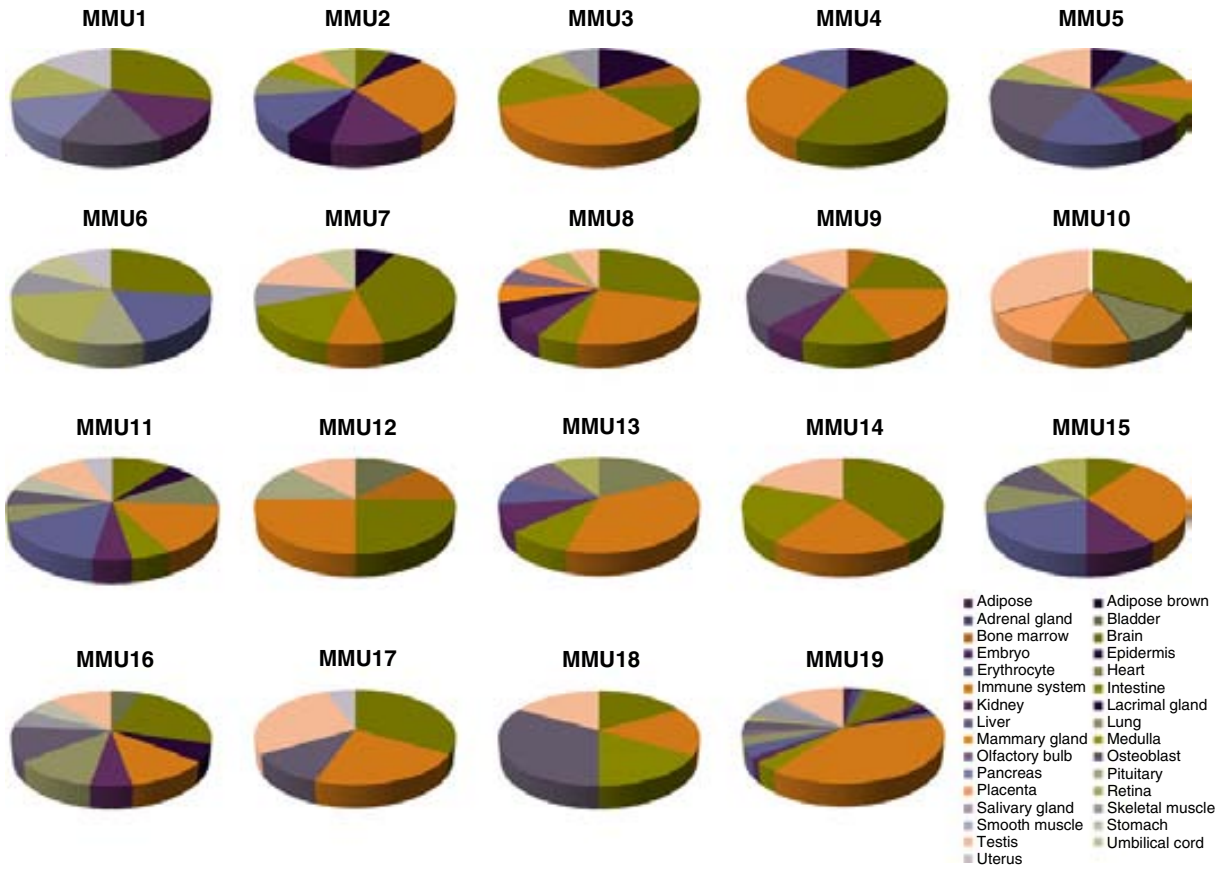


Fig. 2 Distribution of tissue-specific genes in the pericentromeric region of all mouse chromosomes

genes as a result of chromosomal rearrangements close to highly repetitive sequences. This phenomenon is known as "position effect variegation" (PEV) and was initially described in *Drosophila* (Muller 1930; Zhimulev 1988). The first evidence of PEV in mouse came with the description of the Cattanach X-autosome translocation (Cattanach 1974, 1975). This was subsequently similarly noted in transgenic mice when the transgene is located close to the centromere (Dobie et al. 1996; Opsahl et al. 2002) or the telomere (Pedram et al. 2006). In these instances, gene silencing may be attributed to changes in chromatin conformation such as alterations of histone N-terminal tails by deacetylation and methylation (Pedram et al. 2006). Importantly, it has been shown that gene variegation can extend over 4 Mbp to 5 Mbp from the centromeric heterochromatin in mice (Dobie et al. 1996). However, all mouse chromosomes with the exception of MMU19 have a low density of genes in the pericentromeric region and are therefore safeguarded from possible negative PEV effects. Conversely, the gene-rich MMU19 would probably be significantly impacted by PEV reducing the viability of fusions involving this chromosome in wild populations. Its detection in the Madeira metacentric population (Britton-Davidian et al. 2000, 2005) and its retrieval by introgressive breeding in laboratory mice (Evans et al. 1967; White and Tjio 1968; Redi and Capanna 1988) suggest that while translocations involving MMU19 can occur, they are fixed in wild populations at very low frequencies. It is plausible that the fusions Rb2.19 and Rb11.19 found in the Madeira population have escaped PEV. How this has occurred is a matter of speculation at this stage, but certainly, the molecular characterization of this mice population would provide new insights into the role of gene expression in chromosome evolution.

In this regard, considerable progress has been made in determining some of the major molecular features of chromosomal evolution. Among others,

these include findings that the evolutionary breakpoints are not randomly distributed (Bourque et al. 2004; Zhao et al. 2004; Ruiz-Herrera et al. 2006; Larkin et al. 2009) but tend rather to concentrate in intergenic regions (Lemaitre et al. 2009), avoiding accidental gene silencing/disruption, and that telomere attrition and/or centromeric breakage followed by unequal recombination among highly repetitive sequences may facilitate Rb fusion (Slijepcevic 1998; Garagna et al. 2001; Ruiz-Herrera et al. 2008). Our results complement these observations by showing that there is a selection against breakpoint formation in regions rich in genes necessary to maintain basic cellular functions. This finding has important implications for understanding the constraints operating on chromosomal reorganization and provides a plausible explanation for the structural bias in the chromosomal features of wild mice populations.

Acknowledgements Financial support from Ministerio de Ciencia y Tecnologia and the Universitat Autònoma de Barcelona (Ph.D. fellowship to M.F.) are gratefully acknowledged. T.J.R. is funded by a grant from the South African National Research Foundation.

References

Baker RJ, Bickham JW (1986) Speciation by monobranchial centric fusions. Proc Natl Acad Sci USA 83:8245–8248

Bourque G, Pevzner PA, Tesler G (2004) Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res 14:507–516

Britton-Davidian J, Nadeau JH, Croset H, Thaler L (1989) Genetic differentiation and origin of Robertsonian populations of the house mouse (*Mus musculus domesticus*). Genet Res 53:29–44

Britton-Davidian J, Catalan J, Ramalhinho MD et al (2000) Rapid chromosomal evolution in island mice. Nature 403:158

Britton-Davidian J, Catalan J, Ramalhinho MDG et al (2005) Chromosomal phylogeny of Robertsonian races of the house mouse on the island of Madeira: testing between alternative mutational processes. Genet Res Camb 86:171–183

Butte AJ, Dzau VJ, Glueck SB (2001) Further defining housekeeping, or "maintenance", genes. Focus on "a compendium of gene expression in normal human tissues". Physiol Genomics 7:95–96

Cattanach BM (1974) Position effect variegation in the mouse. Genet Res 23:291–306

Cattanach BM (1975) Control of chromosome inactivation. Annu Rev Genet 9:1–18

Dobie KW, Leet M, Fantest JA et al (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. Proc Natl Acad Sci USA 93:6659–6664

Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. Trends Genet 19:362–365

Evans EP, Lyon MF, Daglish M (1967) A mouse translocation giving a metacentric marker chromosome. Cytogenet Cell Genet 6:105–119

Freilich S, Massingham T, Bhattacharyya S et al (2005) Relationship between the tissue-specificity of mouse gene expression and the evolutionary origin and function of the proteins. Genome Biol 6:R56

Garagna S, Broccoli D, Redi CA, Searle JB, Cooke HJ, Cappana E (1995) Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite sequences DNA in the pericentromeric area. Chromosoma 103:685–692

Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA (2001) Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. Proc Natl Acad Sci U S A 98:171–175

Garagna S, Zuccotti M, Capanna E, Redi CA (2002) High resolution organization of mouse telomeric and pericentromeric DNA. Cytogenet Genome Res 96:125–129

Gazave E, Catalan J, Ramalhinho MD et al (2003) The non-random occurrence of Robertsonian fusion in the house mouse. Genet Res 81:33–42

Kalitsis P, Griffiths B, Choo KH (2006) Mouse telomeric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A 103:8786–8791

King M (1993) Species evolution: the role of chromosome change. Cambridge University Press

Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19:770–777

Lemaitre C, Zaghloul L, Sagot MF et al (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics 10:335

Nachman MW, Searle JB (1995) Why is the house mouse karyotype so variable? Trends Ecol Evol 10:397–402

Muller HJ (1930) Types of visible variations induced by X-rays in *Drosophila*. J Genet 22:299–334

Opsahl ML, McClenaghan M, Springbett A et al (2002) Multiple effects of genetic background on variegated transgene expression in mice. Genetics 160:1107–1112

Pedram M, Sprung CN, Gao Q, Lo AWI, Reynolds GE, Murnane JP (2006) Telomere position effect and silencing of transgenes near telomeres in the mouse. Mol Cell Biol 26:1865–1878

Pialek J, Hauffe HC, Searle JB (2005) Chromosomal variation in the house mouse. Biol J Linn Soc 84:535–563

Qumsiyeh MB (1994) Evolution of number and morphology of mammalian chromosomes. J Hered 85:455–465

Redi CA, Capanna E (1988) Robertsonian heterozygotes in the house mouse and the fate of their germ cells. In: Liss AR (ed) The cytogenetics of mammalian autosomal rearrangements. pp 315–359

Rieseberg LH (2001) Chromosomal rearrangements and speciation. Trends Ecol Evol 16:351–358

Ruiz-Herrera A, Castresana J, Robinson TJ (2006) Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol 7:R115

Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E (2008) Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res 122:219–228

Singer GAC, Lloyd AT, Huminiecki LB, Wolfe KH (2005) Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. Mol Biol Evol 22:767–775

Slijepcevic P (1998) Telomeres and mechanisms of Robertsonian fusion. Chromosoma 107:136–140

Su AI, Wiltshire T, Batalov S et al (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101:6062–6067

Vinogradov AE, Anatskaya OV (2007) Organismal complexity, cell differentiation and gene expression: human over mouse. Nucleic Acids Res 35:6350–6356

Wang Y, Rekaya R (2009) Comprehensive analysis of gene expression evolution between humans and mice. Evol Bioinform Online 5:81–90

White MJD (1978) Modes of speciation. Freeman, San Francisco

White BJ, Tjio JH (1968) A mouse translocation with 38 and 39 chromosomes but normal NF. Hereditas 58:284

Williams EJB, Hurst LD (2002) Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. J Mol Evol 54:511–518

Wu C, Orozco C, Boyer J et al (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 10:R130

Zhao S, Shetty J, Hou L et al (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14:1851–1860

Zhimulev IF (1988) Polytene chromosomes, heterochromatin, and position effect variegation. Adv Genet 37:1–555

Zhu J, He F, Song S, Wang J, Yu J (2008) How many human genes can be defined as housekeeping with current expression data? BMC Genomics 9:172

TREBALL 4

Recombination rates and genomic shuffling in human and chimpanzee – a new twist in chromosomal speciation theory

Farré M and Ruiz-Herrera A

Molecular Biology and Evolution (2012): under review

Índex d'impacte (2010): 5,510

ÀREA: Evolutionary biology, Genetics & Heredity, QUARTIL 1

Recombination rates and genomic shuffling in human and chimpanzee - a new twist in chromosomal speciation theory

RESEARCH ARTICLE

Marta Farré[1] and Aurora Ruiz-Herrera[1,2*]

[1] Departament de Biologia Cel•lular, Fisiologia i Immunologia. Universitat Autònoma de Barcelona. Campus UAB, 08193, Cerdanyola del Vallès, Barcelona, Spain.
[2] Institut de Biotecnologia i Biomedicina (IBB), Campus UAB, 08193, Cerdanyola del Vallès, Barcelona, Spain.

***Corresponding author**
Aurora Ruiz-Herrera
Departament de Biologia Cel•lular, Fisiologia i Immunologia
Universitat Autònoma de Barcelona
Campus UAB, 08193, Cerdanyola del Vallès, Barcelona, Spain
E-mail: aurora.ruizherrera@uab.cat
Tel: +34 93 5812015
Fax: +34 93 5813357

Key words: recombination, chromosome speciation, human, hemiplasy, reorganizations**,** chimpanzee.

List of abbreviations: DSBs: Double-Strand Breaks; HSBs: Homologous Synteny Blocks; EBRs: Evolutionary Breakpoint Regions.

ABSTRACT

A long-standing question in evolutionary biology concerns the effect of recombination in shaping the genomic architecture of organisms and, more in particular, how this impacts on the speciation process. Despite efforts employed in the last decade, the role of chromosomal reorganizations in the human-chimpanzee speciation process remains unresolved.

Through whole-genome comparisons we have analyzed the genome-wide impact of genomic shuffling in the distribution of human recombination rates during the human-chimpanzee speciation process. We have constructed a highly-refined map of the reorganizations and evolutionary breakpoint regions in the human genome when compared to the chimpanzee based on orthologous genes and genome sequence alignments. The analysis of the most recent human recombination map inferred from genome-wide SNP data revealed that standardized recombination rate was significantly higher in collinear than in rearranged chromosomes. Importantly, inverted regions presented significantly lower recombination rates than collinear and non-inverted regions, independent of the lineage in which they become fixed. In order to explain our observations, we propose a model where chromosomal reorganizations (heterokaryotypes for inversions) would persist in the human-chimp ancestral population as floating polymorphisms, inducing recombination suppression that is still detected in the human genome.

Our observations have important implications for the chromosomal speciation theory, providing new evidences for the contribution of inversions in suppressing recombination in mammals.

INTRODUCTION

Reorganization (shuffling) of the genomic landscape plays an important role in the evolutionary processes as well as in the development of inherited diseases and carcinogenesis. Traditionally it has been argued that chromosomal reorganization may contribute to speciation due to the underdominant fitness effects associated with meiotic abnormalities, and the creation of unbalanced gametes in chromosomal heterozygotes (White et al. 1978). But this model has important limitations given that "underdominance" is likely to occur in small, inbred populations, or when rearrangements are weakly underdominant individually, but strongly underdominant in combination (King 1993; White et al. 1978) and are difficult to test in natural populations. More recently, a number of related studies have proposed an alternative explanation by which chromosomal rearrangements could reduce gene flow and potentially contribute to speciation by the suppression of recombination (for example, Noor et al. 2001; Rieseberg 2001). According to this "suppressed recombination" model, chromosome rearrangements could have a minimal influence on fitness, but would suppress recombination leading to the reduction of gene flow across genomic regions and to the accumulation of incompatibilities. Recombination provides physical connections between homologues during the first meiotic division, contributing to correct chromosomal segregation. Although recombination can occur at the somatic level (such as V(D)J recombination produced in the immune system), only those recombination events occurring in the germ line are relevant for the speciation process. Recombination introduces inheritable new chromosomal variants that can become fixed with a probability that depends on various population genetic parameters (i.e., frequency, effective population size, among others), contributing, in the long term, to the formation of new species. Few empirical data are available that address the mechanisms by which new chromosomal variants are fixed in populations of mammalian species, and how recombination influences chromosomal speciation and *vice versa*. In this regard, two models [Kirkpatrick-Barton model (Kirkpatrick and Barton 2006) and Navarro-Barton model (Navarro and Barton 2003)] have been formulated in order to explain (i) how, under divergent selection, chromosomal rearrangements can be fixed in two parapatric populations (in the presence of gene flow) and (ii) by which mechanisms these contribute to speciation (revised in Faria and Navarro 2010). In both case, these formulations require a reduction of recombination between heterokaryotyes (chromosomes displaying alternate forms for rearrangments) as a crucial factor for speciation in parapatry.

High-resolution recombination maps have been inferred from high-density single-nucleotide polymorphism (SNP) data using linkage disequilibrium (LD) patterns. Recombination maps using LD analysis or/and sperm typing are now available for a variety of species (International HapMap Consortium et al. 2007; Megens et al. 2009; Qanbari et al. 2010; Rogers et al. 2006; Wu, Getun, Bois 2010). While LD analysis provides indirect evidence of inferring the recombination process, sperm typing directly detects recombinant DNA molecules. Nevertheless, LD-based maps estimate the location of recombination events in the progeny (see (Lynn, Ashley, Hassold 2004) for a review) and reflect the integration of population-level processes over several generations. Among others, this provides an historical view of recombination events, incorporating data on population growth and natural selection, among others (Clark, Wang, Matise 2010). A third methodology consists of analyzing the recombination process as it occurs through the *in situ* immunolocalization of recombination proteins (i.e. MLH1) on germ cells (Lynn et al. 2002). Recombination maps using this (direct) approach have been made in humans (Codina-Pascual et al. 2006; Sun et al. 2005), non-human primates (Garcia-Cruz et al. 2011; Hassold et al. 2009) and several mammalian species (Dumont and Payseur 2011; Froenicke et al. 2002). Although all three approaches result in differences in spatial resolution and rely on different units of measurement (axis length for the cytological measurements vs. DNA length for the genetic analysis), the correspondance in the overall patterns of recombination can be appreciated in homologous chromosomes from different species (Garcia-Cruz et al. 2011). It has been argued that rates of recombination might vary considerably between species when comparing high-resolution (kb) recombination maps (Paigen and Petkov 2010; Ptak et al. 2005) but these differences disappear at a broader scale (Mbp) (revised in Smukowski and Noor 2011). Moreover, recent studies suggest that there is a phylogenetic effect in recombination rates, indicating that closely related species tend to have similar average rates of recombination (Dumont and Payseur 2008; Dumont and Payseur 2011; Garcia-Cruz et al. 2011).

If genomic shuffling is affecting evolutionary and speciation processes, through the mechanical shearing at evolutionary breakpoints, how does this reorganization impact meiotic recombination? The assumption that some chromosomal regions have been reused during the mammalian chromosomal evolution (Larkin et al. 2009; Murphy et al. 2005; Ruiz-Herrera, Castresana, Robinson 2006) has lead evolutionary biologists to investigate whether there is any particular DNA configuration or composition underpinning genomic instability. In this sense, it has been reported that breakpoint regions co-localize with fragile sites (Ruiz-Herrera et al. 2005)

and are enriched in repetitive elements, such as tandem repeats (Farre et al. 2011; Ruiz-Herrera, Castresana, Robinson 2006), segmental duplications (Bailey and Eichler 2006; Kehrer-Sawatzki and Cooper 2008), and transposable elements (Bourque 2009; Caceres et al. 1999; Carbone et al. 2009; Delprat et al. 2009; Farre et al. 2011; Longo et al. 2009). However, few empirical data focus on the relationship between evolutionary breakpoint regions and recombination rates. Initial studies in *Drosophila* have described a strong reduction of recombination around inversion breakpoints and within the reorganization itself (Navarro et al. 1997). But the question whether this pattern also holds for mammals (in our study human and chimpanzee) remains unanswered despite the efforts in the last decade (Marques-Bonet and Navarro 2005; Marques-Bonet et al. 2007; Navarro and Barton 2003; Vallender and Lahn 2004; Zhang et al. 2004). Here we analyze the recombination rate in homologous synteny blocks (HSBs; i.e. regions where the gene order has been conserved among species) and evolutionary breakpoint regions (EBRs; i.e. regions where the synteny has been disrupted due to reorganizations - see Ruiz-Herrera et al 2006) in the human genome when compared to the chimpanzee genome by taking advantage of the most recent human recombination map inferred from genome-wide SNP data at a resolution of 10 kilobases (kb) (Kong et al. 2010). Moreover, we determine whether chromosomal reorganizations (i.e. inversions) may have had a genome-wide impact in the distribution of human recombination rates. Overall, our data provide compelling evidence for the existence of low recombination rates within genomic regions that have been rearranged in the chromosomal evolution of human and chimpanzee.

MATERIAL AND METHODS

Whole-genome comparisons and evolutionary breakpoint definition

We obtained human (hg18), chimpanzee (pantro2) and orangutan (PPYG2) orthologous genes from Biomart and downloaded the masked genome sequences from Ensembl v64 database. In order to detect the evolutionary breakpoint regions (EBRs) and homologous synteny blocks (HSBs) between human and chimpanzee whole-genome sequences, we applied two recently described algorithms: SyntenyTracker (Donthu et al. 2009) and Cassis (Baudet et al. 2010). The former approach relies on the detection of HSBs among different species genomes. Based on the genomic positions of orthologous genes, this algorithm establishes temporary synteny blocks and merges neighboring blocks spaced less than a given threshold and having the same orientation. We used the default parameters proposed by the authors (Donthu et

al. 2009) and set different thresholds (250 Kbp, 500 Kbp and 1 Mbp), obtaining the best performance at 1 Mbp. Then, we defined the EBRs as a bypass product of contiguous HSBs coordinates. Cassis (Baudet et al. 2010), on the other hand, is specially designed to define breakpoint regions. The algorithm establishes the putative location of EBRs using the position of orthologous genes as markers and then by means of sequence alignment, more accurately defines the EBR coordinates. The Cassis algorithm was run using default parameters. The same approach was used to define HSBs and EBRs between the human and orangutan genomes.

