Whole Genome Duplications Shaped the Receptor Tyrosine Kinase Repertoire of Jawed Vertebrates

Frédéric G. Brunet[1,*], Jean-Nicolas Volff[1], and Manfred Schartl[2,3]

[1]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, UMR5242 CNRS, Université Claude Bernard Lyon I, Lyon, France

[2]Physiologische Chemie, Biozentrum, University of Würzburg, Am Hubland, and Comprehensive Cancer Center, University Clinic Würzburg, Würzburg, Germany

[3]Texas Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, USA

*Corresponding author: E-mail: frederic.brunet@ens-lyon.fr.

Abstract

The receptor tyrosine kinase (RTK) gene family, involved primarily in cell growth and differentiation, comprises proteins with a common enzymatic tyrosine kinase intracellular domain adjacent to a transmembrane region. The amino-terminal portion of RTKs is extracellular and made of different domains, the combination of which characterizes each of the 20 RTK subfamilies among mammals. We analyzed a total of 7,376 RTK sequences among 143 vertebrate species to provide here the first comprehensive census of the jawed vertebrate repertoire. We ascertained the 58 genes previously described in the human and mouse genomes and established their phylogenetic relationships. We also identified five additional RTKs amounting to a total of 63 genes in jawed vertebrates. We found that the vertebrate RTK gene family has been shaped by the two successive rounds of whole genome duplications (WGD) called 1R and 2R (1R/2R) that occurred at the base of the vertebrates. In addition, the Vegfr and Ephrin receptor subfamilies were expanded by single gene duplications. In teleost fish, 23 additional RTK genes have been retained after another expansion through the fish-specific third round (3R) of WGD. Several lineage-specific gene losses were observed. For instance, birds have lost three RTKs, and different genes are missing in several fish sublineages. The RTK gene family presents an unusual high gene retention rate from the vertebrate WGDs (58.75% after 1R/2R, 64.4% after 3R), resulting in an expansion that might be correlated with the evolution of complexity of vertebrate cellular communication and intracellular signaling.

Key words: receptor tyrosine kinase, vertebrates, deuterostomes, whole genome duplications.

Introduction

Protein kinases (PKs), which are estimated to functionally interact with up to 30% of all human proteins, constitute one of the largest super-families among eukaryotic proteins (Hanks 2003). With a domain of 250–300 amino acids (aa) in length, these enzymes phosphorylate-specific tyrosine, serine, or threonine residues in substrate proteins using the gamma phosphate of adenosine triphosphate (Lemmon and Schlessinger 2010). Up to 2,500 PK-encoding genes have been described in plants (Lehti-Shiu and Shiu 2012), and 518 are found in human (Manning et al. 2002). Those that specifically phosphorylate tyrosine residues are the protein tyrosine kinases (PTKs) that can be subdivided into cytoplasmic non-receptor proteins (CTKs) that relay intracellular signals, and receptor tyrosine kinases (RTKs) that transduce extracellular signals to the cytoplasm. RTKs constitute a large gene family of

cell-surface receptors that activate several downstream signaling cascades with major roles in the development and maintenance of homeostasis, growth, cellular differentiation, and apoptosis in multicellular organisms. Mutations in many of these genes or their deregulation lead to a wide spectrum of pathologies including cancer (Zwick et al. 2001; Gschwind et al. 2004; Schlessinger 2014). Up to 58 RTK genes have been described in the human and mouse genomes (Robinson et al. 2000).

All RTKs have in common an intracellular C-terminal tyrosine kinase domain (TKD) connected to an alpha helical transmembrane domain. RTKs can be subdivided into 20 subfamilies. Among them, the Lmr (Lmtk/Aatk) and Styk1/Nok subfamilies have been identified as members of the RTK family despite having only a short amino-terminus domain on the outer cellular surface, which consists of a

signal peptide-like sequence (Liu et al. 2004; Ding et al. 2012; Inoue et al. 2014). Each of the other 18 subfamilies has a distinct N-terminal part characterized by modular structural domains exposed to the outer cell surface. Ligand binding to the extracellular domain induces receptor homo- or hetero-dimerization. RTKs propagate an intracellular response by phosphorylation of intracellular target proteins that promote various cell functions through the activation of several signaling pathways (Lemmon and Schlessinger 2010; Annenkov 2014). As an example, Erbb receptors regulate through Akt, Fak, Mapk, Src, and other pathways cell proliferation, growth, shape, differentiation, migration, apoptosis, and cell motility (Hubbard and Miller 2007; Yarden and Pines 2012).

Gene and whole genome duplications, which increase gene copy number at different scales, are considered to be important contributors to the evolution of organisms (Bridges 1936; Stephens 1951; Ohno 1970). WGD is a way to favor evolutionary changes, with duplicated genes evolving faster than singletons (Kondrashov et al. 2002; Jaillon et al. 2009). Gain and loss of members among gene families have been correlated with the acceleration of gene evolution (Chen et al. 2010).

