Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 353 (2005) 147 – 154
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals

Tomàs Marques-Bonet, Arcadi Navarro*

Unitat de Biologia Evolutiva Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Spain

Received 10 February 2005; received in revised form 25 April 2005; accepted 10 May 2005
Available online 13 June 2005
Received by M. Batzer

Abstract

Evolutionary rates are not uniformly distributed across the genome. Knowledge about the biological causes of this observation is still incomplete, but its exploration has provided valuable insight into the genomical, historical and demographical variables that influence rates of genetic divergence. Recent studies suggest a possible association between chromosomal rearrangements and regions of greater divergence, but evidence is limited and contradictory. Here, we test the hypothesis of a relationship between chromosomal rearrangements and higher rates of molecular evolution by studying the genomic distribution of divergence between 12 000 human–mouse orthologous genes. Our results clearly show that genes located in genomic regions that have been highly rearranged between the two species present higher rates of synonymous (0.7686 vs. 0.7076) and non-synonymous substitution (0.1014 vs. 0.0871), and that synonymous substitution rates are higher in genes close to the breakpoints of individual rearrangements. The many potential causes of such striking are discussed, particularly in the light of speciation models suggesting that chromosomal rearrangements may have contributed to some of the speciation processes along the human and mouse lineages. Still, there are other possible causes and further research is needed to properly explore them.

Keywords: Segmental duplications; Molecular evolution; Mammals; Human; Mouse

1. Introduction

The well-known fact that evolutionary rates are not uniformly distributed across the genome (Wolfe et al., 1989; Wolfe and Sharp, 1993) has recently been the subject of renewed interest due to the availability of almost complete genome sequences (Koop, 1995; Paabo, 2003). Comprehensive descriptions of varying rates of molecular evolution in different chromosomes, chromosomal regions, genes, and even different nucleotides have become massively available (Matassi et al., 1999; Bernardi, 2000; Lercher et al., 2001). Such observations have provided insight into the genomical, historical and demographical variables that influence rates of molecular evolution (Paabo, 2003). Over the last two decades, we have learned, for example, that genomic regions with higher GC content experience higher rates of divergence in mammals (Matassi et al., 1999; Castresana, 2002; Ebersberger et al., 2002); that linked genes evolve at similar rates (Williams and Hurst, 2000; Lercher et al., 2001; Williams and Hurst, 2002); that genes are frequently associated with CpG rich islands (Bernardi, 2000; Hardison et al., 2003); that recombination does modify mutational rates (Nekrutenko and Li, 2000; Hellmann et al., 2003); that duplicated genes evolve faster (Lynch and Conery, 2000; Jordan et al., 2004) and, of course, that distinct selective pressures upon different codons, genes or genomic regions affect their rates of evolution (Bernardi, 2001; Fay and Wu, 2003).

This complex scenario, in which several, mutually compatible, evolutionary mechanisms generate a mosaic genome, has been recently enriched by an assembly of hypotheses related to speciation (Wu, 2001). According to models of parapatric speciation, genomic regions involved

Abbreviations: Myrs, million years; Mb, mega bases; cM, centiMorgans; ANCOVA, analysis of covariance.

* Corresponding author. Tel.: +34 93 5422841; fax: +34 93 5422802.
E-mail address: arcadi.navarro@upf.edu (A. Navarro).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.05.007

in the speciation processes splitting one ancestral species into two or more reproductively isolated daughter species, might become isolated earlier relative to other regions because they contain allelic variants that are incompatible between populations. Thus, gene flow will stop in these regions whereas it continues in the rest of the genome. When the resulting species are compared, such regions might present higher rates of evolution because they have been diverging for a longer time (Wu, 2001; Navarro and Barton, 2003a; Osada and Wu, 2005). Chromosomal rearrangements have been shown to trigger speciation processes by acting as genetic barriers to gene flow because they preclude recombination between chromosomes bearing different arrangements and, thus, facilitate the accumulation of incompatible allelic variants (Noor et al., 2001b; Rieseberg, 2001; Navarro and Barton, 2003a). Therefore, genomic regions that have undergone rearrangements could potentially be important contributors to the observation of varying rates of molecular evolution in different parts of a genome. In particular, the reduction of recombination is strongest, first, in regions around rearrangement breakpoints and, second, within the rearrangements themselves (Navarro et al., 1997; Rozas et al., 2001) and, thus, the reduction of gene flow should be stronger in there.