Once the EBRs were defined in the human genome, we grouped human chromosomes into rearranged (those affected by reorganizations) or collinear (if they were conserved), using the orangutan as an outgroup. We then divided the rearranged chromosomes into two additional genomic regions: i) inverted or ii) non-inverted, if they were affected by the reorganization o not, respectively. Finally, we recognized macro-rearrangements as those where the inverted regions spanned > 1.4 Mbp and micro-rearrangements if they spanned < 1.4 Mbp.

Recombination rates

Genetic maps for the human genome were extracted from Kong and collaborators (Kong et al. 2010). These authors inferred genomic recombination rates from high-density single-nucleotide polymorphism (SNP) data using linkage disequilibrium (LD) patterns with a resolution of non-overlapping windows of 10 kilobases (Kbp). Recombination rate data is estimated for 2.4 Gbp of the human genome, excluding chromosome X and the 5 Mbp at the ends of each autosomal chromosome. For each window, the standardized recombination rate is calculated as a fraction of the genetic distance divided by the overall average distance (Kong et al 2010). Genomic regions with a recombination rate $\geq$ 10 were considered as "hotspots", whereas regions with a recombination rate equal to 0 were considered "coldspots". Using in-house Perl scripts and whole-genome comparisons with the human genome, we merged the coordinates of recombination rate windows with the positions of EBRs and the different types of regions detected (collinear, non-inverted and inverted).

Statistical analysis

Statistical analyses were performed using the JMP v7 package. Given that the genomic distribution of recombination rates did not followed a normal distribution we applied non-parametric analysis (Mann-Whitney U or Kruskal-Wallis tests) in order to assess the differential recombination rates between EBRs/HSBs, inverted/non-inverted/collinear regions and

collinear/rearranged chromosomes. We applied the Bonferroni correction when necessary. Fischer's exact was applied test to compare the distribution of recombination "hotspots" and "coldspots" in EBRs and HSBs.

RESULTS

Whole-genome comparisons between human and chimpanzee genomes

A total of 17,360 orthologous genes between the human and chimpanzee genomes and 16,409 orthologous genes between the human and orangutan were used in our estimations of evolutionary breakpoint regions (EBRs). Regarding human/chimpanzee pair-wise analysis we identified 43 homologous synteny blocks (HSBs) and 28 EBRs (with a median length of 93.3 Kbp; range 238 bp - 7.9 Mbp) using SyntenyTracker, .In contrast the Cassis algorithm detected 38 EBRs. Although Cassis detected the same 28 EBRs defined by SyntenyTracker, it narrowed the regions (ranging from 5 bp to 3.1 Mbp with a median size of 11.2 Kbp in length; Additional File 1). This shows that Cassis provided improved resolution in defining EBRs, and therefore we used the 38 EBRs detected by Cassis for further analysis.

We compared our results to previously published comparisons of human and chimpanzee genomes (Feuk et al. 2005; Kehrer-Sawatzki and Cooper 2008) in order to establish a reliable EBRs dataset. All the macro-reorganizations described by Kehrer-Sawatzki and Cooper (2008) affecting human chromosomes 1, 4, 5, 9, 12, 15, 16 and 17, except for the inversion breakpoints of human chromosome 18, which proved to be in regions rich in repetitive sequences, were refined by the analysis. We used the coordinates defined by cytogenetic studies (Kehrer-Sawatzki and Cooper 2008) in the case of human chromosome 18. A comparison of our dataset with the EBRs experimentally validated by Feuk and collaborators (2005) showed that 9 of the 38 EBRs detected by Cassis had not previously been identified (Additional File 1). Finally, we filtered our estimated EBRs dataset retaining EBRs that met the following criteria: i) they were part of the macro-rearrangements experimentally validated by fluorescent *in situ* hybridization (FISH) (Kehrer-Sawatzki and Cooper 2008), ii) they were identified using SyntenyTracker and/or iii) they were experimentally validated by Feuk and coworkers (2005). After the filtering process we obtained a final dataset comprising 37 EBRs, ranging from 5 bp to 171 Mbp with a median length of 9.8 Kbp (Table 1). Overall, we confirmed and refined the breakpoints involved in nine large inversions affecting homologous chromosomes 1, 4, 5, 9, 12, 15, 16, 17 and 18, in

addition to the fusion responsible for human chromosome 2. Additionally, we detected four indels (insertion/deletions) (three of them in chromosome 2 and one in chromosome 10) and 8 micro-inversions (less than 1.4 Mbp) affecting chromosomes 1, 7, 10, 19, X and Y (Table 1).

Taking the reorganizations and the EBRs detected in our study into consideration we conclude that human chromosomes 3, 6, 8, 11, 13, 14, 20, 21 and 22 are considered collinear when compared to chimpanzee genome, whereas human chromosomes 1, 2, 4, 5, 7, 9, 10, 12, 15, 16, 17, 18, 19, X and Y are rearranged. We subsequently divided the rearranged chromosomes into regions that have suffered micro- or macro-rearrangements. Human chromosomes 4, 5, 9, 12, 15, 16, 17 and 18 were affected by macro-rearrangements, whereas human chromosomes 7, 10, 19, X and Y only by micro-rearrangements; only human chromosomes 1 and 2 show both types of reorganizations. As a whole, the macro-rearrangements encompassed 318 Mbp of the whole human genome (ranging from 1.48 to 77.36 Mbp, with a median size of 40 Mbp), whereas the micro-rearrangements spanned 10.8 Mbp, ranging from 12.5 to 919.3 Kbp. Finally, we divided the rearranged chromosomes into regions considered as inverted and non-inverted (i.e., if they were included or excluded in the reorganization). Overall, inverted regions encompassed 328.57 Mbp of the human genome, while non-inverted regions represented 1.67 Gbp of the human genome.

Once we had refined the breakpoints between human and chimpanzee, we proceeded to date the reorganizations detected based on previous reports (Kehrer-Sawatzki and Cooper 2008; Yunis and Prakash 1982) (Figure 1). It is known that human chromosomes 6, 20, 21 and 22 have been maintained as collinear orthologous in the great apes since common ancestry. The ancestral form of orthologous chromosomes 3 and 11 are conserved in orangutan, but suffered an inversion that has been fixed in the human-chimpanzee-gorilla ancestor. The orangutan also presents the ancestral forms for orthologous chromosomes 7 and 10, each of which suffered different inversions at different speciation nodes (Fig. 1). Finally, human chromosomes 1, 2, 4, 5, 9, 12, 15, 16, 17 and 18 are rearranged between human and chimpanzee; inversions affecting human chromosome 1 and 18 have occurred only in the human lineage (i.e are autapomorphies), whereas the rest (on orthologous 4, 5, 9, 12, 15, 16 and 17) have become fixed in the lineage leading to chimpanzee (Kehrer-Sawatzki and Cooper 2008).

Human recombination rates and chromosomal reorganizations

We analyzed the standardized human recombination maps described by Kong and collaborators (2010) in order to study the association between recombination rate and genomic

shuffling in human and chimpanzee genomes. Comparable data for the chimpanzee whole genome sequence are not available. We used the sex averaged (female and male) recombination map which provides a total of 4,006 "hotspots" with standardized recombination rate (SRR) ≥ 10. A first analysis showed that SRR is not homogeneously distributed among human chromosomes (Kruskal-Wallis, $p < 0.0001$). The lowest average SRR (0.856) was on human chromosome 9, while the highest (1.559) was on human chromosome 22 (Table 2). We found a strong negative correlation between recombination rate and chromosome size (Spearman's $\rho = -0.915$, $p < 0.0001$) suggesting that smaller chromosomes have a higher recombination rates than do larger chromosomes (Table 2). These results corroborate previous observations in mammals that show how larger chromosomes tend to accumulate larger numbers of crossovers (COs), and that each chromosome arm generally presents at least one CO (Sun et al. 2005). Small chromosomes, on the other hand, are expected to have higher recombination rates in order to ensure the resolution of, at least, one CO, and therefore guarantee a correct disjunction during meiosis. Moreover, we observed that recombination rate was not uniformly distributed across each human chromosome, which presented clusters of "hotspots" and "coldspots" along chromosomal regions (Fig. 2 and Additional File 2). This follows the non-randomly distribution of COs found in other species (Petes 2001).

We compared SRR between collinear and rearranged chromosomes to assess whether the distribution of recombination rate is affected by chromosomal reorganizations. First, the average recombination rate was estimated by considering all the 10 kb-windows for each chromosome as a whole. This was found to be significantly higher in collinear (0.975) than in rearranged (0.944) chromosomes (Mann-Whitney's U, $p < 0.0001$) (Table 3). We then classified rearranged chromosomes into genomic regions affected (inverted) or not affected (non-inverted) by rearrangement. Importantly, inverted regions present significantly lower recombination rates (0.715) than do collinear (0.975) and non-inverted regions (1.001) (Kruskal-Wallis, $p < 0.0001$), showing a possible effect of suppression of recombination within inverted regions (Table 3). Finally, we considered the length of the region involved in rearrangements by grouping each inverted region into micro- or macro-rearrangements. In doing so, we detected a significantly lower recombination rates in genomic regions within macro- (0.713) rather than within micro-rearrangements (0.976) and non-rearranged regions (0.996) (Kruskal-Wallis, $p < 0.0001$), suggesting that macro-rearrangements have a stronger impact on reducing recombination rate than do micro-rearrangements (Table 4). When analyzing each rearranged human chromosome separately, a striking pattern emerged whereby regions affected by inversions showed lower

SRR than did non-inverted regions (Fig. 2 and Additional file 2). Moreover, when the size of the inverted genomic region is considered, macro-reorganizations have lower recombination rates than micro-rearrangements (Additional File 2), observations consistent with data obtained by the whole genome analysis.

Given that all the macro-reorganizations described involved the centromere (pericentric inversions), we tested if the suppression of recombination that we observed within reorganized areas was due to the low recombination rate characteristic of pericentromeric regions (Kong et al. 2010). To do so, we simulated rearrangements in collinear chromosomes (3, 6, 8, 11, 13, 14 and 20) with the median size of the macro-rearrangements described (40 Mbp). We excluded human chromosomes 21 and 22, since their total sizes are 48 Mbp and 51 Mbp, respectively, and, simulated analysis, would of necessity, include almost the entire chromosome. Although recombination rate in the simulated reorganizations was lower (0.793) than in non-inverted regions (1.01), it was significantly higher than in macro-reorganizations (0.713; Mann-Whitney's U, $p = 0.0021$). These data suggest that centromere position has an influence in the reduction of the recombination rates, but the effect is not strong enough to explain the reduction of recombination rate observed in reorganized chromosomes.

Interestingly, we observed the same pattern for sex specific recombination maps (male and female, Kong et al. 2010). In both sexes, the observed recombination rate was greater in collinear than in rearranged chromosomes, which, in turn, exhibited lower recombination rate in inverted regions than in non-inverted regions. In addition, macro-rearrangements showed lower recombination rates than micro-rearrangements and collinear chromosomes in both sexes (0.788, 0.958 and 0.986 for the female recombination map and 0.771, 0.931 and 0.989 for the male recombination map).

Given that evolutionary breakpoint regions tend to have higher divergence rates than other regions in the genome (Marques-Bonet and Navarro 2005; Navarro et al. 1997), and divergence rate strongly correlates with recombination (Hellmann et al. 2003), we decided to compare recombination rates between EBRs and HSBs in the human genome. Although not statistically significant (Mann-Whitney's U, $p = 0.078$), we nonetheless found a lower recombination rate in EBRs (0.492) than in HSBs (0.962). Remarkably, none of the evolutionary breakpoints detected show recombination "hotspots" (regions with SRR higher than 10) but they contained significantly less "coldspots" than did the HSBs (Fischer's Exact test, $p < 0.0001$) suggesting that this tendency may be of relevance. We also analyzed the recombination rate in the genomic regions surrounding the breakpoints. To do so, we utilized a "breakpoint-edge"

(BP-edge) that spanned a region 100 Kbp upstream or downstream from the breakpoint coordinates. This showed that EBR are surrounded by regions of high recombination (SRR=0.877), although the differences between EBRs and BP-edge were not statistically significant (Mann-Whitney's U, p =0.634).

We did not expect to find any effect on recombination rates for inversions that have been fixed in the chimpanzee lineage (i.e. chromosomes 4, 5, 9, 12, 15, 16 and 17) but, surprisingly, we detected lower SRR not only in macro-rearrangements affecting human chromosomes 1 and 18 (human specific inversions), but also human chromosomes 2, 4, 5, 9, 12, 16 and 17 (Fig. 1). To determine whether this striking pattern was conserved in chromosomal macro-rearrangements involved at other speciation nodes during great apes evolution, we analyzed human chromosomes 3, 7, 10 and 11, since the human and chimpanzee forms represent the derivative state whereas orangutan has retained the ancestral form. To do so, we established the macro-inversions affecting these chromosomes between human and orangutan genome using the Cassis software. We found a significant lower SRR in the rearranged regions of these chromosomes compared to non-rearranged region (Additional File 2). However, this decrease of recombination rate was less marked than those detected in chromosomes involved in the human-chimpanzee speciation node (Kruskal-Wallis, $p < 0.0001$). More importantly, we found the highest SRR (1.103) in the human chromosomes that have been unaltered since the common ancestry of great apes (chromosomes 6, 20, 21 and 22), an intermediate SRR (0.828) in those chromosomes involved in the node characterizing the orangutan divergence (chromosomes 3, 7, 10 and 11) and the lowest SRR (0.715) in chromosomes involved in the human-chimpanzee split (1, 2, 4, 5, 9, 12, 16, 17 and 18) (Kruskal-Wallis, $p< 0.001$). These data suggest that the reduction of recombination rate is still traceable in those human chromosomes where the derivative form was fixed before the human-chimpanzee-orangutan divergence (~10 Mya; Hobolth et al 2011). Moreover, this reduction of recombination is markedly lower in human chromosomes rearranged during the human-chimpanzee speciation node (~4 Mya; Hobolth et al. 2011).

DISCUSSION

Refinement of chromosomal reorganizations between human and chimpanzee genomes

The seminal work of Yunis and Prakash (1982) was pioneering in defining the chromosomal rearrangements that differentiate humans from chimpanzees since divergence

from their common ancestor ≈ 4 million years ago (Hobolth et al. 2011). These differences included nine macro-inversions and one fusion of two ancestral hominoid chromosomes to produce the modern form of human chromosome 2. The inversions affecting human chromosome 1 and 18 have occurred in the human lineage, whereas the rest (on orthologs 4, 5, 9, 12, 15, 16 and 17) occurred in the lineage leading to chimpanzee (Kehrer-Sawatzki and Cooper 2008). Since this initial comparative study, much of the effort has been directed at defining and characterizing the genomic regions (evolutionary breakpoints) and the mechanisms involved in these reorganizations using cytogenetic approaches and/or gene sequence comparisons (Kehrer-Sawatzki and Cooper 2008 and references therein). However, the release of the completed chimpanzee genome provided the wherewithal to undertake a variety of comparative genome-wide studies, which, among others, revealed the presence of abundant putative micro-reorganizations (inversions and indels ranging from 51 bp to 4 Mbp in size) that were beyond the resolution provided by cytogenetic methods (Feuk et al. 2005; Szamalek et al. 2006).

In this context we provide a high-resolution map based on orthologous genes and genomic sequences of the reorganized and evolutionary breakpoint regions in the human genome compared to that of the chimpanzee. Previous comparative studies using genomic sequence alignments between human and chimpanzee have found large numbers of putative rearrangements, including 71 inversions encompassing 2 or more genes (Szamalek et al. 2007) as well as 1576 inversions ranging from 23 bp to 62 Mb (Feuk et al. 2005). Our approximation has been more conservative given that we considered only those reorganizations that have been experimentally confirmed by different approaches (*in silico*, FISH and/or PCR analysis). In this way we have been able to refine the evolutionary breakpoint regions of the inversions affecting human chromosome 1, 4, 5, 12, 16, 17 and 18 (Table 1). In the case of human chromosomes 9 and 15 we were not able to provide more detail than previous cytogenetic studies (Kehrer-Sawatzki et al. 2005; Locke et al. 2003), probably due to the highly complex nature of the breakpoints (presence of segmental duplications flanking the breakpoint). Additionally, we detected four indels (three of them in chromosome 2 and one in chromosome 10) and 8 micro-inversions affecting chromosomes 1, 7, 10, 19 X and Y (Table 1). These data allowed us to analyze the impact of genome reshuffling on the distribution of human recombination rates and more accurately define those human genomic regions involved in structural rearrangement of chromosomes.

Recombination rates and chromosomal reorganizations

It is known that chromosomal inversions prevent recombination in heterokaryotypes due to mechanistic problems during meiosis (Brown et al. 2012). In instances where recombination is suppressed, the formation of genetically unbalanced gametes would be prevented and therefore, the reproductive fitness of the species would not be affected. Direct and indirect evidences of suppressed recombination within rearranged segments have been reported in the literature (Brown and O'Neill 2010; Faria and Navarro 2010). Direct evidence includes the analysis of recombination in the gametes and/or the offspring when reorganization (inversions and/or translocations) occurs. Data supporting recombination suppression by inversions has been provided by early cytogenetic studies in mammals, especially rodents (Ashley et al. 1981; Greenbaum and Reed 1984; Hale 1986) and *Drosophila* (Navarro and Ruiz 1997; Navarro et al. 1997). Hale (1986) described heterosynapsis (asynapsis) and, therefore, suppression of chiasmata (chromosomal configurations resulted from meiotic crossovers, COs) formation within heterozygous pericentric inversions in the Sitka deer mouse (*Peromyscus sitkenssi*) as a mechanism for the maintenance of pericentric inversion polymorphisms in wild populations. Borodin (Borodin et al. 2008) detected a reduction in MLH1 foci (a meiotic protein that marks crossovers) in translocated chromosomes of the common shrew (*Sorex araneus*), whereas two independent studies (Castiglia and Capanna 2002; Dumas and Britton-Davidian 2002) have reported a reduction in chiasmata number in house mice (*Mus musculus domesticus*) with Robertsonian (Rb) translocations.

On the other hand, indirect evidence for the suppression of recombination has included the analysis of rates of genetic divergence between rearranged and collinear chromosomes (Navarro and Barton 2003). High rates of sequence divergence detected in genes located at, or near, chromosomal rearrangements have been interpreted as indirect evidence of chromosomal speciation through suppressed recombination (Marques-Bonet et al. 2007). This latter approach has been used in several studies on *Drosophila* (Brown et al. 2004; Kulathinal et al. 2008), *Helianthus sp.* (sunflower) (Rieseberg et al. 1995; Rieseberg et al. 1999), Solanaceae (Rieseberg and Willis 2007) and *Anopheles* (Besansky et al. 2003; Michel et al. 2006). In mammals, Franchini (Franchini et al. 2010) and Yanninc (Yannic et al. 2009) detected reduced gene flow within the reorganized regions (Rb translocations in heterokaryotypes) in house mouse and shrew populations, respectively, probably as a consequence of a fall-off in recombination around the centromeric regions. However, the study of the human and chimpanzee (Navarro and Barton 2003) has provided contradictory results so far (Marques-Bonet and Navarro 2005; Marques-

Bonet et al. 2007; Vallender and Lahn 2004; Zhang et al. 2004) and demonstration of recombination suppression in these species remains elusive.

Our own study represents a departure from those conducted previously in that it relies on the use of a recent and high-resolution (10kb) genome-wide map of recombination rates in the human to refine genome reshuffling between human and chimpanzee. These recombination rates are not directly quantified from gametes, but inferred from genome-wide SNP data from a human population of 38,167 individuals (Kong et al. 2010). Using this approach we provide evidences of a reduction of recombination within genomic regions that have been implicated in the chromosomal evolution between human and chimpanzee and propose a model where inversions might have persisted in the heterozygous state long enough for the impact on recombination rates still to be detected in the human genome.

Reduced recombination rates within reorganized genomic regions

When initially proposed, the "suppressed recombination" model was considered as a compelling hypothesis to explain the contribution of large genome reshuffling in the formation of new species (Noor et al. 2001; Rieseberg 2001). Under this assumption, chromosome rearrangements in heterokaryotypes have a minimal influence on fitness, but would rather suppress recombination thus contributing to a reduction of gene flow across genomic regions and the accumulation of gene incompatibilities. Most subsequent studies have used sequence divergence (patterns of nucleotide differentiation) between species as an indirect estimation of recombination. But this approximation has its limitations, and the interpretation of amino acid divergence, as an effect of recombination, can be problematic (reviewed in Noor and Bennett 2009). As an example, Bullaughey and collaborators (2008) found no correlation between either broad- or fine-scale rates of recombination and rates of protein evolution (measured by dN/dS ratios) between human, chimpanzee and rhesus macaque, suggesting that additional parameters should be considered.

