Research article

High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution

Laurent Schibler*[1], Anne Roig[1], Marie-Françoise Mahe[1], Pascal Laurent[1], Hélène Hayes[1], François Rodolphe[2] and Edmond P Cribiu[1]

Address: [1]Laboratoire de Génétique biochimique et de Cytogénétique, Département de Génétique Animale, Institut National de la Recherche Agronomique (INRA), Centre de Recherche de Jouy, 78352 Jouy-en-Josas, Cedex, France and [2]Mathématique, informatique et génome, Institut National de la Recherche Agronomique (INRA), Centre de Recherche de Jouy, 78352 Jouy-en-Josas Cedex, France

Email: Laurent Schibler* - Laurent.Schibler@jouy.inra.fr; Anne Roig - anne.roig@jouy.inra.fr; Marie-Françoise Mahe - marie-francoise.mahe@jouy.inra.fr; Pascal Laurent - pascal.laurent@jouy.inra.fr; Hélène Hayes - helene.hayes@jouy.inra.fr; François Rodolphe - Francois.Rodolphe@jouy.inra.fr; Edmond P Cribiu - Edmond-Paul.Cribiu@jouy.inra.fr

* Corresponding author

Abstract

Background: Comparative mapping provides new insights into the evolutionary history of genomes. In particular, recent studies in mammals have suggested a role for segmental duplication in genome evolution. In some species such as Drosophila or maize, transposable elements (TEs) have been shown to be involved in chromosomal rearrangements. In this work, we have explored the presence of interspersed repeats in regions of chromosomal rearrangements, using an updated high-resolution integrated comparative map among cattle, man and mouse.

Results: The bovine, human and mouse comparative autosomal map has been constructed using data from bovine genetic and physical maps and from FISH-mapping studies. We confirm most previous results but also reveal some discrepancies. A total of 211 conserved segments have been identified between cattle and man, of which 33 are new segments and 72 correspond to extended, previously known segments. The resulting map covers 91% and 90% of the human and bovine genomes, respectively. Analysis of breakpoint regions revealed a high density of species-specific interspersed repeats in the human and mouse genomes.

Conclusion: Analysis of the breakpoint regions has revealed specific repeat density patterns, suggesting that TEs may have played a significant role in chromosome evolution and genome plasticity. However, we cannot rule out that repeats and breakpoints accumulate independently in the few same regions where modifications are better tolerated. Likewise, we cannot ascertain whether increased TE density is the cause or the consequence of chromosome rearrangements. Nevertheless, the identification of high density repeat clusters combined with a well-documented repeat phylogeny should highlight probable breakpoints, and permit their precise dating. Combining new statistical models taking the present information into account should help reconstruct ancestral karyotypes.

Background

Comparative mapping represents a major approach in providing new insights into dynamics of genome evolution. Since the pioneer studies that showed linkage conservation among genomes [1,2], genome comparisons have been carried out in about 30 mammalian species [3]. Although Fluorescent In Situ Hybridization (FISH) and painting analyses have contributed significantly to the identification of conserved syntenies among species [4], the number of ordered gene maps is still too small to draw fully meaningful inferences on chromosomal evolution. In cattle and goats, we had previously reported [5] a high level of intra-chromosomal rearrangements and the existence of preferential breakpoints over the whole genome, which have been confirmed by Radiation Hybrid (RH) mapping data [6-19]. Recently, two high-resolution human-bovine comparative maps have been reported, based on data from the human sequence and bovine RH panels [6,7]. These studies have improved the genome-wide comparative coverage by ~20% between man and cattle and identified 195 and 161 segments with conserved gene order, respectively. The availability of whole genome sequences for man, mouse and rat and in draft format for eight other mammalian species has made it possible to evaluate the rates of chromosomal evolution and to detect segmental duplication in most of the primate-specific breakpoint regions [8]. In addition, Everts-van der Wind et al [9] have analyzed the evolution of centromere and telomere positions and the gene content within evolutionary breakpoint regions in cattle versus man.

Here, we have explored the relationships between chromosomal rearrangements and the density of interspersed repeats, since these represent a prevalent feature of mammalian genomes. Indeed, all mammalian genomes present essentially the same four classes of transposable elements (TEs): autonomous long interspersed nucleotide elements (LINEs), LINE-dependent RNA-derived short interspersed nucleotide elements (SINEs), retrovirus-like elements with long terminal repeats (LTRs such as endogenous retroviruses ERVs and MaLRs) and DNA transposons (see [10-13] for review). The age and history of these repeats have been inferred from phylogenic analyses (for review see [14,15]), suggesting that most mammalian TEs are related and thus can be divided into lineage-specific repeats (inserted after the divergence of the studied species) and ancestral repeats (already present in a common ancestor). Moreover, TEs can undergo broad bursts of amplification in a lineage specific way, potentially leading to speciation as suggested in Primates [16].

In order to carry out this study, first we have produced an updated version of the bovine physical map constructed in our laboratory [17], and have used it to extend and refine the bovine comparative map with man and mouse. This has led to a high resolution comparative map integrated with the most recent bovine genetic map [18] and FISH data [19]. Second, based on this comparative map, we have identified and examined the intervals between segments of conserved gene order (evolutionary breakpoint regions) among the cattle, human and mouse genomes. We have found that these breakpoint regions are enriched in lineage specific repeats: Alu repeats in man and SINEs in mouse. Since previous studies have suggested that TEs could be involved in chromosomal rearrangements, we propose that the increased density of TEs and especially SINEs in evolutionary breakpoint regions has a role in mammalian chromosome evolution and we discuss the recent findings of Murphy *et al* [8] on the implication of segmental duplication.

Results

Updated bovine physical map

Our first autosomal bovine physical map reported in 2004 [17] comprised 6615 contigs of which 747 were anchored to the cattle genome. In the present work, we have anchored 350 additional contigs by PCR-screening of 400 new well-distributed microsatellites. Moreover, we have produced 53,550 BAC end sequences (BES) from 26,935 INRA BAC clones in collaboration with the Genoscope and collected 28,468 BES from the CHORI-240 library via Genbank. BLAST analyses of these BES produced 38,469 valid hits, of which 27,676 correspond to known or predicted human genes and 10,793 are located outside genes. BLAST statistics are similar in all cases, i.e. 88% mean identity percentage, $1.6 \ 10^{-07}$ mean e-value and 190 bp mean homology size. As a result, 9,323 INRA and 4,318 CHORI-240 bovine clones have been anchored to the human sequence.