During the early evolution of vertebrates, two successive rounds of whole genome duplications (WGD), called 1R and 2R (1R/2R), have highly influenced the organization and the gene content of genomes (Ohno 1999; Hughes 1999; Li et al. 2001; McLysaght et al. 2002; Dehal and Boore 2005; Kasahara 2007; Holland et al. 2008; Hughes and Liberles 2008; Putnam et al. 2008; Smith et al. 2013). Of note, an alternative evolutionary model has been recently proposed, in which the second round of WGD would rather correspond to several successive events of large segmental duplications (Smith and Keinath 2015). After WGDs, genomes returned progressively to a diploid state, but a certain amount of genes were kept as duplicates, which led to the formation and the extension of many gene families (Blomme et al. 2006). As such, several genes that are present in one copy in non-vertebrate species can be found with two to four copies for instance in tetrapods, depending on the extent of gene duplicate retention during the rediploidization process. In addition, subsequent small-scale duplications (SSD) and further WGD events have expanded the gene repertoire in a lineage-specific manner within vertebrates. For instance, a "teleost fish-specific" genome duplication called the 3R occurred at the basis of the teleost fish lineage (Meyer and Schartl 1999; Woods et al. 2000; Taylor et al. 2003; Christoffels et al. 2004; Jaillon et al. 2004; Naruse et al. 2004; Brunet et al. 2006; Kassahn et al. 2009). Another, more recent round of WGD has been characterized in salmonids (Berthelot et al. 2014). At the genome level, the footprints of WGDs are represented by the presence of multiple large duplicated regions of similar evolutionary age called the paralogons. WGD-duplicated chromosomes and genes are termed "ohnologs" in reference to Susumu Ohno, a pioneer of the genome duplication hypothesis (Ohno 1970, 1999; Wolfe 2000; Turunen et al. 2009).

The age of the teleost-specific 3R WGD has been estimated using different methods. Based on the analysis of gene content in various fish species at different key taxonomic positions, this event is supposed to have taken place very early at the base of the teleost species, with estimations ranging from 225 to 316 mya (Hoegg et al. 2004; Crow et al. 2006; Hurley et al. 2007; Douard et al. 2008; Ogino et al. 2009; Santini et al. 2009). Hence, the teleost-specific WGD event occurred much more recently than the 1R/2R WGDs that took place more than 500 mya before the lamprey/gnathostome split (Holland et al. 2008; Smith et al. 2013; Decatur et al. 2013).

Phylogenetic analyses of some RTK genes have been previously done (Hanks et al. 1988; Gu and Gu 2003; Leveugle et al. 2004) for the Egfr subfamily (Volff and Schartl 2003; Gómez et al. 2004; Liu et al. 2013), the Vegfr subfamily (He et al. 2014), and the Fgfr subfamily (Suga et al. 1999). Some RTK genes are known to have two copies in teleost fish genomes including *fgfr1* (Rohner et al. 2009), *kit/csf1r*, and genes of the Pdgfr subfamily (Mellgren and Johnson, 2005; Braasch et al. 2006; Siegel et al. 2007), as well as RTKs with immunoglobulin-like domains (Grassot et al. 2006). A more systematic survey of RTKs among tetrapods and teleosts was recently presented (Schartl et al. 2015). However, a complete evolutionary analysis of the full repertoire of RTK genes in vertebrate genomes and beyond has not been provided yet.

A large amount of genes within the RTK family in human, mouse, and other vertebrates raise several main questions concerning their evolutionary dynamics. First, what are the evolutionary relationships between members within subfamilies and between different subfamilies? Inherent to this question is the relatedness between RTK and CTK genes. Second, to what extend have the several rounds of WGDs and/or more local events of gene duplications contributed to the expansion of the RTK gene family? Third, did lineage-specific RTK gene loss occur during vertebrate evolution after WGDs? We tackled these questions through both synteny analyses and molecular phylogeny reconstructions particularly using the intracellular TK-conserved domain. This allowed us to analyze the evolutionary origin and dynamics of the RTK gene repertoire among jawed vertebrates. We show that the RTK gene family has been massively expanded through WGDs in vertebrates, with the 3R WGD having increased for more than one-third the number of RTK genes in teleost fish.

Materials and Methods

Species and Databases

Protein sequences of the RTK families were retrieved in vertebrates from four databases: Ensembl (v70; ensembl.org), Uniprot (SwissProt + TrEMBL; uniprot.org), nr-prot (NCBI; ncbi.nlm.nih.gov), and the elephant shark genome project

(esharkgenome.imcb.a-star.edu.sg/; Venkatesh et al. 2014). Sequences from representative species were aligned using MAFFT with automatic search of the most appropriate algorithm (Katoh and Standley 2013). In a first screening set, we selected proteins that aligned best over the whole protein sequence for each RTK genes. For each species and for each gene, ad hoc scripts and manual selection were used to keep the best annotated sequences in those alignments. To ascertain the number of RTK genes in the genome of the common ancestor of the vertebrates (non-duplicated), as well as to root the built phylogenies, we also used sequences from non-vertebrate deuterostomes, belonging either to the superphylum Chordata, that is, sea squirts (*Ciona intestinalis* and/or *C. savignyi*: Urochordata/Tunicata) and amphioxus (*Branchiostoma floridae*: Cephalochordata), or to the superphylum Ambulacraria, that is, purple sea urchin (*Strongylocentrotus purpuratus*: Echinodermata) and acorn worm (*Saccoglossus kowalevskii*: Hemichordata) (Satoh et al. 2014). At the end, 5,181 sequences were collected (supplementary table S1, Supplementary Material online). From this dataset, we added 2,466 sequence entries from the Genomicus database (Louis et al. 2012; supplementary table S1, Supplementary Material online). At the time of analysis, data from myxine and lamprey were too incomplete to be included in the study.