When considering the human genome as a whole, we found that recombination rate was significantly higher in collinear (0.975) than in rearranged (0.944) chromosomes (Mann-Whitney's U, $p < 0.0001$). Moreover, those genomic regions within the macro-reorganizations (historically detected by cytogenetic studies) have a significantly lower recombination rate (0.713) than both micro-rearrangements (0.976) and non-rearranged regions (0.996) (Kruskal-Wallis, $p < 0.0001$). These data support the existence of a possible suppression of recombination effect associated with reorganized chromosomal regions—this being more substantial in large

inversions. Previous studies by Navarro and Barton (2003) compared the recombination rates (cM/Mbp) in collinear and rearranged chromosomes between human and chimpanzees using an earlier version of the human recombination map (Kong et al. 2002). Although no statistical differences were found, they noted a tendency for rearranged chromosomes to show a reduced recombination rate (1.10) compared to collinear chromosomes (1.17). Later studies, however, have shown different trends (Marques-Bonet et al. 2007; Szamalek et al. 2007; Zhang et al. 2004), underscoring the uncertainty surrounding recombination rates. Here we have shown that the effects of genome reshuffling on the distribution of recombination rates can be assessed when combining an accurate delineation of the chromosomal reorganizations and a high-resolution standardized recombination map. We detected not only a lower recombination rate within rearranged genomic regions, but also found that regions not affected by the reorganization in rearranged chromosomes (non-inverted regions) presented significantly higher recombination rates (1.001) than do collinear (0.975) chromosomes (Kruskal-Wallis, $p < 0.0001$). This pattern is not unexpected given that at least one CO per pair of homologous chromosomes is necessary to ensure proper disjunction. Therefore, chromosomes affected by rearrangements showed that the non-inverted regions accumulate recombination events that are absent within the inverted region —the so-called "inter- and intrachromosomal effect" — thus explaining the global increase of recombination rate in regions outside inversions (Schultz and Redfield 1951; Sturtevant 1919).

More importantly, our observations validate the relevance of chromosomal reorganizations in the speciation process thus providing support for the "suppressed recombination" model. Our model follows two lines of evidence. Kirkpatrick and Barton (2006) have suggested that selection could favor reorganizations (i.e. inversions) that reduce recombination of alleles involved in local adaptation. This situation would, eventually, contribute to the fixation of chromosomal reorganizations in different subpopulations in parapatry (connected by gene flow). On the other hand, it has been proposed that chromosomal reorganizations can occasionally survive as polymorphic states for considerable lengths of time, although this would depend on historical variables including effective population size and spatial population structure. Termed hemiplasy (Avise and Robinson 2008), this hypothesis suggests that derived chromosomal rearrangements may have persisted as polymorphisms across multiple speciation nodes (Robinson et al. 2008; Robinson and Ropiquet 2011). This has been the case for chiropteran and afrotherian species (Robinson et al. 2008), Perissodactyla (Trifonov et al. 2008), Rodentia (Badenhorst et al. 2011), Bovidae (Robinson and Ropiquet 2011) and this

is also probably true for primates (Dutrillaux and Couturier 1981; Rumpler et al. 2008). In fact, incomplete lineage sorting (when a gene tree is topologically inconsistent with the species tree) has been detected in the human-chimpanzee-gorilla species phylogeny in genome-wide studies (Chen and Li 2001; Hobolth et al. 2011; Patterson et al. 2006).

Here we propose a model where chromosomal polymorphisms (e.g. heterokaryotypes for inversions) persisted in the ancestral human-chimp population eventually becoming fixed in one of the descendant lineages (Figure 3). A recent analysis has provided support for a more recent human-chimp speciation event than previously reported (~ 4 million years ago) and a common ancestor with an effective population size of ~50,000 (Hobolth et al. 2011). Under such conditions it is plausible that inversions could have been maintained as heterokaryotypes in the human-chimp ancestral population. This would result in recombination suppression within the reorganized genomic regions involved and this suppression would persist up to the present population gradually returning to the same levels observed in ancestral collinear chromosomes. This interpretation is supported by our data given that those genomic regions contained within the inversions characterizing the human-chimpanzee-orangutan speciation node presented lower recombination rates (0.828) than ancestral collinear chromosomes (1.103) but higher than the case of inversions that have been fixed after the human-chimpanzee divergence (0.715). Therefore, chromosomal forms would have different recombination rates according to the speciation node where they had become fixed. We detect lower SRR in recently rearranged chromosomes (human-chimpanzee node, ~4Mya), intermediate SRR in those fixed in the human-chimpanzee-orangutan node (~10 Mya) and higher SRR in those chromosomes that maintained the ancestral form for great apes. Then, from an ancestral human-chimp population characterized with persisting floating heterokaryotypes, seven inversions (affecting chromosomes 4, 5, 9, 12, 16 and 17) have become fixed in the lineage leading to chimpanzees, whereas two inversions (affecting chromosomes 1 and 18) have been fixed in the lineage leading to humans. We found lower recombination rates in human chromosomes that have been rearranged and fixed in human lineage, and also in those human chromosomes that maintained the ancestral form and the reorganization has been fixed in chimpanzee lineage (Fig. 3). Therefore, the reduction of suppression within inversions that took place while ancestral human-chimpanzee population was polymorphic is still traceable in the human genome. At this point there is no available information on the genome-wide recombination landscape in the chimpanzee, but our results show that regions of the human genome that were involved in reorganizations still persist as regions with low recombination rates. Moreover, our model of

persisting heterokaryotypes fits with previous results on gene-expression divergence between human and chimpanzee (Marques-Bonet et al. 2004). The maintenance of the polymorphic state could increase the time of suppressed recombination, which, in turns, could explain gene-expression divergence in both lineages. It is also possible that the persistence of polymorphism in the ancestral population vary for each rearrangement, so they could exhibit quite different divergence times (Noor and Bennett 2009) thus explaining the contradictory results obtained in previous studies (Marques-Bonet and Navarro 2005; Marques-Bonet et al. 2007; Navarro and Barton 2003; Vallender and Lahn 2004; Zhang et al. 2004).

We are aware that our use of recombination rates in the modern human genome as a reflection of its historical evolution is not without criticism. Although fine-scale rates may be partially conserved among species (Myers et al. 2005; Ptak et al. 2005), it has been reported that some "hotspot" locations seem to differ between human and chimpanzee when studying particular regions of the genome (Ptak et al. 2005; Wall et al. 2003; Winckler et al. 2005). Whether this tendency holds across genomes needs further validation given that only a small portion of the chimpanzee genome (chromosome 22) has been analyzed (Ptak et al. 2005). That said, the chromosomal reorganizations that are affecting the observed reduction of recombination correspond to large genomic regions (a total of 318 Mbp of the human genome) where changes in the distribution of specific "hotpots" would probably have a minor impact on the average rate of recombination. Our observations validate the relevance of chromosomal reorganizations in the speciation process thus providing support for the "suppressed recombination" model in the most famous speciation event: human and chimpanzee.

ACKNOWLEDGMENTS

We thank TJ Robinson, C. Gilbert, M. Garcia-Caldés and M. Ponsà for insightful discussions and comments on the manuscript. Financial support was received from Ministerio de Ciencia y Tecnologia (CGL-2010-20170) and the Universitat Autònoma de Barcelona (PhD fellowship to M. Farré). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTERESTS

The authors have declared that no conflicts of interests exist.

AUTHORS CONTRIBUTION

Conceived and designed the experiments: MF, ARH. Performed the experiments: MF. Analyzed the data: MF, ARH. Wrote the paper: MF, ARH.

LITERATURE CITED

Ashley T, Moses MJ, Solari AJ. 1981. Fine structure and behaviour of a pericentric inversion in the sand rat, psammomys obesus. J. Cell. Sci. 50:105-119.

Avise JC, Robinson TJ. 2008. Hemiplasy: A new term in the lexicon of phylogenetics. Syst. Biol. 57:503-507.

Badenhorst D, Dobigny G, Adega F, Chaves R, O'Brien PC, Ferguson-Smith MA, Waters PD, Robinson TJ. 2011. Chromosomal evolution in rattini (muridae, rodentia). Chromosome Res. 19:709-727.

Bailey JA, Eichler EE. 2006. Primate segmental duplications: Crucibles of evolution, diversity and disease. Nat. Rev. Genet. 7:552-564.

Baudet C, Lemaitre C, Dias Z, Gautier C, Tannier E, Sagot MF. 2010. Cassis: Detection of genomic rearrangement breakpoints. Bioinformatics 26:1897-1898.

Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, Fontenille D, Toure Y, Sagnon N. 2003. Semipermeable species boundaries between anopheles gambiae and anopheles arabiensis: Evidence from multilocus DNA sequence variation. Proc. Natl. Acad. Sci. U. S. A. 100:10818-10823.

Borodin PM, Karamysheva TV, Belonogova NM, Torgasheva AA, Rubtsov NB, Searle JB. 2008. Recombination map of the common shrew, sorex araneus (eulipotyphla, mammalia). Genetics 178:621-632.

Bourque G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. Curr. Opin. Genet. Dev. 19:607-612.

Brown JD, O'Neill RJ. 2010. Chromosomes, conflict, and epigenetics: Chromosomal speciation revisited. Annu. Rev. Genomics Hum. Genet. 11:291-316.

Brown JD, Mitchell SE, O'Neill RJ. 2012. Making a long story short: Noncoding RNAs and chromosome change. Heredity (Edinb) 108:42-49.

Brown KM, Burk LM, Henagan LM, Noor MA. 2004. A test of the chromosomal rearrangement model of speciation in drosophila pseudoobscura. Evolution 58:1856-1860.

Bullaughey K, Przeworski M, Coop G. 2008. No effect of recombination on the efficacy of natural selection in primates. Genome Res. 18:544-554.

Caceres M, Ranz JM, Barbadilla A, Long M, Ruiz A. 1999. Generation of a widespread drosophila inversion by a transposable element. Science 285:415-418.

Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, Kim SK, Wall JD, Martin D, Jurka J et al. (x co-authors. 2009. Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. PLoS Genet. 5:e1000538.

Castiglia R, Capanna E. 2002. Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of mus musculus domesticus. Genetica 114:35-40.

Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. 68:444-456.

Clark AG, Wang X, Matise T. 2010. Contrasting methods of quantifying fine structure of human recombination. Annu. Rev. Genomics Hum. Genet. 11:45-64.

Codina-Pascual M, Campillo M, Kraus J, Speicher MR, Egozcue J, Navarro J, Benet J. 2006. Crossover frequency and synaptonemal complex length: Their variability and effects on human male meiosis. Mol. Hum. Reprod. 12:123-133.

Delprat A, Negre B, Puig M, Ruiz A. 2009. The transposon galileo generates natural chromosomal inversions in drosophila by ectopic recombination. PLoS One 4:e7883.

Donthu R, Lewin HA, Larkin DM. 2009. SyntenyTracker: A tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. BMC Res. Notes 2:148.

Dumas D, Britton-Davidian J. 2002. Chromosomal rearrangements and evolution of recombination: Comparison of chiasma distribution patterns in standard and robertsonian populations of the house mouse. Genetics 162:1355-1366.

Dumont BL, Payseur BA. 2011. Evolution of the genomic recombination rate in murid rodents. Genetics 187:643-657.

Dumont BL, Payseur BA. 2008. Evolution of the genomic rate of recombination in mammals. Evolution 62:276-294.

Dutrillaux B, Couturier J. 1981. The ancestral karyotype of platyrrhine monkeys. Cytogenet. Cell Genet. 30:232-242.

Faria R, Navarro A. 2010. Chromosomal speciation revisited: Rearranging theory with pieces of evidence. Trends Ecol. Evol. 25:660-669.

Farre M, Bosch M, Lopez-Giraldez F, Ponsa M, Ruiz-Herrera A. 2011. Assessing the role of tandem repeats in shaping the genomic architecture of great apes. PLoS One 6:e27239.

Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G, Khaja R, Scherer SW. 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet. 1:e56.

Franchini P, Colangelo P, Solano E, Capanna E, Verheyen E, Castiglia R. 2010. Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse mus musculus domesticus. Evolution 64:2020-2032.

Froenicke L, Anderson LK, Wienberg J, Ashley T. 2002. Male mouse recombination maps for each autosome identified by chromosome painting. Am. J. Hum. Genet. 71:1353-1368.

Garcia-Cruz R, Pacheco S, Brieno MA, Steinberg ER, Mudry MD, Ruiz-Herrera A, Garcia-Caldes M. 2011. A comparative study of the recombination pattern in three species of platyrrhini monkeys (primates). Chromosoma 120:521-530.

Greenbaum IF, Reed MJ. 1984. Evidence for heterosynaptic pairing of the inverted segment in pericentric inversion heterozygotes of the deer mouse (peromyscus maniculatus). Cytogenet. Cell Genet. 38:106-111.

Hale DW. 1986. Heterosynapsis and suppression of chiasmata within heterozygous pericentric inversions of the sitka deer mouse. Chromosoma 94:425-432.

Hassold T, Hansen T, Hunt P, VandeVoort C. 2009. Cytological studies of recombination in rhesus males. Cytogenet. Genome Res. 124:132-138.

Hellmann I, Ebersberger I, Ptak SE, Paabo S, Przeworski M. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. Am. J. Hum. Genet. 72:1527-1535.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21:349-356.

International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P et al. (x co-authors. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-861.

Kehrer-Sawatzki H, Cooper DN. 2008. Molecular mechanisms of chromosomal rearrangement during primate evolution. Chromosome Res. 16:41-56.

Kehrer-Sawatzki H, Szamalek JM, Tanzer S, Platzer M, Hameister H. 2005. Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. Genomics 85:542-550.

King M. 1993. Species evolution. the role of chromosome change. Cambridge University Press .

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419-434.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT et al. (x co-authors. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467:1099-1103.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G et al. (x co-authors. 2002. A high-resolution recombination map of the human genome. Nat. Genet. 31:241-247.

Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. 2008. Fine-scale mapping of recombination rate in drosophila refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. U. S. A. 105:10051-10056.

Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res. 19:770-777.

Locke DP, Archidiacono N, Misceo D, cardone MF, Deschamps S, Roe B, Rocchi M, Eichler EE. 2003. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplications cluster. Genome Biology 4:50.

Longo MS, Carone DM, NISC Comparative Sequencing Program, Green ED, O'Neill MJ, O'Neill RJ. 2009. Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. BMC Genomics 10:334.

Lynn A, Ashley T, Hassold T. 2004. Variation in human meiotic recombination. Annu. Rev. Genomics Hum. Genet. 5:317-349.

Lynn A, Koehler KE, Judis L, Chan ER, Cherry JP, Schwartz S, Seftel A, Hunt PA, Hassold TJ. 2002. Covariation of synaptonemal complex length and mammalian meiotic exchange rates. Science 296:2222-2225.

Marques-Bonet T, Navarro A. 2005. Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals. Gene 353:147-154.

Marques-Bonet T, Caceres M, Bertranpetit J, Preuss TM, Thomas JW, Navarro A. 2004. Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. Trends Genet. 20:524-529.

Marques-Bonet T, Sanchez-Ruiz J, Armengol L, Khaja R, Bertranpetit J, Lopez-Bigas N, Rocchi M, Gazave E, Navarro A. 2007. On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees. Genome Biol. 8:R230.

Megens HJ, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving R, Vereijken A, Silva P, Muir WM, Cheng HH et al. (x co-authors. 2009. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. BMC Genet. 10:86.

Michel AP, Grushko O, Guelbeogo WM, Lobo NF, Sagnon N, Costantini C, Besansky NJ. 2006. Divergence with gene flow in anopheles funestus from the sudan savanna of burkina faso, west africa. Genetics 173:1389-1395.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L et al. (x co-authors. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613-617.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321-324.

Navarro A, Barton NH. 2003. Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. Science 300:321-324.

Navarro A, Ruiz A. 1997. On the fertility effects of pericentric inversions. Genetics 147:931-933.

Navarro A, Betran E, Barbadilla A, Ruiz A. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146:695-709.

Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? examining the role of restricted recombination in maintaining species. Heredity (Edinb) 103:439-444.

Noor MA, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. U. S. A. 98:12084-12088.

Paigen K, Petkov P. 2010. Mammalian recombination hot spots: Properties, control and evolution. Nat. Rev. Genet. 11:221-233.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103-1108.

Petes TD. 2001. Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. 2:360-369.

Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. Nat. Genet. 37:429-434.

Qanbari S, Hansen M, Weigend S, Preisinger R, Simianer H. 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens. BMC Genet. 11:103.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16:351-358.

Rieseberg LH, Willis JH. 2007. Plant speciation. Science 317:910-914.

Rieseberg LH, Whitton J, Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713-727.

Rieseberg LH, Linder CR, Seiler GJ. 1995. Chromosomal and genic barriers to introgression in helianthus. Genetics 141:1163-1171.

Robinson TJ, Ropiquet A. 2011. Examination of hemiplasy, homoplasy and phylogenetic discordance in chromosomal evolution of the bovidae. Syst. Biol. 60:439-450.

Robinson TJ, Ruiz-Herrera A, Avise JC. 2008. Hemiplasy and homoplasy in the karyotypic phylogenies of mammals. Proc. Natl. Acad. Sci. U. S. A. 105:14477-14481.

Rogers J, Garcia R, Shelledy W, Kaplan J, Arya A, Johnson Z, Bergstrom M, Novakowski L, Nair P, Vinson A et al. (x co-authors. 2006. An initial genetic linkage map of the rhesus macaque (macaca mulatta) genome using human microsatellite loci. Genomics 87:30-38.

Ruiz-Herrera A, Castresana J, Robinson TJ. 2006. Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol. 7:R115.

Ruiz-Herrera A, Garcia F, Giulotto E, Attolini C, Egozcue J, Ponsa M, Garcia M. 2005. Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal telomeric sequences in primates. Cytogenet. Genome Res. 108:234-247.

Rumpler Y, Warter S, Hauwy M, Fausser JL, Roos C, Zinner D. 2008. Comparing chromosomal and mitochondrial phylogenies of sportive lemurs (genus lepilemur, primates). Chromosome Res. 16:1143-1158.

Schultz J, Redfield H. 1951. Interchromosomal effects on crossing over in drosophila. Cold Spring Harb. Symp. Quant. Biol. 16:175-197.

Smukowski CS, Noor MA. 2011. Recombination rate variation in closely related species. Heredity (Edinb) 107:496-508.

Sturtevant AH. 1919. Inherited linkage variation in the second chromosome. Contributions to Genetics of Drosophila Melanogaster :305-341.

Sun F, Trpkov K, Rademaker A, Ko E, Martin RH. 2005. Variation in meiotic recombination frequencies among human males. Hum. Genet. 116:172-178.

Szamalek JM, Cooper DN, Hoegel J, Hameister H, Kehrer-Sawatzki H. 2007. Chromosomal speciation of humans and chimpanzees revisited: Studies of DNA divergence within inverted regions. Cytogenet. Genome Res. 116:53-60.

Szamalek JM, Goidts V, Cooper DN, Hameister H, Kehrer-Sawatzki H. 2006. Characterization of the human lineage-specific pericentric inversion that distinguishes human chromosome 1 from the homologous chromosomes of the great apes. Hum. Genet. 120:126-138.

Trifonov VA, Stanyon R, Nesterenko AI, Fu B, Perelman PL, O'Brien PC, Stone G, Rubtsova NV, Houck ML, Robinson TJ et al. (x co-authors. 2008. Multidirectional cross-species painting illuminates the history of karyotypic evolution in perissodactyla. Chromosome Res. 16:89-107.

Vallender EJ, Lahn BT. 2004. Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. Genomics 84:757-761.

Wall JD, Frisse LA, Hudson RR, Di Rienzo A. 2003. Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. Am. J. Hum. Genet. 73:1330-1340.

White BJ, Crandall C, Raveche ES, Hjio JH. 1978. Laboratory mice carrying three pairs of robertsonian translocations: Establishment of a strain and analysis of meiotic segregation. Cytogenet. Cell Genet. 21:113-138.

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P et al. (x co-authors. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308:107-111.

Wu ZK, Getun IV, Bois PR. 2010. Anatomy of mouse recombination hot spots. Nucleic Acids Res. 38:2346-2354.

Yannic G, Basset P, Hausser J. 2009. Chromosomal rearrangements and gene flow over time in an inter-specific hybrid zone of the sorex araneus group. Heredity (Edinb) 102:616-625.

Yunis JJ, Prakash O. 1982. The origin of man: A chromosomal pictorial legacy. Science 215:1525-1530.

Zhang J, Wang X, Podlaha O. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. Genome Res. 14:845-851.