To date, the updated version of our physical map comprises 5,081 contigs ranging from 150 kb to 5.5 Mb in size (600 kb on average). Among these, 860 are anchored to the bovine genome by 1,763 markers, 3,444 to the human sequence and 774 to the genome of both species. These 774 contigs comprise 1121 bovine markers and 2,127 hits on the human genome. Data are available through the cattle webFPC [20] or the BovMap database [21].

An integrated high-resolution comparative autosomal map for cattle

To construct this comparative autosomal map, first we used the 774 contigs anchored to both the bovine and human genomes. Then, we added 598 FISH localizations (131 microsatellites and 467 genes from [19], 213 of which are also mapped to the contigs). Since contigs contain several BES, they are anchored to human genome with 4290 positive BLASTs. Although the Btau 2.0 whole genome assembly was released in the course of this work,

it was not used to draw meaningful conclusions for the following reasons: (1) blasting large marker sets, which had not been used to build Btau 2.0 assembly, showed that it contained partial and sometimes erroneous data (markers detected in more than one scaffold or at several positions in the same scaffold and markers from different bovine chromosomes present in the same scaffolds) and (2) since the Btau 2.0 scaffolds were built based mainly on syntenic relations with the human genome, there may be a bias towards a reduced number of rearrangements between these two species. Thus, the current bovine assembly should be exploited with care and we only used it to anchor 1519 additional microsatellites to the human sequence.

As a result, our comparative map comprises 2721 loci anchored to both the bovine and human genomes [see additional file 1 and additional file 2].

According to the bovine genetic map, our physical map covers 90% of the bovine autosomes and only BTA21 has a coverage below 80%. Bovine orthologous regions have been identified for about 91% of the human autosomes, leaving 166 human segments (~2 Mb on average) without comparative mapping information.

Our results reveal 68 conserved synteny segments between bovine and human autosomes, in good agreement with previous studies [6,19]. If gene order is taken into account, we identify 211 segments, of which 198 include at least two markers. Thus, our work has extended or refined 72 known segments with conserved gene order and identified 33 new segments. Only nine previously reported segments based on single FISH localizations have not been confirmed or remain inaccurate in the present study (gene names and cytogenetic locations indicated in red in additional file 1: PPCB, NAB1, BMPR1B, TXK, GM2A, GALT, PIGR, PIM1, ACPP).

Analysis of evolutionary breakpoints

Evolutionary breakpoint regions between two species are defined as the interval between two segments of conserved gene order. They are classified into four types, based on their likely origin (see Material and Methods and Figure 1). Repeat densities have been computed for each class of repeats and type of breakpoints and compared to the densities along the human or mouse chromosomes. Repeat densities were not computed along the bovine genome because of flaws in the bovine genome sequence. Indeed, bovine scaffolds are in general smaller than the corresponding BAC contig lengths and contain many large gaps. Scaffolds contain only one third to one half of the sequence of the corresponding genomic region. This is consistent with assembly statistics describing that ~34% of the bovine genome are covered by anchored scaffolds.

Moreover, typical errors in draft genome sequences include misassemblies of repeated sequences, collapses of repeated regions, making such sequences difficult to use for repeat content analysis.

The one-sample Kolmogorov-Smirnov test, based on the repeat density, showed an increased density of several interspersed sequences in certain regions. P-values are given in Figure 2 and Figure 3 [see also Additional file 3].

We have detected a highly significant increased density of SINE Alu repeats in the human sequences corresponding to HSA and MMU breakpoints but not to the BTA breakpoints (see Figure 4 for an example). A Principal Components Analysis (Figure 5) suggests that most of the HSA breakpoints are enriched in various repeats: three HSA breakpoints contain only a high density of ancient Alu repeats, four only a high density of young Alu repeats and four display an elevated frequency in L1 repeats (L1MB3, L1MB7 and L1MC5).

The same analysis performed on MMU breakpoints shows a strong correlation between all repeat densities (data not shown). Analysis of the mouse sequences reveals an increased density of some rodent specific SINEs B1, B2, B4 and ID in MMU breakpoints, whereas the BTA and HSA breakpoints do not show any difference with the whole mouse genome. Moreover, 23 MMU breakpoints are enriched only in ERV-L (MT2A, MT2B B or RMER10) or ERV-K RMER17C. Five of these MMU breakpoints also delineate breakpoint regions in rat, suggesting that they occurred in the *Mus* lineage.

Human sequences of 28 BTA breakpoints display an Alu repeat content typical of HSA breakpoints, with an increase mainly in the oldest Alu repeats (FRAM, FLAM, AluJ, AluSx). Likewise, a repeat content typical from MMU breakpoints is also observed in the orthologous mouse sequences.

