Mamm Genome (2014) 25:618–635
DOI 10.1007/s00335-014-9537-8

A high resolution map of mammalian X chromosome fragile regions assessed by large-scale comparative genomics

Carlos Fernando Prada · Paul Laissue

Received: 30 May 2014 / Accepted: 14 July 2014 / Published online: 3 August 2014
© Springer Science+Business Media New York 2014

Abstract Chromosomal evolution involves multiple changes at structural and numerical levels. These changes, which are related to the variation of the gene number and their location, can be tracked by the identification of syntenic blocks (SB). First reports proposed that ~180–280 SB might be shared by mouse and human species. More recently, further studies including additional genomes have identified up to ~1,400 SB during the evolution of eutherian species. A considerable number of studies regarding the X chromosome's structure and evolution have been undertaken because of its extraordinary biological impact on reproductive fitness and speciation. Some have identified evolutionary breakpoint regions and fragile sites at specific locations in the human X chromosome. However, mapping these regions to date has involved using low-to-moderate resolution techniques. Such scenario might be related to underestimating their total number and giving an inaccurate location. The present study included using a combination of bioinformatics methods for identifying, at base-pair level, chromosomal rearrangements occurring during X chromosome evolution in 13 mammalian species. A comparative technique using four different algorithms was used for optimizing the detection of hotspot regions in the human X chromosome. We identified a significant interspecific variation in SB size which was related to genetic information gain regarding the human X chromosome. We found that human hotspot regions were enriched by LINE-1 and Alu transposable elements, which may have led to intraspecific chromosome rearrangement events. New fragile regions located in the human X chromosome have also been postulated. We estimate that the high resolution map of X chromosome fragile sites presented here constitutes useful data concerning future studies on mammalian evolution and human disease.

Introduction

The first version of the human genome led to the beginning of the post-genomic era, which has led to a better understanding of distinct aspects of biology and medicine, such as evolution, genomic architecture, genetics, physiology, and pathology (Chimpanzee and Analysis 2005; Lander et al. 2001; Mouse Genome Sequencing et al. 2002; Venter et al. 2001). Remarkable advances in technology have been made since then, thereby enabling efficient, full-length, genome sequencing. Nucleotide sequences for at least 29 complete mammalian genomes are currently available in public databases, 13 of which have been totally assembled in chromosomal structures (Lindblad-Toh et al. 2011). Bioinformatics tools have also evolved, enabling simultaneous analysis of considerable amounts of data. Large genomic regions can currently be aligned at base-pair (bp) level for studying interspecific chromosomal rearrangements occurring during evolution (Bourque et al. 2005; Darling et al. 2004). This approach has been used for studying the potential pathogenic consequences of human phenotypes, such as infertility, developmental disorders,

Electronic supplementary material The online version of this article (doi:10.1007/s00335-014-9537-8) contains supplementary material, which is available to authorized users.

C. F. Prada · P. Laissue (✉)
Unidad de Genética, Grupo GENIUROS, Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario,
Carrera 24 N° 63C-69, Bogotá, Colombia
e-mail: paul.laissue@urosario.edu.co

C. F. Prada
e-mail: cafepra@hotmail.com

and cancer (Darai-Ramqvist et al. 2008; Mitelman et al. 2013; Murphy et al. 2005). Chromosomal evolution has been widely studied by the identification and conservation of syntenic blocks (SB) across species.

The first studies in this area proposed that ~180–280 SB might be shared by mouse and human species (Nadeau and Taylor 1984; Pevzner and Tesler 2003a). More recently, further studies which have included additional genomes (e.g., rat, chicken, dog, cow, and pig) have identified up to ~1,400 SB during the evolution of eutherian species (Bourque et al. 2005; Larkin et al. 2009; Ma et al. 2006; Murphy et al. 2005). However, it is considered nowadays that such calculations may have been underestimated due to low resolution maps, which may have led to an inaccurate identification of chromosome rearrangements and the potential misunderstanding of some evolutionary mechanisms (Attie et al. 2011; Copeland et al. 1993; DeBry and Seldin 1996; Gregory et al. 2002; Mouse Genome Sequencing et al. 2002).

Fragile regions or fragile sites have been described by cytogenetic techniques as specific chromosomal regions that exhibit an increased frequency of gaps (breaks) during in vitro cell exposure to DNA replication stress (Sutherland 1977). Distinct hypotheses have been proposed from a mechanistic point of view for explaining chromosome evolution, such as the "fragile breakage'' model (FBM) (Nadeau and Taylor 1984; Pevzner and Tesler 2003b). FBM postulates that genomes consist of mosaics of fragile and solid segments. These regions may have a high (hotspots) or low propensity to rearrangements. Studies on mammalian breakpoint co-location and other genomic features (e.g., segmental duplications, repeated elements, experimental fragile sites, high GC content, and CpG island density) have provided evidence in favor of the FBM (Armengol et al. 2003; Bailey et al. 2004; Gordon et al. 2007; Ruiz-Herrera et al. 2005; Ruiz-Herrera and Robinson 2007; Schibler et al. 2006). Some of these DNA characteristics, such as segmental duplications and/or repetitive elements, have been associated with regions, which are often reused during mammalian chromosome evolution (named evolutionary breakpoint regions, EBR). This is consistent with the hypothesis that EBRs are evolutionarily unstable genomic regions, promoting chromosome rearrangement via non-allelic homologous recombination (NAHR) (Bailey et al. 2004; Murphy et al. 2005; Schibler et al. 2006). Identifying new EBRs would thus contribute toward understanding chromosome evolution and breakpoint reuse-related processes, as well as identifying potential predisposing factors concerning human disease. Human EBR location is currently based on a methodology used to define SB, involving orthologous genes and BAC-end sequences (Larkin et al. 2009; Ma et al. 2006; Ruiz-Herrera et al. 2006). However, this approach may be slightly inaccurate compared to other methodologies, such as multiple genomic alignments permitting genomic analysis at bp level. Large chromosomal regions in humans which are prone to breakage when subjected to replication stress (common fragile sites, CFS) have been related to EBRs. For instance, genomic instability (a hallmark of cancer biology) has been related to some rearrangement hotspots preferentially occurring at CFSs (Lukusa and Fryns 2008; Ma et al. 2012).

A considerable number of studies have been undertaken regarding X chromosome structure and evolution because of their extraordinary biological impact on reproductive fitness and speciation. Several authors have demonstrated a constant traffic between the X chromosome and autosomes during the Drosophila's and mammalian evolution (Cheng and Disteche 2006; Vibranovski et al. 2009a, b; Emerson et al. 2004; Zhang et al. 2010). Some reports have identified EBRs and six fragile sites have been located at specific locations in the human X chromosome: *FRAXB*, *FRAXC*, *FRAXD*, *FRAXA*, *FRAXE*, and *FRAXF* (Ruiz-Herrera et al. 2006; Schwartz et al. 2006). *FRAXA* and *FRAXE* have been associated with mental retardation, tremor/ataxia syndrome, and Parkinson's disease (Brouwer et al. 2009; Costa et al. 2011; Debacker and Kooy 2007). However, as for other genomic regions, CFS and EBR mapping on the X chromosome has involved using low-to-moderate resolution techniques. This might have been related to underestimating their total number and inaccurate location.

In the present study, we have used a combination of bioinformatics methods to identify, at bp level, chromosomal rearrangements present during X chromosome evolution in 13 mammalian species. A comparative technique involving four different algorithms was used to optimize detecting hotspot regions in the human X chromosome. A significant interspecific variation in SB size was found, which was related to a gain in genetic information regarding the human X chromosome. It was found that human hotspot regions were enriched by LINE-1 and Alu transposable elements (TE) which might have led to intraspecific chromosome rearrangement events. This also led to postulating new fragile regions located in the human X chromosome. The high resolution map of X chromosome fragile sites presented here constitutes useful data for future studies concerned with mammalian evolution and human disease.

Results

Interspecific SB size variation

Aligning 13 homologous mammalian X chromosomes revealed 39 SB which were consecutively numbered from p to q chromosomal regions (Fig. 1). X chromosome length ranged from 111.7 Mb (rabbit) to 171 Mb (mouse) (Supplementary Table S1). Average SB size was 3.5 Mb,