FIGURES

Figure 1. Evolutionary history of human chromosomes superimposed on the phylogeny of great apes. Black lines within the phylogenetic tree represent the ancestral state of the chromosomes, whereas dark grey and light grey lines represents the rearranged forms. Orangutan maintains the ancestral form for orthologous chromosomes 3 and 11 while human, chimpanzee and gorilla forms are derived. Orthologous chromosomes 1, 2 and 18 have been rearranged in the lineage leading to humans, whereas orthologous chromosomes 4, 9, 15, 16 and 17 are rearranged in the lineage leading to chimpanzee. Ancestral chromosome 5 has been maintained in orangutan and human but has suffered 2 independent inversions in chimpanzee and gorilla, respectively. Chromosome 7 has suffered one inversion, which has been fixed in gorilla, and another inversion has fixed in the lineage leading to human and chimpanzee. Chromosome 10 underwent one inversion, that was fixed in human and chimpanzee, and a new inversion fixed in gorilla. Finally, chromosome 12 has maintained the ancestral form in humans and orangutans but has undergone an inversion that is fixed in chimpanzee and gorilla, therefore, the polymorphic state has persisted across multiple speciation nodes (gorilla-human-chimpanzee and human-chimp).

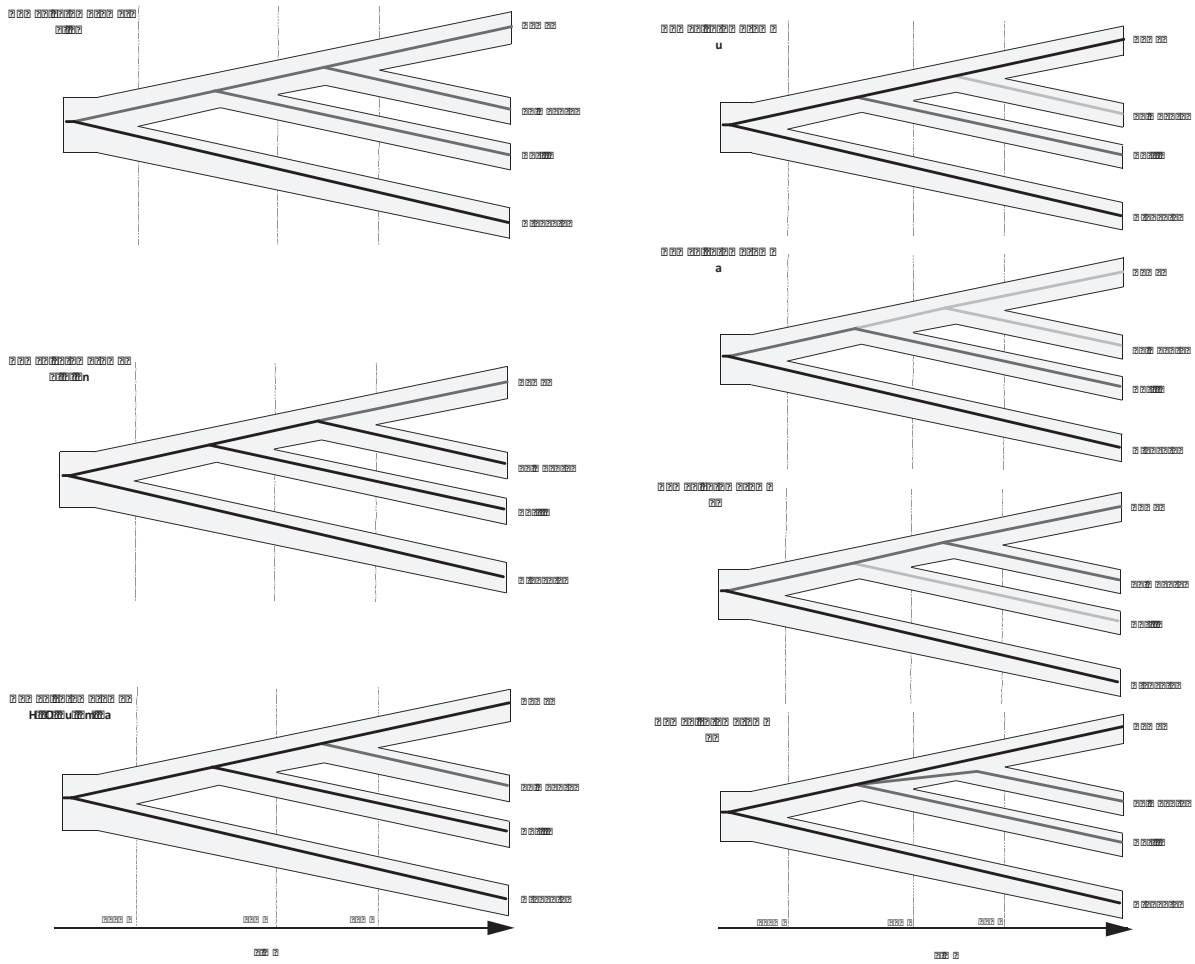


Figure 2. **Standardized recombination rate (SRR) along the human chromosome 4.** SRR (y-axes) are shown as needles across non-overlapping windows of 10 kb in the whole chromosomal length (x-axes). The genomic region affected by an inversion is depicted in grey, whereas non-inverted regions are showed in black. White rectangle indicates the centromere. Average recombination rate for each region is shown in numbers in the x-axes.

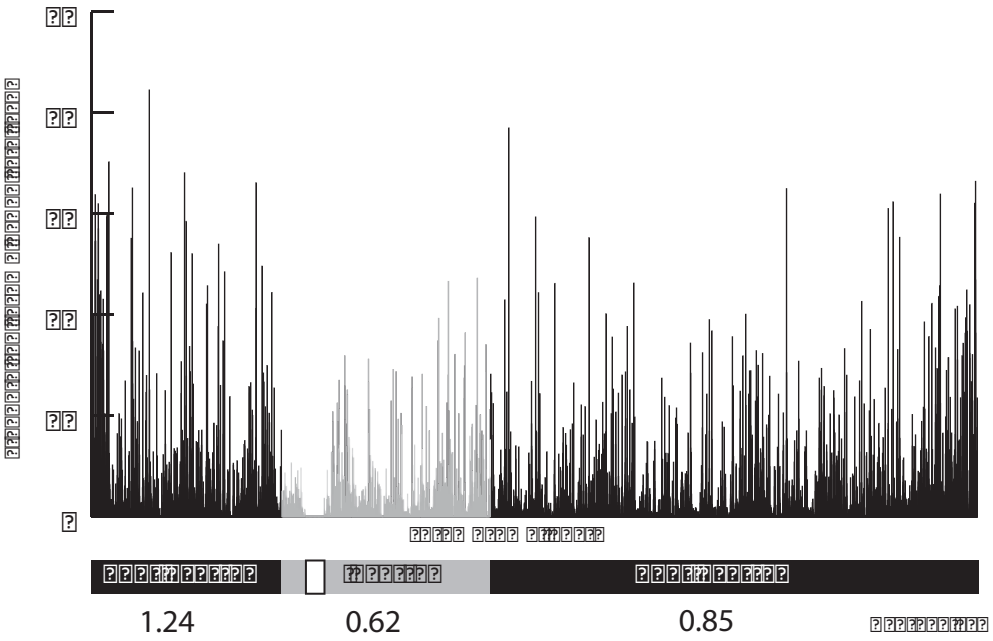


Figure 3. Schematic outlining a scenario where chromosomal polymorphisms would persist in an ancestral human-chimp population. In an ancestral population, an initial chromosome rearrangement (inversion) occurs. Following hybridization between carries with both chromosomal forms, a heterokaryotypic F_1 hybrid results. The inverted chromosomal form would be maintained as a floating polymorphism during ~ 6 myr of coalescence time since the last common node for human-chimp-orangutan (estimated at 10 mya; Hobolth et al. 2011), and the human-chimp node (estimated at 4 mya; Hobolth et al. 2011). During this period, the inversion would behave as a porous barrier, reducing recombination within the inverted region (in grey) of homologous chromosomes. Eventually some of the inverted chromosomal forms could be fixed in the lineage leading to chimpanzees, whereas others could be fixed in the human lineage. Whatever the case, genomic regions that were involved in the reorganizations would persist as regions with low recombination rates in the present populations.

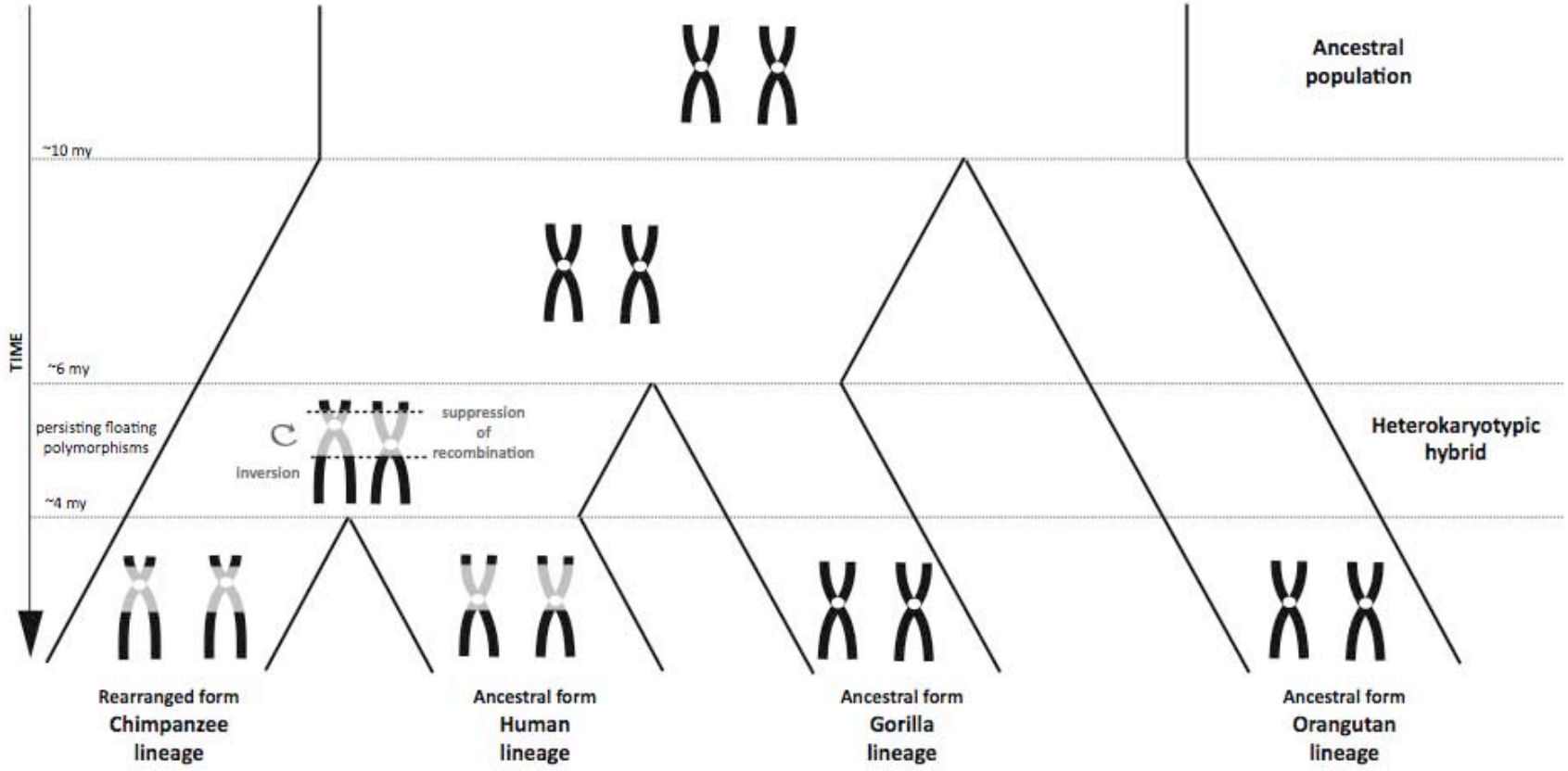

Ancestral
population
~10 my
~6 my
~4 my
TIME
persisting floating
polymorphisms
inversion
suppression
of
recombination
Heterokaryotypic
hybrid
Rearranged form
Chimpanzee
lineage
Ancestral form
Human
lineage
Ancestral form
Gorilla
lineage
Ancestral form
Orangutan
lineage

Table 1. Breakpoints of the reorganizations detected between the human and chimpanzee genomes. The data correspond to the human chromosome involved in reorganization: start and end positions of the breakpoints, length of the evolutionary regions and type of reorganizations detected (macro- and micro-reorganization) is shown.

HSA chr	Start	End	Length (bp)	Type of rearrangment
1	110,210,397	110,218,053	7,656	macro
	147,737,500	147,908,988	171,488	macro
	205,922,669	206,074,582	151,913	micro
	206,332,100	206,482,531	150,431	micro
2	3,049,226	3,061,801	12,575	micro
	3,529,313	3,583,081	53,768	micro
	5,018,789	5,163,731	144,942	micro
	114,328,435	114,382,002	53,567	macro
4	44,810,366	44,818,018	7,652	macro
	85,958,259	85,963,014	4,755	macro
5	18,552,882	18,555,128	2,246	macro
	95,921,427	95,921,474	47	macro
7	39,663,596	39,673,431	9,835	micro
	44,072,917	44,076,368	3,451	micro
9	46,947,614	47,060,133	112,519	macro
	88,801,017	88,930,700	129,683	macro
10	46,764,052	46,861,824	97,772	micro
	51,504,215	51,505,370	1,155	micro
	51,914,036	51,917,603	3,567	micro
	81,259,753	81,310,866	51,113	micro
	81,969,374	82,018,118	48,744	micro
12	20,963,042	20,963,853	811	macro
	68,381,348	68,404,608	23,260	macro
15	28,957,462	28,962,654	5,192	macro
16	34,195,797	34,197,035	1,238	macro
	46,498,803	46,500,515	1,712	macro
17	7,930,990	7,930,995	5	macro
	47,615,108	47,621,198	6,090	macro
18	18,510,899	18,646,538	135,639	macro
19	36,739,338	36,822,850	83,512	micro
	37,742,181	37,828,007	85,826	micro
X	52,612,063	52,684,105	72,042	micro
	52,871,611	52,878,711	7,100	micro
	71,998,917	72,000,811	1,894	micro
	72,208,690	72,222,380	13,690	micro
Y	14,366,181	14,372,744	6,563	micro
	18,747,930	18,756,681	8,751	micro

Table 2. Average values of the recombination rates (SRR**) in each human chromosome.** Recombination rates are significantly different among chromosomes (Kruskal-Wallis, $p < 0.0001$) and are negatively correlated with chromosomal size (Spearman's $\rho = -0.915$, $p < 0.0001$).

HSA chr	Mean SRR	Standard Error
1	0.899	0.016
2	0.856	0.015
3	0.881	0.018
4	0.883	0.018
5	0.895	0.019
6	0.859	0.018
7	0.910	0.020
8	0.910	0.023
9	0.856	0.021
10	0.978	0.023
11	0.915	0.021
12	0.948	0.021
13	1.031	0.032
14	1.055	0.033
15	1.198	0.035
16	1.140	0.035
17	1.277	0.037
18	1.165	0.041
19	1.140	0.037
20	1.275	0.046
21	1.506	0.081
22	1.559	0.074

Table 3. Comparison of means recombination rates in each type of chromosome and region (inverted, non-inverted or collinear). Rearranged chromosomes exhibited a lower recombination rate than do collinear chromosomes (Mann-Whitney's U, $p < 0.0001$). Recombination rate is significantly lower in inverted regions compared to collinear and non-inverted regions (Kruskal-Wallis test $p < 0.0001$).

Type of region	Mean SRR	Standard Error
Collinear	0.975	0.009
Rearranged	0.944	0.006
Collinear	0.975	0.009
Non-inverted	1.001	0.007
Inverted	0.715	0.012

Table 4. Comparison of means of recombination rate in regions that have suffered macro- and micro-rearrangements. Recombination rate is lower in regions affected by macro-rearrangements compared to those affected by micro-rearrangements or those that are not rearranged (Kruskal-Wallis, $p < 0.0001$).

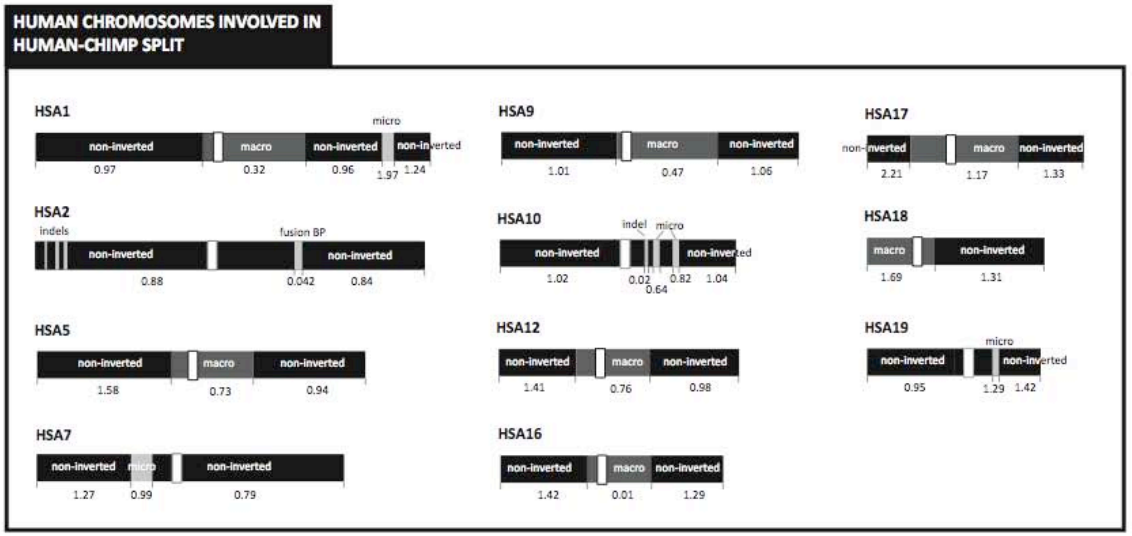
Type of rearrangement	Mean SRR	Standard Error
Non rearranged	0.996	0.005
Micro-rearrangement	0.976	0.108
Macro-rearrangement	0.713	0.012

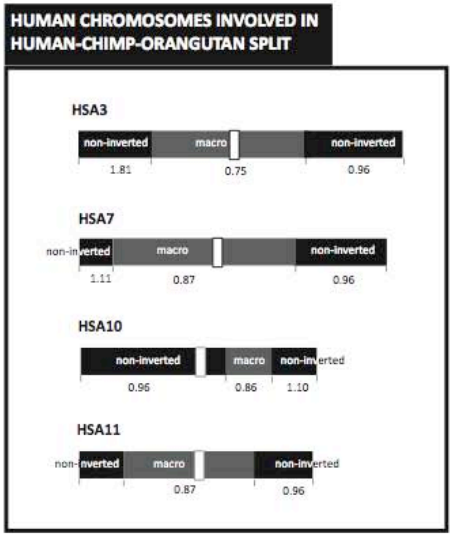
SUPPLEMENTARY MATERIAL

Additional File 1. Evolutionary breakpoint regions between human and chimpanzee genomes. Comparison of evolutionary breakpoint regions detected by Cassis and SyntenyTracker with previously published data (Feuk et al. 2005; Kehrer-Sawatzki and Cooper 2008 and references therein).

HSA_chr	CASSIS DATA			Appears in Synteny Tracker?	KEHRER-SAWATZKI & COOPER 2008			FEUK et al 2005
	START	END	LENGTH (bp)		BAC	BAC_start (hg18)	BAC_end (hg18)	
1	110.210.397	110.218.053	7.656	✔	RP11-439A17	120.747.157	120.936.695	✔
1	147.412.900	148.795.315	1.382.415	✔	RP11-495P10	147.737.500	147.908.988	✔
1	205.922.669	206.074.582	151.913	✖				✔
1	206.332.100	206.482.531	150.431	✖				✖
2	3.049.226	3.061.801	12.575	✖				✔
2	3.529.313	3.583.081	53.768	✖				✔
2	5.018.789	5.163.731	144.942	✖				✖
2	114.328.435	114.382.002	53.567	✔				✖
4	44.810.366	44.818.018	7.652	✔	RP11-779N22	44.815.150	44.834.253	✔
4	85.958.259	85.963.014	4.755	✔	RP11-8N8	85.816.998	86.000.969	✔
5	18.552.882	18.555.128	2.246	✔	RP11-35A11	18.430.127	18.568.970	✔
5	95.921.427	95.921.474	47	✔	RP11-432G16	95.827.191	95.900.731	✔
7	39.663.596	39.673.431	9.835	✔				✔
7	44.072.917	44.076.368	3.451	✔				✔
8	12.372.819	12.400.736	27.917	✔				✖
9	-	-	-	✖	RP11-259A5	46.947.614	47.060.133	✖
9	-	-	-	✖	RP11-507D14	88.801.017	88.930.700	✖
10	46.764.052	46.861.824	97.772	✖				✔
10	51.504.215	51.505.370	1.155	✖				✔
10	51.914.036	51.917.603	3.567	✖				✔
10	81.259.753	81.310.866	51.113	✖				✖
10	81.969.374	82.018.118	48.744	✔				✖
12	20.963.042	20.963.853	811	✔	RP11-80N2	20.944.218	21.117.820	✔
12	68.381.348	68.404.608	23.260	✔	RP3-491B7	66.593.873	68.695.640	✔
15	28.957.462	28.962.654	5.192	✖				✔
16	34.195.797	34.197.035	1.238	✖	CTD-2144E22	34.173.151	34.341.696	✖
16	46.498.803	46.500.515	1.712	✖	RP11-696P19	46.385.802	46.508.594	✖
17	7.930.990	7.930.995	5	✔	RP1-179H24	7.927.649	7.939.128	✖
17	47.615.108	47.621.198	6.090	✔	RP5-1029K10	47.587.437	47.749.643	✔
18	15.366.244	18.520.344	3.154.100	✔	RP11-666N19	18.510.899	18.646.538	✔
18	-	-	-	✖	RP11-683L23	46.421	151.961	
19	36.739.338	36.822.850	83.512	✖				✔
19	37.742.181	37.828.007	85.826	✖				✔
X	52.612.063	52.684.105	72.042	✖				✔
X	52.871.611	52.878.711	7.100	✖				✔
X	71.998.917	72.000.811	1.894	✖				✔
X	72.208.690	72.222.380	13.690	✖				✔
Y	14.366.181	14.372.744	6.563	✖				--
Y	18.747.930	18.756.681	8.751	✖				--

Additional File 2. Average recombination rate for the different genomic regions (inverted and non-inverted) in rearranged chromosomes. Regions affected by macro-rearrangement are marked in dark grey and micro-rearrangement in light grey. White rectangles indicate centromeres.





