Title
Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes?

Permalink
https://escholarship.org/uc/item/3k2029zg

Journal
Genome Research, 16(3)

ISSN
1088-9051

Authors
Froenicke, Lutz
Caldés, Montserrat Garcia
Graphodatsky, Alexander
et al.

Publication Date
2006-03-01

Peer reviewed
Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes?

Lutz Froenicke, Montserrat Garcia Caldés, Alexander Graphodatsky, Stefan Müller, Leslie A. Lyons, Terence J. Robinson, Marianne Volleth, Fengtang Yang and Johannes Wienberg

*Genome Res.* 2006 16: 306-310
Access the most recent version at doi:10.1101/gr.3955206

**Supplementary data**

“Supplemental Research Data”
http://www.genome.org/cgi/content/full/16/3/306/DC1

**References**

This article cites 28 articles, 9 of which can be accessed free at:
http://www.genome.org/cgi/content/full/16/3/306#References

Article cited in:
http://www.genome.org/cgi/content/full/16/3/306#otherarticles

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

**Notes**

To subscribe to *Genome Research* go to:
http://www.genome.org/subscriptions/

© 2006 Cold Spring Harbor Laboratory Press
Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes?

Lutz Froenicke,1,9 Montserrat Garcia Caldés,2 Alexander Graphodatsky,3 Stefan Müller,4 Leslie A. Lyons,1 Terence J. Robinson,5 Marianne Volleth,6 Fengtang Yang,7 and Johannes Wienberg4,8

1Department of Population Health & Reproduction, School of Veterinary Medicine, University of California Davis, Davis, California 95616, USA; 2Department de Biología Cellular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain; 3ASG–Institute of Cytology and Genetics, SB Russian Academy of Sciences, Novosibirsk, 630090, Russia; 4Department Biology II, Human Genetics, Ludwig-Maximilians-University Munich, Martinsried, 82152, Germany; 5Evolutionary Genomics Group, Department of Botany & Zoology, University of Stellenbosch, 7602, South Africa; 6Department of Human Genetics, Otto-von-Guericke-University Magdeburg, Magdeburg, 39120, Germany; 7The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; 8Institute of Human Genetics, GSF-National Research Center for Environment and Health, Neuherberg, 85764, Germany

“Excavating” ancestral genomes

The recent release of the chicken genome sequence (Hillier et al. 2004) provided exciting news for the comparative genomics community as it allows insights into the early evolution of the human genome. A bird species can now be used as an outgroup to model early mammalian genome organization and reshuffling. The genome sequence data have already been incorporated in a computational analysis of chicken, mouse, rat, and human genome sequences for the reconstruction of the ancestral genome organization of both a mammalian ancestor as well as a murid rodent ancestor (Hillier et al. 2004; Bourque et al. 2005). This bioinformatic effort joins a molecular cytogenetic model (Richard et al. 2003; Yang et al. 2003; Robinson et al. 2004; Svartman et al. 2004; Wienberg 2004; Froenicke 2005) as the second global approach to explore the architecture of the ancestral eutherian karyotype—a fundamental question in comparative genomics. Since both models use the human genome as reference, they are readily comparable. Surprisingly, however, they share few similarities. Only two small autosomes and the sex chromosomes of the hypothesized ancestral karyotypes are common to both. Unfortunately, given its significance, neither the extent of these differences nor their impact on comparative genomics have been discussed by Bourque and colleagues (2005). In an attempt to redress this, we compare the two methods of ancestral genome reconstruction, verify the resulting models, and discuss reasons for their apparent divergence.