Phylogenetic Analyses

Phylogenetic analyses were done using the Maximum-Likelihood method (PhyML v2.4.4; Guindon et al. 2010) with 1,000 bootstrap iterations each. ML trees were run using the WAG matrix for aa substitutions with estimated gamma distribution parameter. Full protein sequence alignments are provided in the following website: http://igfl.ens-lyon.fr/equipes/j.-n.-volff-fish-evolutionary-genomics/brunet_additional_data/Brunet_RTK_Additional_Data_Alignments.tar.gz/view

Synteny Analyses

In addition to phylogenetic analyses, orthology relationships between genes were verified using Genomicus (Louis et al. 2012), which allows the visualization of a conserved synteny in the near vicinity of a gene of interest. When phylogenies showed patterns compatible with WGD, information about gene locations onto chromosomes were used and paralogy relationships were determined based on the synteny information. To this end, we used both the Genomicus synteny viewer (Louis et al. 2012) and designed a tool to visualize the synteny information for orthology and paralogy at the chromosome level as in Jaillon et al. (2004). Synteny analyses were done for fish species with chromosomal information: the zebrafish *Danio rerio* (version Zv8) (Cypriniformes, Ostariophysi) and three Percomorpha (Acanthomorpha) including the medaka *Oryzias latipes* (v. MEDAKA1), the stickleback *Gasterosteus*

aculeatus (v. BROADS1), and the green spotted puffer *Tetraodon nigroviridis* (v. TETRAODON8).

Intron and Phase Position Analysis in RTKs and CTKs

Intron gains and losses are very rare events exposed to selective pressures stronger than aa substitutions and can, therefore, be used as evolutionary informative parameters (de Souza et al. 1998) as long as their positions along the region have not been too much drifted apart by shrinkage or extension of sequences. Intron occurrences and phase positions in the human RTK and CTK genes were retrieved from the Ensembl database. Sequences were aligned using clustal omega (Sievers and Higgins 2014). An ad hoc program was designed to insert the phase positions in each of the aligned proteins. The Neighbor-Joining (BioNJ) method (Lajoie et al. 2007) implemented in SeaView (Gouy et al. 2010) was used to make a phylogeny based only on the observed intron phases inside the TK domain alignment.

Results

Evolutionary Relationships among Vertebrate Tyrosine Kinases

We first investigated the evolutionary relationships among the known 58 RTKs and 32 CTKs in human (Robinson et al. 2000; Lemmon and Schlessinger 2010) through phylogenetic analysis of the TK domain (fig. 1). Neither the RTKs nor the CTKs formed clearly distinct monophyletic groups, and many links between RTK and CTK subfamilies were only supported by low bootstrap values. Hence, the phylogeny obtained was not robust enough to establish with certainty a scenario explaining the evolutionary switch between RTKs and CTKs through the gain versus loss of extracellular and transmembrane domains. Neither synteny analysis nor the study of intron phase and position along the TK domain revealed significant additional information to assess relationships between RTKs and CTKs (fig. 1, supplementary figs. S1 and S2, Supplementary Material online).

Within RTKs, some of the 20 subfamilies are grouped together in the TK domain molecular phylogeny with significant bootstrap values (fig. 1). Phylogenetic clustering was found for the Tie/Fgfr/Ret/Vegfr/Pdgfr, Met/Ryk/Tam, Alk/Ros1/Insr, and Ddr/Ror/Musk/Trk subfamilies, as well as between Lmr and Styk1 but with a lower bootstrap value. The Ephrin receptors (Eph, subdivided in EphA and EphB) and the Erbb subfamily formed distinct subgroups. Ptk7 was not particularly related to any other RTK.

The RTK subfamily clustering obtained using the TK domain sequence was confirmed by a phylogenetic analysis based on the intron characteristics (supplementary figs. S1 and S2, Supplementary Material online). In addition, 13 out of the 20 RTK subfamilies, including the Tie/Fgfr/Ret/Vegfr/Pdgfr, Met/Ryk/Tam, and Alk/Ros1/Insr groups as well as Lmr and
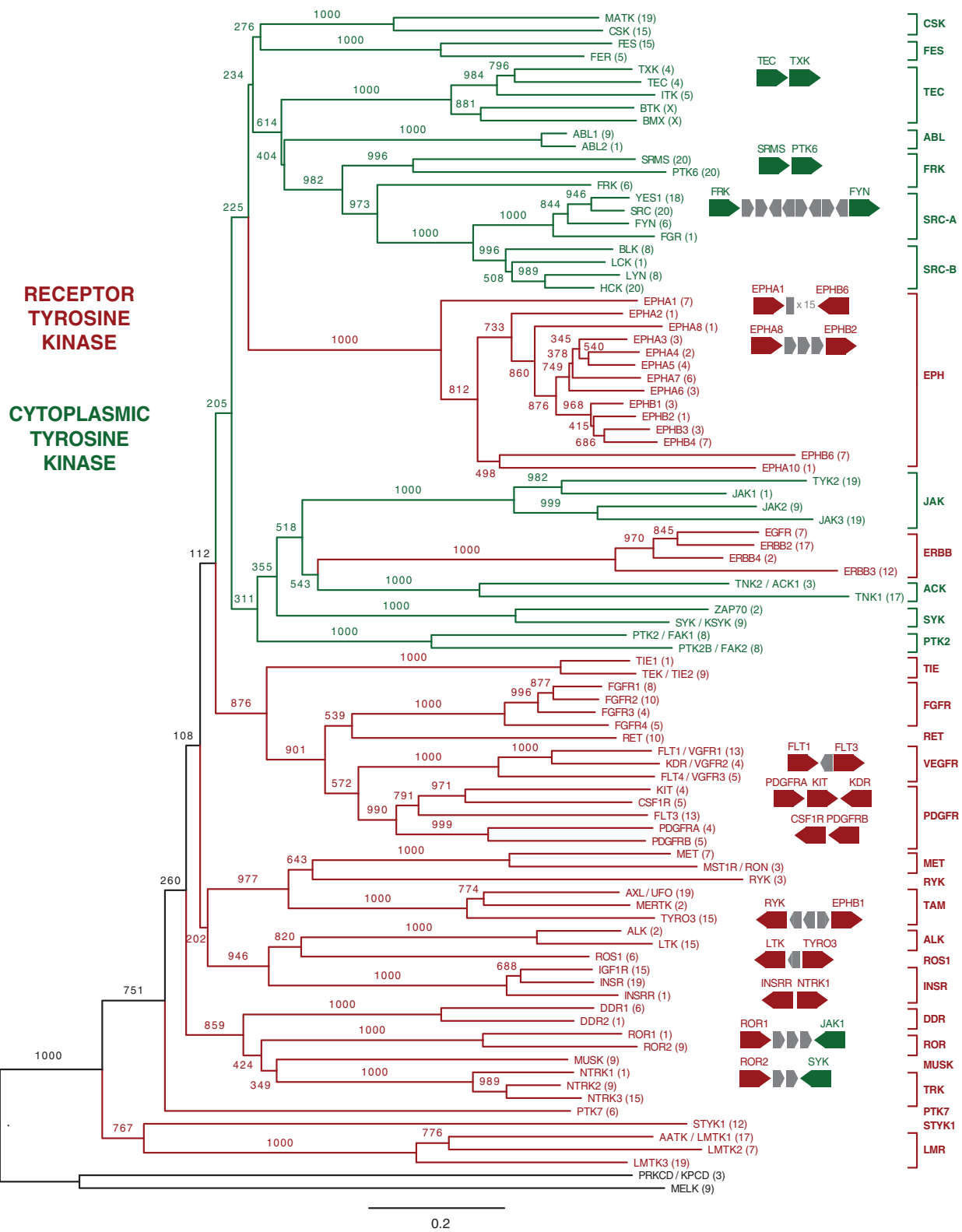
Fig. 1.—Phylogenetic analysis by Maximum-Likelihood of all trans-membrane receptor tyrosine kinases (in red) and cytoplasmic non-receptor tyrosine kinases (in green) described in human. Alignment is based on the common tyrosine kinase domain and these proteins are rooted by PRKCD and MELK kinases. Bootstrap replicates: 1,000. Name of gene families are given on the right column of the tree. Genes that are in close vicinity in the human genome are also indicated on the right using the same color code; other interspaced genes are in grey.