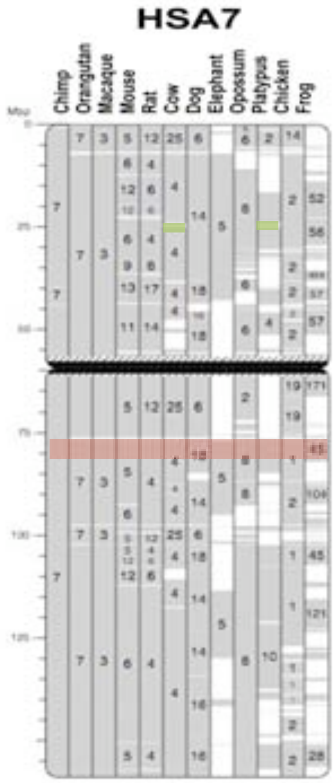
4 DISCUSSIÓ

.

4.1 COMPARACIÓ DE LES TÈCNIQUES EMPRADES PER A L'ESTUDI DE LES REORGANITZACIONS CROMOSÒMIQUES

Hem vist que per a poder estudiar les regions de sintènia conservades així com les reorganitzacions cromosòmiques ocorregudes entre espècies es poden utilitzar diferents aproximacions metodològiques, resultant en diferents graus de resolució. Aquestes aproximacions engloben des de tècniques citogenètiques de bandeig cromosòmic (Bandes G i R) i pintat cromosòmic (Zoo-FISH) fins a reconstruccions *in silico* de les reorganitzacions, passant per tècniques moleculars com els mapes de lligament i els mapes híbrids de radiació (veure secció 1.1.1). Tot i la gran utilitat de les aproximacions experimentals, en aquesta tesi ens hem centrat en tres metodologies bioinformàtiques diferents per a reconstruir els blocs sintènics (HSBs) i les regions de trencament evolutiu (EBRs) entre els genomes de diferents espècies de mamífers i vertebrats: (i) Mòdul Ensembl-Compara, (ii) SyntenyTracker (Donthu i col, 2009) i (iii) CASSIS (Baudet i col, 2010).

La metodologia emprada per la primera aproximació (Mòdul Ensembl-Compara) es divideix en dues etapes. En una primera fase, partint de l'alineament de la seqüència dels genomes de dues espècies mitjançant l'algoritme BlastZ-net (Schwartz i col, 2003) s'agrupen les regions alineades que es troben a una distància inferior de 200 Kpb formant els blocs sintènics provisionals. Posteriorment, en una segona etapa s'agrupen les regions sintèniques obtingudes en la primera fase sempre i quan no hi hagi més de 2 regions no alineades entre elles i estiguin a una distància inferior a 3Mpb. Per tant, l'objectiu d'aquest mètode és la detecció de HSBs entre dues espècies basant-se en la seqüència nucleotídica. Pel que fa al SyntenyTracker, l'algoritme parteix de les posicions dels gens ortòlegs entre dues espècies per a crear els blocs sintènics provisionals. Seguidament, s'agrupen els blocs obtinguts entre ells en funció de la distància que els separa donat un llindar fixat per l'usuari (que pot anar des de 250 Kpb fins a 1 Mpb) (Fig. 13). Per tant, igual que el Mòdul Ensembl-Compara, també es focalitza en la detecció de HSBs entre dues espècies. L'avantatge d'aquest


HSA7
Mbp
Chimp
Orangutan
Macaque
Mouse
Rat
Cow
Dog
Elephant
Opossum
Platypus
Chicken
Frog

Taula 1. Comparació dels mètodes emprats per a detectar reorganitzacions entre humà i ximpanzé.

Mètode	HSBs		EBRs	
	Nº	Rang de mida (pb)	Nº	Rang de mida (pb)
Ensembl-Compara v64	108	102.321-109.628.017	87	1-91.977.771
SyntenyTracker	39	2.730.871-197.532.312	15	13.495-4.150.943
CASSIS	66	1.257-199.446.826	38	5-3.905.248

Així doncs, la utilització d'un o altre algoritme en els nostres treballs ha vingut definida pels objectius proposats. En el cas de la reconstrucció del cariotip ancestral (Treball 1), l'objectiu era establir els blocs sintènics entre el màxim nombre d'espècies de vertebrats per tal de determinar les regions que s'han conservat al llarg de l'evolució d'aquestes espècies. És per això que vam utilitzar el SyntenyTracker, ja que ens permetia obtenir HSBs de mida gran basats en la posició de gens ortòlegs. Tenint en compte que la fracció codificant del genoma està més conservada que la no-codificant [en humà només entre un 0,3 i un 2% de la fracció no codificant està conservada (Lander i col, 2001;Dermitzakis i col, 2005)], escollir un mètode que només es basés en l'alineament de la seqüència de DNA, i no en els gens ortòlegs, podria haver donat lloc a falsos positius en espècies filogenèticament llunyanes, com ara humà i gall (~300 Ma) o humà i granota (~350 Ma), ja que s'haurien d'haver emprat paràmetres menys restrictius per tal d'alinear aquestes seqüències.

En canvi, els objectius de la nostra aproximació a l'estudi de les regions de trencament evolutiu (EBRs) van ser, per una banda acotar amb la màxima resolució aquestes regions per estudiar-ne la distribució i el seu contingut (Treball 2) així com determinar l'efecte de la reorganització que delimiten sobre la taxa de recombinació (Treball 4). Per tant, un mètode que es focalitzés en l'acotament de les EBRs com és el CASSIS era el més adequat en tots dos casos. Malauradament, a l'inici del Treball 2 aquest mètode no havia estat descrit, i per tant, vam utilitzar l'implementat per l'Ensembl-Compara. L'ús d'aquest algoritme ens va permetre obtenir blocs sintènics grans i EBRs amb un rang de mida ampli. Per tal de evitar l'anàlisi de falsos positius o regions amb *gaps* de seqüència, vam seguir l'aproximació d'estudis anteriors (Ruiz-Herrera i col,

2006; Larkin i col, 2009) i vam desestimar aquelles EBRs més grans de 4 Mpb. Tot i així, vam poder determinar un número elevat de EBRs. De forma addicional vam estimar que més del 71% del genoma humà està conservat al comparar-lo amb les 10 espècies estudiades (Farré i col, 2011). Pel que respecta al Treball 4, necessitàvem una resolució de les EBRs més alta, ja que volíem comparar les EBRs amb la distribució de taxes de recombinació en el genoma humà en finestres de 10 kpb (Kong i col, 2010). Per aquesta raó vam utilitzar el CASSIS per a definir les EBRs entre el genoma humà i el de ximpanzé, que ens va ajudar a acotar les EBRs amb un alt grau de resolució (entre 5 pb i 1.8 Mpb).

Cal remarcar que en tots els treballs presentats en aquesta tesi les dades obtingudes a partir dels tres algoritmes han passat per un procés de curació manual per tal d'evitar la introducció de falsos positius en els estudis. En el cas del treball 1, els HSBs obtinguts entre totes les espècies es van corroborar amb dades prèvies en aquells casos on s'havien descrit les homologies cromosòmiques per tècniques citogenètiques (Bandes G, Zoo-FISH). En el treball 2, vam redefinir les EBRs més grans de 4Mpb com a *gaps* i no les vam incloure en l'estudi. I finalment, en el treball 4, vam verificar les regions definides amb l'algortime CASSIS comparant-les amb dades obtingudes fent servir el SyntenyTracker i dades d'estudis anteriors tant experimentals com bioinformàtics (Kehrer-Sawatzki i Cooper, 2008; Feuk i col, 2005).

Tot i que hem sigut conservadors a l'hora de descriure els HSBs i les EBRs, hem de tenir en compte que tots els mètodes anteriors es basen en la localització dels gens ortòlegs i en la seqüència genòmica de referència tant d'humà com de les altres espècies. Aquesta aproximació pot provocar un biaix en el resultats obtinguts ja que en tots els casos, la seqüència de referència representa un únic individu i per tant un únic haplotip i pot ser que no sigui el majoritari per a totes les espècies. A més a més, en l'assemblatge dels genomes en algunes de les espècies utilitzades en aquest estudi no s'han emprat marcadors com ara BACs ni mapes híbrids de radiació per a poder establir l'ordre dels fragments seqüenciats, resultant en reconstruccions errònies dels genomes de referència a partir dels quals treballem. És per això que seria interessant no només estudiar les reorganitzacions cromosòmiques a nivell bioinformàtic o molecular, sinó que utilitzar una aproximació mixta com ara les tècniques moleculars

basades en la seqüenciació de nova generació (*Next Generation Sequencing*, NGS) que permetran millorar la resolució de les tècniques moleculars i fer estudis utilitzant genomes sencers per a validar les reorganitzacions descrites a nivell bioinformàtic. Actualment hi ha dues metodologies que permeten detectar delecions, insercions, inversions, translocacions i duplicacions entre els genomes complerts de dues espècies o dos individus de la mateixa espècie: *paired-end mapping* i *split read* (Alkan i col 2011; Le Scouarnec i Gribble, 2012). El *paired-end mapping* parteix de la seqüenciació dels extrems d'un mateix fragment de DNA (*paired-ends*) i el mapatge d'aquests extrems en el genoma de referència d'una espècie. Els extrems del mateix fragment haurien de mapar en el genoma de referència a una distància concreta, corresponent al fragment que engloben, i quan no és així, això indica que hi ha una reorganització en aquesta regió (Korbel i col, 2007). En canvi, el *split read* es basa en l'alineament de fragments seqüenciats (*reads*) amb el genoma de referència; les reorganitzacions es poden determinar quan el *read* no s'alinea completament amb el genoma de referència i es troba dividit (Mills i col, 2006). La majoria d'estudis fent servir aquests dos mètodes han detectat reorganitzacions entre individus de la mateixa espècie, sobretot en humans (Mills i col, 2006; Korbel i col, 2007; Uddin i col, 2011, entre d'altres). Però en pocs casos s'han emprat per a analitzar les reorganitzacions cromosòmiques entre dues espècies (Griffin i col, 2008; Gazave i col, 2011). El desenvolupament de nous algoritmes i l'ús d'aquestes tècniques en un futur pròxim ens permetrà detectar amb més precisió les regions genòmiques reorganitzades entre espècies.

4.2 AVANÇOS EN LA RECONSTRUCCIÓ DE CARIOTIPS ANCESTRALS DELS VERTEBRATS

Tradicionalment, els cariotips ancestrals dels diferents grups de mamífers s'han establert mitjançant aproximacions citogenètiques amb la tècnica de Zoo-FISH (p.e. Froenicke, 2005; veure apartat 1.1.2). El principal problema d'aquesta tècnica però és

que no és aplicable en estudis on es comparen espècies filogenèticament llunyanes, com ho són el grup dels euteris i dels metateris (que van divergir fa 190 Ma; Meredith i col, 2011). Aquest fet, juntament amb la creixent disponibilitat dels genomes de moltes espècies representants de clades diferents, ha fet que actualment s'estiguin aplicant tècniques bioinformàtiques per a definir el cariotip de nodes ancestrals, com ara els amniotes i els vertebrats.

El treball 1 d'aquesta tesi s'emmarca dins d'aquesta aproximació metodològica. Partint dels HSBs obtinguts utilitzant el SyntenyTracker hem proposat el cariotip ancestral de mamífers, de tetràpodes i de vertebrats. Per fer-ho, hem emprat els genomes de diferent espècies: humà, ximpanzé, orangutan, macaco, ratolí, rata, vaca i gos com a representants de *Boreoeutheria*, el genoma de l'armadillo com a part dels *Xenarthra*, el genoma de l'elefant i el tenrec com a part dels *Afrotheria*; tots ells inclosos dins del clade de Placentalia o *Eutheria*. Hem inclòs el genoma de l'opòssum o sariga com a marsupial per a englobar tots els *Theria*; i l'ornitorinc com a monotrema dins dels *Prototheria*, que juntament amb els *Theria* formen el clade de mamífers (Fig. 14). A més a més hem analitzat el genoma del gall com a espècie *outgroup* de mamífers, que juntament amb la resta forma part del clade d'amniotes; i finalment, el genoma de la granota, seqüenciat recentment (Hellsten i col, 2010), per tal de tenir un representant del grup dels amfibis i poder reconstruir el cariotip ancestral dels tetràpodes (Fig. 14).

Gràcies a la nostra aproximació hem pogut refinar el cariotip ancestral de Placentaris (*Placental Ancestral Karyotype*, PAK) utilitzant com a *outgroup* els genomes de gall i opòssum. Proposem un PAK que estaria format per n = 23 cromosomes, dels quals 10 es mantenen com un cromosoma *in toto* en el genoma humà: HSA 1, 5, 6, 9, 11, 13, 17, 18, 20 i X. De forma addicional, hem definit set associacions sintèniques (4q/8p/4pq, 21/3, 14/15, 10p/12pq/22qt, 19q/16q, 16p/7a i 12qt/22q) i sis fragments cromosòmics humans que formaven cromosomes ancestrals sencers: 2q, 7b, 2p-q13, 10q, 8p i 19p (veure Fig. 2a, treball 1). Tenint en compte que tant els cromosomes 13 i 18, els fragments cromosòmics corresponents als cromosomes humans 10q, 19p i 8q com les associacions sintèniques 4q/8p/4pq, 21/3, 14/15, 10p/12pq/22qt, 19q/16q, 16p/7a i 12qt/22q estan presents en gall i/o opòssum, podem dir que són simplesiomorfies

(caixa 2), és a dir, formes ancestrals d'aquests caràcters compartides per tots els placentaris, teris i, fins i tot, amniotes. En canvi, els cromosomes del PAK corresponents als cromosomes humans 1, 5, 6, 9, 11, 17, 20 i X, més els fragments cromosòmics 7b, 2p-q13 i 2qt no estan presents com a blocs independents en els genomes de gall ni opòssum, per tant es consideren caràcters sinapomòrfics (CAIXA 2), és a dir, formes derivades comuns en el clade dels placentaris que marquen el seu origen monofilètic. Estudis anteriors proposaven un PAK amb n=22 cromosomes, ja que definien l'associació sintènica 1/19p (Yang i col 2003; Graphodatsky i col 2011) perquè es trobava en espècies del grup *Afrotheria*. Nosaltres no l'hem inclosa al no trobar-se ni en gall ni en opòssum, i per tant, molt possiblement aquesta associació cromosòmica pot ser una sinapomorfia del clade *Afrotheria*. Kemkemer i col·laboradors (2009) tampoc van incloure l'associació sintènica 1/19p en el PAK ja que no la van detectar en el genomes del opòssum ni del gall fent servir la tècnica de E-painting.

A l'incloure el genoma de gall i de granota com a espècies *outgroup* hem pogut anar més enllà del cariotip ancestral de mamífers. Malgrat que el genoma del gall es va seqüenciar l'any 2004 i està assemblat en cromosomes de manera fiable (Hillier i col, 2004), el genoma de granota no està reconstruït en cromosomes, només està representat per fragments del genoma (*scaffolds* no assemblats) d'una mida mitjana de 1,57 Mpb (Hellsten i col 2010); per tant, només hem pogut analitzar les associacions sintèniques presents en les regions assemblades en *scaffolds,* subestimant les associacions sintèniques que englobin més d'un *scaffold.* En gall hem trobat les associacions sintèniques corresponents als cromosomes humans: 3p/21, 8p/4pq, 7a/16p, 12qt/22q, 12pq/22qt, 14/15 i 16q/19q. I en granota les mateixes excepte la 16q/19q. Per tant, podem dir que aquestes associacions sintèniques són precedents a la separació fa aproximadament 360 Ma dels amniotes i els amfibis i molt possiblement representen simplesiomorfies en el grup dels tetràpodes. En canvi, l'associació sintènica 16q/19q és característica dels amniotes i per tant, és una forma derivada o sinapomorfia d'aquest clade.

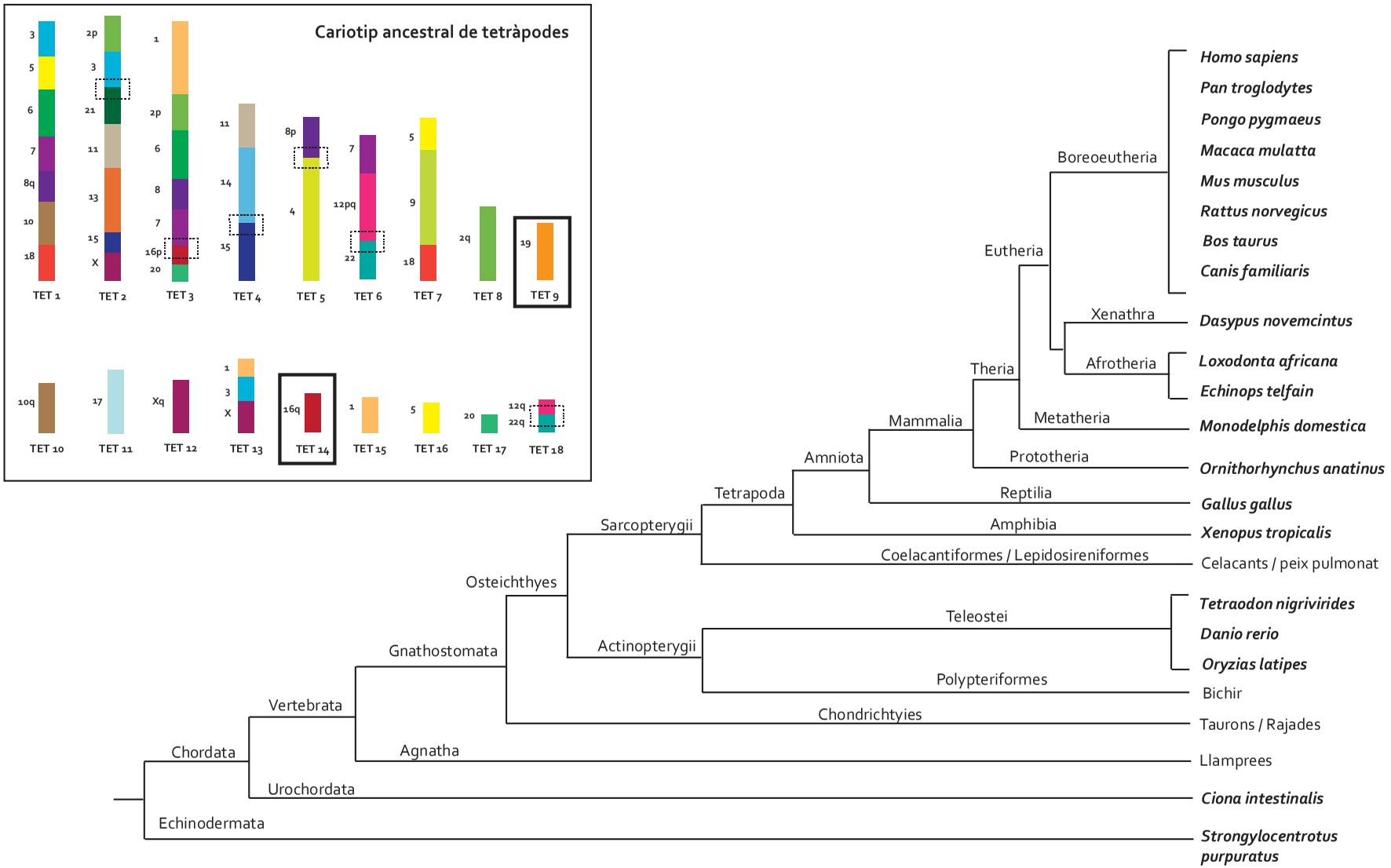


Figura 14. Relacions taxonòmiques de les espècies del Phylum Animalia, d'acord amb International Code of Zoological Nomenclature (1999). En negreta aquelles espècies incloses en els treballs de Nakatani i col. (2007), Kohn i col (2006) i el nostre treball 1. Les espècies que no estan en negreta formen part de llinatges a tenir en compte a l'hora de definir cariotips ancestrals de nodes més profunds. En el quadre hi ha representat el cariotip ancestral de tetràpodes; a l'esquerra de cada cromosoma ancestral de tetràpodes (TET) es troben les homologies amb els cromosomes humans. Hem refinat el cariotip proposat per Kohn i col (2006), les diferències estan emmarcades en línies continues i les associacions sintèniques confirmades per la nostra aproximació en línies discontinues.