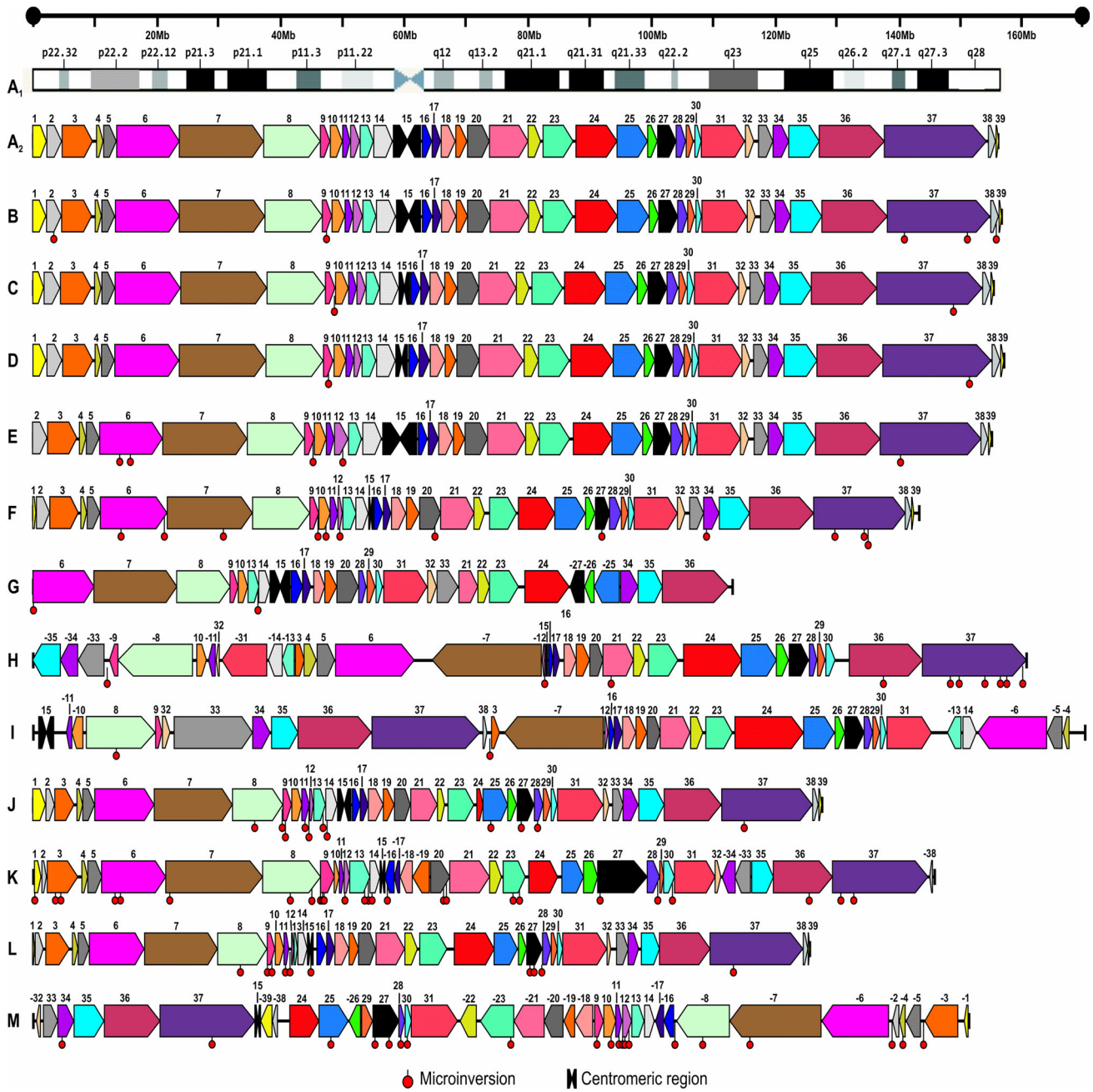

Fig. 1 Synteny map scheme for 13 X chromosomes. The *black numbers* indicate syntenic segments ($n = 39$) which were conserved among 13 mammalian species. X chromosomes are orientated from p (*short arm*) to q (*long arm*). The inverted orientation of syntenic blocks is indicated by a (-) sign. Genomes from *top* to *bottom*: human X chromosome ideogram (*A1*) and syntenic block ordination (*A2*); chimpanzee (*B*); gorilla (*C*); orangutan (*D*); macaque (*E*); marmoset (*F*); rabbit (*G*); rat (*H*); mouse (*I*); dog (*J*); pig (*K*); horse (*L*); cow (*M*). *Red circles* represent micro-inversions (Color figure online)

ranging from 0.85 Mb (SB4) to 15.2 Mb (SB37). Covariance analysis of SB length revealed high variability among mammalian species (Fig. 2; Supplementary Table S2). The length of six SB (6, 19, 31, 34, 35, and 36: group 1) displayed minimum variance (homogeneous blocks, covariance (CV) $\leq$10 %). Ten SB (7, 8, 9, 18, 20, 22, 23, 25, 26, and 30: group 2) were classified as moderately homogeneous blocks (CV = 10.1–20 %). The remaining 23 SB (group 3) had high variance regarding their length (heterogeneous blocks, CV >20 %). Rodents (especially rabbit) had large (heterogeneous) length variations regarding SB (1–5, 11, 12, 24, 33, and 37). Conversely, SB length did not have a significant variation in other species, such as anthropoid primates. The study of genetic information loss and gain for each human SB showed that most of them gained an average of 292.9 Kb (Supplementary Table S3;

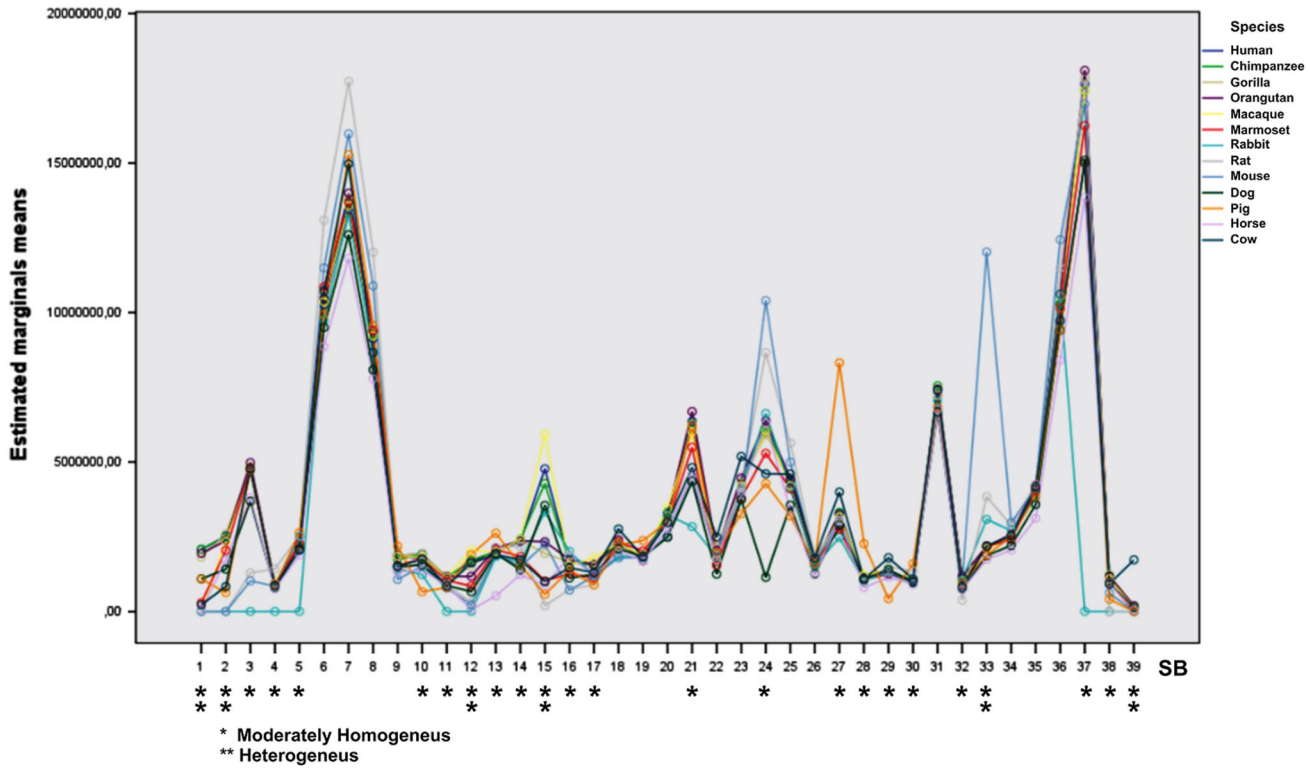

Fig. 2 Marginal means estimated for each SB. The *colored lines* represent marginal means variation regarding SB size for each species. *Unmarked* homogeneous blocks (CV ≤10 %, group 1). *Asterisk* moderately homogeneous blocks (CV = 10.1–20 %, group 2). *Double asterisk* heterogeneous blocks (CV >20 %, group 3) (Color figure online)

Fig. 3a, Supplementary Figure S1). Sixty-four percent (25/39) of SB had a gain in genetic information, ranging from 28.2 Kb (SB18) to 2.5 Mb (SB37). Twenty-six percent of SB (14/39) had a loss of genetic information, ranging from 927 bp (SB34) to 976 Kb (SB33). Comparing human SB length to that of other species showed that 75 % (9/12) of them had gained genetic information (Supplementary Table S3, Fig. 3b). A loss of genetic information ranging from 3.8 to 29.5 Kb was exclusively observed when human SB were compared to those for chimpanzee, orangutan and mouse.

Chromosomal rearrangements

A total of 252 chromosomal events were identified and classified into 4 categories: deletions, insertions, non-homologous regions (non-HR: considered regions having less than 10 % nucleotide identity), and inversions (Fig. 4). Deletions were found at a frequency of 12.3 % (31/252) (Supplementary Table S4, deletion section), having 2.1 Mb length on average. Rodents (rabbit, rat and mouse) had the most deletions (45 %) compared to the human genome. Rodents shared three SB deletions (1, 2, and 39) (Fig. 1; Supplementary Table S4). Five SB (3, 4, 5, 37, and 38) were lost exclusively in rabbit. One of these was a 3 Mb X-to-17 chromosome translocation (Table S4, deletion section). Nine large deletions encompassing 3.8 Mb were recorded in pig (429 Kb average). Compared to the human genome, rabbit had the largest loss of genetic information since seven large deletions (4.9 Mb average) were observed. A few deletions (1.4 Mb average) were observed in dog ($n = 3$), horse ($n = 3$) and cow ($n = 1$). Deletions were uncommon in anthropoid primates, only being observed in macaque (2.3 Mb total deletion length). SB1 was completely deleted in this species, but partially present in marmoset, pig, horse, and cow (Supplementary Table S4, deletion section).

Insertions occurred with 3.5 % (9/252) frequency. Each had an average 459 kb length (Supplementary Table S4, insertion section). Pig had the most insertions ($n = 4$; 653 Kb length, 163 Kb/insertion) having a gain of three genes (*ENSSSCG00000012105-07*, *ENSSSCG00000012240*, and *ENSSSCG00000012331* genes). Mouse carries two insertions. One encompasses ~1 Mb, consisting of 23 genes (*Wdr44* to *GM4764* genes) and the other 2.2 Mb (*Hs6st2* to *Mospd1* genes). One insertion was identified in rabbit (104 Kb), dog (39 Kb) and horse (14 Kb) (Supplementary Figure S2; Table S4).