Comparative chromosome painting

The cytogenetic model was developed over the last 10 years and is now based on comparative chromosome painting data from >80 eutherian species (including 50 primates). Comparative chromosome painting (or Zoo-FISH) allows a fast generation of large-scale comparative genome maps in Placentalia (Wienberg et al. 1990; Scherthan et al. 1994). It uses cross-species fluorescence in situ hybridization (FISH) using human or other chromosome-specific DNA sequences as painting probes. Zoo-FISH identifies three kinds of genomic data: (1) evolutionarily conserved chromosomes (Chowdhary et al. 1998), (2) conserved chromosomal segments (O’Brien and Graves 1990), and (3) syntenic associations (Chowdhary et al. 1998). Syntenic associations are comprised of adjacent conserved segments that display homology to two different human chromosomes. The “algorithm” used in the reconstruction of an ancestral chromosome form is based on a cladistic analysis of ancestral versus derived features using appropriate outgroup species (Hennig 1966). A particular chromosome form is considered ancestral if the trait is not only found within a given taxon but also in more distantly related species that serve as outgroups. For example, the homologs of human chromosomes 3 and 21 are also independent in great apes but make up a single chromosome in Prosimians and New World monkeys. This association is further found in all analyzed nonprimate mammals (e.g., see Fig. 2 below). Thus, the association of these two chromosomes represents the ancestral (i.e., pliosiomorphic) form for primates, while human and apes show the evolutionarily derived state. Since all analyzed mammalian orders also comprise species with conserved karyotypes, the analysis of ancestral eutherian karyotypes can be achieved by comparing the data in character tables as have been published by several groups (e.g., Richard et al. 2003; Froenicke 2005). In taxa with highly rearranged genomes such as New World monkeys and gibbons, ancestral karyotypes were reconstructed by treating chromosome rearrangements as discrete characters in parsimony analyses using the PAUP (Phylogenetic Analysis Using Parsimony) software package (e.g., Müller et al. 2003; for review, see Dobigny et al. 2004). Wherever possible, the orthologies of such identified ancestral chromosome forms are verified by comparisons to other available genome data (gene-mapping data, reciprocal Zoo-FISH, and recently genome sequence alignments).

In spite of the limitations of the technique (intrachromosomal rearrangements usually cannot be identified), Zoo-FISH...
has clearly demonstrated that species within most eutherian clades have retained rather conserved karyotypes, extending even to some rodents (Fig. 1A,B). In fact, the average rate of chromosomal exchanges characterizing their evolutionary history is estimated at ∼1.9 rearrangements per 10 million years (Froenicke 2005), although some notable exceptions do exist. Vastly accelerated rates of chromosomal evolution have been encountered within murid rodents, caniform carnivores, and gibbons (Wienberg 2004).

The molecular cytogenetic model

The first model of an ancestral eutherian karyotype was proposed as early as 1998 on the basis of Zoo-FISH data from seven non-primate mammalian species (Chowdhary et al. 1998). The present consensus model incorporates data from 14 out of 19 placental mammalian orders (Fig. 1A) and reflects the independent cladistic analyses of several groups (Richard et al. 2003; Yang et al. 2003; Robinson et al. 2004; Svartman et al. 2004; Wienberg 2004; Froenicke 2005). Figure 1A depicts the ancestral chromosomal complement of two superordinal clades, the Boreoeutheria and Euarchontoglires, both with a chromosome number of 2n = 46. According to the cytogenetic data, the ancestral Placentalia karyotype was either identical or closely similar to this reconstruction. Our confidence in the model is founded on the following: (1) Reciprocal chromosome painting (i.e., the heterologous hybridization of chromosome-specific probes of two species in both directions) has shown that the proposed ancestral syntenic associations are made up of orthologous segments. They are thus likely to represent true homologies and not homoplasies. (2) One species (aardvark), which belongs to the presumed oldest placental clade, the Afrotheria, has retained all of the ancestral segments suggested by cross-species chromosome painting (Yang et al. 2003). (3) The comparisons of Zoo-FISH maps with genome sequence alignments for the species pairs human/dog (Sargan et al. 2000), and mouse/rat (Stanyon et al. 1999; Nilsson et al. 2001) reveal a high degree of concordance between the two approaches of genome analyses, even for very derived genomes (Supplemental Fig. 1).

Reconstructions using the Multiple Genome Rearrangement algorithm

One of the greatest challenges in handling the vast amount of data produced by the genome projects is the development of algorithms that can trace the evolutionary process of genome reorganization based on DNA sequence or gene order informa-
tion. The most frequently used algorithm (Multiple Genome Rearrangement, MGR) (Bourque and Pevzner 2002) explains the evolutionary changes between the genomes by attempting to calculate the minimum number of rearrangements between synteny blocks (conserved homologous segments with a mostly conserved gene order). This method has also been applied to model the ancestral mammalian genome (Hillier et al. 2004; Bourque et al. 2005) based on the identification of 586 synteny blocks among chicken, mouse, rat, and human genome sequences (referred to here as the “bioinformatic model”). According to the current molecular phylogenetic tree (Amrine-Madsen et al. 2003; Margulies et al. 2005), however, this model should correctly be referred to as a model of an “Euarchontoglires ancestor,” and not as the “mammalian ancestor,” since it considers only data from a single superordinal eutherian clade and the chicken.