Moreover, a search for centromeric or telomeric satellites in the human sequence has shown that 24 breakpoint regions contain at least two satellite repeats. About 71% of these breakpoints involve two different chromosomes and may thus be considered as interchromosomal breakpoint regions, whereas 29% may be intrachromosomal breakpoint regions [see additional file 3]. Although not statistically significant, satellites may be slightly over-represented in interchromosomal breakpoints. About 11% and 4% of interchromosomal and intrachromosomal breakpoint regions contain satellite repeats, respectively.
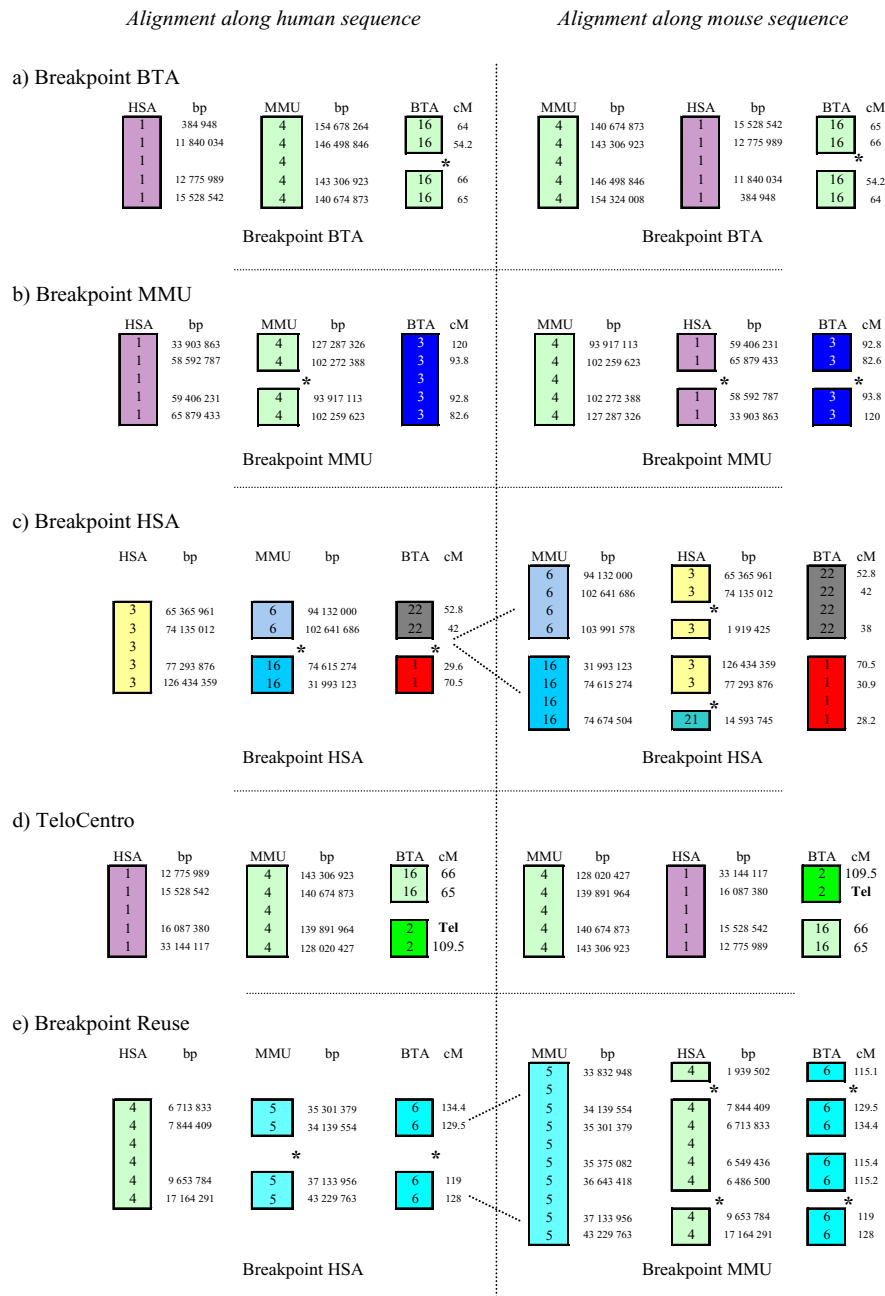
Figure 1
Principle of the classification of breakpoints. Classification of breakpoints is exemplified by real data. Chromosomes for each species are drawn as colored boxes, with chromosome numbers indicated in the boxes. Positions along the genome (base pairs or cM) are written on the right. Breaks in boxes and * indicate evolutionary breakpoints. Genomes are drawn aligned along the human and mouse sequences. a) The breakpoint occurs solely in the bovine genome, in both genome alignments. This situation is classified as a BTA breakpoint. b) The breakpoint involves two mouse chromosome segments based on the human alignment. When aligned on the mouse genome, these segments define breakpoints in both human and bovine genomes. This situation is classified as a MMU breakpoint. c) The breakpoint involves both the bovine and the mouse genomes based on the human alignment. When aligned along the mouse genome, these regions define breakpoints only in the human genome. This situation is classified as a HSA breakpoint. d) The breakpoint involves either telomeres or centromeres. This situation is classified as TeloCentro. e) The breakpoint is classified in a different way according to the human and mouse alignments. This situation is classified as a breakpoint Reuse.

		Breakpoint class			
		BTA	**HSA**	**MMU**	**Atypical BTA**
SINEs	Alu AluJb	1.6E-02	4.6E-07	1.3E-05	7.4E-11
	Alu AluJo	1.2E-03	2.2E-08	2.2E-05	9.7E-13
	Alu AluSc	2.6E-03	4.0E-05	1.3E-04	1.2E-08
	Alu AluSg	7.4E-03	1.8E-05	2.7E-05	3.6E-09
	Alu AluSg/x	1.7E-03	3.3E-06	7.0E-04	1.5E-11
	Alu AluSp	2.1E-01	2.2E-06	2.6E-05	1.9E-07
	Alu AluSq	8.2E-02	8.7E-06	1.6E-04	1.1E-11
	Alu AluSx	3.0E-02	3.7E-07	1.1E-05	2.7E-14
	Alu AluY	9.6E-02	1.2E-05	1.3E-04	5.0E-10
	Alu FLAM_A	2.3E-02	4.4E-04	3.2E-05	6.0E-08
	Alu FLAM_C	1.2E-02	1.5E-05	7.1E-07	1.9E-09
	Alu FRAM	1.7E-01	6.9E-05	2.9E-04	1.3E-08
LINEs	L1 L1MB3	9.6E-02	6.7E-05	5.4E-04	1.6E-03
	L1 L1MB7	1.3E-02	6.8E-05	7.3E-03	3.8E-03
	L1 L1MC5	3.0E-02	7.9E-05	3.2E-03	2.6E-03

Figure 2
Results of the one-sample Kolmogorov-Smirnov test in the human sequence. P-values for the "greater" hypothesis are shown for each class of breakpoints and for repeats with significant results. Highly significant p-values are indicated with a light grey background.

Discussion
Discrepancies between different bovine comparative maps
The autosomal comparative map reported here is in good agreement with previous studies based either on FISH or RH mapping of genes. Most discrepancies are due to differences between framework maps, especially in that reported in 2004 by Everts-van der Wind *et al*, which was built with a low resolution RH_{5000} panel and a reduced number of backbone microsatellites. Erroneous identification of orthologous sequences may also introduce some errors in all studies and contribute to the observed discrepancies. Moreover nine cattle FISH localizations have been found to be inconsistent, probably due to false orthologous gene pairs or mapping errors.

