Comparative Biochemistry and Physiology, Part C 163 (2014) 24–36

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part C

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

Connectivity of vertebrate genomes: *Paired-related homeobox* (*Prrx*) genes in spotted gar, basal teleosts, and tetrapods ☆

Ingo Braasch [a], Yann Guiguen [b], Ryan Loker [a], John H. Letaw [a,1], Allyse Ferrara [c], Julien Bobe [b], John H. Postlethwait [a,*]

[a] Institute of Neuroscience, University of Oregon, Eugene, 97403-1254 OR, USA
[b] INRA, UR1037 LPGP, Campus de Beaulieu, F-35000 Rennes, France
[c] Department of Biological Sciences, Nicholls State University, Thibodaux, LA 70310, USA

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form 23 January 2014
Accepted 24 January 2014
Available online 30 January 2014

Keywords:
Prrx1
Prrx2
Genome duplication
Ohnolog
RNA-Seq
Paired appendages
Fin/limb bud
Craniofacial

ABSTRACT

Teleost fish are important models for human biology, health, and disease. Because genome duplication in a teleost ancestor (TGD) impacts the evolution of teleost genome structure and gene repertoires, we must discriminate gene functions that are shared and ancestral from those that are lineage-specific in teleosts or tetrapods to accurately apply inferences from teleost disease models to human health. Generalizations must account both for the TGD and for divergent evolution between teleosts and tetrapods after the likely two rounds of genome duplication shared by all vertebrates. Progress in sequencing techniques provides new opportunities to generate genomic and transcriptomic information from a broad range of phylogenetically informative taxa that facilitate detailed understanding of gene family and gene function evolution.

We illustrate here the use of new sequence resources from spotted gar (*Lepisosteus oculatus*), a rayfin fish that diverged from teleosts before the TGD, as well as RNA-Seq data from gar and multiple teleost lineages to reconstruct the evolution of the *Paired-related homeobox* (*Prrx*) transcription factor gene family, which is involved in the development of mesoderm and neural crest-derived mesenchyme. We show that for *Prrx* genes, the spotted gar genome and gene expression patterns mimic mammals better than teleosts do. Analyses force the seemingly paradoxical conclusion that regulatory mechanisms for the limb expression domains of *Prrx* genes existed before the evolution of paired appendages. Detailed evolutionary analyses like those reported here are required to identify fish species most similar to the human genome to optimally connect fish models to human gene functions in health and disease.

1. Introduction

Several teleost fish species, most prominently zebrafish, medaka, platyfish, stickleback, and killifish, are used as model species for human development, physiology, health, and disease (reviewed in Schartl, 2013). Despite their advantages as laboratory research models – such as easy husbandry, high fertility, external fertilization, embryo transparency, tractable genetics, and amenability for high-throughput drug screens – interpretations of results obtained from teleosts species are challenged by ~900 Ma of independent evolution that separates teleosts from the *condition humaine*: Since the last fish-like bony vertebrate (euteleostome) ancestor that lived ~450 Ma ago (Hedges et al., 2006), the lobefin vertebrate (sarcopterygian) lineage that led to tetrapods, mammals and later humans has evolved independently of the rayfin vertebrate (actinopterygian) lineage, including teleost fishes, which have significantly remodeled their morphology and, importantly, their genomes since the euteleostome ancestor.

Two rounds of vertebrate genome duplication (VGD1 and VGD2; Fig. 1) likely occurred at the stem of the vertebrate branch (Dehal and Boore, 2005; Putnam et al., 2008) and, despite the overall high level of conservation of genes between teleosts and tetrapods (Howe et al., 2013), important differences characterize the arrangement of specific gene families in both lineages. The lineages of teleosts and tetrapods have divergently retained gene duplicates from whole genome duplications, a class of paralogs generally termed 'ohnologs' (Wolfe, 2001). For example, a central regulator of pluripotency in mammals, *Pou5f1* (also known as *Oct3* or *Oct3/4*), was lost in the rayfin fish lineage (Frankenberg and Renfree, 2013). The tetrapod genome, on the other hand, has undergone substantial loss of ancestral genes during the

☆ This paper is based on a presentation given at the 6th Aquatic Annual Models of Human Disease Conference, hosted by the University of Wisconsin-Milwaukee (June 30–July 3, 2013).

