Molecular Phylogenetics and Evolution 57 (2010) 836–848

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Unraveling ancient segmental duplication events in human genome by phylogenetic analysis of multigene families residing on HOX-cluster paralogons

Amir Ali Abbasi *

National Center for Bioinformatics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

ARTICLE INFO

Article history:
Received 17 April 2010
Revised 21 July 2010
Accepted 27 July 2010
Available online 7 August 2010

Keywords:
Vertebrates
Segmental duplications
Whole genome duplications
HOX-paralogons
Multigene families
Phylogeny
Co-duplications

ABSTRACT

Background: Vertebrate genomes contain extensive intra-genomic conserved synteny, which is the presence of similar set of genes on two or more chromosomes (paralogons). The existence of these paralogons has led to the proposal that vertebrate genome was structured by one or more rounds of ancient whole genome duplications (2R hypothesis).

Results: The 2R hypothesis was tested by phylogenetic analysis of gene families residing on human HOX-bearing chromosomes (HOX-cluster paralogons). These results revealed that, based on their duplication history, 23 gene families with representation on three or four of the human HOX-bearing chromosomes can be partitioned into four discrete co-duplicated groups. The distinct genes within each co-duplicated group share the same evolutionary history and are duplicated in concert with each other, while the constituent genes of two different co-duplicated groups do not share their evolutionary history and are not duplicated simultaneously. These co-duplicated groups are large constituting members from 3 to 8 gene families and suggest that human HOX-cluster paralogons were shaped by ancient segmental duplications (SDs) and rearrangement events that occurred at least as early as before the divergence of bony fishes and tetrapods.

Conclusions: Based on the recovery of ancient SD events in this analysis and given the widespread evidence in favor of the fact that recent SD events played a pivotal role in changing genome architecture of primates and other recently diverged animals, it is concluded that a more realistic model of ancient vertebrate genome evolutionary history can be deduced by tracing the evolutionary trajectory of the genomes of recently diverged vertebrate species.

1. Introduction

To explain species diversification and major morphological and developmental transitions in vertebrate evolution, it has been proposed that two rounds of whole genome duplications (WGDs) occurred at the root of vertebrate history (Ohno, 1970, 1973). This contention famously known as 2R hypothesis has been the subject of intense debate over the past couple of decades (Skrabanek and Wolfe, 1998; Martin, 1999, 2001; Hughes and Friedman, 2003; Furlong and Holland, 2004; Donoghue and Purnell, 2005; Kasahara, 2007; Abbasi, 2008, 2010; Abbasi et al., 2009; Furlong and Holland, 2002). Among those evidences that have been put forward in favor of ancient octaploidy, the most popular one is the occurrence of paralogons/paralogy regions in the vertebrate genomes: the distinct chromosomal regions within a genome that contains a set of similar genes (Larhammar et al., 2002; McLysaght et al., 2002; Hokamp et al., 2003; Lundin et al., 2003; Furlong and Holland (2002)). In particular, fourfold paralogy regions in human genome, notably on HSA 1/6/9/19, HSA 4/5/8/10, HSA 1/2/8/10 and the HOX-bearing chromosomes HSA 2/7/12/17 are considered to be shaped directly by vertebrate octaploidy (Hokamp et al., 2003; Lundin et al., 2003).

The investigation of genomic sequence data from an expanding range of vertebrate and invertebrate species casts serious doubt on the plausibility of 2R hypothesis (Abbasi, 2008). It has been suggested that elucidating intra-genomic synteny by comparing global physical organization of genes (map self-comparison) does not constitute the evidence for the mechanism of origin of paralogons (Abbasi, 2008). Therefore, sheer map distribution of subset of human genes should not be taken as evidence in favor of WGDs. Instead it has been suggested that such patterns supports 2R only if some conditions are met; evolutionary history of the gene families constituting paralogons should suggest that majority of them duplicated in the early history of vertebrate lineage (Hughes, 1998; Hughes et al., 2001; Martin, 2001; Abbasi and Grzeschik, 2007); families that duplicated within the time window of invertebrates–vertebrates and bony fishes–tetrapods split should reveal consistencies in their tree topologies (Hughes et al., 2001; Martin, 2001; Abbasi and Grzeschik, 2007); ideally the phylogenetic trees

* Tel.: +92 051 9064 4066/3034, mobile: +92 03348959354.
E-mail address: abbasiam@qau.edu.pk

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.07.021

of quadruplicated families (under 2R assumption) should exhibit the topology of the form (AB)(CD), i.e. two clusters of two genes (Hughes, 1999; Martin, 2001).

The human HOX-bearing chromosomes (Hsa 2, 7, 12 and 17) contain extensive intra-genomic quadruplicate synteny centered on HOX clusters. The fact that two or more paralogs of numerous gene families are linked with HOX genes has been taken as evidence that these paralogous gene sets along with the linked HOX clusters might have arisen simultaneously through two rounds of block/whole chromosome or even whole genome duplication events (Larhammar et al., 2002; Hokamp et al., 2003; Lundin et al., 2003; Sundstrom et al., 2009). To test this assumption, by employing broader taxonomic sampling Abbasi and Grzeschik, 2007) conducted a thorough phylogenetic analysis of 11 gene families sharing members on three or more of the human chromosomes 2, 7, 12, and 17, the chromosomes that bear HOX clusters. The results provided strong evidence against the 2R hypothesis (Abbasi and Grzeschik, 2007).