Breakpoint regions display a high density of lineage specific repeats
We have shown an association between the density of interspersed repeats and the evolutionary breakpoints in the human and mouse genomes. Interestingly, the repeat increase in these regions concerns only repeats having spread after the divergence of these species. Moreover, for a given species, only breakpoint regions having occurred within this species are enriched in repeats and more surprisingly, in species specific repeats, except for MMU breakpoints that appear to be Alu rich in the human sequence. This latter result agrees with the strong correlation observed between the local SINE density in mouse and the Alu density in orthologous loci in man [15]. Common biological factors governing the insertion and/or

retention of repeats in particular regions or the phylogeny of SINE repeats (Alu and B1 repeats derive from a common ancestor 7SL RNA) could explain these correlations.

Conversely, human sequences spanning regions of BTA breakpoints do not reveal any enrichment in Alu repeats, which suggests that they may be associated with repeats having different insertional characteristics. Bovine lineage specific SINEs or LINEs Bov-tA, Bov-A2 or Bov-B could be good candidates for such repeats and this hypothesis should be assessed after completion of the bovine whole genome sequence.

Moreover, 28 BTA breakpoints show a typical HSA and MMU repeat content in the corresponding human and mouse sequences. If we assume that Rodents diverged from a Primate-Artiodactyl clade, about 50% of the human rearrangements should have also occurred independently in mice, suggesting a very high frequency of convergent evolution. If one rejects this hypothesis, then Artiodactyls may be more distant from human than are rodents. Thus, this analysis of the repeat content in evolutionary breakpoint regions supports the recent hypothesis that Artiodactyls diverged from a Primate-Rodent clade [22-24].

Since evolutionary breakpoints appear to be enriched in specific interspersed repeats, widespread activity of TEs could be involved in genome evolution. Indeed, they provide material for DNA mispairing and thus can lead to genomic instability and rearrangements. A key role of TEs in intrachromosomal rearrangements (large deletions, translocations, inversions) has been evidenced in *Drosophila* [25-31], Mosquito [32-34], Man [32-34] and Maize [35]. Likewise, in human tumors, the high density of repetitive DNA in a given region provides hot spots for homologous recombination and mediates translocation processes [36].

Moreover, Alu mispairing could have generated many segmental duplications [37,38] that are found in most primate-specific breakpoint regions, suggesting their role in chromosomal rearrangements [8,39-41].

However, it is possible that repeats and breakpoints accumulate independently but in the same region where modifications are better tolerated. High repeat densities could also be a consequence of breakpoints as suggested by Dobigny *et* al [42] i.e. major regulation pathway of repatterned genomes could be transiently inefficient, thus resulting in conditions suitable for TE amplification.

An overall model of genome evolution
Genomes may evolve cyclically under the influence of three main factors: transposon normal activity, bursts of

		Breakpoint class			
		BTA	**HSA**	**MMU**	**Atypical BTA**
SINEs	Alu B1_Mm	3.41E-01	5.63E-02	1.15E-06	3.41E-05
	Alu B1_Mur1	1.00E-02	1.76E-01	7.06E-08	6.60E-06
	Alu B1_Mur2	1.38E-01	2.26E-02	5.07E-06	1.24E-06
	Alu B1_Mur3	2.86E-01	7.47E-02	2.39E-08	3.38E-07
	Alu B1_Mur4	1.78E-01	1.91E-02	4.15E-07	1.24E-06
	Alu B1_Mus1	4.77E-01	7.81E-04	1.48E-09	1.71E-07
	Alu B1_Mus2	2.27E-02	3.33E-03	9.69E-09	1.73E-08
	Alu B1F	1.56E-02	1.28E-02	6.08E-04	9.92E-06
	Alu B1F1	7.27E-02	1.62E-02	3.43E-04	3.55E-05
	Alu B1F2	2.95E-02	1.70E-01	1.56E-03	1.71E-05
	Alu PB1	3.17E-02	2.56E-02	4.85E-06	5.60E-07
	Alu PB1D10	4.15E-02	6.35E-02	9.60E-06	2.34E-07
	Alu PB1D7	5.66E-04	2.86E-02	1.67E-03	8.59E-07
	Alu PB1D9	5.52E-02	1.46E-01	1.42E-07	2.69E-06
	B2 B2_Mm1a	5.97E-02	2.37E-02	1.53E-07	1.05E-06
	B2 B2_Mm1t	2.91E-01	9.82E-02	8.23E-08	3.26E-06
	B2 B2_Mm2	1.73E-01	7.17E-02	2.19E-08	1.05E-07
	B2 B3	9.04E-02	4.19E-02	1.20E-05	3.39E-07
	B4 ID_B1	1.78E-01	2.04E-02	2.30E-05	2.29E-05
	B4 RSINE1	3.18E-02	1.73E-02	9.91E-04	2.15E-06
	ID ID	4.26E-02	2.34E-02	1.70E-02	1.79E-05
	ID ID4	3.99E-02	1.11E-01	1.64E-04	4.43E-05
	ID ID4_	3.32E-02	2.36E-02	1.28E-03	2.82E-06
LTRs	ERVK RMER17C	2.41E-01	4.60E-01	7.47E-06	1.23E-01
	ERVL MT2B	3.91E-02	1.40E-01	2.29E-05	1.59E-03