Dos estudis anteriors havien definit el cariotip d'amniotes i de vertebrats, analitzant el genomes de diferent espècies (Kohn i col. 2006; Nakatani i col. 2006). Kohn i col·laboradors (2006) empraven el genoma humà, el de gall, el del peix globus (pufferfish, *Tetraodon nigrovirides*), el del peix zebra (zebrafish, *Danio rerio*) i el de medaka (*Oryzias latipes*) per a definir el cariotip ancestral de tetràpodes i de vertebrats. Tot i que els autors proposen un cariotip ancestral per als tetràpodes, no van incloure cap espècie del grup d'amfibis, per tant, en aquest cas seria més adequat avocar per la definició del cariotip ancestral d'amniotes (aus i mamífers) (Fig. 14). Si tenim en compte aquest fet, el cariotip ancestral proposat estaria constituït per n=18 cromosomes i les associacions sintèniques que presenten són les mateixes proposades per el nostre estudi, excepte per la associació 1/19p. Probablement, Kohn i col·laboradors (2006) van incloure aquesta associació ja que es basaven en el cariotip ancestral de placentaris proposat per Yang i col·laboradors (2003). Per el que fa al cariotip ancestral de vertebrats, Kohn i col·laboradors (2006) proposen un complement format per n= 11, tot i que més correctament seria el cariotip ancestral de sarcopterigis i actinopterigis, ja que no inclouen cap espècie representativa d'agnats (per exemple la llamprea) ni de condrictis (com el tauró). Pel que respecta al segon estudi, Nakatani i col·laboradors (2006) van incloure les espècies d'humà, ratolí, gos, gall i peix globus, i van emprar com a *outgroups* dels vertebrats les espècies *Ciona intestinalis* (un urocordat) i l'eriçó lila (purple sea urchin, *Strongylocentrotus purpuratus).* Així van definir un cariotip ancestral dels amniotes de n=26 cromosomes i dels vertebrats de n=10-13 cromosomes. Aquests autors postulen que després de l'aparició dels primers vertebrats, va donar-se una duplicació de tot el genoma sencer i posteriors reorganitzacions cromosòmiques, donant lloc als gnatostomats (vertebrats amb mandíbula) amb un cariotip ancestral de n=40. Aquest grup va patir fusions i reorganitzacions addicionals i va desembocar en l'ancestre de sarcopterigis i actinopterigis amb un cariotip ancestral de n=31, d'acord amb el que proposen Voss i col·laboradors (2011) a l'estudiar els grups de lligament de la granota, el gall i la salamandra. A més a més Nakatani i col·laboradors (2006) proposen que en el llinatge dels metateris (opòssum) s'han donat moltes fusions; això és consistent amb els resultats obtinguts en el treball 2 (Farré i col, 2011; Figura 3) on vam determinar una taxa de reorganització per aquesta espècie de 1 EBR/milió d'anys (Ma). Així mateix,

Kohn i col·laboradors (2006) proposen que en placentaris hi ha hagut una intensa reestructuració del genoma, consistent amb les nostres dades, ja que trobem taxes de reorganització que arriben a 1,95 EBRs/Ma en el llinatge de ratolí (Farré i col, 2011; Figura 3).

Per tant, el nostre treball ha representat un avançament en l'estudi dels cariotips ancestrals ja que hem inclòs un representant de tetràpodes no amniotes (granota) per així poder determinar les simplesiomorfies i sinapomorfies del llinatge de tetràpodes (Taula 2 i Fig. 14). Gràcies a la inclusió del genoma de granota, nosaltres hem pogut clarificar el cariotip ancestral d'amniotes i de tetràpodes, proposant un cariotip molt semblant al de Kohn i col·laboradors (2006), amb un número haploide pròxim als 18 cromosomes (Fig. 14). Tot i així és important remarcar que cal incloure més espècies representatives de cada llinatge dins dels tetràpodes (espècies addicionals de rèptils i amfibis) i dels vertebrats (varis ordres de peixos sarcopterigis i actinopterigis com també espècies de taurons i llamprees) per a determinar el cariotip ancestral d'aquests llinatges. És per això que els resultats esperats del projecte internacional de seqüenciació del genoma de 10.000 espècies de vertebrats i el seu correcte assemblatge (G10K Community Scientists, 2009), proporcionarà una eina molt important per avançar en aquesta àrea de la genòmica comparativa (Lewin i col, 2009).

Taula 2. Associacions sintèniques del cariotip ancestral de tetràpodes descrites per Kohn i col·laboradors (2006), Nakatani i col·laboradors (2007) i el nostre treball 1.

Associació sintènica	Kohn 2006	Nakatani 2007	Treball 1
3/21	SI	NO	SI
4/8p	SI	NO	SI
7a/16p	SI	SI	SI
12q/22q	SI	SI	SI
12pq/22qt	SI	SI	SI
14/15	SI	SI	SI
1/19p	SI	NO	NO
16q/19q	SI	NO	NO

4.3 PUNTS DE TRENCAMENT EVOLUTIUS: CAUSES I CONSEQÜÈNCIES DE LA SEVA DISTRIBUCIÓ

Hem vist com la reconstrucció de cariotips ancestrals ha permès comprovar que la major part del genoma de mamífers està altament conservat; hi ha grans blocs sintènics i cromosomes sencers que s'han mantingut al llarg de l'evolució (Kohn i col, 2006; Nakatani i col, 2007; Ruiz-Herrera i col, 2012). Però aquest treballs també han posat de manifest que hi ha regions que han patit reorganitzacions, delimitades per les anomenades regions de trencament evolutiu (EBRs), com també ha ajudat a poder determinar quins tipus de reorganitzacions es donen en cada llinatge (Burt i col, 1999; Bourque i col, 2005). La localització d'aquestes regions de trencament evolutives en els genomes de diferents espècies així com el seu origen evolutiu és objecte de debat en l'actualitat. De fet, la presència de EBRs s'ha relacionat amb seqüències repetitives del genoma, com ara duplicacions segmentals (Bailey i col, 2006; Carbone i col, 2006; Kehrer-Sawatzki i Cooper, 2008), elements mòbils (Cáceres i col, 1999; Carbone i col, 2009; Larkin i col, 2009; Longo i col, 2009) i repeticions en tàndem (Ruiz-Herrera i col, 2006). També s'ha relacionat la seva distribució amb la presència de gens i el possible efecte de la selecció natural per evitar la fixació de reorganitzacions deletèries (Peng i col, 2006; Becker i Lenhard, 2007; Lemaitre i col, 2009), i, més recentment amb l'estructura tridimensional de la cromatina dins del nucli interfàsic (Lemaitre i col, 2009; Veron i col, 2011). Tot i aquests estudis, els possibles mecanismes i els factors causants de la distribució de les EBRs no estan clars. És per això que ens vam plantejar estudiar el paper de diferent factors que podrien determinar la distribució de les EBRs, com ara (i) les repeticions en tàndem i altres seqüències repetitives (Treball 2, Farré i col, 2011), (ii) l'existència de constrenyiment funcional per al manteniment de HSBs i l'impediment de la fixació d'algunes reorganitzacions cromosòmiques (Treball 3, Ruiz-Herrera i col, 2011) i, finalment, (iii) l'efecte de les reorganitzacions en la distribució de la recombinació meiòtica (Treball 4, Farré i Ruiz-Herrera, 2012).

4.3.1 *Fragile breakage model*: la distribució depèn de la seqüència de DNA

L'any 2003, Pevzner i Tesler van proposar el model de trencaments fràgils (*fragile breakage model*), on exposaven que els trencaments evolutius es donaven en llocs concrets del genoma i no a l'atzar com prèviament s'havia indicat (Nadeau i Taylor, 1986). A més a més, van veure que una fracció d'aquests trencaments havia estat reutilitzada al llarg de l'evolució, és a dir, s'havien donat trencaments en la mateixa regió genòmica en espècies de llinatges diferents. Aquests dos fets impliquen l'existència de regions del genoma que tenen tendència a reorganitzar-se. Per intentar explicar aquest fenomen i la seva relació amb l'estructura genòmica del mamífers ens vam plantejar estudiar la distribució de les EBRs comparant el genoma humà amb espècies representatives dels amniotes, així com estudiar la seva relació amb repeticions en tàndem i els elements mòbils (Farré i col, 2011). Aquests tipus de seqüències repetitives s'han relacionat amb la formació de DSBs (Bacolla i col, 2008; Zhao i col, 2010) i s'han proposat com a desencadenants de mecanismes com ara la NAHR (Cordeaux i Batzer, 2009) (veure CAIXA 3) però la seva implicació en mecanismes evolutius no és del tot coneguda.

Així, en el treball 2, comparant els genomes d'humà, ximpanzé, orangutan, macaco, ratolí, rata, vaca, gos, cavall, opòssum i gall, vam estudiar la distribució de les EBRs en el genoma humà en un context evolutiu i vam veure que les EBRs no es distribuïen de manera homogènia, algunes cromosomes acumulaven més EBRs independentment de la seva mida. A més a més vam determinar en un 20,3% el grau de reutilització de les EBRs al llarg de l'evolució (Fig. 13). És a dir, gairebé un quart de les EBRs que vam detectar entre aquestes espècies havien estat utilitzades en varis llinatges que no descendien d'un avantpassat comú proper. Aquest percentatge és semblant al descrit per Murphy i col·laboradors (2005b) però molt superior al 7% trobat en estudis posteriors (Ma i col, 2006; Larkin i col, 2009). Aquestes diferències poden ser degudes a diferències en el número d'espècies que s'han inclòs en l'estudi com també al grau de resolució amb el que s'han descrit les EBRs i els HSBs. S'ha vist que a una resolució major (és a dir, EBRs més petites) el grau de reutilització és menor i es passa a parlar d'agrupació o *clustering* dels punts de trencament en una mateixa regió genòmica

(Attie i col, 2011). Tenint en compte que en el treball 2 vam utilitzar l'algoritme del mòdul Ensembl-Compara que descriu EBRs de mida més gran (Taula 1), no descartem la possibilitat que haguem sobreestimat el percentatge de reutilització. Tot i així, tant si parlem de reutilització estricta com de *clustering* de punts de trencament evolutius, els nostres resultats demostren la presència en el genoma humà de regions re-utilitzades (Fig. 13). A més a més, proposem que les regions que hi estan involucrades han de tenir unes característiques determinades (ja sigui seqüencies específiques, ja sigui estructura tridimensional que adopta la cromatina) que facin que tinguin tendència a patir DSBs i reorganitzar-se.

De fet, els nostres resultats mostren que les EBRs acumulen més repeticions en tàndem que els HSBs , tant pel que fa al número de *loci* com a la quantitat de bases implicades en repeticions (Farré i col, 2011). Hem observat que 5 motius rics en AT (CA, AT, AAAT, TC, CAAA i AAAG) representen més del 30% de totes les repeticions en tàndem dels genomes d'humà, ximpanzé, orangutan i macaco, i que el motiu de repetició AAAT és el més abundant en les EBRs. Aquestes dades s'emmirallen amb les publicades recentment per Payseur i col·laboradors (2011), on comparant dos genomes humans, determinen que motius rics en AT són els més freqüents en el genoma d'aquesta espècie.

Així mateix hem vist que només les EBRs específiques de primats són riques en elements *Alu*, com també que el microsatèl·lit AAAT està associat a elements *Alu* en els punts de trencament evolutiu de primats. Fa aproximadament 40 Ma va haver-hi una explosió d'aquests elements transponibles en els genomes de primats (Shen i col, 1991) i per tant probablement aquests elements es van inserir en els punts de reorganització d'aquest llinatge. Els nostres resultats també s'adiuen amb estudis previs on es demostrava que les EBRs del genoma humà al comparar-lo amb el ratolí eren riques en *Alu* (Schibler i col, 2006). Tradicionalment s'ha proposat que l'associació AAAT-*Alu* pot ser deguda a que els *Alu* són font de microsatèl·lits rics en AT generats a partir de la seva cua poli-A o de la seva regió intermitja rica en As (Nadir i col, 1996; Roy-Engel i col, 2002; Cordeaux i Batzer, 2009; Kelkar i col, 2011), però el fet que nosaltres trobéssim aquesta associació en EBRs de primats i que en aquest tipus d'EBRs hi haguessin menys loci AAAT que en la resta d'EBRs ens va portar a

plantejar la hipòtesi que el motiu AAAT podia ser una diana d'inserció d'*Alu*s, ja que aquesta seqüència és molt semblant al motiu canònic d'inserció (5'-TTAAA-3', Jurka, 1997). Seguint aquestes evidencies, Levy i col·laboradors (2009) ja mostraven que els *Alu* podien inserir-se en dianes diferents del motiu canònic però sempre en seqüències riques en AT. En aquesta línia, Kvikstad i Makova (2010) van demostrar que les famílies d'*Alu* més joves es trobaven en regions riques en AT. Per tant, podem postular que en les EBRs no només hi ha un enriquiment en repeticions en tàndem, com ja havia estat demostrat pel nostre grup (Ruiz-Herrera i col, 2006), sinó que alguns motius poden ser dianes d'inserció d'elements mòbils i/o formar estructures del DNA diferents de la canònica (Zhao i col, 2010), donant lloc a DSBs i inestabilitat genòmica.

Per tant, les nostres dades, juntament amb estudis anteriors en mamífers (Ruiz-Herrera i col, 2006; Larkin i col, 2009; Longo i col, 2009; Girirajan i col, 2009), han posat de manifest que les EBRs es localitzen en regions riques en elements mòbils, duplicacions segmentals i repeticions en tàndem. A més a més, von Grottus i col·laboradors (2010) van veure que l'organització genòmica del gènere *Drosophila* s'explicava gràcies a l'existència de regions fràgils i no per pressions selectives. Per tant, tots aquests estudis apunten a la seqüència de DNA com a possible responsable de la distribució de les EBRs en el genoma. Però aquesta heterogeneïtat en els resultats indica la existència d'altres factors, no només la seqüència nucleotídica, que poden afectar la presència i/o distribució de les EBRs al llarg del genoma.

4.3.2 *Intergenic Breakage model*: constrenyiment funcional

Una altre proposta per a explicar la distribució de les EBRs va ser la de Peng i col·laboradors l'any 2006 (*Intergenic Breakage model*). Seguint aquest model, les EBRs no es trobarien en regions fràgils, sinó que probablement es donen trencaments en tot el genoma però la majoria no es fixen ja que són deleteris per la cèl·lula. Només aquells trencaments que es troben en regions intergèniques o que no afecten a l'expressió gènica escaparan de la selecció purificadora i podran ser detectats (Becker i Lenhard, 2007; Kikuta i col, 2007). Diversos estudis han donat suport a aquest model

per explicar la distribució de EBRs en el genomes del mamífers. Larkin i col·laboradors (2009) van descriure els HSBs entre humà i 9 espècies (ximpanzé, macaco, ratolí, rata, porc, vaca, gos, opòssum i gall) definint els HSBs entre múltiples espècies (msHSBs) que s'havien conservat durant 310 Ma. Van veure que aquests msHSBs contenien xarxes gèniques encarregades de processos de desenvolupament embrionari i tissular comunes en totes les espècies estudiades i, per tant, van proposar que el manteniment d'aquests grans blocs estava regulat degut al constrenyiment funcional. Així mateix, Lemaitre i col·laboradors (2009) van demostrar que les EBRs entre el genoma humà, ximpanzé, macaco, gos, ratolí i rata es localitzaven en regions intergèniques i no en regions gèniques. Van mostrar que les EBRs es trobaven en regions genòmiques amb una alta densitat gènica però poc sovint interrompien els gens d'aquestes regions (d'acord amb Kemkemer i col, 2009). En línia amb aquestes evidencies, recentment s'ha posat de manifest que un 4% del genoma humà està sota constrenyiment al comparar-lo amb els genomes de 28 espècies d'altres mamífers (Lindblad-Toh i col, 2011). Els autors han observat que la major part d'aquest constrenyiment afecta regions codificants, introns i regions intergèniques. D'aquestes últimes, s'ha detectat que el constrenyiment és més elevat en promotors de gens involucrats en desenvolupament i funcions cel·lulars bàsiques, en canvi, han detectat poc constrenyiment en promotors de gens relacionats amb el sistema immune, reproducció i percepció. Aquestes observacions podrien enllaçar amb el *Intergenic Breakage model*: disrupcions o trencaments en les regions amb un elevat constrenyiment evolutiu podrien tenir efectes nocius per a l'organisme i la progènie.

Paral·lelament, estudis en cèl·lules somàtiques han demostrat que les reorganitzacions cromosòmiques estan associades al desenvolupament de varis tipus de càncer i malalties humanes ja que poden interrompre, eliminar i duplicar gens o afectar les regions reguladores d'aquests gens modificant-ne el nivell d'expressió (Shaw i Lupski, 2004; Hurles i col, 2008; Chen i col, 2010). Aquests resultat es poden emmarcar en un context evolutiu ja que varis estudis han relacionat els punts de trencament de reorganitzacions cromosòmiques associades a càncer o malalties humanes amb les EBRs (Ruiz-Herrera i col. 2005; Murphy i col. 2005b; Ruiz-Herrera i Robinson, 2008; Larkin i col, 2009) apuntant a que els mecanismes implicats en

aquestes reorganitzacions poden ser comuns. A nivell evolutiu s'han estudiat les reorganitzacions cromosòmiques entre varies espècies i s'ha vist que en alguns casos sí que interrompen gens (per exemple, la inversió del cromosoma 2 entre *Drosophila buzzatti* i *D. mojavensis;* Calvete i col, 2012), però sovint afecten el nivell d'expressió de gens propers (per exemple, silenciant un gen degut a la formació d'un nou RNAi; Puig i col, 2004). De fet, al acotar els punts de trencament de les reorganitzacions cromosòmiques entre humà i ximpanzé s'ha vist que no afecten a cap gen (Kehrer-Sawatzki i Cooper, 2007) però estan associades a un increment de la diferència d'expressió en gens del còrtex cerebral (Marques-Bonet i col, 2004).

En el nostre treball 3, hem estudiat les reorganitzacions cromosòmiques que es donen en poblacions naturals de ratolí domèstic (*Mus musculus domesticus*; Gazave i col, 2003; Pialek i col, 2005). Aquesta és una espècie model per a estudiar l'evolució cromosòmica. Tot i tenir un cariotip estàndard de 2n = 40 cromosomes acrocèntrics, s'han detectat una àmplia diversitat de números diploides en poblacions naturals (des de 22 a 40), que es deuen a fusions Robertsonianes (Rb; fusió de dos cromosomes acrocèntrics per a formar un metacèntric) (Gazave i col, 2003; Pialek i col, 2005). Tot i aquesta gran diversitat, no tots els cromosomes de ratolí contribueixen en la mateixa freqüència a les translocacions Rb, per exemple, el cromosoma 19 no s'ha trobat involucrat en Rb en poblacions naturals de ratolí (Pialek i col, 2005). Ja que les Rb són reorganitzacions que alteren les regions pericentromèriques dels cromosomes afectats, vam estudiar la distribució de gens en aquestes regions en tots els cromosomes de ratolí per tal d'explicar les diferents freqüències de translocació de cada cromosoma. Vam veure que a la regió pericentròmerica del cromosoma 19 de ratolí hi havia una elevada quantitat de gens, fins a 10x més gens que els altres cromosomes sovint implicats en translocacions Rb, com ara el cromosoma 11 o 17. A l'estudiar l'ontologia dels gens vam detectar que la majoria dels gens d'aquesta regió estan relacionats amb el manteniment cel·lular (*housekeeping genes*), i presenta gens que s'expressen en un ventall més ampli de teixits (Treball 3, Figura 2). Aquestes dades concorden amb el model de constrenyiment funcional i de trencaments intergènics, ja que disrupcions sintèniques en aquesta regió podrien provocar una

davallada d'expressió de gens *housekeeping* o una inactivació d'aquests gens, desembocant en canvis deleteris per a la cèl·lula i la seva la progènie.