Comparing the two models

Both approaches reconstruct evolutionary genomic change by identifying the most parsimonious number of rearrangements of ancestral building blocks, albeit on vastly different scales. However, the molecular cytogenetic and bioinformatic models suggest widely divergent ancestral karyotypes. While the chromosome numbers of 1n = 23 and 1n = 21, respectively, are similar, the numbers of conserved segments (cytogenetics: 32; bioinformatics: 56) and the numbers of syntenic associations (cytogenetics: 8; bioinformatics: 29) display great differences. Of the 29 proposed syntenic associations suggested by the bioinformatic model, only four are shared with the eight ancestral syntenic associations identified by Zoo-FISH. Although it could be argued that the difference in the numbers of conserved segments is simply a reflection of the increased resolution provided by DNA sequence comparisons, this explanation is only applicable to a very small number of the segments. According to the Supplemental data (Bourque et al. 2005), the majority of the conserved segments are well above the resolution limit of Zoo-FISH (~4 Mb). Thus, these syntenic associations and conserved segments should have been observed by chromosome painting in at least a subset of the mammalian species analyzed so far.

The dog genome sequence test

The release of the dog draft genome sequence (Lindblad-Toh et al. 2005) and the availability of human and dog genome sequence alignments through the ENSEMBL browser (http://www.ensembl.org; assemblies NCBI 35 and CanFam1.0) provide an excellent opportunity for testing the two putative ancestral genome models. Syntenic associations can be expected to be the best conserved markers suitable for a comparison of genome sequence alignments, synteny block maps, and Zoo-FISH maps. From the 29 syntenic associations suggested in the bioinformatic ancestral model, 17 are absent in the dog genome, and six are present but have a different gene content in either one or both participating syntenic blocks (HSA 1/16, 1/17, 2/9, 2/18, 4/6, 10/11) and are therefore not homologous. The homology of two associations (HSA 5/9 and 1/18) in the dog cannot be verified because of a lack of sequence alignments in the respective regions. The four remaining associations that are present in both the dog genome and the bioinformatic model (HSA 3/21, 4/8, 12a/22a, 12b/22b) are also part of the cytogenetic model (Figs. 1A and 2). The dog genome shares two further associations exclusively with the cytogenetic model (HSA 7/16, 14/15). In summary, six out of eight ancestral associations identified by the cytogenetic model can be retrieved in alignments of human and dog sequences. In contrast, none of 25 syntenic associations exclusively delineated by the bioinformatic model can be confirmed by such comparisons.

Conservation of ancestral Amniota syntenies

The alignment of chicken/human genome sequences (http://www.ensembl.org; assemblies NCBI 35 and WASHUC1) (Birney et al. 2004) enables for the first time an outgroup comparison of the molecular cytogenetic genome data (Figs. 1A and 2; Supplemental Fig. 2). The cytogenetic model differs by only eight syntenic associations from the human karyotype. Seven out of these eight proposed ancestral placental associations can be recovered in the chicken genome sequence (Figs. 1A and 2; Supplemental Fig. 2). Only the HSA10/12 syntenic association seems to be missing in the chicken genome (Fig. 1A). Despite the presence of bigger gaps in the alignments, it can be shown for five of the seven associations that they are, indeed, made up of orthologous conserved segments by comparing the chicken/human and dog/human genome sequence alignments (http://www.ensembl.org) (Fig. 2). Exceptions are HSA14/15 and 16/19. This concordance provides strong corroboration of the cytogenetic models of ancestral Eutherian karyotypes. Moreover, they extend the time...
frame for the conservation of the following larger segments of the ancestral eutherian genome to the ancestral Amniota genome itself (>350 million years ago, [Mya]). These are HSA4, 5p, 6p21.2-qter, 8q, 10q, 12p2ter-q/22qter, 12qter/22cen, 15 qter, 18p+qter, 16q/19q, and 21. Additionally, human/chicken genome sequence alignments suggest the conservation of chromosomes HSA13, 14, 16p, 19p, 20q, Xp, and Xq, although in these instances the alignments display larger gaps.