* Corresponding author at: Institute of Neuroscience, University of Oregon, 1425 E. 13th Avenue, Eugene, OR 97403, USA. Tel.: +1 541 346 4538.
E-mail addresses: ibraasch@uoneuro.uoregon.edu (I. Braasch), yann.guiguen@rennes.inra.fr (Y. Guiguen), loker@uoregon.edu (R. Loker), letaw@ohsu.edu (J.H. Letaw), allyse.ferrara@nicholls.edu (A. Ferrara), Julien.Bobe@rennes.inra.fr (J. Bobe), jpostle@uoneuro.uoregon.edu (J.H. Postlethwait).
[1] Present address: Oregon National Primate Research Center, Division of Neuroscience, Oregon Health and Science University, 505 NW 185th Ave, Beaverton, OR 97006, USA.

1532-0456/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cbpc.2014.01.005

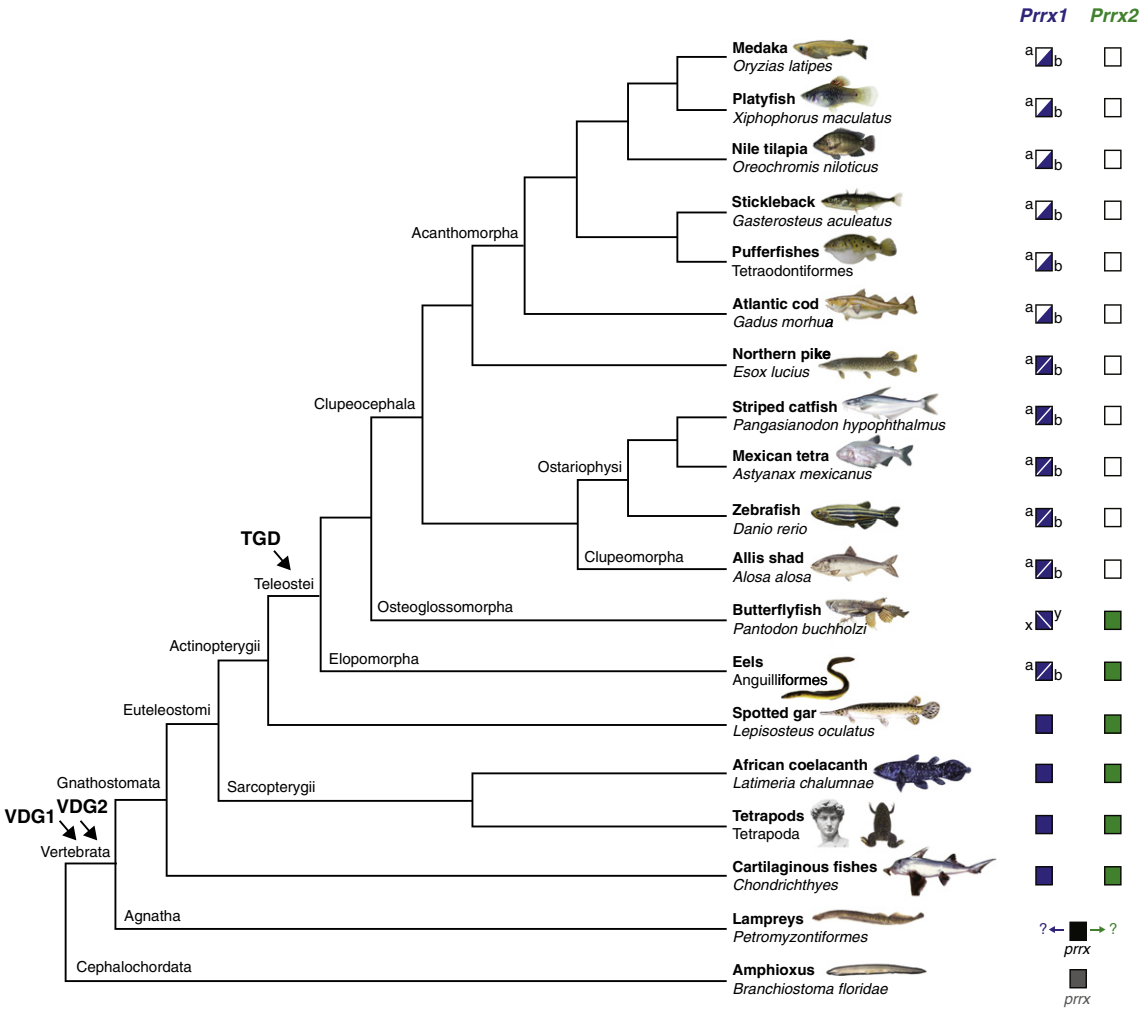

Fig. 1. Cladogram showing phylogenetic relationships among vertebrates analyzed in the present study. Tree topology was adopted from Near et al. (2012). VGD1/2: vertebrate genome duplication 1/2; TGD: teleost genome duplication. The results of our *Prrx* gene surveys are shown to the right. Presence/absence of genes is indicated by colored/white boxes. Within teleosts, a and b refer to the presence/absence of the a- and b-paralogs of *prrx1*, with the special case of x/y referring to the two *prrx1* paralogs of from butterflyfish, for which the orthology to other teleost *prrx1* genes remains unclear. The relationship of the single lamprey *prrx* gene to other vertebrate *prrx* genes also remains unresolved (see text for further information).

water-to-land transition as well, for example *actinodin* and *fgf24* (Amemiya et al., 2013).

In addition to these instances of 'ohnologs gone missing' (Postlethwait, 2007) leading to the divergence of gene contents in teleost and tetrapod genomes, an additional round of whole genome duplication occurred in an ancestor of the teleost lineage, the teleost genome duplication or TGD; 12–24% of genes have been retained as two paralogous genes in teleosts compared to one gene in tetrapods (reviewed in Braasch and Postlethwait, 2012). TGD paralogs (also termed co-orthologs), have often, like VGD1 and VGD2 ohnologs, changed functions by mechanisms such as subfunctionalization, i.e. the partitioning of ancestral gene functions among duplicates, and/or neofunctionalization, i.e. the acquisition of new gene functions in one or both co-orthologs (Force et al., 1999; Postlethwait et al., 2004). These differences in gene repertoires and gene functions between teleosts and tetrapods can make it difficult to transfer knowledge obtained in a teleost model species to the human condition.