Styk1 are all joined by one phase-2 intron, suggesting a common origin of all these genes. Styk1 is more particularly associated with the Vegfr/Pdgfr/Ret/Fgfr/Tie subfamilies and shares with them three introns. In this group of genes, all but the Tie genes are characterized by a TK domain subdivided in two portions (Lemmon and Schlessinger 2010). The *Lmr* gene shares common introns with *Alk/Ros1/Insr*. With their tiny extracellular domain compared with the long and distinct ones characterizing all other RTKs, Lmr, and Styk1 may not be considered as *bona fide* RTKs, and their phylogenetic position in the TK domain tree suggests that they are indeed divergent (fig. 1; the human protein kinome database, www.kinase.com/human/kinome; Manning et al. 2002). However, they have a transmembrane domain and their kinase domain shares strong sequence similarity with other RTKs. Intron-based homologies suggest that Lmr and Styk1 derive from RTKs; the shortness of their extracellular region probably reflects secondary reductions or losses of the extracellular receptor domain.

Within the Eph subfamily, EphA and EphB, which were not clearly separated by the molecular phylogeny of the TK domain sequence, could be distinguished by one intron (*EphA* has an additional phase-1 intron compared with the *EphB*).

The RTK Repertoire of Jawed Vertebrates Has Been Extensively Shaped by Ancestral WGDs

We analyzed 7,376 RTK genes (supplementary table S1, Supplementary Material online) that were retrieved from 143 species covering all major clades among jawed vertebrates. Since only few RTKs were described in some species, we concentrated on the 47 species with the largest set of RTKs (fig. 2), but added information from the other species for confirmation (supplementary table S1, Supplementary Material online). Particularly, related species were analyzed when available to differentiate true lineage-specific absence of a gene from incomplete genome assemblies or gene annotations. Evolutionary relationships between genes were assessed by protein sequence phylogeny (supplementary fig. S3, Supplementary Material online) and synteny analysis (supplementary fig. S4, Supplementary Material online). A total of 63 RTK genes representing 20 subfamilies were found in non-teleost-jawed vertebrate species (including spotted gar and elephant shark).

To assess the vertebrate-specific evolution of the RTK repertoire, we screened the available non-vertebrate deuterostome genomes, that is, amphioxus, sea squirt, acorn worm, and sea urchin species, using human RTK genes as queries. With the exception of Eph (several genes in non-vertebrates) and Pdgfr (no clear orthologous sequence in non-vertebrates, but see below), 18 out of 20 RTK subfamilies generally had a single representative gene among non-vertebrate deuterostomes (fig. 3 and supplementary fig. S5, Supplementary Material online). For Met, Tie and Eph, lineage-specific duplications were found in the sea squirt and/or the amphioxus.