A total of 63 non-HRs (25 % of all chromosomal rearrangements) were detected. Most were found in anthropoid

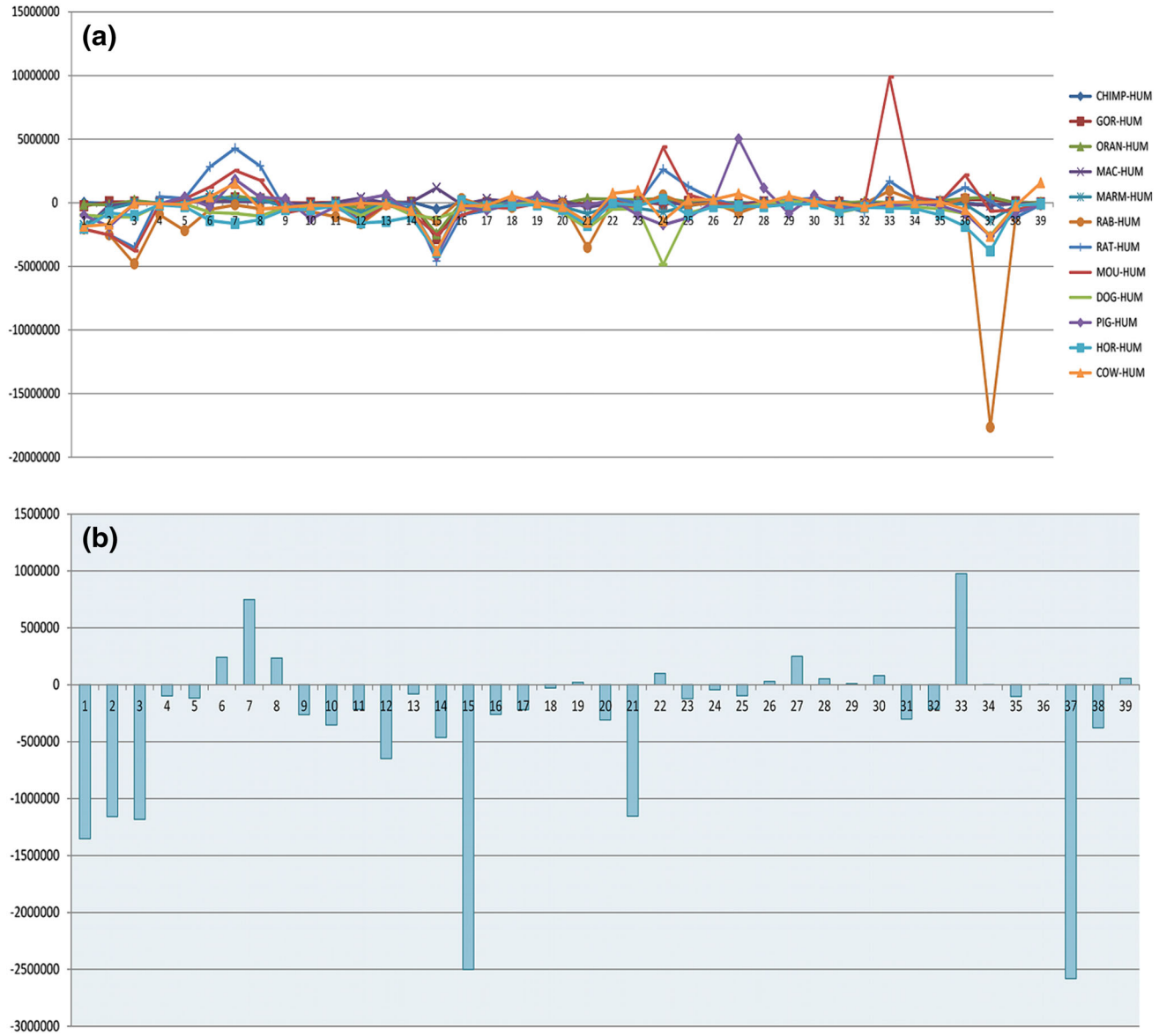

Fig. 3 SB gain and loss of genetic information. **a** Paired analysis by species, **b** comparative analysis between human and other species. *Negative* and *positive* values indicate gain and loss of genetic information in the human lineage, respectively

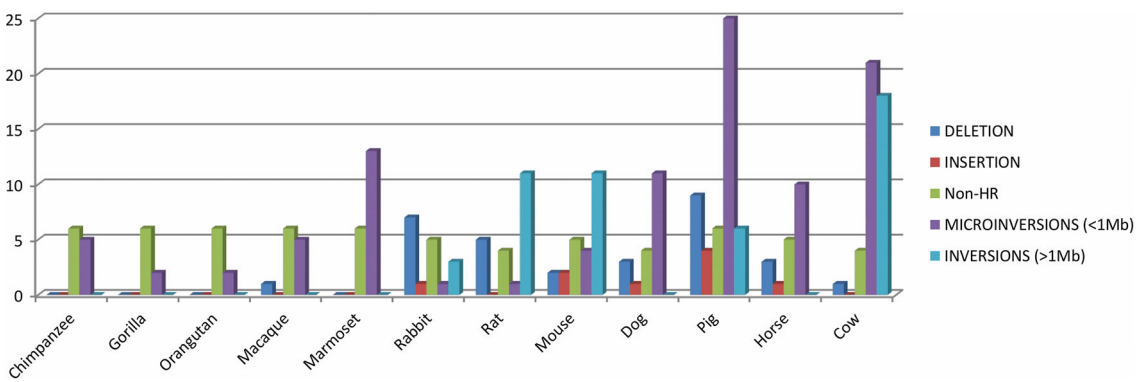

Fig. 4 Chromosomal events identified in mammalian X chromosomes. Deletions, insertions, low nucleotide identity regions, non-HR, micro-inversions, and inversions identified by species. The human X chromosome was taken as reference

primates (30/93, 47.6 %). Taking the human genome as reference, five of six non-HRs were located at identical coordinates in all anthropoid primates. These regions carry *SSX, GAGE,* and *XAGE* genes (Table S4, non-HR section). Rodents, dog, horse, and cow shared four non-HRs at the same genomic location. Pig displayed four exclusive non-HRs (270 Kb on average), including one deletion (*PHCDH11X*) and one duplication (*RAB9B*) (Supplementary Table S4, non-HR section).

One hundred and forty-nine inversions (59.1 % of all chromosomal rearrangements) were detected (Fig. 4). 67.1 % (100/149) were classified as micro-inversions, since they encompassed less than 1 Mb. Average micro-inversion size was 241.4 Kb, ranging from 3 Kb to 891 Kb. Anthropoid primates had presented 27 micro-inversions, 13 (48 %) being present in marmoset. A unique 27 Kb micro-inversion (human coordinates: X: 46,809,122–46,836,795) was shared by chimpanzee, orangutan, macaque, and marmoset. Supplementary Table S4 (micro-inversion section) shows micro-inversion distribution in all species. Micro-inversions were concentrated at specific SB in pig, cow, dog, horse, and rodents (6, 9, 10, 12 and 37). Complete and partial deletions (12 in pig, 3 in dog, 3 in cow, and 1 in horse) and gene duplications (2 in pig) caused by micro-inversions were identified.

Based on SB position, the minimum number of large inversions (>1 Mb) needed to transform the ancestral mammalian X chromosome into any other mammalian X chromosome was determined. There were 49 large inversions (Supplementary Table S4, total rearrangement section).

Testing inversion phylogeny

Matrix inversion distance was obtained from consecutive SB order and direction. A phylogeny of inversions in mammalian X chromosomes, which included three ancestral chromosomes (Table 1; Fig. 5), was then proposed using GRIMM and UniMoG software. An ancestral mammalian X chromosome (A1) sequence was predicted. Such structure was then searched for the inversions necessary to transform the X chromosome from one species into another. GRIMM, HP, and the inversion model identified 44 chromosomal inversions (X1–X44). Forty-three chromosomal inversions (X1–X43) were identified using the restricted DCJ model. A minimum of 16 (X1–X17 for the DCJ model) or 17 inversions (X1–X17 for GRIMM software, HP and inversion model) were required to transform A1 to the cow X chromosome. Three chromosome inversions (X18–X20 from the DCJ model, X19–X21 from GRIMM software, HP and inversion model) were necessary to transform A1 to the ancestral structure shared by rat and mouse (A2). Seven A2 inversions produced mouse (X21–X27 for DCJ model or X22–X28 for GRIMM software, HP and inversion model) and rat X chromosomes (X28–X34 for the DCJ model or X29–X35 for GRIMM software, HP and inversion model) (Supplementary Figures S1, S2). A3 was the ancestral anthropoid primate, dog, horse, rabbit, and pig X chromosome which originated from a unique inversion (X17 for the DCJ model or X18 for GRIMM software, HP and inversion model). Rabbit and pig X chromosomes were generated from A3 by three (X35–X37 DCJ model or X36–X38 for GRIMM software, HP and inversions model) and six inversions (X38–X43 for the restricted DCJ model or X39–X44 for GRIMM software, HP and inversion model) (Supplementary Figure S3).

Fragile regions in the X chromosome

One hundred and sixty-four regions were identified adjacent to 82 inversion breakpoints (IB). Based on the human SB composition (Fig. 1), 117, 108, 121, and 120 distinct reused regions were detected (at least two in the same region) using GRIMM, DCJ, HP and inversion model, respectively. Forty-four regions had 0 to 1 IB (Supplementary Table S5). Thirty fragile regions (ranging from 0.79 Kb to 2.2 Mb in length) were identified in the human X chromosome (Table 2; Fig. 6). Ten chromosome regions were identified which had less than 2 breakpoints (Fig. 6). These findings enabled human X chromosome fragile regions identified in the present study to be compared to those reported in the pertinent literature (Fig. 7; Supplementary Table S6).

A total of 10,274 transposable elements (TE) were detected in fragile regions (354.2 copies/region average and 1.64 TE/Kb density) (Supplementary Table S7). 34 % were L1/LINE-like and 19 % Alu/LINE-like elements while 11 % were classified as hAT/TIR-like elements (other types of TE did not exceed 8 %). Over 99.9 % of the detected TEs could be considered as being ineffective or incomplete copies which might lack the ability to transpose. Fifty-nine TEs displayed >95 % nucleotide identity with the consensus sequences available at the Repbase database. Five of these copies (L1HS-like) contained an open reading frame encoding a transposase enzyme. No correlations ($r^2 < 0.1$) were observed between reuse number and TE type, fragile region size or TE/Kb density.