Until recently, genomic sequence information was restricted to a few teleost model species, i.e. zebrafish (Howe et al., 2013), medaka (Kasahara et al., 2007), stickleback (Jones et al., 2012), and pufferfishes (Aparicio et al., 2002; Jaillon et al., 2004), but progress in sequencing techniques has enabled the relatively cheap and fast generation of genomic and transcriptomic sequence data from numerous fish species.

The present study illustrates how the availability of sequence information from a wide range of phylogenetically informative fish species can improve our understanding of gene function evolution among vertebrates, thereby better informing the suitability of teleost models for the analysis of specific gene functions and associated diseases. To this end, we take advantage of genomic sequence information from the spotted gar (*Lepisosteus oculatus*), a member of the closest living sister lineage to the teleosts, and the holosteans (gars and bowfin), which diverged from teleosts before the TGD (Amores et al., 2011) (Fig. 1). The spotted gar offers a unique opportunity to study gene functions in a non-teleost, unduplicated rayfin species that is suitable for gene function analysis in a laboratory environment (Amores et al., 2011; Braasch and Postlethwait, 2012). Other, even more basally diverging lineage of rayfin fish such as Polypteriformes (bichirs and reedfish, the most basal living rayfin group; Raincrow et al., 2011), as well as Acipenseriformes (sturgeons and paddlefish), would be helpful for the analysis of rayfin gene functions, but genomic resources for these lineage are currently lacking and, in the case of Acipenseriformes, are further complicated by multiple, lineage-specific polyploidization events (Braasch and Postlethwait, 2012; Crow et al., 2012). Basally diverging teleost lineages, such as elopomorphs and osteoglossomorphs (Fig. 1), on the other hand, offer the opportunity to gain better understanding of the evolutionary mechanisms leading to the divergence of gene repertoires and