Figure 3
Results of the one-sample Kolmogorov-Smirnov test in the mouse sequence. P-values for the "greater" hypothesis are shown for each class of breakpoints and for repeats with significant results. Highly significant p-values are indicated with a light grey background.
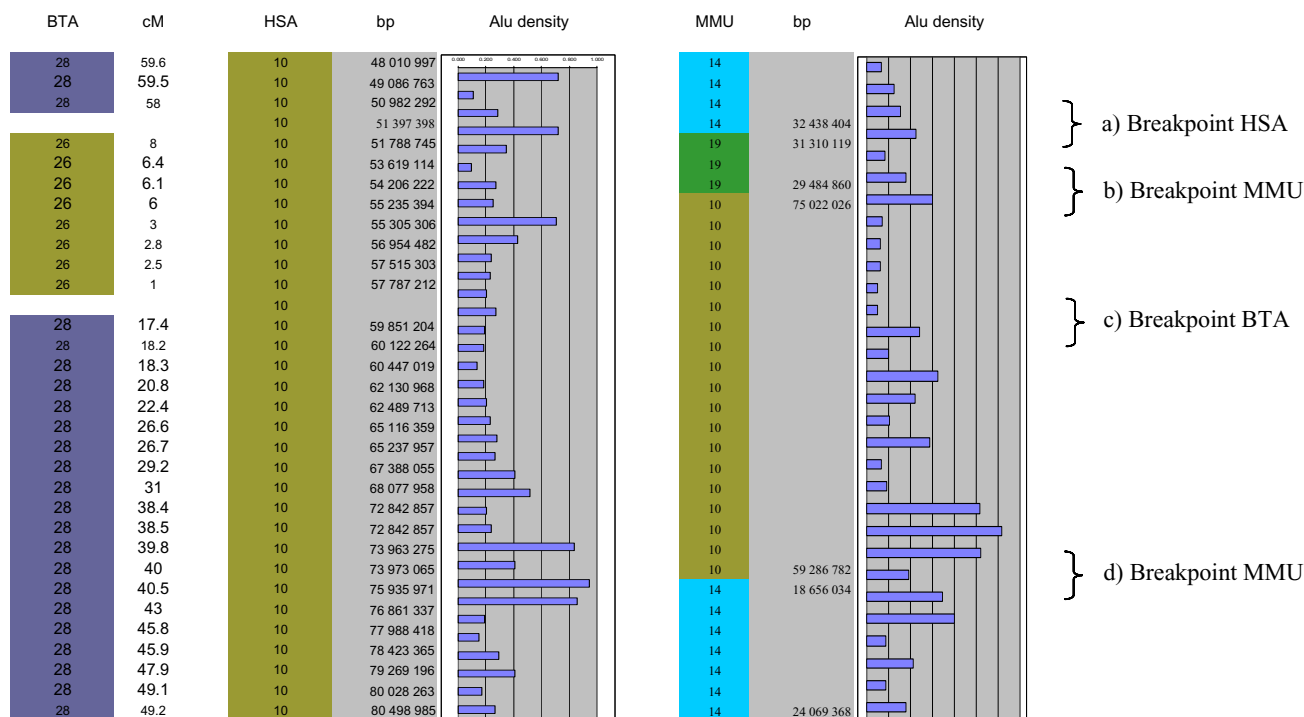
Figure 4
Examples of increased TE density. The figure presents the comparative bovine/human/mouse map for a segment of HSA10. a) Alu density is increased in the human genome in the vicinity of an HSA breakpoint. b) Alu density is increased in the mouse genome, in the vicinity of a MMU breakpoint. c) No repeat density increase is observed in the human or mouse genome in the vicinity of a BTA breakpoint region. d) Alu density is increased in both human and mouse genomes in a MMU breakpoint, probably due to the strong correlation observed between the local SINE density in mouse and the Alu density in orthologous loci in man.

transposon activity and dynamics of centromeres and telomeres.

(i) Normal transposon activity may be responsible for most small intrachromosomal rearrangements (small duplications, deletions or inversions... for review, see [43]).

(ii) Bursts of transposon activity result in a dramatic increase of the local repeat density and thus, may induce large intrachromosomal rearrangements directly or indirectly by segmental duplications. For example, Alu repeats and rodent SINEs could have given rise to most of the intrachromosomal rearrangements during a middle term period (80–50 Million years (Myr)), while during a more recent period (4–10 Myr), ERVs could have caused at least 18 rearrangements in the ancestor of the Muridae and five in the mouse.

(iii) Chromosome fusions or fissions may have produced most interchromosomal rearrangements, as suggested by the proportion of interchromosomal breakpoints containing centromeric and telomeric satellites (present study) and by early cytogenetic analyses [44]. Likewise, a recent study has revealed a clustering of centromere and telomere sequences at the site of evolutionary breakages [8]. Telomeres may have mediated or enhanced these rearrangements, as demonstrated for subtelomeres [45]. It has been shown that centromeres can be repositioned or created *de novo* and thus they could be involved in stabilizing interchromosomal rearrangements.

This model agrees both with the random breakage and the 'fragile breakage' models proposed by Nadeau [2] and Pevzner [46,47], respectively. On the one hand, rearrangements may occur in repeat-rich regions that are roughly randomly distributed along the genome, even if SINEs are preferentially located in GC-rich regions whereas LINEs are essentially present in AT-rich regions. On the other hand, the high density of repeats could promote both hot spots of chromosomal evolution [48] and segmental duplications, in accordance with the fragile breakpoint model. Moreover, the mammalian genome tolerates insertion and accumulation of repeats only in a limited
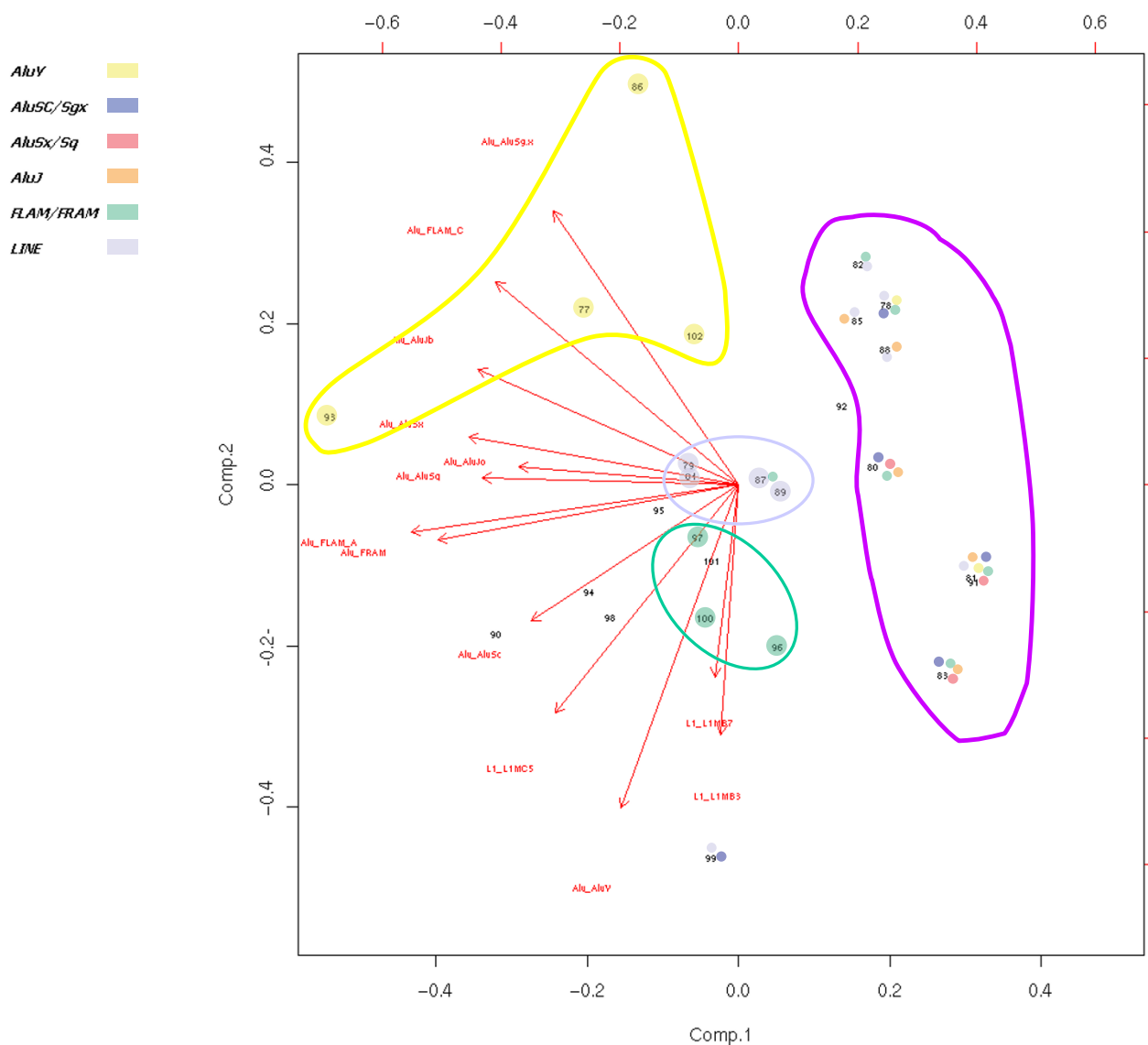
Figure 5
Principal Components Analysis of human breakpoint repeat content. Breakpoint region numbers are plotted against the two major component axes, arrows indicating the contribution of each repeat. High enrichment in repeats is indicated for each breakpoint by colored spots.