Discussion

X chromosome SB in mammals

Previous studies concerning a reduced number of vertebrate species have reported 11–17 SB located in the X chromosome (Bourque et al. 2005; Pevzner and Tesler 2003a). The first studies revealed an average 9.6 Mb size for human and

Table 1 Inversion distance matrix in mammals

	Hu	Ch	Go	Or	Mq	Ma	Ra	Mo	Ra	Do	Pi	Ho	Co
Human (Hu)	–	0	0	0	0	0	3	11	11	0	6	0	**17**/18[a]
Chimpanzee (Ch)	0	–	0	0	0	0	3	11	11	0	6	0	**17**/18[a]
Gorilla (Go)	0	0	–	0	0	0	3	11	11	0	6	0	**17**/18[a]
Orangutan (Or)	0	0	0	–	0	0	3	11	11	0	6	0	**17**/18[a]
Macaque (Mq)	0	0	0	0	–	0	3	11	11	0	6	0	**17**/18[a]
Marmoset (Ma)	0	0	0	0	0	–	3	11	11	0	6	0	**17**/18[a]
Rabbit (Ra)	3	3	3	3	3	3	–	14	14	3	9	3	**20**/21[a]
Mouse (Mo)	11	11	11	11	11	11	14	–	14	11	17	11	23
Rat (Ra)	11	11	11	11	11	11	14	14	–	11	17	11	**25**/27[a]
Dog (Do)	0	0	0	0	0	0	3	11	11	–	6	0	**17**/18[a]
Pig (Pi)	6	6	6	6	6	6	9	17	17	6	–	6	**22**/23[a]
Horse (Ho)	0	0	0	0	0	0	3	11	11	0	6	–	**17**/18[a]
Cow (Co)	18	18	18	18	18	18	21	23	27	18	23	18	–

[a] Numbers in bold show rearrangements reported by restricted DCJ model. Numbers which have not been highlighted represent rearrangements estimated by GRIMM software, HP and inversion models. Even though the same number of events was indicated, they may not necessarily be shown in the same order or the same type of event

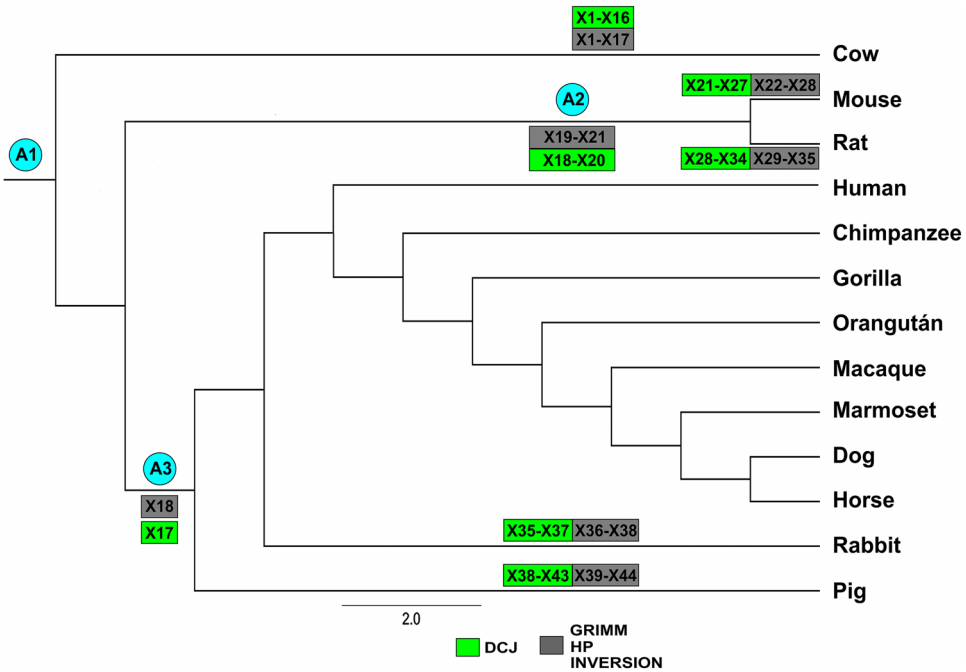

Fig. 5 X chromosome phylogeny of inversions. Phylogenetic tree obtained using GRIMM and UniMoG software. Inversion changes in phylogeny are shown in *bold*. *Blue circles* show hypothetical ancestral ordination. *Green rectangles* represent inversions produced by the restricted DCJ model. *Gray rectangles* represent inversions found using GRIMM, HP and inversion models (Color figure online)

8.5 Mb for mouse SB. Furthermore, an average 668 Kb in human and 458 Kb in mouse were identified for breakpoint regions (BPR) (Pevzner and Tesler 2003a). These results which reported large SB suggested further hidden regions. More recently Bourque et al. (2005) detected 17 SB in human (average 3.2 Mb), mouse (2.8 Mb), and chicken X chromosome (1.2 Mb). They also reported average BPR size: 1.8 Mb in human, 1.4 Mb in mouse, 1.6 Mb in rat, and 395 Kb in chicken. The SB generated in such work were mainly based on orthologous gene location combined with paired alignment. This methodology (tending toward the identification of large SB) led to the detection of a reduced number of long BPRs.

A combination of bioinformatics methods in the present study has been used to determine the number of SB located in 13 mammalian species' X chromosome.

Thirty-nine SB were identified, ranging from 0.85 to 15.2 Mb in length. Average SB size was 3.8 Mb in human and mouse and 3.5 Mb in the other species. It was estimated that BPR in mammals encompassed 210 kb on average. These results significantly differed from those reported in previous studies. This might have been due to MAUVE's high sensitivity for detecting SB structure. Indeed, this software represents the first sequence system which integrates an analysis of large-scale evolutionary events with traditional multiple sequence alignment

Table 2 Fragile regions located on the human X chromosome

Region	Coordinates	Size in bp	Intergene region
1	X:0–30,513	30,513	Telomeric region
2	X:1,762,694–2,135,866	373,172	ASMT-DHRSX
3	X:4,659,472–5,115,200	455,728	PRKX-NLGN4X
4	X:9,917,948–10,411,034	493,086	SHROOM2-MID1
5	X:13,421,823–13,568,985	147,162	ATXN3L-EGFL6
6	X:23,804,997–23,851,764	46,767	SAT1-APOO
7	X:37,301,730–37,302,526	796	PRRG1-LANCL3
8	X:46,432,037–46,434,653	2,616	ZNF674-CHST7
9	X:48,273,689–48,315,998	42,309	SSX4B-SLC38A5
10	X:51,253,261–51,483,700	230,439	NUDT11-GSPT2
11	X:53,024,906–53,074,777	49,871	FAM156A-GPR173
12	X:57,516,741–57,617,343	100,602	FAAH2-ZXDB
13	X:62,519,964–62,564,676	44,712	CBX1P1-SPIN4
14	X:64,256,246–64,625,979	369,733	ZC4H2-TLE1P1
15	X:65,489,465–65,812,766	323,301	HEPH-EDA2R
16	X:67,946,711–68,047,313	100,602	STARD8-EFNB1
17	X:73,073,661–73,402,982	329,321	XIST-BMP2KL
18	X:79,700,763–79,924,323	223,560	FAM46D-BRWD3
19	X:80,554,843–82,755,224	2,200,381	SH3BGRL-POU3F4
20	X:86,962,912–87,995,314	1,032,402	RPSAP9-CPXCR1
21	X:93,764,789–94,014,015	249,226	FAM133A-DIAPH2
22	X:98,374,387–98,382,503	8,116	DIAPH2-PCH19
23	X:104,331,409–104,341,330	9,921	IL1RAPL2 (intron 2)
24	X:105,455,019–105,853,987	398,968	MUM1L1-CXorf57
25	X:107,036,569–107,068,383	31,814	TSC22D3-MID2
26	X:114,541,809–114,794,604	252,795	LUZP4-PLS3
27	X:116,057,222–116,998,293	941,071	CXorf61-KLHL13
28	X:125,877,163–125,880,720	3,557	DCAF12L1-CXorf64
29	X:153,796,807–153,902,568	105,761	IKBKG-GAB3
30	X:155,240,626–155,270,560	29,934	IL9R-telomeric region

methods. This software facilitates the identification and alignment of both conserved regulatory and hypervariable intergene regions (Darling et al. 2010). The high sensitivity for locating SB reported here is also related to the significant number of genomes used for multiple alignment at bp level. Previous comparisons have been made between a limited amount of related genomes, which has resulted in lower quantity BPR. Our analysis was based on current genome versions which contain less annotation errors and gaps which might have affected chromosome structure/length estimation.

Concerning interspecific SB variation, significant differences in average size were observed ($p < 0.0001$) between covariance groups (groups 1–3). For instance, group 1 average SB size was 6 Mb while this was 4.3 Mb for group 2 and 2.5 Mb for group 3. These features indicated that short SB tend to have great interspecific variability during mammalian chromosome X evolution, while large SB tend to be more homogeneous. Bourque et al. (2005) reported an average of 10.7 genes/SB (ranging from 3 to 120) in the human genome. Their observations argued in favor of the fact that short SB have lower gene density regarding intergene regions, which could explain the significant variability in size among species. Human/mouse and human/rat sequence comparisons showed that all orthologous regions could be aligned in a pair-wise fashion and that ~40 % of the human genome could be aligned with rodent genomes (Gibbs et al. 2004; Mouse Genome Sequencing et al. 2002). Moreover, some short SB may become completely or partially deleted in the course of evolution (Pevzner and Tesler 2003a, b). These findings supported the SB variability (especially in short SB) observed in the present study. Our results showed that some short SB (1, 15 and 39) located in telomeric and centromeric regions had interspecific size variability. Indeed, it has been shown that the large number of highly