Vertebrate-specific expansion was observed for most RTK families (fig. 3). Only five RTK subfamilies out of 20 contain one single gene: *Ryk*, *Ros1*, *Musk*, *Ptk7*, and *Ret*. Another five subfamilies are constituted of two genes: Met (*Met/Mst1r*), Alk (*Alk/Ltk*), Ror (*Ror1/Ror2*), Tie (*Tie1/Tek*), and Styk1 (*Styk1/Styk1*-like). Four subfamilies are formed by three genes: Insr (*Insr/Igfr/Insrr*), Ddr (*Ddr1/Ddr2/Ddr2*-like), Trk (*Ntrk1/Ntrk2/Ntrk3*), and Lmr (*Aatk/Lmtk2/Lmtk3*). Four subfamilies have retained the full set of four genes: Erbb (*Egfr/Erbb2/Erbb3/Erbb4*), Tam (*Tyro3/Axl/Axl*-like/*Mertk*), Vegfr (*Flt1/Kdr/Kdr*-like/*Flt4*), and Fgfr (*Fgfr1/Fgfr2/Fgfr3/Fgfr4*). The Pdgfr subfamily is composed of five genes and is subdivided in two monophyletic groups made of *Csf1r/Kit/Flt3* and *PdgfrA/PdgfrB* (fig. 3). Finally, the Eph subfamily contains as many as 15 genes (see below). The observed expansion of the RTK repertoire at the base of the vertebrates is consistent with the involvement of the 1R/2R-WGDs. Up to four copies have been maintained in about 75% of the RTK subfamilies, depending on the rate of gene retention after rediploidization.

Taken together, without considering the Eph gene subfamily, which has a rather complex evolutionary history, and *Insrr*, which might be the result of a segmental duplication (see below), and if we assume that the Vegfr/Pdgfr subfamilies have been formed from three pre-WGD genes (see below), the retention rate of RTK genes after 1R/2R is 58.75% [47/(20*2*2)]. As a comparison, 31 CTK genes have been kept after 1R/2R duplications of 11 (or 12) subfamily progenitors, with a retention rate of 70.45% (or 64.5%).

After 1R/2R-WGDs, lineage-specific gene losses contributing to differences in RTK repertoire occurred in different groups of vertebrates. *Axll* was lost in tetrapods, *EphA4l* and *Styk1l* in amniotes (sauropsids and mammals), and *Ddr2l* in mammals. The fourth member of the Vegfr family, that we named *Kdrl* (*Kdr*-like) according to the synteny data and the phylogeny of the Vegfr subfamily, was not maintained in eutherians. More restricted gene losses were also detected (for example, the *Ltk* gene in Carnivora).

Two RTK Subfamilies Have Also Evolved by Ancestral Local Duplications

The Pdgfr (*Flt3/Kit/Csf1r*, *PdgfrA/PdgfrB*) and Vegfr (*Flt1/Kdr/Flt4*) subfamilies are phylogenetically related and very similar in structure, with five and seven immunoglobulin (Ig) domains characterizing the extracellular part of the proteins (Robinson et al. 2000; Lemmon and Schlessinger 2010). Strikingly, genes from the Pdgfr subfamily are clustered with genes from the Vegfr subfamily in human and other vertebrates. *Csf1r/PdgfrB* are organized in tandem, and *Kdr/Kit/PdgfrA* are clustered without any other intervening gene (fig. 4). *Flt1* and *Flt3* are neighbors and only separated by one (non-PTK) gene (*Pan3*). *Flt4* is on the same chromosome as the *Csf1r–PdgfrB* tandem.
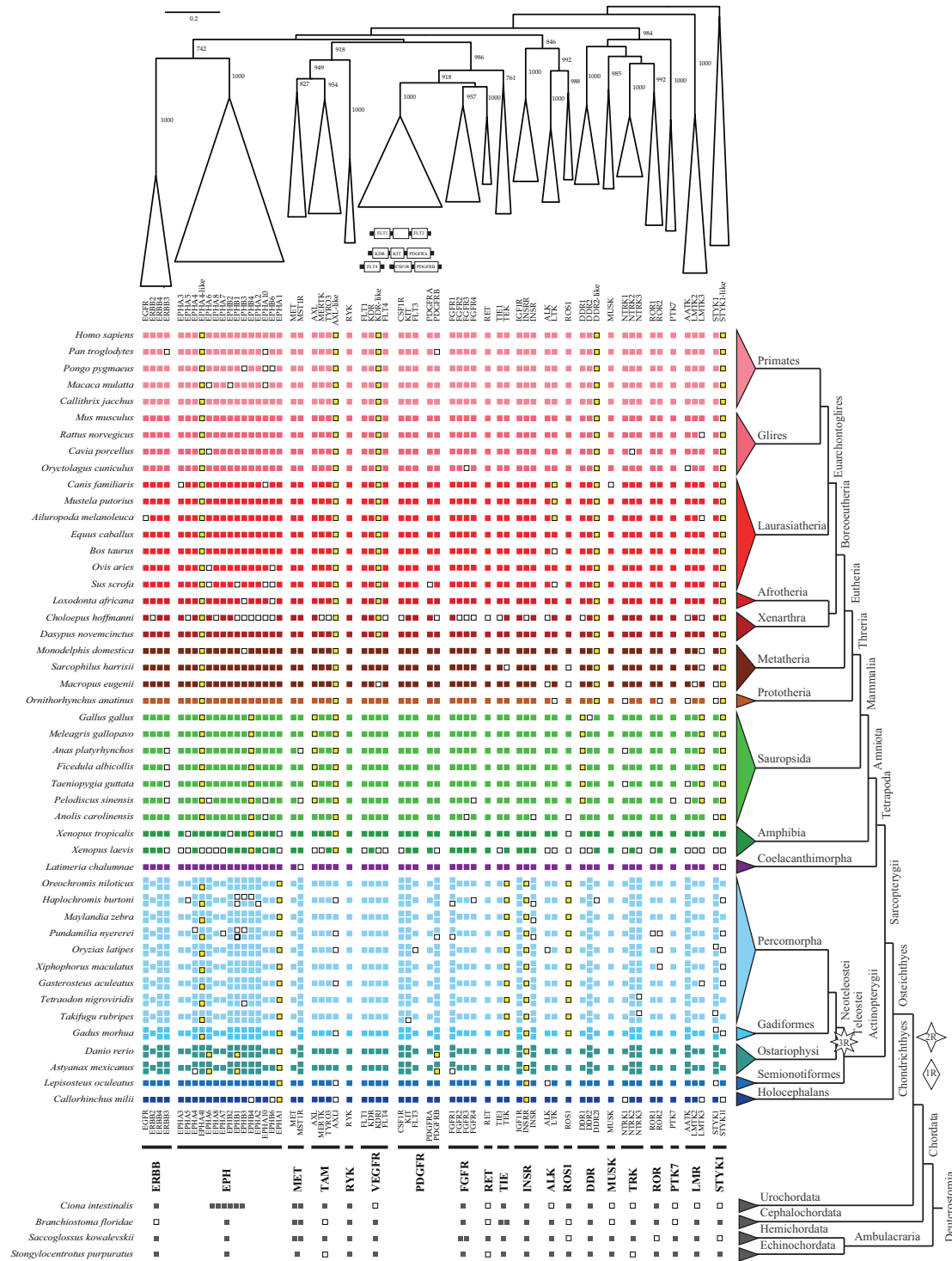
Fig. 2.—Schematic representation of the occurrence of receptor tyrosine kinases in 47 representative vertebrates species. A white box is used when no sequence, even partial, was found in a species, which is named on the left. Yellow boxes refer to lack of a gene in a taxonomic group of species, which are shown on the right. Other plain colored boxes are used when a sequence, even partial was found. Duplicated genes from the teleost specific whole genome duplication are shown by double-squares. The top phylogeny refers to a Maximum-Likelihood phylogenetic analysis of the RTKs found in human and in other vertebrate species for those that were lost in mammals (*ephA4*-like, *axl*-like, *ddr2*-like, *kdr*-like, and *styk1*-like, see details in supplementary fig. S3, Supplementary Material online). All collapsed subfamilies are rooted by non-vertebrate deuterostomes and have a significant bootstrap value. Linkage of the genes tandemly duplicated in the vegfr and pdgfr subfamilies are represented.