Així doncs, els nostres resultats, conjuntament amb estudis previs, apunten a un paper important de la selecció en la localització de les EBRs en el genoma dels mamífers: les reorganitzacions poden tenir un efecte indirecte sobre l'expressió gènica i per això es trobarien seleccionades negativament, fixant-se només aquelles que no siguin deletèries per l'organisme.

4.3.3 Fixació de reorganitzacions cromosòmiques: paper de la recombinació meiòtica

Tant si el model de fragilitat o el model intergènic són correctes, el que demostren les dades publicades fins al moment és la presència de regions del genoma que es conserven i d'altres que han patit reorganitzacions de forma recurrent en diferents llinatges que s'han fixat al llarg de l'evolució. Aquestes reorganitzacions poden tenir un valor adaptatiu (Kirkpatrick i Barton, 2006) o facilitar l'acumulació d'incompatibilitats gèniques entre espècies (Rieseberg, 2001; Noor i col, 2001; Navarro i Barton, 2003), desembocant en un procés d'especiació. Però els mecanismes pels quals les reorganitzacions cromosòmiques donen lloc a noves espècies (models d'especiació cromosòmica, veure secció 1.3.2.2) són objecte de gran controvèrsia (Brown i O'Neill, 2010).

En el treball 4, ens vam proposar estudiar la relació entre les reorganitzacions cromosòmiques i la recombinació meiòtica, i si aquesta ha tingut algun efecte en la fixació de les reorganitzacions, utilitzant com a model el procés d'especiació entre humà i ximpanzé. Per fer-ho, vam emprar l'últim mapa de recombinació del genoma humà (Kong i col, 2010), juntament amb una acurada delimitació de les reorganitzacions cromosòmiques i les EBRs que homologuen el genoma humà i el de ximpanzé. Primer de tot, hem vist que tot i no ser estadísticament significatiu (p-valor = 0,078) la taxa de recombinació en les EBRs era més baixa (0,492) que en els HSBs

(0,962), concordant amb dades d'un estudi previ fet en *Drosophila* (Navarro i col, 1997). En segon lloc, hem pogut comprovar que els cromosomes implicats en inversions presenten una taxa de recombinació (0,944) inferior que els cromosomes no reorganitzats (0,975) (p-valor < 0,0001). En tercer lloc, les regions invertides presenten una taxa de recombinació inferior a les regions no afectades per la inversió dins del mateix cromosoma (invertides: 0,715; no-invertides: 1,001; p-valor < 0,0001). A més a més, hem vist que la supressió de recombinació encara no s'ha recuperat en les formes actuals dels cromosomes que van patir una reorganització en el node d'especiació entre humans i ximpanzés. Per tal d'explicar aquest fenomen hem proposat el *model d'heterocariotips flotants* sota el qual reorganitzacions cromosòmiques (en el nostre cas, inversions) podrien persistir com a heterocariotips (híbrids entre la forma ancestral i la reorganitzada) durant el node d'especiació entre humà i ximpanzé (fa aproximadament 4 Ma; Hobolth i col, 2011) degut a un procés d'hemiplasia (Avise i Robinson, 2008). Aquest fenomen va donar lloc a una supressió de recombinació en les regions reorganitzades en el node ancestral d'humà i de ximpanzé durant un temps suficient per a evitar el flux genètic i l'augment de divergència entre les regions afectades. Aquesta baixada de recombinació encara és detectable en les regions del genoma humà afectades per la reorganització, tant en les que mantenen la forma ancestral com les que mantenen la forma reorganitzada (veure Fig. 3, Treball 4). Aquest model concorda amb la datació de les reorganitzacions cromosòmiques entre humà i ximpanzé (Szamalek i col, 2006b), ja que se sap que aquestes reorganitzacions són precedents a la separació de ximpanzé i bonobo (0.86-2 Ma; Yoder i Yang, 2000; Won i Hey, 2005). Per tant, els heterocariotips s'haurien d'haver mantingut entre 2-3 Ma, fet viable sota el fenomen d'hemiplasia. És durant aquest temps on s'hauria d'haver donat la supressió de recombinació en les regions reorganitzades, i que explicaria per què en les espècies actuals s'observa una taxa de recombinació més baixa en les regions que s'havien mantingut polimòrfiques en el node ancestral. A més a més, el manteniment dels heterocariotips durant aquest temps pot haver provocat una barrera al flux gènic en els cromosomes reorganitzats (Rieseberg i Livingstone, 2003). Tot això podria incrementar l'acumulació de diferències genètiques entre les regions afectades.

Aquesta falta d'homogeneïtzació degut a la supressió de recombinació en les regions reorganitzades podria explicar perquè, en el treball 2, vam detectar un perfil de repeticions en tàndem diferent en els cromosomes de primats reorganitzats i, en canvi, el perfil s'havia mantingut en els cromosomes que conservaven la forma ancestral (veure Fig. 4, Treball 2). Aquest fet concorda amb les dades de Buschiazzo i Gemmell (2010) a l'estudiar els microsatèl·lits del genoma humà ja que van veure que no havien patit canvis ni a nivell de seqüència ni en número de repeticions en aquelles regions conservades entre humà, ximpanzé, macaco, ratolí, rata, conill, gos, vaca, armadillo, elefant, tenrec, opòssum, gall, granota, peix zebra i dos peixos globus (pufferfish i fugu). Un altre treball que relaciona els elements repetitius i la taxa de recombinació és el de Kvikstad i Makova (2010). Aquests autors van demostrar que les seqüències *Alu* s'inserien en regions amb baixa taxa de recombinació. Això unifica les nostres dades, ja que hem vist que les EBRs de primats són riques en elements *Alu* (treball 2) i al mateix temps mostren una taxa de recombinació molt baixa respecte a les HSBs (treball 4).

4.3.4 Quin model de distribució de EBRs és l'adequat?

Pot semblar que les dades que existeixen fins al moment sobre la distribució genòmica de les EBRs són contradictòries. Per una banda, hi ha estudis que afavoreixen el model de trencaments fràgils (Ruiz-Herrera i col, 2006; Longo i col, 2009; Girirajan i col, 2009; von Grottus i col, 2010); en canvi, d'altres es decanten pel model de trencaments intergènics (Peng i col, 2006; Becker i Lenhard, 2007; Lemaitre i col, 2009). Però, de fet, aquests dos models no són excloents i és per això que nosaltres ens decantem per un model integrador: *model integrador de la distribució de les EBRs*. Seguint aquest model els trencaments del DNA es produeixen en regions genòmiques amb certa susceptibilitat a reorganitzar-se però només es fixaran aquells que no involucrin canvis estructurals o d'expressió gènica en gens implicats en el desenvolupament i manteniment cel·lular. Aquest fet explicaria perquè Larkin i col·laboradors (2009) van trobar un enriquiment de gens relacionats amb

desenvolupament embrionari en els msHSBs i, en canvi, gens associats a processos adaptatius en les EBRs, fenomen que podria ser degut a un procés de selecció positiva en aquestes regions. Se sap que moltes EBRs es troben en regions riques en duplicacions segmentals (Bailey i col, 2004; Murphy i col, 2005b; Kehrer-Sawatzki i Cooper, 2008; Larkin i col, 2009) i s'ha proposat que un possible origen de les reorganitzacions cromosòmiques sigui mitjançant la recombinació ectòpica entre aquestes duplicacions segmentals (veure CAIXA 3). A més a més, algunes d'aquestes duplicacions segmentals estan associades a canvis en el número de còpies (*copy number variants*, CNV) en un llinatge concret (Marques-Bonet i col, 2009b; Gazave i col, 2011), que si afecten a regions codificants poden haver-se fixat gràcies a un procés de selecció positiva (Johnson i col, 2001; Marques-Bonet i col, 2009a). Per tant, la variació en el número de copies d'aquestes regions pot ser potencialment necessària per a l'evolució de caràcters adaptatius específics en cada llinatge.

Per altre banda, el *model integrador de distribució de EBRs* concorda amb el concepte de veïnat funcional (*functional neighborhood*) proposat recentment per Al-Shahrour (2010), segons el qual hi hauria un constrenyiment sobre la funció d'un clúster de gens però no sobre els gens en si mateixos, és a dir, espècies filogenèticament llunyanes conserven clústers de gens amb les mateixes funcions tot i que no són sempre gens ortòlegs entre ells. Aquest estudi també mostra com en aquests blocs funcionals hi ha un increment de disrupcions sintèniques o EBRs, sobretot en aquells que contenen menys gens ortòlegs. Per tant, aquest manteniment de la funció i no de gens en particular es podria explicar per l'existència de regions amb predisposició a patir trencaments en aquests blocs funcionals.

Per últim, el model integrador que proposem també és compatible amb el *Turnover Fragile Breakage Model* (TFBM) proposat per Alekseyev i Pevzner (2010), ja que seguint aquest model, les regions fràgils del genoma serien fenòmens transitoris que apareixerien seguint un cicle continu de naixement-mort. És a dir, aquestes regions fràgils no són constants al llarg l'evolució, sinó que apareixen i desapareixen durant el procés evolutiu en diferents llinatges. Així mateix, qualsevol tret associat a un procés de naixement-mort, com ho són els microsatèl·lits (Kelkar i col, 2011) o les duplicacions segmentals (Zhao i Bourque, 2009), podria influenciar la distribució de les

EBRs. A més a més, s'ha vist que els microsatèl·lits pateixen un constrenyiment en la variació de la llargada més acusat en les regions codificants i reguladores dels gens (Payseur i col, 2011); cosa que unifica el model integrador proposat i el TFBM.

En un futur i per tal de seguir avançant en l'estudi de la distribució de les EBRs s'hauran de tenir en compte altres factors que puguin influir-hi. No podem deixar de banda el paper de l'estat de la cromatina per afavorir l'accessibilitat de la maquinària necessària per a produir i reparar els trencaments de doble cadena del DNA. Ja se sap que determinades modificacions de les histones [acetilacions (ac) o metilacions (me)] en residus concrets (H3K9ac, H3K27ac o H3K4me3) són marcadors de conformacions obertes de la cromatina relacionades amb regions de transcripció activa (revisat a Campos i Reinberg, 2009). A més a més es coneix que algunes modificacions epigenètiques associades a conformacions obertes són riques en gens i illes CpG (Terrenoire i col, 2010) i estan caracteritzades per nivells baixos de metilació del DNA (Gilbert i col, 2005; Bird, 1986). Així mateix, s'ha proposat que canvis en la metilació dels *Alus* presents en les EBRs descrites entre humà i gibó (*Nomascus leucogenys*) podrien explicar les nombroses reorganitzacions patides en el llinatge de gibons (Carbone i col, 2009). També s'haurà de considerar l'estructura tridimensional del nucli, ja que per a que es doni una reorganització concreta, les regions implicades en aquesta reorganització haurien d'estar properes físicament en l'ancestre. Recentment, Veron i col·laboradors (2010) fent servir les dades obtingudes per la tècnica de Hi-C de detecció d'interaccions entre regions genòmiques (Lieberman-Aiden i col, 2009), han demostrat *in silico* que *loci* llunyans en el genoma humà però propers en el genoma de ratolí (i que, per tant, havien estat afectats per una reorganització cromosòmica) estan propers en el nucli interfàsic en humà. A més a més, un estudi fet en cèl·lules de ratolí ha posat de manifest que les regions involucrades en translocacions es troben més properes en el nucli interfàsic (Zhang i col, 2012). Totes aquestes evidencies apunten a un paper important i encara per a explorar tant de l'estat epigenètic, de la conformació de la cromatina com de l'estructura tridimensional del nucli en la distribució de les EBRs.

Hem vist que l'estudi de la distribució dels punts de trencament evolutius ens ha permès avançar en el coneixement dels mecanismes d'evolució genòmica. Així doncs,

podem plantejar un model integrador segons el qual els punts de trencament evolutiu es localitzen en regions riques en repeticions en tàndem i elements transponibles i, per tant, regions fràgils del genoma; però només aquelles reorganitzacions que no modifiquin l'expressió de gens considerats imprescindibles per al correcte desenvolupament de l'organisme es fixaran en les següents generacions.

5 CONCLUSIONS

.

CONCLUSIONS

L'estudi de les reorganitzacions cromosòmiques en els genomes de mamífers dut a terme en aquesta tesi ha conduit a les següents conclusions:

1. La comparació de diferents algoritmes per a la determinació de EBRs i HSBs ens ha permès decidir la aproximació *in silico* que millor s'adapta als objectius plantejats. El SyntenyTracker és l'algoritme més adequat per a establir blocs sintènics grans entre espècies allunyades filogenèticament; en canvi, el CASSIS és el millor per a detectar i acotar les regions de trencament evolutiu.
2. Els genomes dels mamífers estan molt conservats. Hem proposat un cariotip ancestral de placentaris de n=23, format per 10 cromosomes que es conserven *in toto* en el genoma humà (1, 5, 6, 9, 11, 13, 17, 18, 20 i X), sis fragments cromosòmics humans (2q, 7b, 2p-q13, 10q, 8p i 19p) i set associacions sintèniques (4q/8p/4pq, 3/21, 14/15, 10p/12pq/22qt, 19q/16q i 12qt/22q).
3. Les associacions sintèniques 3p/21, 8p/4pq, 7a/16p, 12qt/22q, 12pq/22qt, i 14/15 estan presents tant en gall com en granota, per tant són simplesiomorfies del clade de tetràpodes. En canvi, l'associació sintènica 16q/19q no es troba en granota, molt possiblement perquè és una sinapomorfia del clade d'amniotes. I pel que respecte a l'associació sintènica 10/12/22 probablement és una sinapomorfia del clade de mamífers.
4. L'establiment de HSBs entre el genoma d'humà i 10 espècies d'amniotes (ximpanzé, macaco, orangutan, ratolí, rata, vaca, gos, cavall, opòssum i gall) ens ha permès observar una distribució no homogènia de les EBRs en el genoma humà. A més a més hem estimat que un 20.3% de les EBRs detectades han estat reutilitzades en diferents llinatges al llarg de l'evolució.
5. Hem detectat una acumulació de repeticions en tàndem en les regions de punts de trencament evolutius, sobretot de microsatèl·lits rics en AT. El motiu AAAT està associat a elements mòbils *Alu* en les regions de trencament evolutiu (EBRs) de primats. Proposem que els dos tipus de seqüències repetitives podrien ser font d'inestabilitat genòmica i per tant, la seva

acumulació en EBRs pot explicar per què certes regions genòmiques estan predisposades a patir trencaments de doble cadena de DNA i reorganitzar-se, donant suport al *Fragile Breakage Model*.

6. L'elevada presència de gens *housekeeping* en la regió pericentromèrica del cromosoma 19 de ratolí domèstic pot explicar perquè aquest cromosoma no es troba involucrat en fusions Robertsonianes en poblacions natural. La fixació de reorganitzacions que afectin a aquesta regió del cromosoma 19 es veurà disminuïda per l'efecte de la selecció purificadora. Aquest fet concorda amb el *Intergenic Breakage Model*.
7. Gràcies a la precisa delimitació de les regions de trencament evolutiu entre el genoma humà i el de ximpanzé hem determinat que la taxa de recombinació és més baixa en les regions afectades per reorganitzacions cromosòmiques que en les no reorganitzades. Aquestes dades indiquen que el model d'especiació cromosòmica per supressió de recombinació podria haver jugat un paper en el procés d'especiació d'humà i ximpanzé.
8. Hem formulat el *model d'heterocariotips flotants* que permet explicar la baixa taxa de recombinació en les formes actuals dels cromosomes que van patir una reorganització en el node d'especiació entre humans i ximpanzés. Segons aquest model, les reorganitzacions cromosòmiques van persistir com a heterocariotips durant el node d'especiació d'aquestes espècies degut a un procés d'hemiplasia creant una barrera al flux gènic en les regions reorganitzades.
9. Per tal d'unificar les nostres dades hem proposat el *model integrador de distribució de les EBRs*, segons el qual les EBRs es localitzarien en regions riques en repeticions en tàndem i elements transponibles i, per tant, regions fràgils del genoma; però només aquelles reorganitzacions que no modifiquin l'expressió de gens considerats imprescindibles per al correcte desenvolupament de l'organisme es fixaran en les següents generacions.

6 BIBLIOGRAFIA

.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Ce t zGA t R GBIABF t zBIa t vA a t vR a t 4zzFAJ eIIA t v4, I ThTtn T h h r m T h h n-ohT T r ohT T- hr o hr o b A z a doost 2 8 mMIssJNdm

IaG4 t v B t I4 te A 4 t 4Faa t O Aa t IA 4 t 4IO t z a t I 4 I FG I n -hT c n e h - u n o Th b z A z Mf f t 1 AmMI- DJ- f m

AFR Mo I Th 8. R emheho -h h -T d h t Th o r T o h 8. f. . . a n- m -oe ob 4 doof t 8. . IDuf JDr Nm

IG 4z I T o o hn -Ta nhT T- n r -hT h T men o ohTR h h oe -o h T a he T- b IFa Idooout A. mMIDuJrNm

I44 IA t BA t 4e a t 41 vaJ FA z t Azv4 t 4FAIO t vG FA t F BI t vG FA t IGFA t 4Ia t IBG4 I r T T r Tt hT n eh T-ch ocT- Tc n a o ho T d To n-hTo - n m T T- o+ ob AFR a doof t 83 mdIIMf- JMf om

4 BF zaQ t 4FAFe t zA, FA I Th a no-c T nch-ce ah r -hT h ob FG , zF A z doMMMyIddm

4 BF zaQ t A t FIA, t FR A ACF t IGv e t 4GR A t OGR IaB e t GD G e t 4vCFe t 4vaFAJ R IzB t v4 B, t FIAaFA I n T Th hn Tu -hT T nh T-o c hht b B4FR FaFR a doo- t 8z mdIIdDMdr Nm

4 BF zaQ t A t 4GR A t i 4IA t 4vCFe t IGA t IGv e t v t F4FIIe t E t FIAaFA t 4vaFAJ R IzB I h e n -a h r n c-h T- o-r o T- hn n nT ahn R ee T nh hoh n m T T-o hT-h e ch T- -n b , zF A z AFR a doodt 3 z mMNIIMs r JMNum

4, I - o -dh -n Toeho hTo -h -T hR -n Toeho t T-t -nh hoh h m T T-ob 4A a A z dooot 8z mMoIINDMJND- m

4 F4 t IFG t R R t vGR A t vGR A t IzB t v44, t 4vA t 4IGv 4 t AA zz I r nch- Th ou - o ob vGII I a a doort Apm z a Iaav II ssdJ- m

4IIA t F 4zaFA t R az t IA G t IGFA t 4FFIIR Aa t 4F A A t 4 vaB e t IAaO , t 44 t IA zFA t G IFz t FGD 4 t a A t v4z I h Th h e n -a o-r o -d T T T -r n c T - n e -hTo hn a T Th ah r -hTb AFR I a doo- t 3IMD- m

vzz A B t A IBzIA 4 t a A t 4 A t BR I I h e n -a nh hoh e T-T h T r -hoh e T-o 8t3 T T n T- n oe ob , zF A z AFR a doost 8. AmMdIIMfs JM Nm

-----

G I -nhocT eo b T or eer o ohTh o -d -T -nhuc hr o e n T-n Ta n ohTo h nh hr o b B4FR FaFR Mf- Dt 3 ymDII NduJNsdm

Gaz4FR t AO I T ah r -hT c Th- o-n -c r n -T b FG IFG eFG doMot 1 gmMdIId- oNJd- MDm

Gaz4FR t vG 4 t IGaFA t AO I c h Th - T c o o emha a T - - T n-n T nh -n n o o-n nhr eo b FG IFG eFG doort 1 ymfIIdouf JdoD- m

44Ia I emha e nd o T T- h Th b B B aIa doorm B AAa, Ge AI z z AIe 4aIz, m

aa t vG 4 t 4zFAI t G v I4F I h e n -a nh hoh eo h T h-nhe nh T-o nh co o r nr o T e-h co T n- s n - i o- o c t b , zF A z AFR a doMMC 8Apr mMIINdJuom

[illegible] a L[illegible]R[illegible] 2 4[illegible]2[illegible]B[illegible]O[illegible]LIR[illegible] O[illegible] A[illegible]I[illegible] IO[illegible]R[illegible] O[illegible]R[illegible] 44[illegible] BAAL[illegible] Ow[illegible]R[illegible] A fa[illegible] A[illegible]f[illegible] A[illegible]s [illegible] a[illegible] A[illegible]f[illegible] t [illegible] A[illegible] G G [illegible] RAG [illegible] gDDD[illegible] tpu brJbhi Mol t, [illegible]

[illegible] 1[illegible] 4e[illegible] J[illegible] A[illegible] f[illegible] A[illegible] f[illegible] e[illegible] f [illegible] A[illegible] f[illegible] r [illegible] i f[illegible] IA[illegible] 11 [illegible] k1[illegible] htt c[illegible] wd ubrJobc MoDi , [illegible]

[illegible] O4[illegible] R[illegible] O[illegible]R[illegible] IO[illegible] B I14[illegible] R[illegible] Iw[illegible] 2 L[illegible] FI4ARAH[illegible] H[illegible]B[illegible] RLA[illegible] 24R[illegible] G [illegible] B2[illegible] B[illegible] I[illegible] AQ J[illegible] f[illegible] f[illegible] A[illegible] s [illegible] A[illegible] I[illegible] R[illegible] ht gt [illegible] hwy uc DobrJYl l MYl Y, [illegible]