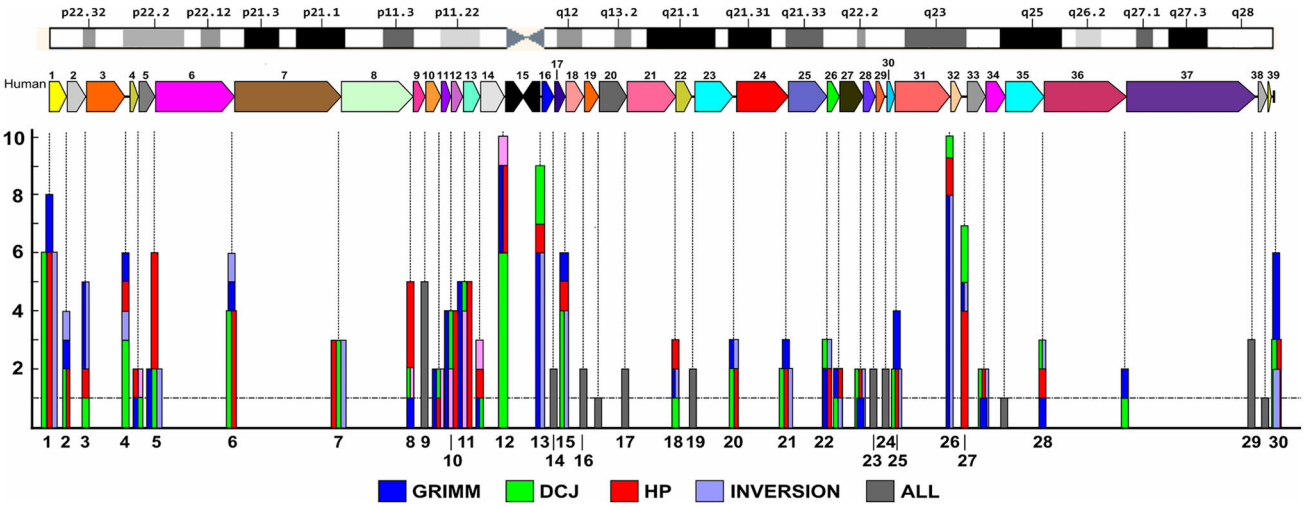

Fig. 6 Fragile regions detected in the human X chromosome. *Top* human X chromosome ideogram with cytological bands. SB on the human X chromosome are shown as *colored arrows*. *Bottom color bars* represent the number of breakpoints by region. *Blue bars* represent breakpoint reuse displayed by GRIMM software. Restricted DCJ model: *green bars*; HP: *red bars*; inversion model: *lilac bars*. All models: *gray bars*. Non-reuse breakpoints are those having average lower than 2 values (Color figure online)

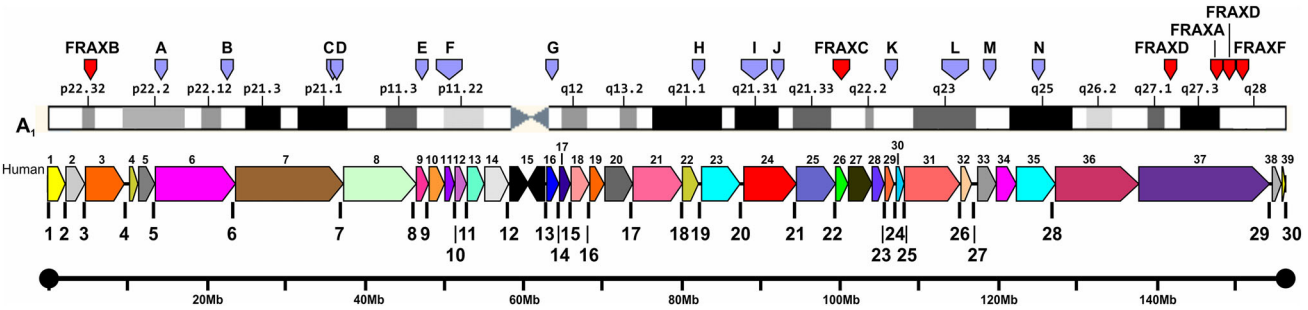

Fig. 7 Fragile regions on the human X chromosome. *Top* human X chromosome ideogram with cytological bands. *Red arrows* show previously reported human fragile regions. A–N (*blue arrows*) represent SB reported by Ruiz-Herrera et al. (2006). Fragile regions identified in the present paper are numbered from 1 to 30 (*bold*) (Color figure online)

repeated and rapidly evolving heterochromatin DNA sequences (e.g., centromeric and telomeric) have been related to methodological limitations (sequencing) thereby leading to potentially inaccurate assembling into specific genomes (DeBaryshe and Pardue 2011). Such telomeric region analysis in distinct species has revealed a high insertion rate for particular types of TE (DeBaryshe and Pardue 2011; George et al. 2006).

Analyzing species' gain or loss of genetic information (for each SB) revealed an average 292.9 Kb gain in the human X chromosome. As previously suggested, this feature may have been due to a high number of segmental (SD) and gene duplications (Eichler 2001). In fact, human sequences harbor an excess of large and complex interspersed SDs involving substantial mutational consequences related to evolution and disease, compared to other mammalian genomes (Bailey and Eichler 2006; Gazave et al. 2011; Marques-Bonet et al. 2009). For instance, human and great-ape lineages have undergone a surge of genomic duplications over the last 10 million years (Marques-Bonet et al. 2009). It has been proposed that large expanses of genomic sequences adjacent to telomeric and pericentromeric regions in the human genome have almost solely emerged through segmental duplication events during primate evolution (Horvath et al. 2000; Samonte and Eichler 2002).

In agreement with this assumption, it was found that only three species had increased X chromosome size when compared to human sequences, chimpanzee (3.8 Kb), orangutan (16.1 Kb), and mouse (29.5 Kb). Such variability was conferred by a restricted number of SB. For example, mouse SB33 contained 9.8 Mb more than the corresponding region in the human genome (Fig. 3). Waterston et al., suggested that dozens of gene family expansions at particular locations have occurred in the mouse lineage, most of which involve genes related to reproduction, immunity, and olfaction (Mouse Genome Sequencing et al. 2002). These findings suggested that

some physiological systems in rodents have been the focus of extensive specific lineage adaptation.

X chromosomal rearrangements in mammals

The present work revealed that deletions of contiguous SB represent a small proportion of rearrangements (31/252). Rodents had most deletions (45 %) compared to other mammalian species. These regions, which are particularly large (average 2.1 Mb), are restricted to certain locations (SB 1, 2, 38, and 39) in the X chromosome, mainly being homologous to human peritelomeric regions. The rabbit genome also had exclusive deleted regions (SB: 3–5 and 37) when compared to those for human, mouse, and rat. Such interspecific variability between human and rodents can be explained by some lineages having a high rate of chromosome rearrangement (especially non-functional DNA). Compared to humans, pig, dog and mouse had smaller genomes (7–14 %) (Mouse Genome Sequencing et al. 2002). However, there was significant homology between human and other species, such as pig. At nucleotide level, the swine genome was 3 times more similar to human than mouse (Archibald et al. 2010; Humphray et al. 2007).

Concerning insertions, our results revealed that they accounted for a small proportion of all rearrangements (9/252). Such inserts covered short regions (average 459 Kb). Our results led to two large insertions being identified which were exclusive to the mouse genome. One of them (1.69 Mb) contains 23 genes associated with a series of duplication events in the *Slx* gene family (Supplementary Table S4). Slx, Sxl1, and Sly proteins are known to have antagonistic effects on sex chromosome expression in developing sperm and skew the offspring sex-ratio in favor of females *cf* males (Cocquet et al. 2012). These findings supported their crucial role during mice speciation. Another insertion (2.2 Mb) was detected in an intergene region. A gain of three genes having an unknown function was identified in pig (Supplementary Table S4), which may also have been related to speciation. Other small insertions (average size 52 Kb) were also identified, which were located in intergenic regions, were shared by rabbit, dog, and horse. This feature might have been related to these species' specific critical physiological and/or molecular characteristics. Interestingly, a 3 Mb region was identified in rabbit which was translocated from the X to the 17 chromosome. This fragment was homologous to the human X:49,965,119–53,032,610 region and contains the *BMP15* gene which plays a crucial role during mammalian reproduction (Shimasaki et al. 2004). Indeed, mutations of this gene in humans have been associated with premature ovarian failure etiology (Di Pasquale et al. 2004; Laissue et al. 2008).

Non-HRs regions had an average 547 Kb length and they were most frequently present (47.6 %) in anthropoid primates. Multiple sequence comparison of these regions in human and other mammalian species revealed low homology. However, paired analysis (excluding the human genome) showed over 40 % increase in nucleotide identity. Genetic analysis of these regions demonstrated that they mainly consisted of gene families (*SSX, GAGE, XAGE*) located in tandem. These genes have been classified as cancer-testis (CT) antigens. Such proteins are exclusively expressed in adult testis germ cells as well as during fetal testis and ovary development. CT antigens represent a group of tumor-associated genes containing more than 130 RefSeq annotated transcripts (Caballero and Chen 2009; Hofmann et al. 2008; Scanlan et al. 2002; Simpson et al. 2005). Furthermore, it has been predicted that $\sim$10 % of X-linked genes are CT antigen type (Ross et al. 2005). According to our analysis, the increased number of gene copies in this family argues in favor of human speciation-related duplication (Supplementary Table S4). CT gene deletions and potential deregulation might thus be involved in human cancer development and may partly explain why human species display increased susceptibility to developing cancer. Non-HR in pig showed similar behavior since deletion of *PHCDH11X* and *RAB9B* genes was identified in two regions in particular (X: 76,402,730–77,127,799 bp and X: 82,384,943–82,422,971 bp). Molecular studies have reported a significant association between *PCDH11X* SNPs and late-onset (>65 years) Alzheimer's disease (LOAD) (Carrasquillo et al. 2009; Lambert et al. 2009). *RAB9B* has similarly been associated with the Chediak–Higashi syndrome (Davies et al. 1997; Seki et al. 2000). A total of 100 micro-inversions were detected in the present work in the mammalian X chromosome (241.4 Kb average size). Previous studies involving whole-genome comparative analysis have described 3,170 micro-inversions occurring between human and mouse SB (Pevzner and Tesler 2003a). Although most of these chromosomal rearrangements probably played a major role during genome evolution, some of them might reflect artifacts issued from inaccurate assemblies (Pevzner and Tesler 2003a, b).

The present study found that micro-inversions were mainly located on five SB (6, 9, 10, 12, and 37) and distributed among some species, such as anthropoid primates ($n = 27$) and pig ($n = 25$). As mentioned above, such differential behavior could be explained by SB nucleotide composition and by artifacts being sequenced during genome assembly. Similarly to large chromosomal inversions, micro-inversions could originate from two major mechanisms: non-allelic homologous recombination (NAHR) and non-homologous end-joining (NHEJ). NAHR can occur between TEs, SDs or short repeated sequences (Caceres et al. 2007; Coulibaly et al. 2007; Delprat et al. 2009;

Kupiec and Petes 1988; Lim and Simmons 1994; Richards et al. 2005) whereas NHEJ produces inverted duplications on both sides of an inverted segment (Furuta et al. 2011; Ranz et al. 2007; Sonoda et al. 2006).