number of regions as indicated by the clustering of some repeats in orthologous regions from different species. Repeat families may have sequentially gathered in the same regions, potentially leading to successive rearrangements. Such regions may thus appear as 'fragile' and could explain that nearly 20% of the breakpoints are reused during mammalian evolution [8]. In addition, some regions could be involved in evolutionary breakpoints in one species and could lead to disease in another, as suggested by the close relationship between evolutionary breakpoints and neoplasia-associated human chromosome sites [5].

Conclusion
In this paper, we report a high-resolution integrated bovine comparative map with man and mouse based on microsatellites, INRA bovine contigs, BAC End sequencing and previous FISH data. This comparative physical map should be useful for the assembly of the bovine whole genome sequence.

Moreover, it has allowed us to analyze the DNA content in breakpoint regions to reveal specific repeat density patterns. Thus, 'junk DNA' may have played a significant role

in chromosome evolution and genome plasticity. Periodic phases of intrachromosomal rearrangements (due to interspersed repeats), followed by interchromosomal rearrangements (providing new opportunities for later intrachromosomal recombinations) could be the driving force of genome evolution. Interestingly, amplification of TEs has been described as a response to stress factors either of environmental origin or genomic nature (see [49] for review). In our model, TE mediated chromosome evolution, leading to new gene regulations and expressions, could thus represent an inventive mechanism to speed up fitness to new environmental conditions, leading to speciation.

However, more studies are required to ascertain whether increased TE density is the cause or the consequence of chromosome rearrangements. Moreover, we cannot rule out that repeats and breakpoints accumulate independently in the few same regions where modifications are better tolerated. Nevertheless, high density repeat clusters combined with a well-documented repeat phylogeny should highlight probable breakpoints, and permit their precise dating. New statistical models of evolution should be developed, taking into account not only the parsimony principle but also the present information in order to help reconstruct ancestral karyotypes.

Methods
Anchoring the physical map
The INRA bovine BAC library was screened by PCR using published microsatellite primer pairs (see BovMap database [21]) as previously described [17].

In collaboration with the Genoscope, we produced 53,550 BES from 26,935 INRA BAC clones and we collected another 28,468 BES from the CHORI-240 library via Genbank. Similarities with the human sequence (NCBI built 35) were searched for by BLASTN after repeat-masking ([50]) We used an expectation value E of 10^{-5} as the significance threshold for comparisons, since this value was shown to provide 95% accuracy in identifying orthologs [51]. Coordinates in the human sequence were then obtained for each retained hit using GoldenPath [52]. After localizing microsatellites in the bovine scaffolds by BLASTN analysis, the position of the bovine scaffolds on the human sequence were determined using BLASTN with parameters set to optimize detection of distant homologies (Ensembl BLASTView: W9, M1, N-1, Q2, R1).

Segment identification and integration of FISH and bibliographic data
After aligning the contigs along the bovine genome using the Ihara and USDA97 linkage maps [18,53], microsatel-

lite BLAST results and FISH-mapping data were integrated into this framework map.

The boundaries of conserved segments between the bovine and human genomes were defined according to a parsimonious interpretation. Discrepancies in microsatellite and FISH-mapped gene locations between backbone maps were resolved by choosing the location that minimized rearrangements between the bovine and the human genomes.

Searching for particular sequence features in breakpoint regions
Evolutionary breakpoint regions between two species were defined as the interval between two segments of conserved gene order. As illustrated in the left part of Figure 1, breakpoint regions were classified into four categories, based on the human alignment:

- BTA breakpoints, having probably taken place in the bovine lineage, when a perfect co-linearity is observed between human and mouse maps, but not with the bovine map. (Figure 1a: the breakpoint region is identified only in cattle).

- MMU breakpoints, having probably taken place in the mouse lineage, when a perfect co-linearity is observed between human and bovine maps, but not with the mouse map (Figure 1b: the breakpoint region is detected only in mouse).

- HSA breakpoints, having probably taken place in the human lineage, when no co-linearity is observed among human, mouse or bovine maps. (Figure 1c: the breakpoint region is identified both in cattle and mouse).