In this study, I extend our previous work and have identified 11 more HOX linked gene families with paralogs linked to at least three of the four human HOX clusters (Fig. 1 and Table 1). By exploiting the current accessibility of an immense amount of protein data from an expanding range of vertebrate and invertebrate species from interspersed time points, a robust and thorough phylogenetic analysis of these families was performed. Given the phylogenetic data, the topology comparison approach (Hughes et al., 2001; Martin, 2001; Abbasi and Grzeschik, 2007) was used to test which genes have duplicated concurrently with each other and with the linked HOX clusters at the base of vertebrate lineage. Together with our previous results, the present analysis indicate that triplicate or quadruplicate conserved synteny seen on the human HOX cluster bearing chromosomes is not the outcome of two rounds of whole chromosome or whole genome duplications. Instead, these data showed that HOX-cluster paralogons were structured by segmental duplications and rearrangement events that occurred at different time points during early evolution of vertebrate lineage.

2. Materials and methods

2.1. Dataset

Genes from 11 families were included in the analysis (Table 1). The chromosomal location of human gene families was obtained from Ensembl genome browser (Hubbard et al., 2002), three of these families have members on each of the human HOX-bearing chromosomes while eight have their members on at least three of those chromosomes (Table 1 and Fig. 1). Information about the molecular functions (Table 1) of selected gene families was retrieved from GeneReports available at SOURCE (Diehn et al., 2003).

The closest putative orthologous sequences of human proteins in other species were obtained from Orthologue Prediction at Ensembl (Hubbard et al., 2002). To enrich these gene families with sequences from those organisms for which the sequence information was not available at Ensembl, BLASTP (Altschul et al., 1990) search was carried out against the protein database available at National Centre for Biotechnology Information (Johnson et al., 2008) and the Joint Genome Institute (http://www.jgi.doe.gov/). Because the focus of this study was to identify the duplications events which had occurred during vertebrate evolution, only blast hits giving a higher score than the sequence of available invertebrate ancestral sequences were retained. Further confirmation of ancestral–descendents relationship among putative orthologs was done through clustering of homologous proteins within phylogenetic trees. Sequences whose position within a tree was sharply in conflict with the uncontested animal phylogeny were excluded from analysis. Final dataset includes, 51 sequences from 22 animal

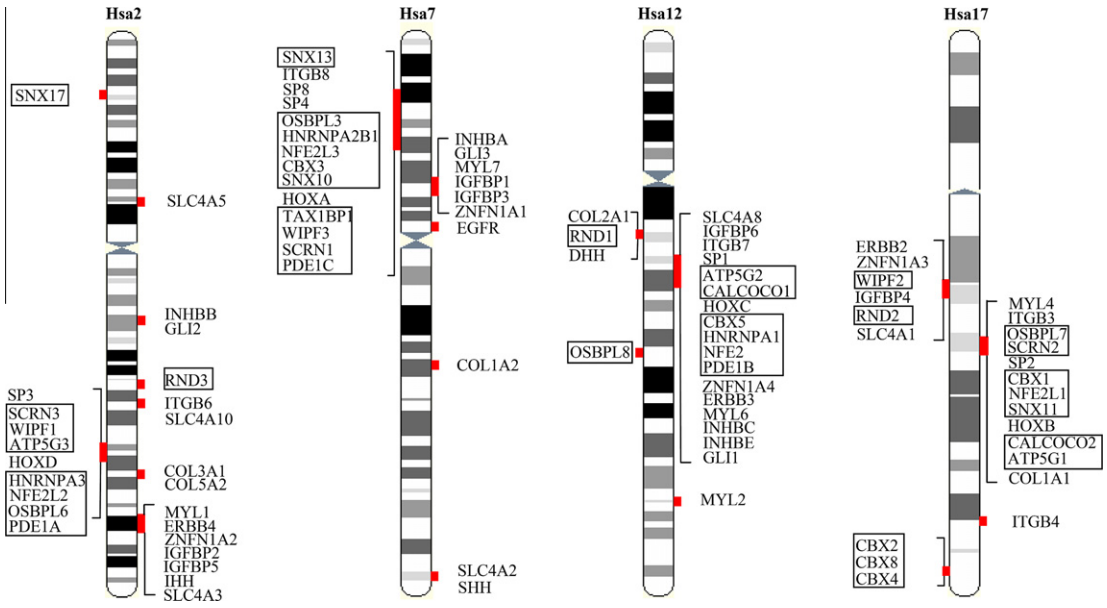

Fig. 1. Gene families with members on at least three of the human HOX-bearing chromosomes 2, 7, 12 and 17. Restricted location of members of many of these gene families near the HOX clusters suggests that HOX-cluster paralogons might have been shaped by two rounds of block/whole chromosome duplication. SLC4, solute carrier family 4; INHB, inhibins; GLI, glioma-associated oncogene homolog belonging to kruppel family; ITGB, integrin β chains; SP, transcription factor Sp; HOX, homeobox; COL, collagens; MYL, myosin light chains; EGFR/ERBB, epidermal growth factor receptor/erythroblastoma; ZNFN1A, zinc finger protein, subfamily 1A; IGFBP, insulin-like growth factor-binding protein; HH, hedgehog; SNX, sorting nexin; SCRN, secernin; WIPF, WAS/WASL interacting protein family; HNRNPA, Heterogeneous nuclear ribonucleoprotein A; NFE2, Nuclear factor erythroid-derived 2; OSBPL, Oxysterol binding protein like; PDE1, Phosphodiesterase 1, Calmodulin-dependent; CBX, Chromobox homolog; RND, Rho family GTPase; ATP5G, ATP synthase, H^+ transporting, mitochondrial F0 complex; CALCOCO1, Calcium binding and coiled-coil domain. Genes analyzed in this study are enclosed within rectangles, whereas the histories of other genes (not enclosed in rectangles) were presented in our previous data (Abbasi and Grzeschik, 2007). None of the features of this figure are drawn to scale.