Molecular Phylogenetics and Evolution 94 (2016) 95–100

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Phylogenomic analysis reveals ancient segmental duplications in the human genome ☆

Madiha Hafeez [1], Madiha Shabbir [1], Fouzia Altaf [1], Amir Ali Abbasi *

National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

ARTICLE INFO

Article history:
Received 21 April 2015
Revised 4 August 2015
Accepted 21 August 2015
Available online 2 September 2015

Keywords:
FGFR
Phylogeny
Gene duplications
Whole genome duplications
Vertebrates
Human chromosomes
Paralogon

ABSTRACT

Evolution of organismal complexity and origin of novelties during vertebrate history has been widely explored in context of both regulation of gene expression and gene duplication events. Ohno (1970) for the first time put forward the idea of two rounds whole genome duplication events as the most plausible explanation for evolutionarizing the vertebrate lineage (2R hypothesis). To test the validity of 2R hypothesis, a robust phylogenomic analysis of multigene families with triplicated or quadruplicated representation on human FGFR bearing chromosomes (4/5/8/10) was performed. Topology comparison approach categorized members of 80 families into five distinct co-duplicated groups. Genes belonging to one co-duplicated group are duplicated concurrently, whereas genes of two different co-duplicated groups do not share their duplication history and have not duplicated in congruency. Our findings contradict the 2R model and are indicative of small-scale duplications and rearrangements that cover the entire span of animal's history.

1. Introduction

One of the major focuses of evolutionary developmental biologists is to unveil the genetic underpinnings of major changes in organismal design and the origin of novelties during evolutionary history of animals. A large body of evolutionary developmental (*evo-devo*) studies has revealed that differences among closely and distantly related animal taxa are associated with differences in the spatial and temporal aspects of gene expression regulation during development (Villar et al., 2014). Still others consider increase in gene number by duplication as a principle mechanism responsible for increase in organismal complexity and diversity (Ohno, 1973; Van de Peer et al., 2009). Evolution of organismal complexity and origin of novelties during vertebrate history has been widely explored in context of both small and large scale gene duplication events (Abbasi, 2008; Rogers and Gibbs, 2014). The first instance was in the year 1970 when Susumu Ohno put forward the idea of whole genome duplications (WGD) as the most plausible explanation of evolution of form during early history of vertebrates (Ohno, 1970). This notion popularly theorized as "2R hypothesis" (two rounds of WGDs) has been widely debated (Abbasi, 2010a). Among substantial evidence adduced in favor of ancient vertebrate polyploidy (genome duplications), the most widely cited suggests the existence of paralogons or paralogous genomic segments in vertebrate genomes: homologous chromosomal segments within the genome sharing similar sets of genes (Hwang et al., 2013; Putnam et al., 2008). Precisely, the occurrence of four potential quadruplicated regions, notably on Homo sapiens autosome (Hsa) 1/6/9/19, Hsa 4/5/8/10, Hsa 1/2/8/10 and the HOX-cluster bearing chromosomes Hsa 2/7/12/17, are considered to be structured by two rounds of polyploidy (Furlong and Holland, 2002). However, it is alternatively hypothesized that the profusion of paralogy regions on human chromosomes is due to a high rate of local duplications, translocations and genomic rearrangement events that occurred at widely different time points during early vertebrate history, disputing the central tenet of Ohno's idea (Hughes et al., 2001).

Over the years, several human paralogons are being surveyed to determine the mechanistic of embarkation of these vertebrate specific paralogy regions. In this regard, HOX and FGFR clusters are most intensively studied paralogons and are considered to have coevolved through 2R-WGD, collectively contributing toward higher complexity in vertebrates (Coulier et al., 1997). The earlier studies on these paralogons incorporated data from sparse sampling of genes and were limited to few species, leaving this longstanding mystery yet to be solved (Abbasi, 2008). To dig deeper into the debate, our group is continuously putting the

☆ This paper was edited by the Associate Editor Austin Hughes.

* Corresponding author.
E-mail address: abbasiam@qau.edu.pk (A.A. Abbasi).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ympev.2015.08.019
1055-7903/

efforts in assembling and dating the primordial duplication events that molded animal lineage (Abbasi, 2010b; Abbasi and Hanif, 2012; Ajmal et al., 2014; Ambreen et al., 2014; Asrar et al., 2013). Previously we analyzed the evolutionary histories of 21 multi gene families (93 genes) with triplicated or quadruplicated distribution on human FGFR bearing chromosomes (4/5/8/10). To perform the robust phylogenetic analysis, 1494 sequences were obtained from diverge range of vertebrate and invertebrate species. These data was not substantiated in the favor of en bloc duplications, instead, it appeared that paralogy blocks residing on human FGFR bearing paralogon are the consequence of small-scale duplication events that spread across the entire history of animals (Abbasi, 2010b; Abbasi and Grzeschik, 2007; Ajmal et al., 2014; Ambreen et al., 2014; Asrar et al., 2013; Martin, 2001).