[illegible] IO[illegible] IO[illegible] I R[illegible] k[illegible] R[illegible] I1AR[illegible] AR4[illegible] A L[illegible] R[illegible] R4[illegible] R[illegible] 4IAR[illegible] O[illegible] BI[illegible] L[illegible] R[illegible] RAG [illegible] J[illegible] i [illegible] A[illegible] f r [illegible] t [illegible] f[illegible] A[illegible] r [illegible] i [illegible] i [illegible] A [illegible] f r [illegible] A f[illegible] A[illegible] f[illegible] r [illegible] i f[illegible] 42 [illegible] htt i [illegible] l hw [illegible] YDc MogY, [illegible]

[illegible] A[illegible] AQ B[illegible] 24B[illegible] IO[illegible] a L1[illegible] BI[illegible] 2F[illegible] IQ R[illegible] J[illegible] f[illegible] A f[illegible] ff[illegible] A[illegible] i [illegible] g[illegible] - [illegible] g [illegible] A[illegible] i f[illegible] i [illegible] f[illegible] A[illegible] f[illegible] A[illegible] i f[illegible] f[illegible] s [illegible] A[illegible] f[illegible] RAG [illegible] 1[illegible] ht ggm[illegible] wcu l rJl i DMcY, [illegible]

[illegible] 2 O1[illegible] G I4e[illegible] LI1[illegible] kO[illegible] M[illegible] G I4B[illegible] J[illegible] i [illegible] f[illegible] f[illegible] f Ai [illegible] fi A[illegible] r [illegible] A[illegible] f[illegible] i [illegible] R[illegible] 1[illegible] R[illegible] 4[illegible] htt brn[illegible] wl ucrJhl bMi c, [illegible]

[illegible]

[illegible] 2[illegible] B[illegible] I[illegible] R[illegible] 4[illegible] RkAR[illegible] RAO[illegible] A[illegible] R[illegible] O[illegible] 1BI[illegible] G [illegible] J[illegible] f Ai [illegible] f[illegible] A[illegible] AA[illegible] f [illegible] A[illegible] f[illegible] t [illegible] A[illegible] f[illegible] A[illegible] 4O[illegible] I[illegible] gDDhm[illegible] y bu gbrJbYggNbYgc, [illegible]

[illegible] 2 L[illegible] J[illegible] i [illegible] ff[illegible] A[illegible] f[illegible] - t [illegible] f[illegible] r [illegible] r [illegible] f[illegible] f[illegible] A[illegible] f[illegible] A f A[illegible] A[illegible] 4O[illegible] I[illegible] gDDom[illegible] bl ucrJgbohMgboo, [illegible]

[illegible]

[illegible] R[illegible] k[illegible] l A2w[illegible] O[illegible] LR[illegible] J[illegible] a [illegible] i [illegible] a f A[illegible] A[illegible] A[illegible] f[illegible] t - [illegible] t [illegible] ccg[illegible] A[illegible] s [illegible] t [illegible] A r [illegible] A[illegible] f[illegible] t [illegible] O[illegible] gDDom[illegible] yy ul rJl ocMbi , [illegible]

[illegible] B [illegible] M[illegible] a [illegible] 4eLI[illegible] AAF[illegible] J[illegible] i [illegible] A[illegible] A[illegible] A[illegible] AA[illegible] f[illegible] i A[illegible] A[illegible] f[illegible] r [illegible] i f[illegible] B AG A1AG [illegible] 1[illegible] htt br[illegible] tmu grJi gMY, [illegible]

[illegible] B [illegible] M[illegible] a [illegible] 4eLI[illegible] AAF[illegible] J[illegible] A f[illegible] f[illegible] A[illegible] f[illegible] r [illegible] i f[illegible] f[illegible] i [illegible] e[illegible] f [illegible] A[illegible] i [illegible] a[illegible] - [illegible] A[illegible] 2G [illegible] 24[illegible] 4 htt om[illegible] wy uhrJDDMgl t, [illegible]

[illegible] O[illegible] L[illegible] 4[illegible] BI[illegible] AG AR4[illegible] LAH[illegible] J[illegible] ff[illegible] A[illegible] A[illegible] f[illegible] e[illegible] s [illegible] A[illegible] f[illegible] f[illegible] A[illegible] r [illegible] A[illegible] f[illegible] i [illegible] RAG [illegible] 1 ht ggm[illegible] wcu ghrJht l bMt i b, [illegible]

[illegible] G L[illegible] G [illegible] ABR[illegible] AAF[illegible] A[illegible] RI[illegible] L[illegible] A[illegible] O[illegible] G [illegible] I14[illegible] B [illegible] M[illegible] a [illegible] 4eLI[illegible] J[illegible] t f[illegible] t [illegible] A[illegible] fs [illegible] A[illegible] f r [illegible] A f[illegible] A[illegible] f[illegible] A[illegible] r [illegible] s [illegible] f [illegible] f[illegible] A[illegible] i [illegible] r [illegible] f [illegible] i A[illegible] A t [illegible] ft [illegible] r [illegible] i f[illegible] HAO[illegible] IAO[illegible] htt Dm[illegible] bJbi , [illegible]

[illegible] R4[illegible] 41[illegible] B[illegible] IR I[illegible] B1[illegible] IO[illegible] 211O[illegible] J[illegible] r [illegible] i f[illegible] o[illegible] i [illegible] A[illegible] e[illegible] i [illegible] f[illegible] g [illegible] f[illegible] g[illegible] A[illegible] AA[illegible] f[illegible] f[illegible] i [illegible] i [illegible] A[illegible] 4O[illegible] I[illegible] htt l [illegible] tppu ht rJggi bi Mggi bD, [illegible]

[illegible] IL24[illegible] FO[illegible] R4[illegible] H [illegible] 4IO AH[illegible] AG I1[illegible] e2L[illegible] R[illegible] 14 AG [illegible] G [illegible] R[illegible] L[illegible] O[illegible] R[illegible] G A[illegible] O[illegible] Aa [illegible] BI1O[illegible] IR[illegible] e[illegible] AR[illegible] A2 [illegible] IR[illegible] O[illegible] R[illegible] 1 R[illegible] 41[illegible] BI11[illegible] BI11[illegible] A2[illegible] B[illegible] AO[illegible] O[illegible] R[illegible] RB[illegible] L[illegible] J[illegible] A[illegible] i [illegible] f[illegible] A t [illegible] i f[illegible] A[illegible] f[illegible] A r [illegible] t [illegible] f[illegible] t [illegible] r [illegible] A f[illegible] A[illegible] f[illegible] RAG [illegible] 1[illegible] htt om[illegible] tdu crJci cMcc, [illegible]

[illegible] RAG [illegible] u: : 5 [illegible]

[illegible] RAG [illegible] u: : 2 [illegible]

[illegible] u: : 5 [illegible]

[illegible] u: : 5 [illegible]

[illegible] RAG [illegible] u: : 5 [illegible]

[illegible] u: : 5 [illegible]

[illegible] RAG [illegible] u: : 5 [illegible]

[illegible]

[illegible] RAG [illegible] u: : 2 [illegible]

[illegible] u: : 5 [illegible]

[illegible] u: : 5 [illegible]

[illegible] u: : 5 [illegible]

[illegible] RAG [illegible]

[illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]

[illegible]

[illegible]

[illegible]

[illegible] (oo, :Gn3 2N6ov 56. a

[illegible] (oo5 :2ga5-v) 6a

[illegible] (oo5 :2. . η 52N6-(v-6--a

[illegible] (oo3 :1vN53v365a

[illegible] (oo9 :Gun 2N-. 3v, 909a

[illegible] (oo) :2. 2r( ) 2No, 5v. o, 9a

[illegible]

[illegible] (oo. :Gun, 2N-3v) -. a

[illegible] (oo5 :22r62N6o3v6, 9a

[illegible] (oo9 :fGn) 2N696v-(-a

[illegible] . . . a [illegible]

[illegible] (oo5 :g. . a(6-a

[illegible] (oo, :2: n-2N3, v539a

[illegible] . . . :2uGr(2N, 5v-(-a

[illegible] (oo9 :2: n92N, 55v, ,) , a

[illegible] (oo(:3: η v) 2N(5v((-a

[illegible] . . 3 :G1r(2N3, v(3) a

[illegible] (ooo :2un, 2N3) v) 3. a

[illegible] (oo(:2Gn 2N555v, 5)) a

[illegible] (oo. :2G: η v(2N3) v) 9a

[illegible] (oo3 :2gr62N6. v3-) a

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] FG[illegible]VB[illegible], [illegible] G[illegible]5[illegible], [illegible] I[illegible]5[illegible], [illegible] R [illegible] IoV[illegible]5[illegible] H[illegible] A[illegible] o[illegible] pr [illegible] -[illegible] vm[illegible] m[illegible] pm[illegible] a[illegible] p[illegible] m[illegible] p[illegible] v[illegible] A[illegible] p[illegible] o[illegible] pA[illegible] m[illegible] m[illegible] pt [illegible] oA p[illegible] o[illegible] r[illegible] oA[illegible] oA p[illegible] o[illegible] n[illegible] O[illegible] B5FR FoFR [illegible] d[illegible] saas, [illegible] HSR(fH[illegible] C66K[illegible] C46r[illegible]

I FA[illegible] 5FWBooo[illegible], [illegible] oB[illegible] o5A[illegible] 5[illegible], [illegible] AM[illegible] H[illegible] o[illegible] o[illegible] A[illegible] m[illegible] Ap[illegible] a[illegible] pA[illegible] A[illegible] npo[illegible] pm[illegible] o[illegible] A[illegible] p[illegible] m[illegible] Ao[illegible] e[illegible] pA[illegible] p[illegible] am[illegible] oAmA [illegible] O[illegible] AFR [illegible] d[illegible] sama, [illegible] nSP) fHma) CKma4(r[illegible]

[illegible] Fod[illegible], [illegible] oR [illegible], [illegible] oW[illegible], [illegible] oG[illegible], [illegible] hAFG[illegible] o[illegible], [illegible] A5h[illegible], [illegible] o[illegible], [illegible] G[illegible] 5[illegible], [illegible] R IVB[illegible] H[illegible] o[illegible] A[illegible] r[illegible] oA[illegible] AmA[illegible] m[illegible] v[illegible] mm[illegible] A[illegible] Z[illegible] am[illegible] A[illegible] Z[illegible] o[illegible] p[illegible] pA[illegible] A[illegible] npo[illegible] oA[illegible] AmA[illegible] n[illegible] O[illegible] AFR [illegible] d[illegible] samm, [illegible] nHHm: a(Km: ms r[illegible]

[illegible]

[illegible] 5V[illegible] 5[illegible], [illegible] oe h[illegible], [illegible] oV5JG[illegible] ot[illegible], [illegible] oR G[illegible] 5[illegible] H[illegible] pA[illegible] A[illegible] A[illegible] a[illegible] o[illegible] vpA[illegible] p[illegible] m[illegible] Ao[illegible] oA[illegible] AmA[illegible] r A[illegible] ap[illegible] A[illegible] A[illegible] p[illegible] ao[illegible] Ao[illegible] m[illegible] F[illegible] oAm[illegible] m[illegible] O[illegible] hVF[illegible] A[illegible] V[illegible] AFR [illegible] d[illegible] saa2, [illegible] HSbPmK: fHm46Ks a: r[illegible]

[illegible] V[illegible] A[illegible], [illegible] IBF5[illegible] o[illegible], [illegible] 5VBh[illegible], [illegible] I[illegible] oe [illegible], [illegible] IoBI[illegible] CI[illegible], [illegible] ooLL[illegible], [illegible] 5Fe A[illegible], [illegible] R [illegible] o[illegible] CI[illegible], [illegible] IMF[illegible] o[illegible] BIK[illegible] Ih[illegible] CIV[illegible], [illegible], [illegible] FG[illegible] H[illegible] o[illegible] pA[illegible] v[illegible] o[illegible] A[illegible] p[illegible] o[illegible] p[illegible] A[illegible] A[illegible] p[illegible] Zhss[illegible] o[illegible] n[illegible] O[illegible] V[illegible] A[illegible] Vm444, [illegible] nnPmfHs6K (r[illegible]

[illegible] V[illegible] 5d[illegible], [illegible] F[illegible] I[illegible] Ah[illegible], [illegible] G[illegible], [illegible] F[illegible] IAoFA[illegible] H[illegible] r A[illegible] ap[illegible] A[illegible] ov[illegible] npAov[illegible] A[illegible] -H[illegible] p[illegible] Ao[illegible] m[illegible] A[illegible] p[illegible] n[illegible] O[illegible] F[illegible] A[illegible] saa6, [illegible] nPmfHm2) r[illegible]

[illegible] V[illegible] 5oFA[illegible], [illegible] IA[illegible] G[illegible] K[illegible] FB[illegible], [illegible] I5A[illegible] h[illegible], [illegible] F[illegible] 5o[illegible], [illegible] 5JG[illegible], [illegible] 5e [illegible] G[illegible], [illegible] 5e [illegible] G[illegible], [illegible] IAo[illegible] Fo[illegible] B[illegible] Fod[illegible] AFR [illegible] 1o[illegible] A[illegible] IA[illegible] FAoF5VoR H[illegible] p[illegible] m[illegible] ca[illegible] A[illegible] o[illegible] pr [illegible] vm[illegible] m[illegible] A[illegible] p[illegible] Aam[illegible] A[illegible] O[illegible] Vo5[illegible] saas, [illegible] ynSR(4m2fH2saK2(s r[illegible]

[illegible] BIV[illegible] H[illegible] A[illegible] m[illegible] A[illegible] m [illegible] pA[illegible] O[illegible] t [illegible] A[illegible] pm[illegible] ma[illegible] np[illegible] p[illegible] p[illegible] m[illegible] mv[illegible] po[illegible] A p[illegible] A[illegible] m[illegible] o[illegible] App[illegible] A[illegible] v[illegible] p[illegible] o[illegible] pr [illegible] n[illegible] O[illegible] I[illegible] A[illegible] m4( ), [illegible] HsuP.) m4fHma(2Kma6ar[illegible]

[illegible] I[illegible] A[illegible] 5[illegible] H[illegible] a Ao[illegible] m[illegible] np[illegible] a[illegible] v[illegible] o[illegible] e[illegible] pA[illegible] pA[illegible] oA[illegible] AmA[illegible] m[illegible] m[illegible] pAA[illegible] pA[illegible] a[illegible] omp[illegible] a[illegible] o[illegible] p[illegible] A[illegible] r A[illegible] ap[illegible] A[illegible] O[illegible] hVF[illegible] A[illegible] V[illegible] AFR [illegible] d[illegible] saa2, [illegible] HSbPmK: fHm: 4Km(ar[illegible]

[illegible] I[illegible] A[illegible] 5[illegible], [illegible] V[illegible] AhFA[illegible] H[illegible] A[illegible] o[illegible] pr [illegible] oA[illegible] AmA[illegible] p[illegible] A[illegible] o[illegible] p[illegible] A[illegible] n[illegible] O[illegible] m44), [illegible] uPsK: fH66K4m[illegible]

[illegible] I[illegible] A[illegible] 5[illegible], [illegible] V[illegible] AhFA[illegible], [illegible] o[illegible] B[illegible], [illegible] 5R [illegible] 5[illegible] H[illegible] A[illegible] A[illegible] A[illegible] m[illegible] a[illegible] oA[illegible] AmA[illegible] m[illegible] r[illegible] v[illegible] np[illegible] a[illegible] ma o[illegible] mm[illegible] A[illegible] v[illegible] o[illegible] e[illegible] pA[illegible] t p[illegible] a[illegible] oA[illegible] AmA[illegible] m [illegible] o[illegible] o[illegible] n[illegible] O[illegible] B5FR FoFR [illegible] m44s, [illegible] HSHP2K(fHs(2Ks6ar[illegible]

[illegible] I[illegible] A[illegible] 5[illegible], [illegible] o[illegible] B[illegible], [illegible] V[illegible] AhFA[illegible], [illegible] 5R [illegible] 5[illegible] H[illegible] A[illegible] a[illegible] o[illegible] vpAp[illegible] i A[illegible] A[illegible] v[illegible] A[illegible] o[illegible] p[illegible] m[illegible] v[illegible] oA[illegible] AmA[illegible] np[illegible] a[illegible] ma o[illegible] mm[illegible] A[illegible] v[illegible] o[illegible] e[illegible] pA[illegible] O[illegible] AFR I[illegible] d[illegible] m44a, [illegible] bPsfH: C6K: 2ar[illegible]

[illegible] FA[illegible], [illegible] h[illegible] H[illegible] r [illegible] o[illegible] A a[illegible] pA[illegible] p[illegible] m[illegible] A[illegible] e[illegible] n[illegible] O[illegible] FG[illegible] IFG[illegible] I FG[illegible] saa2, [illegible] nnPsfHs46K: a6r[illegible]

[illegible]

[illegible] A[illegible], [illegible] G[illegible] G[illegible] I [illegible], [illegible] 5[illegible] CR [illegible] A[illegible], [illegible] 5[illegible] IAI[illegible], [illegible] 55IoFA[illegible], [illegible] i[illegible] 5[illegible] A[illegible], [illegible] o[illegible], [illegible] 5[illegible] BF[illegible] VoCh[illegible], [illegible] 5[illegible] ooFAK[illegible] R IVB[illegible], [illegible] F[illegible] IAoFA[illegible] H[illegible] oA[illegible] oA[illegible] AmA[illegible] p[illegible] A[illegible] a[illegible] Z[illegible] o[illegible] r [illegible] o[illegible] Z[illegible] p[illegible] Fma [illegible] oAo[illegible] o[illegible] oAp[illegible] o[illegible] I[illegible] o[illegible] r [illegible] m[illegible] p[illegible] v[illegible] ap[illegible] o[illegible] npo[illegible] ovApv [illegible] O[illegible] 5F[illegible] VG[illegible] I[illegible] saa: , [illegible] HSSP. fHma(sKma((r[illegible]

[illegible] A[illegible], [illegible] 5[illegible] BF[illegible] VoCh[illegible], [illegible] i[illegible] 5[illegible] A[illegible], [illegible] FG[illegible] G[illegible], [illegible] FG[illegible] ACh[illegible], [illegible] 1oI5[illegible], [illegible] 5[illegible] A[illegible], [illegible] 5[illegible] ooFAK[illegible] R IVB[illegible] H[illegible] oA[illegible] oA[illegible] AmA[illegible] p[illegible] a[illegible] p[illegible] m[illegible] p[illegible] npAov[illegible] A[illegible] A[illegible] r A[illegible] ap[illegible] A[illegible] A[illegible] p[illegible] A[illegible] np[illegible] pZ[illegible] A[illegible] a[illegible] O[illegible] B5FR FoFR [illegible] d[illegible] saaa, [illegible] bP2fH:4: KCaCr[illegible]

[illegible] AAI[illegible], [illegible] oo[illegible] V[illegible], [illegible] ooo[illegible] 5[illegible] H[illegible] oA[illegible] AmA[illegible] b[illegible] co[illegible] pm[illegible] At [illegible] Ar [illegible] o[illegible] p[illegible] p[illegible] o-m [illegible] v[illegible] o[illegible] eA[illegible] A[illegible] p[illegible] Ao[illegible] i [illegible] o[illegible] am[illegible] oAa O[illegible] 5[illegible] IVh[illegible] P[illegible] IA[illegible] f[illegible] saa4, [illegible] HSnR(fH(m(K(s 2r[illegible]

[illegible] R [illegible] F [illegible] d [illegible] G [illegible] e [illegible] E [illegible] RL [illegible] RL [illegible] YRL [illegible] Znnndt **ta,** Eaegnvg Dgn2nr [illegible]

[illegible]

[illegible] G [illegible] d [illegible] G [illegible] d [illegible] R [illegible] L [illegible] e [illegible] E [illegible] GRO [illegible] ZnnMdt **ti,** l aev M l Dv l gr [illegible]

[illegible] G [illegible] d [illegible] RF [illegible] d [illegible] R [illegible] d [illegible] R [illegible] d [illegible] L [illegible] F [illegible] G [illegible] d [illegible] O RG [illegible] d [illegible] l F [illegible] d [illegible] L [illegible] d [illegible] I I F [illegible] e [illegible] E [illegible] LL ZngZdt **i o,** l ae2 nv D2 Zgr [illegible]

[illegible] R [illegible] d [illegible] R5 FA5 [illegible] e [illegible] E [illegible] GRO [illegible] Znn2dt **n,** l ae2 : MD2 MZr [illegible]

[illegible] R [illegible] d [illegible] RLL [illegible] d [illegible] G [illegible] d [illegible] A5 o [illegible] e [illegible] s [illegible] s [illegible] E [illegible] LL [illegible] RL [illegible] Zngndt **ma,** gae M D Zr [illegible]