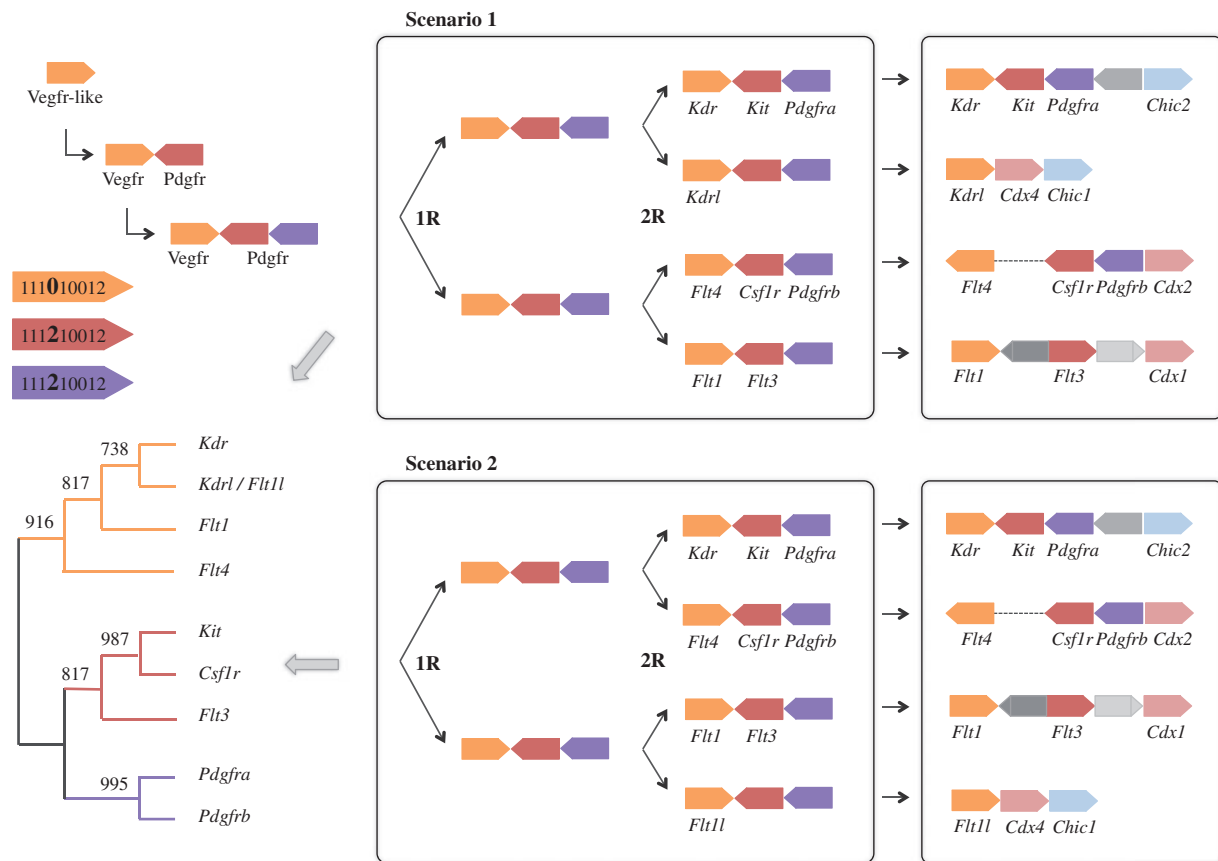FIG. 3.—Phylogenetic analysis by Maximum-Likelihood of all trans-membrane receptor tyrosine kinase genes observed in human (in red) and those lost in human but found in other species (in blue) (*ephA4*-like, *axl*-like, *ddr2*-like, *kdr*-like, and *styk1*-like in red). These genes have been aligned with RTK proteins found in non-vertebrate deuterostomes (in black), which root all RTK subfamilies except the *pdgfr* subfamily. Bootstrap replicates: 1,000. Only major bootstrap values at key phylogenetic nodes were kept.