- TeloCentro, when either telomeres or centromeres are involved in the breakpoint (Figure 1d).

The same classification was also performed based on the mouse alignment (right part of Figure 1). When identical breakpoint regions fell into different categories according to the human or mouse alignments (Figure 1e), they were discarded from the global analysis because they could represent regions of breakpoint reuse. Breakpoints involving centromeres or telomeres were likewise discarded because these regions may be poorly represented in the genome assemblies and rearrangements may have been mediated by centromeres or telomeres, independently of the repeats we wanted to analyze.

Human and mouse chromosome masked sequences (chromFaMasked.zip) and repeat annotations (chromOut.zip) were downloaded from UCSC [54].

The density (number of repeat per Mb of DNA) of each class of repeats was computed for each kind of breakpoint region. Likewise, densities were also computed along each chromosome in sliding windows of different sizes (0.5, 1, 2, 3, 4 and 5 Mb). The proportion of densities, in these sliding windows, greater than the density in each repeat region was then computed for each class of repeats. These calculations were performed using sliding widows of a size range similar to the breakpoint region size.

Enrichment was tested by applying the one-sample Kolmogorov-Smirnov test to these proportions (distribution under the null-hypothesis is uniform on [0,1]) and using the R software.

Classification results, positions on human and mouse sequences, repeat densities and proportions of frequencies are available as supplementary data (break_supp).

Authors' contributions
L. Schibler conceived the study, participated in its design and coordination, built the bovine physical map, carried out bioinformatic and statistical analyses and drafted the manuscript. A. Roig, MF. Mahe and P. Laurent performed the BAC library screening. F. Rodolphe helped to design the statistical analysis and to draft the manuscript. H. Hayes kindly provided FISH data and participated in the revision of the manuscript. EP. Cribiu contributed to the preparation and revision of the manuscript and gave final approval of the version to be submitted.

Additional material

Additional File 1
comparative_supp. This excel file contains the comparative maps aligned along each genome in three different pages and gives details on their genome coverage.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-194-S1.xls]

Additional File 2
Cattlemap_supp. This file contains a graphical version of the bovine map.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-194-S2.doc]

Additional File 3
breakpoint_supp. break_supp.xls details breakpoint analysis results, including breakpoint regions description, proportions of densities observed above the density of each class of repeats....
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-194-S3.xls]

References
1. Ohno S: **Ancient linkage groups and frozen accidents.** *Nature* 1973, **244:**259-262.
2. Nadeau JH, Taylor BA: **Lengths of chromosomal segments conserved since divergence of man and mouse.** *Proceedings of the National Academy of United States of America* 1984, **81:**814-818.
3. O'Brien SJ, Marshall Graves JA: **Report of the committee on comparative gene mapping: London.** *Volume 58*. Edited by: Klinger P. S.Karger; 1991:1151.
4. Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H: **Emerging patterns of comparative genome organization in some mammalian species as revealed by zoo-FISH.** *Genome Res* 1998, **8(6):**577-589.
5. Schibler L, Vaiman D, Oustry A, Giraud-Delville C, Cribiu EP: **Comparative gene mapping: A fine-scale survey of chromosome rearrangements between ruminants and humans.** *Genome Res* 1998, **8(9):**901-915.
6. Everts-van der Wind A, Kata SR, Band MR, Rebeiz M, Larkin DM, Everts RE, Green CA, Liu L, Natarajan S, Goldammer T, Lee JH, McKay S, Womack JE, Lewin HA: **A 1463 Gene Cattle-Human Comparative Map With Anchor Points Defined by Human Genome Sequence Coordinates.** *Genome Res* 2004, **14(7):**1424-1437.
7. Itoh T, Watanabe T, Ihara N, Mariani P, Beattie CW, Sugimoto Y, Takasuga A: **A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci.** *Genomics* 2005, **85(4):**413-424.
8. 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 S J, Pevzner PA, Lewin HA: **Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps.** *Science* 2005, **309(5734):**613-617.
9. Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA: **A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution.** *PNAS* 2005, **102(51):**18526-18531.
10. Deininger PL, Batzer MA: **Mammalian retroelements.** *Genome Res* 2002, **12(10):**1455-1465.
11. Deininger PL, Moran JV, Batzer MA, Kazazian HHJ: **Mobile elements and mammalian genome evolution.** *Curr Opin Genet Dev* 2003, **13(6):**651-658.
12. Kazazian HHJ: **Mobile elements: drivers of genome evolution.** *Science* 2004, **303(5664):**1626-1632.
13. Smit AF: **Interspersed repeats and other mementos of transposable elements in mammalian genomes.** *Curr Opin Genet Dev* 1999, **9(6):**657-663.
14. Consortium IHGS: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409(6822):**860-921.
15. Abril JF, Agarwal P, Alexandersson M, Antonarakis SE, Baertsch R, Berry E, Birney E, Bork P, Bray N, Brent MR, Brown DG, Butler J, Bult C, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Collins FS, Copley RR, Couronne O, Cawley S, Cuff J, Curwen V, Cutts T, Daly M, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dubchak I, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Flicek P, Frankel WN, Fulton LA, Furey TS, Gnerre S, Glusman G, Goldman N, Goodstadt L, Green ED, Gregory S, Guigo R, Hardison RC, Haussler D, Hillier LDW, Hinrichs A, Hlavina W, Hsu F, Hubbard T, Jaffe DB, Kamal M, Karolchik D, Karlsson EK, Kasprzyk A, Keibler E, Kent WJ, Kirby A, Kolbe DL, Korf I, Kulbokas EJIII, Kulp D, Lander ES, Letunic I, Li M, Lindblad-Toh K, Ma B, Maglott DR, Mauceli E, Mesirov JP, Miller W, Mott R, Mullikin JC, Ning Z, Pachter L, Parra G, Pevzner P, Poliakov A, Ponting CP, Potter S, Reymond A, Roskin KM, Sapojnikov V, Schultz J, Schwartz MS, Schwartz S, Searle S, Singer JB, Slater G, Smit A, Stabenau A, Sugnet C, Suyama M, Tesler G, Torrents D, Tromp J, Ucla C, Vinson JP, Wade CM, Weber RJ, Wheeler R, Winter E, Yang SP, Zdobnov EM, Waterston RH, Whelan S, Worley KC, Zody MC: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420(6915):**520-562.