Human and chimpanzee comparative genomics have shown that micro-inversions flanked by SDs result from NAHR between distinct SDs arranged in inverted orientations (Kolb et al. 2009). As well as large chromosome inversions, micro-inversions play an important role in species adaptation and speciation since their recombination-reducing effect keeps alleles together at loci with fitness epistatic effects ("coadaptation hypothesis") (Kirkpatrick and Barton 2006; Navarro et al. 1997). Furthermore, inversions might be fixed in species due to direct mutational effects associated with inversion breakpoints located near or inside genes, which might affect their function and/or expression profile ("position effect" hypothesis) (Sperlich and Pfriem 1986). Our approach also led to identifying 15 micro-inversions (especially in pig), giving rise to gene deletions and duplications (Supplementary Table S4). Regarding this point, some micro-inversions have been associated with encoding region disruption, leading to genetic loss of function, such as *Dpp6* in mice and the *hemophilia A factor VIII* gene in humans (Hough et al. 1998; Lakich et al. 1993). Most micro-inversions detected in the present study contained up to two genes for which breakpoints were located near 5′UTR regions, potentially affecting their regulatory sequences. According to the position effect hypothesis, these features might have implications for gene expression patterns and would place the encoding region in a different regulatory context, as described for the *Drosophila melanogaster Antp73b* gene mutation and the *Antirrhinum majus Niv* gene inversion (Frischer et al. 1986; Lister et al. 1993). A significant number of disease-related position-effect cases have been described in humans (Kleinjan and Lettice 2008; Kleinjan and van Heyningen 2005).

X chromosome evolution in mammals has been well studied. Sex chromosomes have evolved from a pair of autosomes during 300 million years (Ross et al. 2005). Several studies have suggested that the X chromosome might be divided into five evolutionary strata because of step-wise repression of recombination during evolution (Lahn and Page 1999; Ross et al. 2005; Wilson and Makova 2009). The X-conserved region (XCR) consists of the oldest strata (named 1 and 2). The X-added region (XAR) includes younger strata which are shared by primates and rodents. Younger strata (4 and 5) were derived from primates (Graves et al. 2006; Wilson and Makova 2009). According to these observations, the largest and more frequent changes might be easily found in older strata due to the loss of recombination. These features have been corroborated by our data which revealed that anthropoid primates have not large inversions. By contrast, in more distant lineages large X chromosome rearrangements have been detected (e.g., pig, cow, and rodents). It has been demonstrated that the distribution of X-linked gene duplications is correlated with the recombination environment (Zhang et al. 2010). Zhang et al. showed that the X chromosome sequence and expression have dramatically evolved after the split of eutherian mammals and marsupials. Specifically, the X chromosome presented genetic gain which might explain the presence of duplicated genes adjacent to breakpoints of chromosomal rearrangements.

Inversion phylogeny, fragile sites, and evolutionary breakpoints

An in silico method was used in this work for determining inversion phylogenies in mammalian X chromosomes. It was estimated that a minimum of 43–44 inversions (>1 Mb) were required to transform the human X chromosome into any other mammalian X chromosome (Fig. 5). Previous studies have detected up to 7 chromosomal inversions between human and mouse (Bafna and Pevzner 1995; Pevzner and Tesler 2003a). It has been shown that chicken, rat, mouse, and human share 19 inversions (Bourque et al. 2005). The present work's sensitivity for identifying SB (and therefore chromosomal rearrangement) was mainly due to the comparative genomic method (discussed above) and to the significant number of species analyzed. Our results concerning the phylogeny of inversion analysis showed that particular species had a high chromosomal inversion rate. For instance, rat and mouse species had an 11 inversion difference compared to the human X chromosome (Fig. 5). These results corroborated previous reports of an accelerated intra-chromosome rearrangement rate in rodents (Bourque et al. 2005).

Comparing four different models using two programs (GRIMM and UniMog) led to inferring fragile regions located in chromosomes and in entire genomes (with high reliability). Interestingly, although these programs involved using different algorithms, their final results were highly similar. Differences among them allowed us to discard potential false positive and negative findings.

Our approach led to 88 breakpoints and 176 adjacent affected regions being expected to be present on the X chromosome. However, only 82 breakpoints and 164 adjacent regions (108–120 reused) were observed (Supplementary Table S5). These findings indicated that, regardless of inversion sequence, ~ 27 % ($n = 22$) of breakpoints were not reused and that ~ 73 % ($n = 60$) had been reused at least twice during mammalian X chromosome evolution. Inversion breakpoint reuse is a common but enigmatic phenomenon in whole species. Coincident inversion breakpoints

have been observed in *Drosophila*; at least 23 % of inversion breakpoints had been reused, 70 % of them occurring in chromosome 2 (Wasserman 1982). Different mammalian species have shown an increase in breakpoint reuse (Murphy et al. 2005; Pevzner and Tesler 2003a, b; Zhao et al. 2004). Pevzner and Tesler (Pevzner and Tesler 2003a) have estimated that BBRs have been intensively reused in mammalian evolution (supporting FBM and contradicting the random breakage model, RBM) but Ma et al., (2006) revealed low breakpoint inter-reuse (contradicting the FBM). Alekseyev and Pevzner (Alekseyev and Pevzner 2010) recently proposed a turnover fragile breakage model (TFBM) thereby extending the FBM and complying with the multispecies breakpoint reuse test. Our results were consistent with the TFBM as they revealed five regions having high breakpoint reuse concentration or EBR: X:1–13,552,391, X:23,831,317–23,851,764, X:46,434,653–116,057,222, X:136,149,783–136,159,117, and X:153,803,660–155,240,626 (Fig. 6).

A comparative analysis of 9 genomes (human, mouse, rat, cat, cow, dog, pig, horse, and chicken) by Ruiz-Herrera et al. (2006) identified 14 EBR located in the human X chromosome. 12 of them coincided with inversion breakpoint hotspots in the human X chromosome which have been identified in the present work (Fig. 7). A perfect coincidence of two fragile regions (FRAXB and FRAXC) was also identified with those previously reported in the human X chromosome (FRAXA, Xq27.3, FRAXB, Xp22.32, FRAXC, Xq22.1, FRAXD, Xq27.2, FRAXE, Xq28, and FRAXF, Xq28) (Arlt et al. 2003; Knight et al. 1993; Kremer et al. 1991; Parrish et al. 1994; Ritchie et al. 1994; Schwartz et al. 2006; Verkerk et al. 1991) (Fig. 7; Supplementary Table S6). FRAXB, FRAXC, and FRAXD have been classified as CFR while FRAXA, FRAXE, and FRAXF have been defined as RFR (Schwartz et al. 2006). There has been no indication that RFRs are conserved in other species than human until now, which might explain the mismatch with our findings. It has been described that CFRs are conserved in primates and other mammalian species (Arlt et al. 2003). Orthologous sequences have been found in rodents, pig, cow, horse, cat, dog or distinct primate species for some CFRs in human sequences (Elder and Robinson 1989; Fundia et al. 2000; Glover et al. 1998; Helmrich et al. 2006; Krummel et al. 2002; Rodriguez et al. 2002; Ronne 1992, 1995a, b; Ruiz-Herrera et al. 2004, 2005, 2002; Shiraishi et al. 2001; Yang and Long 1993). The conservation of fragile sites across a wide range of species might imply that these regions play a functional role (Schwartz et al. 2006).

Nucleotide sequence analysis of 30 hotspots in the present study revealed an average 268.9 Kb size, having high TE density (Supplementary Table S7). Thirty-four percent of them were L1/LINE-like and 19 % Alu/LINE-like elements; only 0.01 % of TEs located in the fragile regions represented effective copies capable of transposition. According to Ross et al. (2005), interspersed repeats account for 56 % of the euchromatic X chromosome sequence, compared to a 45 % genome average. Furthermore, L1 have been shown to comprise 29 % of the X chromosome sequence and *Alu* elements 8.3 %. L1 represented an average of 17 % of the whole genome and *Alu* 11 % (Ross et al. 2005). The increased presence of this kind of TE in the human X chromosome might thus explain the significant number of chromosomal rearrangements identified in this study.

Interestingly, previous studies have reported that CFS regions are enriched by pre-neoplastic lesion-related *Alu* repeats (Tsantoulis et al. 2008). Our data showed a considerable increased frequency for both types of TE in the human X chromosome's fragile regions. However, no statistical difference ($p = 0.1$) was found between the number of breakpoint reuse and the presence of a specific TE class. Our study also led to quantifying fragile regions' differences in size related to breakpoints that were not reused. Indeed, fragile regions which were reused tended to be larger (268.9 Kb average size) than those which were not reused (23.8 Kb average) (Fig. 6). This might have been due to a "natural" overrepresentation of genomic repeat elements (mainly TEs) in long intergene regions. This feature might facilitate ectopic recombination events.

Conclusions

Taken together, our findings constitute a useful tool for interpreting genomic events which have occurred during mammalian X chromosome evolution. Our results are consistent with the theory concerning the five steps evolution of the X chromosome. They may also be used for studying human disease etiology and evolution itself. Indeed, 30 fragile regions described in the present work matched breakpoints previously related to human disorders, such as premature ovarian failure and cancer (data not shown). The methodology presented here could be expanded to entire genomes to establish a comprehensive map of chromosome evolution dynamics.

Materials and methods

Genomic sequences and multiple alignments

Comparative genomic analysis involved using 13 eutherian mammals' complete X chromosome sequences which are available at the Genbank (NCBI) database: *Homo sapiens* (NC_000023.10), *Pan troglodytes* [NC_006491.3], *Gorilla*

gorilla [FR853103.1], *Pongo abelii* [NC_012614.1], *Macaca mulatta* [NC_007878.1], *Callithrix jacchus* [NC_013918.1], *Oryctolagus cuniculus* [NC_013690.1)], *Rattus norvegicus* [NC_005120.3], *Mus musculus* [NC_000086.7], *Canis lupus familiaris* [NC_006621.2], *Sus scrofa* [NC_010461.4], *Equus caballus* [NC_009175.2], and *Bos Taurus* [NC_009175.2] (Supplementary Table S1). Sequences were aligned using MAUVE aligner software v.2.3. progressive alignment algorithm 1 (Darling et al. 2010). Briefly, MAUVE involves an efficient methodology for constructing multiple whole-genome alignments regarding large-scale evolutionary events, such as rearrangement and inversion. The resulting alignments represented a mosaic of rearranged segments which were conserved among complete genomes, subsets of genomes, or unique genome segments (Darling et al. 2004, 2008). Alignments were made using the following parameters: skip-refinement and seed-weight = 15, total alignment, determining local collinear blocks (LCB) and pairs of LCBs. The option of using seed families in the anchorage and linear genomes was ignored. The input files were phylogenetically organized in a multi FastA file which included all X chromosomes. The human X chromosome was taken as reference for further analysis. Subsequent analysis involved using the backbone output file to identify each LCB's nucleotide coordinates and the corresponding breakpoints for chromosomal rearrangement, thereby enabling LCBs to be located at X chromosome coordinates for each species tested here.