Fig. 4.—Possible scenarios proposed for the evolution of the vegfr and pdgfr subfamilies. SSD in tandem initiated the amplification of this subfamily. The sequences of their intron phases show the SSD events. Scenario 1 supports partially the phylogeny of vegfr, the most ancient genes. Members of other gene families, Cdx and Chic, in synteny with vegfr and pdgfr subfamilies endorse this scenario. The phylogeny of *Kit/Csf1r/Flt3* favors scenario 2. Since the Vegfr are the ancestral genes, and because both syntenic data and phylogeny support the Scenario 1, the gene lost in eutherians should be the named *Kdrl*.

There is only one gene found in non-vertebrate deuterostomes that roots the Vegfr gene subfamily, and none was identified for the Pdgfr subfamily (fig. 3 and supplementary fig. S5, Supplementary Material online). *Pdgfr* genes have the same intron phase and position structure (111210012), which is different by only one change of intron phase and position from the *Vegfr* genes (111010012) (supplementary fig. S1, Supplementary Material online).

Putting together, these observations suggest that the ancestor of the *Pdgfr* genes was duplicated head-to-head in tandem from the ancestor of the *Vegfr* genes, very early at the base of vertebrates, before the two ancestral WGDs (fig. 4). *PdgfrA* and *PdgfrB* genes are more related to each other than to the three other members of the Pdgfr subfamily (figs. 1 and 2). This suggests that the ancestral *PdgfrA/B* gene was generated from another tandem duplication, tail-to-head this time, from the previously duplicated copy. Thus, the Vegfr and Pdgfr subfamilies probably originated from two tandem duplications of a single ancestor (fig. 4).

The Eph (EphA/B) subfamily presents a more complex picture and the evolutionary relationships between member genes are more difficult to disentangle. Up to 15 genes have been detected in vertebrates. In addition, Eph's are also duplicated in other chordate lineages, with at least six copies in the sea squirt *Ciona* and two in amphioxus, and only one Eph gene is present in the more distantly related Ambulacraria superphylum represented by the sea urchin and the acorn worm (figs. 2 and 3 and supplementary fig. S5, Supplementary Material online). This suggests either independent lineage-specific expansion of the Eph gene family in different groups of chordates or the existence of duplications in chordate ancestors prior to the 1R/2R-WGDs. No obvious case of tandem duplication was detected in this subfamily (fig. 1). Within *EphB* genes, which can be clearly distinguished from *EphA* genes by one phase-1 intron, *EphB1/EphB2/EphB3/EphB4* form a strongly supported monophyletic group, in which *EphB6* is not included (fig. 3). Interestingly, the *EphB1/EphB2/EphB3/EphB4* genes are more related to *Ciona* sequences than to vertebrate *EphB6*. This suggests an event of gene duplication before the urochordate/vertebrate split, with subsequent 1R/2R-mediated duplications of the *EphB1/EphB2/EphB3/EphB4* progenitor in vertebrates.

Other genes might also have been generated by local gene duplications. For example, the *EphA1* and *Insrr* genes are found in sarcopterygian species including the coelacanth, but neither in actinopterygians nor in the elephant shark. A simple explanation for this distribution implies local duplication events at the base of sarcopterygians, even if loss of the genes in cartilaginous and ray-finned fishes but maintenance in coelacanth/tetrapods after 1R/2R-WGDs cannot be excluded.

The RTK Repertoire of Teleost Fish Has Been Shaped by the 3R WGD

Teleost fish genomes have been shaped by a third WGD called the 3R. The fish investigated here are two Ostariophysi (the cave fish, a characiform and the zebrafish, a cypriniform) and nine Percomorpha species, which together with the cod belong to the Neoteleostei. The genome of the gar *Lepisosteus oculeatus* was also analyzed. This holostean fish diverged from the fish lineage that includes the teleosts before the 3R WGD, and is expected to define the set of genes that were present before the teleost WGD (Amores et al. 2011).

According to the RTK gene set found in the gar, the coelacanth, and the elephant shark, 61 RTK genes probably existed before the 3R event in the lineage leading to teleosts. Among the 14 Eph genes, eight duplicates were retained in all fish from the 3R-WGD up to the split between Neoteleostei and Ostariophysi (fig. 2). One duplicate of *EphA4l* was then lost in the Neoteleostei, and one duplicate was not maintained in the two representatives of the Ostariophysi for *EphA6b* and *EphB1b*. Among the 47 remaining RTK genes, 15 were duplicated in teleosts. Ros1 and Tek were subsequently lost in the Neoteleostei, and one *PdgfrB* duplicate is lacking in the Ostariophysi.

Taken together, 23 out of 61 RTK genes duplicated by the 3R WGD were kept in a least one major group of teleosts. This makes 84 RTK genes in the representatives of the teleosts used in this analysis, with a retention rate of 68.8% [84/(61*2)]. These values are as high as those obtained for the two ancestral vertebrate 1R/2R-WGDs. Hence, teleosts have a considerably larger repertoire of RTKs than tetrapods.

Discussion

Evolutionary Relationships within the Vertebrate PTK Repertoire

The large protein tyrosine kinase gene family (RTKs and CTKs, the transmembrane receptor tyrosine kinase and the cytosolic tyrosine kinase, respectively) has diversified specifically in animals (Hanks and Hunter 1995), with an initial repertoire that could have played a key role in the origin of metazoans (Suga et al. 2014). An increase in subfamilies has been proposed to correlate with the split between protostomes and deuterostomes (Iwabe et al. 1996; Suga et al. 1997; Srivastava et al. 2010).