16. Kim TM, Hong SJ, Rhyu MG: **Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate.** *J Korean Med Sci* 2004, **19(2):**177-185.

17. Schibler L, Roig A, Mahe MF, Save JC, Gautier M, Taourit S, Boichard D, Eggen A, Cribiu EP: **A first generation bovine BAC-based physical map.** *Genet Sel Evol* 2004, **36:**105-122.

18. Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennett GL, Sugimoto Y: **A comprehensive genetic map of the cattle genome based on 3802 microsatellites.** *Genome Res* 2004, **14(10A):**1987-1998.

19. Hayes H, Elduque C, Gautier M, Schibler L, Cribiu E, Eggen A: **Mapping of 195 genes in cattle and updated comparative map with man, mouse, rat and pig.** *Cytogenet Genome Res* 2003, **102(1-4):**16-24.

20. WebFPC: **[http://locus.jouy.inra.fr/fpc/cattle_bac_map.htm].** .

21. BovMap: **[http://dga.jouy.inra.fr/cgi-bin/lgbc/livestock.pl].** .

22. Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS: **Resolution of the early placental mammal radiation using Bayesian phylogenetics.** *Science* 2001, **294(5550):**2348-2351.

23. Amrine-Madsen H, Koepfli KP, Wayne RK, Springer MS: **A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships.** *Mol Phylogenet Evol* 2003, **28(2):**225-240.

24. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED: **Comparative analyses of multi-species sequences from targeted genomic regions.** *Nature* 2003, **424(6950):**788-793.

25. Casals F, Caceres M, Ruiz A: **The foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of Drosophila buzzatii.** *Mol Biol Evol* 2003, **20(5):**674-685.

26. Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikodvorskaja V, Pyatkov KI, Zhivotovsky LA, Kidwell MG: **Mobile elements and chromosomal evolution in the virilis group of Drosophila.** *Proc Natl Acad Sci U S A* 2000, **97(21):**11337-11342.

27. Ladeveze V, Aulard S, Chaminade N, Periquet G, Lemeunier F: **Hobo transposons causing chromosomal breakpoints.** *Proc Biol Sci* 1998, **265(1402):**1157-1159.

28. Mathiopoulos KD, della Torre A, Santolamazza F, Predazzi V, Petrarca V, Coluzzi M: **Are chromosomal inversions induced by transposable elements? A paradigm from the malaria mosquito Anopheles gambiae.** *Parassitologia* 1999, **41(1-3):**119-123.

29. Lyttle TW, Haymer DS: **The role of the transposable element hobo in the origin of endemic inversions in wild populations of Drosophila melanogaster.** *Genetica* 1992, **86(1-3):**113-126.

30. Lim JK: **Intrachromosomal rearrangements mediated by hobo transposons in Drosophila melanogaster.** *Proc Natl Acad Sci U S A* 1988, **85(23):**9153-9157.

31. 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: **Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution.** *Genome Res* 2005, **15(1):**1-18.

32. Gilbert N, Lutz-Prigge S, Moran JV: **Genomic deletions created upon LINE-1 retrotransposition.** *Cell* 2002, **110(3):**315-325.

33. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD: **Human l1 retrotransposition is associated with genetic instability in vivo.** *Cell* 2002, **110(3):**327-338.

34. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA: **Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages.** *Nucleic Acids Res* 2005, **33(13):**4040-4052.

35. Zhang J, Peterson T: **Genome rearrangements by nonlinear transposons in maize.** *Genetics* 1999, **153(3):**1403-1410.

36. Kolomietz E, Meyn MS, Pandita A, Squire JA: **The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors.** *Genes Chromosomes Cancer* 2002, **35(2):**97-112.

37. Babcock M, Pavlicek A, Spiteri E, Kashork CD, Ioshikhes I, Shaffer LG, Jurka J, Morrow BE: **Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution.** *Genome Res* 2003, **13(12):**2519-2532.

38. Bailey JA, Liu G, Eichler EE: **An Alu transposition model for the origin and expansion of human segmental duplications.** *Am J Hum Genet* 2003, **73(4):**823-834.

39. Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X: **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* 2003, **12(17):**2201-2208.

40. Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N: **Chromosome 6 phylogeny in primates and centromere repositioning.** *Mol Biol Evol* 2003, **20(9):**1506-1512.

41. Stankiewicz P, Park SS, Inoue K, Lupski JR: **The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP.** *Genome Res* 2001, **11(7):**1205-1210.

42. Dobigny G, Ozouf-Costaz C, Waters PD, Bonillo C, Coutanceau JP, Volobouev V: **LINE-1 amplification accompanies explosive genome repatterning in rodents.** *Chromosome Res* 2004, **12(8):**787-793.

43. Gray YH: **It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements.** *Trends Genet* 2000, **16(10):**461-468.

44. Rumpler Y, Dutrillaux B: **Chromosomal evolution and speciation in primates.** *Revis Biol Celular* 1990, **23:**1-112.

45. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ: **Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication.** *Nature* 2005, **437(7055):**94-100.

46. Pevzner P, Tesler G: **Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution.** *Proc Natl Acad Sci U S A* 2003, **100(13):**7672-7677.

47. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D: **Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes.** *Proc Natl Acad Sci U S A* 2003, **100(20):**11484-11489.

48. Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE: **Hotspots of mammalian chromosomal evolution.** *Genome Biol* 2004, **5(4):**R23.

49. Arnault C, Dufournel I: **Genome and stresses: reactions against aggressions, behavior of transposable elements.** *Genetica* 1994, **93(1-3):**149-160.

50. RepeatMasker: **[http://www.repeatmasker.org].** .

51. Band MR, Larson JH, Rebeiz M, Green CA, Heyen DW, Donovan J, Windish R, Steining C, Mahyuddin P, Womack JE, Lewin HA: **An ordered comparative map of the cattle and human genomes.** *Genome Res* 2000, **10(9):**1359-1368.

52. GoldenPath: **[http://genome.ucsc.edu].** .

53. Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP, Lopez-Corrales NL, Beattie CW: **A second-generation linkage map of the bovine genome.** *Genome Res* 1997, **7(3):**235-249.

54. UCSC: **[http://hgdownload.cse.ucsc.edu/downloads.html].** .