In the present study, we extended our previous work (Ajmal et al., 2014) and investigated the evolutionary history of further 59 human multigene families with three or fourfold representation on FGFR bearing chromosomes (Hsa 4/5/8/10). A thorough phylogenetic analysis of these gene families was performed by employing the currently available well-annotated and high-quality finished genomic sequence data using neighbor joining (NJ) and maximum likelihood (ML) methods. In addition, congruency among topologies of 80 phylogenies (59 present data and 21 previous data) was scrutinized to identify the genes which could have duplicated simultaneously at the root of vertebrate history (Abbasi, 2010b; Ajmal et al., 2014) (Fig. 1).

2. Materials and methods

2.1. Dataset

Genes from 59 gene families (in total 447 genes and 4807 amino acid sequences) with threefold or fourfold representations on human FGFR bearing chromosomes were included in the analysis. The chromosomal location of human gene families was obtained from Ensembl genome browser (Hubbard et al., 2002;

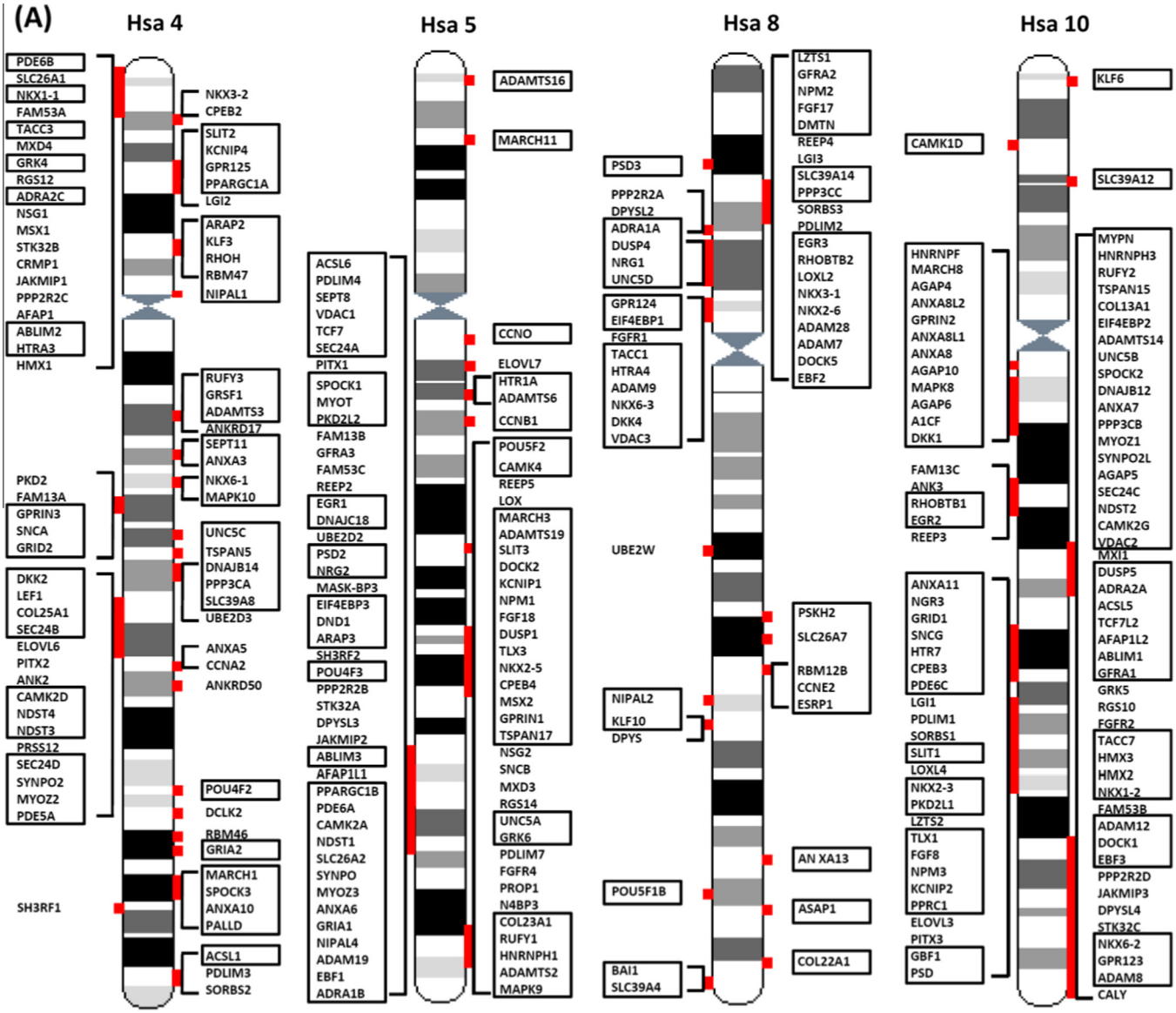

Fig. 1. Phylogenomic analysis of human chromosomes 4, 5, 8 and 10 revealed ancient segmental duplications. (a) Gene families with at least three-fold representation on human FGFR bearing chromosomes 4, 5, 8 and 10 were subjected to rigorous phylogenetic investigation. Genes analyzed in this study are enclosed within rectangles, whereas the histories of other genes (not enclosed in rectangles) were presented in our previous data (Ajmal et al., 2014). (b) The human genes duplicated concurrently lie in respective co-duplicated groups. In each case the percentage bootstrap support is given in parentheses. The connecting bars on the left depict the close physical linkage. *Represents situation where a gene family member is not residing on FGFR bearing chromosomes. None of the features of this figure are drawn to scale.