Detecting rearrangement breakpoints and fragile chromosome regions

Chromosomal rearrangement (inversion, deletion, and insertion) was identified using pairwise comparison. Genes lying adjacent to genomic breakpoints were positioned by comparing LCB coordinates to the available annotation coordinates for each available genome in the Ensembl database (http://www.ensembl.org). These annotations were then used for identifying SB location on all X chromosomes and establish differences in their length between species. The following were used for calculating the minimum number of rearrangement events required to transform one genome into another: GRIMM software (http://grimm.ucsd.edu) which uses Hannenhallís and Pevzner's (HP) algorithms (Hannenhalli and Pevzner 1995) for computing uni-chromosomal and multi-chromosomal genomic distances (Tesler 2002) and UniMoG software (Hilker et al. 2012) (combining five genome rearrangement models: double cut and join (DCJ) (Bergeron et al. 2006), restricted DCJ (Kovac et al. 2011), HP algorithm (Hannenhalli and Pevzner 1995) and inversion and translocation events). Three (restricted DCJ, HP and inversion) at UniMoG were analyzed. Fragile regions in the X chromosomes were identified using GRIMM and UdiMoG software according to inversion breakpoint reuse during evolution. The number of reuses in each breakpoint region was calculated by counting the number of breakpoints issued from each chromosomal inversion located in the same intergene region. The average number of breakpoints in all models was established by region. Fragile regions were characterized by having two or more reuses of breakpoints. RepeatMasker software (http://www.repeatmasker.org) was used for assessing the presence of potential transposable elements (TE) contiguously located near fragile regions, classified according to the RepBase database (Jurka et al. 2005). Geneious software (http://www.geneious.com) was used to analyze the nucleotide sequences located in fragile regions (Drummond 2010). The IBM-SPSS Statistic 20.0 software package (SPSS Inc., Chicago, IL) was used for statistical analysis.

Acknowledgments PL and CFP are funded by the Universidad del Rosario. This study was supported by The Universidad del Rosario (Grant CS/Genetics 2013).

References

Alekseyev MA, Pevzner PA (2010) Comparative genomics reveals birth and death of fragile regions in mammalian evolution. Genome Biol 11:R117

Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen MA, Harlizius B, Lee KT, Milan D, Rogers J, Rothschild MF, Uenishi H, Wang J, Schook LB (2010) Pig genome sequence: analysis and publication strategy. BMC Genom 11:438

Arlt MF, Casper AM, Glover TW (2003) Common fragile sites. Cytogenetic and genome research 100:92–100

Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum Mol Genet 12:2201–2208

Attie O, Darling AE, Yancopoulos S (2011) The rise and fall of breakpoint reuse depending on genome resolution. BMC Bioinformatics 12(Suppl 9):S1

Bafna V, Pevzner P (1995) Sorting permutations by tanspositions. In: Proceedings of the sixth annual ACM-SIAM symposium on Discrete algorithms: Society for Industrial and Applied Mathematics, pp 614-623

Bailey JA, Eichler EE (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet 7:552–564

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE (2004) Hotspots of mammalian chromosomal evolution. Genome Biol 5:R23

Bergeron A, Mixtacki J, Stoye J (2006) A unifying view of genome rearrangements. Algorithms in bioinformatics. Springer, Heidelberg, pp 163–173

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
Brouwer JR, Willemsen R, Oostra BA (2009) The FMR1 gene and fragile X-associated tremor/ataxia syndrome. Am J Med Genet B 150b:782–798
Caballero OL, Chen YT (2009) Cancer/testis (CT) antigens: potential targets for immunotherapy. Cancer Sci 100:2014–2021
Caceres M, Sullivan RT, Thomas JW (2007) A recurrent inversion on the eutherian X chromosome. Proc Natl Acad Sci USA 104:18571–18576
Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Bisceglio GD, Ertekin-Taner N, Crook JE, Dickson DW, Petersen RC, Graff-Radford NR, Younkin SG (2009) Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. Nat Genet 41:192–198
Cheng MK, Disteche CM (2006) A balancing act between the X chromosome and the autosomes. J Biol 5:2
Chimpanzee S, Analysis C (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87
Cocquet J, Ellis PJ, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS (2012) A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet 8:e1002900
Copeland NG, Jenkins NA, Gilbert DJ, Eppig JT, Maltais LJ, Miller JC, Dietrich WF, Weaver A, Lincoln SE, Steen RG et al (1993) A genetic linkage map of the mouse: current applications and future prospects. Science 262:57–66
Costa A, Gao L, Carrillo F, Caceres-Redondo MT, Carballo M, Diaz-Martin J, Gomez-Garre P, Sobrino F, Lucas M, Lopez-Barneo J, Mir P, Pintado E (2011) Intermediate alleles at the FRAXA and FRAXE loci in Parkinson's disease. Parkinsonism Relat Disord 17:281–284
Coulibaly MB, Lobo NF, Fitzpatrick MC, Kern M, Grushko O, Thaner DV, Traore SF, Collins FH, Besansky NJ (2007) Segmental duplication implicated in the genesis of inversion 2Rj of Anopheles gambiae. PLoS ONE 2:e849
Darai-Ramqvist E, Sandlund A, Muller S, Klein G, Imreh S, Kost-Alimova M (2008) Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. Genome Res 18:370–379
Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403
Darling AE, Miklos I, Ragan MA (2008) Dynamics of genome rearrangement in bacterial populations. PLoS Genet 4:e1000128
Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147
Davies JP, Cotter PD, Ioannou YA (1997) Cloning and mapping of human Rab7 and Rab9 cDNA sequences and identification of a Rab9 pseudogene. Genomics 41:131–134
Debacker K, Kooy RF (2007) Fragile sites and human disease. Hum Mol Genet 16(2):R150–158
DeBaryshe PG, Pardue ML (2011) Differential maintenance of DNA sequences in telomeric and centromeric heterochromatin. Genetics 187:51–60
DeBry RW, Seldin MF (1996) Human/mouse homology relationships. Genomics 33:337–351
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
Di Pasquale E, Beck-Peccoz P, Persani L (2004) Hypergonadotropic ovarian failure associated with an inherited mutation of human bone morphogenetic protein-15 (BMP15) gene. Am J Hum Genet 75:106–111
Drummond AJea (2010) Geneious v5.5
Eichler EE (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17:661–669
Elder FF, Robinson TJ (1989) Rodent common fragile sites: are they conserved? Evidence from mouse and rat. Chromosoma 97:459–464
Emerson JJ, Kaessmann H, Betran E, Long M (2004) Extensive gene traffic on the mammalian X chromosome. Science 303:537–540
Frischer LE, Hagen FS, Garber RL (1986) An inversion that disrupts the Antennapedia gene causes abnormal structure and localization of RNAs. Cell 47:1017–1023
Fundia A, Gorostiaga M, Mudry M (2000) Expression of common fragile sites in two Ceboidea species: *Saimiri boliviensis* and *Alouatta caraya* (Primates: Platyrrhini). Genet Sel Evol 32:87–97
Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I (2011) Birth and death of genes linked to chromosomal inversion. Proc Natl Acad Sci USA 108:1501–1506
Gazave E, Darre F, Morcillo-Suarez C, Petit-Marty N, Carreno A, Marigorta UM, Ryder OA, Blancher A, Rocchi M, Bosch E, Baker C, Marques-Bonet T, Eichler EE, Navarro A (2011) Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. Genome Res 21:1626–1639
George JA, DeBaryshe PG, Traverse KL, Celniker SE, Pardue ML (2006) Genomic organization of the Drosophila telomere retrotransposable elements. Genome Res 16:1231–1240
Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, Lopez-Otin C, Puente XS,

Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

Glover TW, Hoge AW, Miller DE, Ascara-Wilke JE, Adam AN, Dagenais SL, Wilke CM, Dierick HA, Beer DG (1998) The murine Fhit gene is highly similar to its human orthologue and maps to a common fragile site region. Cancer Res 58:3409–3414

Gordon L, Yang S, Tran-Gyamfi M, Baggott D, Christensen M, Hamilton A, Crooijmans R, Groenen M, Lucas S, Ovcharenko I, Stubbs L (2007) Comparative analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of gene-rich vertebrate regions. Genome Res 17:1603–1613

Graves JA, Koina E, Sankovic N (2006) How the gene content of human sex chromosomes evolved. Curr Opin Genet Dev 16:219–224

Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, Burridge PW, Cox TV, Fox CA, Hutton RD, Mullenger IR, Phillips KJ, Smith J, Stalker J, Threadgold GJ, Birney E, Wylie K, Chinwalla A, Wallis J, Hillier L, Carter J, Gaige T, Jaeger S, Kremitzki C, Layman D, Maas J, McGrane R, Mead K, Walker R, Jones S, Smith M, Asano J, Bosdet I, Chan S, Chittaranjan S, Chiu R, Fjell C, Fuhrmann D, Girn N, Gray C, Guin R, Hsiao L, Krzywinski M, Kutsche R, Lee SS, Mathewson C, McLeavy C, Messervier S, Ness S, Pandoh P, Prabhu AL, Saeedi P, Smailus D, Spence L, Stott J, Taylor S, Terpstra W, Tsai M, Vardy J, Wye N, Yang G, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Shvartsbeyn A, Gebregeorgis E, Krol M, Russell D, Overton L, Malek JA, Holmes M, Heaney M, Shetty J, Feldblyum T, Nierman WC, Catanese JJ, Hubbard T, Waterston RH, Rogers J, de Jong PJ, Fraser CM, Marra M, McPherson JD, Bentley DR (2002) A physical map of the mouse genome. Nature 418:743–750

Hannenhalli S, Pevzner PA (1995) Transforming men into mice (polynomial algorithm for genomic distance problem). Proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science, 1995, pp. 581–592

Helmrich A, Stout-Weider K, Hermann K, Schrock E, Heiden T (2006) Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. Genome Res 16:1222–1230

Hilker R, Sickinger C, Pedersen CN, Stoye J (2012) UniMoG: a unifying framework for genomic distance calculation and sorting based on DCJ. Bioinformatics (Oxford, England) 28:2509–2511