Although initial surveys seemed to confirm an association between chromosomal rearrangements and regions of greater genic divergence in several species, including humans and chimpanzees (Lu et al., 2003; Navarro and Barton, 2003b; Navarro et al., 2003; Marques-Bonet et al., 2004), other studies found no signs of such relationship in the later species pair (Vallender and Lahn, 2004; Zhang et al., 2004). Moreover, it is still unknown to what extent such an association, if it exists at all, could be attributed only to speciation or to alternative processes. For example, it has been suggested that rearrangements tend to occur or be favored in genomic regions of fast molecular evolution either because they are regions of low functional constraint or because they contain clusters of genes under positive selection (Lu et al., 2003; Navarro and Barton, 2003a). It is also possible that changes in recombinational context associated with rearrangements might move linked regions to regions with a different equilibrium base composition and thus lead to changes in mutation rates (Navarro and Barton, 2003a). In addition, regions harboring rearrangement breakpoints have been shown to be rich in segmental duplications (Bailey et al., 2002; Armengol et al., 2003) which may help to explain their faster molecular evolution rates (Marques-Bonet et al., 2004). Finally, chromosomal rearrangements may have direct positional effects. Indeed, there is experimental evidence that the rearrangements can induce changes in the expression patterns of genes located around their breakpoints (Tanimoto et al., 1999; Phippard et al., 2000; Spitz et al., 2003).

Here we investigate the existence of a relationship between chromosomal rearrangements and faster molecular evolution by means of a comparison of the genomes of humans and mice. After approximately 80 Myrs. of separate evolution these two species differ by more than 350 breakpoints (Pevzner and Tesler, 2003a,b). Besides allowing us to test for the existence of a systematic association pervading the mammal lineage, these features allow to compare the rates of molecular evolution of genes located in regions with different degrees of rearrangement. Also, the fact that high quality complete genomes of the two species are available makes it feasible to control for genomic variants that have been shown to be related to rates of molecular evolution.

2. Materials and methods

Human and mouse orthologous genes and their respective genomic locations were obtained from NCBI's Homologene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene). Genes were gap-aligned using BLASTN. Following standard procedures (Castresana, 2002; Castresana et al., 2004), only the best-hit segment of the alignment was used if the alignment was larger than 150 positions. Several conventional indexes of molecular evolution, such as the number of non-synonymous substitutions per non-synonymous site (K_a), the number synonymous substitutions per silent site (K_s), and its ratio (K_a/K_s) were estimated for about 12 000 orthologous genes using the maximum likelihood method implemented in the package PAML (Yang, 1997). As a first measure to avoid false orthologous gene pairs (that is, gene pairs that are not orthologous, but paralogous generated by gene duplication), only genes with $K_s<2$ and $K_a<0.5$ were used which eliminates ~1000 genes and leaves 10 869 for the analysis. Finally, a surrogate measure of GC content (GC4, the GC content at 4-fold degenerate sites) was also taken for every orthologous pair.

The positions of rearrangement breakpoints where assessed by comparing the relative chromosomal positions of orthologous genes and compared to the synteny blocks described by (Pevzner and Tesler, 2003b). The genes in our dataset were positioned in these synteny blocks with the only difference that we considered that rearrangements <1 Mb also produce synteny changes (i.e., we took micro-rearrangements into account). Using the human genome as the reference, genes were classified according to their location relative to four classes of genomic regions: (a) genes in highly rearranged regions, that is, regions where no synteny block could be determined. They contain fragments from several chromosomes and, thus, they are regions where multiple breakpoints have accumulated; (b) genes in regions containing a translocation breakpoint, that is, synteny changes that separated genomic fragments belonging to different mouse chromosomes and that were not associated with highly rearranged regions; (c) genes in regions containing an inversion breakpoint, that is, synteny changes that separated two regions of the same chromosome