A similar analysis of the syntenic associations present in the bioinformatic model (Bourque et al. 2005, Fig. 2) shows the following: Of the 25 ancestral mammalian associations that are exclusively delineated by the bioinformatic model, eight occur only in the Muridae, and four are limited to the chicken. Six syntenic associations are shared by chicken and Muroidea genomes (HSA 1/16, 2/3, 2/6, 2/15, 5/9, 11/15). However, none of these shared syntenic associations are made up of orthologous syntenic block pairs in both mouse and chicken genomes, and thus they are not truly homologous.

In conclusion, the attempt to verify the two ancestral genome models using additional genome sequence data does not provide support for the current bioinformatic model. The cytogenetic model, as broad-scale as it is compared to sequence data, aligns well with the dog and chicken genome sequence data. Given the taxonomic coverage provided by the cytogenetic data, coupled to the resolution offered by the dog genome sequence comparison, we would suggest that the cytogenetic model provides a far more likely representation of the ancestral Eutherian genome.

Potential causes for the discrepancies of both models

What are the most compelling reasons for the discrepancies between the two models? The major factor is probably the small sample size of just four genomes analyzed by the bioinformatic study. This problem is likely to be aggravated by the presence of mammalian genomic rearrangement hotspots. These have been identified by both bioinformatic and molecular cytogenetic methods (Murphy et al. 2003; Bailey et al. 2004; Ruiz-Herrera et al. 2005). Clearly, the “reuse” of such breakpoint regions (Pevzner and Tesler 2003) complicates the tracing of rearrangements. A reliable determination of the ancestral organization in such cases will only be possible through the analysis of a wide range of species, far exceeding the number of currently available genome sequences.

Other confounding factors should also be considered. One of these might be introduced by the algorithm used in the calculations. According to Bourque et al. (2004), the MGR algorithm does not reconstruct an ancestral genome, but “computes a possible median ancestor.” However, the median between two closely related and highly rearranged Muroidea genomes (Stanyon et al. 1999), the evolutionarily distant chicken genome, and only one conserved genome (human), is unlikely to be a representative of the common ancestor of all mammals. Rather, it will be biased toward the two highly rearranged murid genomes. This hypothesis is consistent with the outcomes of an earlier study that used MGR in the comparative analysis of 114 markers in two conserved genomes (human and domestic cat), and only a single murid (mouse) (Bourque and Pevzner 2002). Although the resulting “ancestral median” of this study is missing four human chromosome homologs, it displays a much higher overall similarity to the cytogenetic ancestral model (nine ancestral chromosomes are in agreement with the cytogenetic model).

Outlook

The reliability of computational reconstructions will certainly be improved by the future inclusion of more species and more sophisticated algorithms. An improved method for the computational reconstruction of genome evolution might need to include some form of correction for significantly increased rates of rearrangements, in particular, evolutionary lineages such as the Muridae. Moreover, in cases in which the relationships of the analyzed species are already established, it might be advantageous to include a system of cladistic checks to allow for a reliable assignment of rearrangements to specific phylogenetic nodes.

Nevertheless, attempts to identify ancestral genomes might ultimately remain approximations. For example, Bourque et al. were able to generate >3000 alternative ancestors that did not increase the genomic distance to extant genomes simply by applying two changes each to their original model (Bourque et al. 2005, Methods section). Presumably, with greater computational effort, there would be far more alternative models conceivable with similar genomic distances and thus with equal claims to being the Euarchontoglires ancestor. The ambiguities entailed in the reported method raise questions about the appropriate presentation of such models. Should the discussion focus on a single model? Would the genomic distance to a model that conforms to the framework established by comparative cytogenetics actually be greater?

There can be no doubt that the computational reconstruction of mammalian genome evolution based on genome sequences holds great promise for understanding the mechanisms of genomic change. However, bioinformatic methods and models, as in any other discipline, need to be verified—the extensive and well-scrutinized molecular cytogenetic data set provides a framework for such a check.

Conclusions

We are of the opinion that the apparent conflict between the cytogenetic and bioinformatic models of the ancestral genome may, in part, reflect limited taxon sampling in the bioinformatic analysis coupled to an algorithm that does not cater for evolutionary rate variation among lineages. It can be anticipated that with the increased availability of genomic data from a wider spectrum of species and the development of more sophisticated algorithms, the bioinformatic and the cytogenetic models will probably converge. However, the most informative and reliable reconstruction of ancestral genomes is likely to emerge from the future integration of both data sets.

References