Hofmann O, Caballero OL, Stevenson BJ, Chen YT, Cohen T, Chua R, Maher CA, Panji S, Schaefer U, Kruger A, Lehvaslaiho M, Carninci P, Hayashizaki Y, Jongeneel CV, Simpson AJ, Old LJ, Hide W (2008) Genome-wide analysis of cancer/testis gene expression. Proc Natl Acad Sci USA 105:20422–20427

Horvath JE, Schwartz S, Eichler EE (2000) The mosaic structure of human pericentromeric DNA: a strategy for characterizing complex regions of the human genome. Genome Res 10:839–852

Hough RB, Lengeling A, Bedian V, Lo C, Bucan M (1998) Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-like protein 6 and causes dysregulation of Kit expression. Proc Natl Acad Sci USA 95:13800–13805

Humphray SJ, Scott CE, Clark R, Marron B, Bender C, Camm N, Davis J, Jenks A, Noon A, Patel M, Sehra H, Yang F, Rogatcheva MB, Milan D, Chardon P, Rohrer G, Nonneman D, de Jong P, Meyers SN, Archibald A, Beever JE, Schook LB, Rogers J (2007) A high utility integrated map of the pig genome. Genome Biol 8:R139

Mitelman F, Johansson B, Mertens F (2013) Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467

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

Kleinjan DA, Lettice LA (2008) Long-range gene control and genetic disease. Adv Genet 61:339–388

Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32

Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME et al (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74:127–134

Kolb J, Chuzhanova NA, Hogel J, Vasquez KM, Cooper DN, Bacolla A, Kehrer-Sawatzki H (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. Chromosom Res 17:469–483

Kovac J, Warren R, Braga MD, Stoye J (2011) Restricted DCJ model: rearrangement problems with chromosome reincorporation. J Comput Biol 18:1231–1241

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711–1714

Krummel KA, Denison SR, Calhoun E, Phillips LA, Smith DI (2002) The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. Genes Chromosom Cancer 34:154–167

Kupiec M, Petes TD (1988) Allelic and ectopic recombination between Ty elements in yeast. Genetics 119:549–559

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964–967

Laissue P, Vinci G, Veitia RA, Fellous M (2008) Recent advances in the study of genes involved in non-syndromic premature ovarian failure. Mol Cell Endocrinol 282:101–111

Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet 5:236–241

Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fievet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossu P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanche H, Dartigues JF, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094–1099

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D,

Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, International Human Genome Sequencing C (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

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

Lim JK, Simmons MJ (1994) Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. BioEssays 16:269–275

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alfoldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Cree A, Dihn HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M (2011) A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478:476–482

Lister C, Jackson D, Martin C (1993) Transposon-induced inversion in Antirrhinum modifies nivea gene expression to give a novel flower color pattern under the control of cycloidearadialis. Plant Cell 5:1541–1553

Lukusa T, Fryns JP (2008) Human chromosome fragility. Biochim Biophys Acta 1779:3–16

Ma J, Zhang L, Suh BB, Raney BJ, Burhans RC, Kent WJ, Blanchette M, Haussler D, Miller W (2006) Reconstructing contiguous regions of an ancestral genome. Genome Res 16:1557–1565

Ma K, Qiu L, Mrasek K, Zhang J, Liehr T, Quintana LG, Li Z (2012) Common fragile sites: genomic hotspots of DNA damage and carcinogenesis. Int J Mol Sci 13:11974–11999

Marques-Bonet T, Girirajan S, Eichler EE (2009) The origins and impact of primate segmental duplications. Trends Genet 25:443–454

Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien SJ, Pevzner PA, Lewin HA (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613–617

Nadeau JH, Taylor BA (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci USA 81:814–818

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

Parrish JE, Oostra BA, Verkerk AJ, Richards CS, Reynolds J, Spikes AS, Shaffer LG, Nelson DL (1994) Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. Nat Genet 8:229–235

Pevzner P, Tesler G (2003a) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res 13:37–45

Pevzner P, Tesler G (2003b) Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc Natl Acad Sci USA 100:7672–7677

Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J, Ashburner M, Bergman CM (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol 5:e152

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang Y, Waldron L, Verduzco D, Clerc-Blankenburg KP, Dubchak I, Noor MA, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. Genome Res 15:1–18

Ritchie RJ, Knight SJ, Hirst MC, Grewal PK, Bobrow M, Cross GS, Davies KE (1994) The cloning of FRAXF: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. Hum Mol Genet 3:2115–2121

Rodriguez V, Llambi S, Postiglioni A, Guevara K, Rincon G, Fernandez G, Mernies B, Arruga MV (2002) Localisation of aphidicolin-induced break points in Holstein-Friesian cattle (Bos taurus) using RBG-banding. Genet Sel Evol 34:649–656

Ronne M (1992) Putative fragile sites in the horse karyotype. Hereditas 117:127–136

Ronne M (1995a) Localization of fragile sites in the karyotype of Felis catus. Hereditas 122:279–283

Ronne M (1995b) Localization of fragile sites in the karyotype of Sus scrofa domestica: present status. Hereditas 122:153–162

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RI, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnin D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen Z, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Cobley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, Deshazo D, Dhami P, Ding Y, Dinh H, Dodsworth S, Draper H, Dugan-Rocha S, Dunham A, Dunn M, Durbin KJ, Dutta I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill R, Glockner G, Gregory SG, Gribble S, Griffiths C, Grocock R, Gu Y, Gwilliam R, Hamilton C, Hart EA, Hawes A, Heath PD, Heitmann K, Hennig S, Hernandez J, Hinzmann B, Ho S, Hoffs M, Howden PJ, Huckle EJ, Hume J, Hunt PJ, Hunt AR, Isherwood J, Jacob L, Johnson D, Jones S, de Jong PJ, Joseph SS, Keenan S, Kelly S, Kershaw JK, Khan Z, Kioschis P, Klages S, Knights AJ, Kosiura A, Kovar-Smith C, Laird GK, Langford C, Lawlor S, Leversha M, Lewis L, Liu W, Lloyd C, Lloyd DM, Loulseged H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma J, Maheshwari M, Matthews LH, McDowall J, McLaren S, McMurray A, Meidl P, Meitinger T, Milne S, Miner G, Mistry SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsiek G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Pelan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridler KA, Schlessinger D, Schueler MG, Sehra HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Sotheran EC, Steingruber HE, Steward CA, Storey R, Swann RM, Swarbreck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevanion S, Tromans AC, d'Urso M, Verduzco D, Villasana D, Waldron L, Wall M, Wang Q, Warren J, Warry GL, Wei X, West A, Whitehead SL, Whiteley MN, Wilkinson JE, Willey DL, Williams G, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang J, Zhou J, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poustka A, Rosenthal A, Lehrach H, Meindl A, Minx PJ, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM, Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulston JE, Durbin R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DR (2005) The DNA sequence of the human X chromosome. Nature 434:325–337

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

Ruiz-Herrera A, Ponsa M, Garcia F, Egozcue J, Garcia M (2002) Fragile sites in human and *Macaca fascicularis* chromosomes are breakpoints in chromosome evolution. Chromosom Res 10:33–44

Ruiz-Herrera A, Garcia F, Fronicke L, Ponsa M, Egozcue J, Caldes MG, Stanyon R (2004) Conservation of aphidicolin-induced fragile sites in Papionini (Primates) species and humans. Chromosom Res 12:683–690

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

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

Samonte RV, Eichler EE (2002) Segmental duplications and the evolution of the primate genome. Nat Rev Genet 3:65–72

Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT (2002) Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. Immunol Rev 188:22–32

Schibler L, Roig A, Mahe MF, Laurent P, Hayes H, Rodolphe F, Cribiu EP (2006) High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. BMC Genom 7:194

Schwartz M, Zlotorynski E, Kerem B (2006) The molecular basis of common and rare fragile sites. Cancer Lett 232:13–26

Seki N, Azuma T, Yoshikawa T, Masuho Y, Muramatsu M, Saito T (2000) cDNA cloning of a new member of the Ras superfamily,

RAB9-like, on the human chromosome Xq22.1-q22.3 region. J Hum Genet 45:318–322

Shimasaki S, Moore RK, Otsuka F, Erickson GF (2004) The bone morphogenetic protein system in mammalian reproduction. Endocr Rev 25:72–101

Shiraishi T, Druck T, Mimori K, Flomenberg J, Berk L, Alder H, Miller W, Huebner K, Croce CM (2001) Sequence conservation at human and mouse orthologous common fragile regions, FRA3B/FHIT and Fra14A2/Fhit. Proc Natl Acad Sci USA 98:5722–5727

Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ (2005) Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer 5:615–625

Sonoda E, Hochegger H, Saberi A, Taniguchi Y, Takeda S (2006) Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. DNA Repair 5:1021–1029

Sperlich D, Pfriem P (1986) Chromosomal polymorphism in natural and experimental populations. Genet Biol Drosoph 3:257–309

Sutherland GR (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. Science 197:265–266

Tesler G (2002) GRIMM: genome rearrangements web server. Bioinformatics (Oxford, England) 18:492–493

Tsantoulis PK, Kotsinas A, Sfikakis PP, Evangelou K, Sideridou M, Levy B, Mo L, Kittas C, Wu XR, Papavassiliou AG, Gorgoulis VG (2008) Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. Oncogene 27:3256–3264

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291:1304–1351

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914

Vibranovski MD, Lopes HF, Karr TL, Long M (2009a) Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet 5:e1000731

Vibranovski MD, Zhang Y, Long M (2009b) General gene movement off the X chromosome in the *Drosophila* genus. Genome Res 19:897–903

Wasserman M (1982) Evolution of the repleta group. Genet Biol Drosoph 3:61–139

Wilson MA, Makova KD (2009) Genomic analyses of sex chromosome evolution. Annu Rev Genomics Hum Genet 10:333–354

Yang MY, Long SE (1993) Folate sensitive common fragile sites in chromosomes of the domestic pig (Sus scrofa). Res Vet Sci 55:231–235

Zhang YE, Vibranovski MD, Landback P, Marais GA, Long M (2010) Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. PLoS Biol 8:e1000494

Zhao S, Shetty J, Hou L, Delcher A, Zhu B, Osoegawa K, de Jong P, Nierman WC, Strausberg RL, Fraser CM (2004) Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14:1851–1860