In our analysis, we could not find any evidence of monophyly for either CTKs or RTKs. Our molecular phylogeny based on the TK domain was not robust enough to precisely assess the ancestral state of the PTK family (CTKs or RTKs) and to determine the dynamics of the evolutionary switch between CTKs or RTKs (single or multiple events of gain vs. loss of transmembrane and extracellular domains).

Within RTKs, we could identify three major phylogenetic groups. Using a combination of protein sequence molecular phylogeny and analysis of intron phases and positions, a large phylogenetic group emerged that included as many as 17 out of 20 RTK families. We could add into this group the Styk1 and Lmr subfamilies, which encode proteins with a very short extracellular domain, suggesting reduction or loss of the classical extracellular domain. The Ephrin receptor and Erbb subfamilies form two more divergent RTK groups, according to previous analyses (e.g., Manning et al. 2002; Suga et al. 2008; Robinson et al. 2010). In our phylogenetic analysis, both subfamilies were nested among the CTKs, albeit with low bootstrap values, suggesting possible evolutionary links with CTKs.

High Retention Rate of Duplicates after WGDs Expanded the Vertebrate RTK Repertoire

The 1R/2R WGDs that occurred at the base of the vertebrates have contributed to a large amplification of RTK genes, with a degree of duplicate retention depending on the RTK subfamily. Subsequently, the 3R had the same effect in teleost fish. In addition, more local events of gene duplications were also involved, as exemplified by the tandem duplication events that led before 1R/2R to the formation of the Pdgfr and Vegfr subfamilies. Thus, without considering the peculiar evolutionary history of the 15 Eph genes, the number of RTK in vertebrates increased from 20 to 48. Including the Eph genes, there is a total of 63 different RTK genes represented in all major vertebrate sublineages. After 3R, the RTK repertoire was even expanded to as many as 84 genes in teleost fish.

Similar retention rates were observed for the RTK genes after 1R/2R in vertebrates (58.75%) and after 3R in teleost fish (68.8%), in the range of what was observed for CTKs after 1R/2R. Various global duplicate retention rates after WGD have been obtained depending on the organism studied: about 8% in yeast (WGD 100 mya; Seoighe and Wolfe 1999), 72% in maize (WGD 11 mya; Ahn and Tanksley 1993; Gaut and Doebley 1997), and an average of 48% for Paramecium species (WGD 320 mya; McGrath et al. 2014). In fish and other vertebrates, duplicate retention was 48% in salmonids (WGD about 96 mya; Berthelot et al. 2014); 50% in castatomid fish (WGD 50 mya, Ferris and Whitt 1979), and 77% in Xenopus (WGD 30 mya; Hughes and Hughes, 1993).

Concerning the vertebrate WGDs we analyzed here, quite similar global rates of retention have been reported for 1R/2R (20–30%) (Makino and McLysaght 2010) and 3R (12–20%) (Postlethwait et al. 2000; Jaillon et al. 2004; Woods et al.

2005; Brunet et al. 2006; Steinke et al. 2006; Kassahn et al. 2009; Inoue et al. 2015). Clearly, RTK (and CTK) genes have a retention rate higher than average for both 1R/2R and 3R. This is consistent with the fact that genes that have a fair retention rate through the two rounds of ancestral vertebrate WGDs are on an average also prone to a high rate of persistence after the teleost 3R WGD (Howe et al. 2013; but see the Adamts genes for a counter-example, Brunet et al. 2015).

High retention rate of RTKs in vertebrates is consistent with the observation that duplicates maintained after WGDs show an excess of transcription factor genes and genes implicated in development and signal transduction in fish (Brunet et al. 2006; Steinke et al. 2006; Hufton et al. 2008; Kassahn et al. 2009), other vertebrates (Bertrand et al. 2004; Blomme et al. 2006), yeast (Davis and Petrov 2005), and plants (Blanc and Wolfe 2004; Seoighe and Gehring 2004; Maere et al. 2005; Paterson et al. 2006). The maintenance of specific gene classes as duplicates has often been associated with increased levels of cell complexity and subsequently with an increase in organismal diversity (Maere et al. 2005; Freeling and Thomas 2006; Sémon and Wolfe 2007; Huminiecki and Heldin 2010). Singh et al. (2012) rather proposed that ohnolog genes prone to autosomal-dominant deleterious mutations should be retained more intensely by purifying selection, RTKs and CTKs being among the top list. Because of their implication in cancers and genetic disorders, they called them "dangerous" genes. However, not all RTKs were retained as duplicates. There are five families with only one gene in the non-teleost vertebrates: *Musk*, *Ptk7*, *Ret*, *Ros1*, and *Ryk* that were noticeably also kept as singletons after the 3R-WGD. Those five genes do not show any obvious common pattern in biological processes differentiating them from other RTKs which might explain why they might be refractory to duplicate retention.

The Vertebrate RTK Repertoire also Evolved through Segmental Gene Duplications

Single gene duplications have also shaped the vertebrate RTK repertoire. One well-described example is the *Xmrk* gene, which is found only in some *Xiphophorus* fish species. This RTK gene, which possesses a melanoma-inducing oncogenic activity in *Xiphophorus* hybrids, has been formed through a segmental duplication of the *Egfrb* gene. This, added to the WGD-mediated expansion of the Erbb subfamily in teleost, increased the number of Erbb genes to eight among several species of the genus *Xiphophorus* (Volff et al. 2003; Gómez et al. 2004). In our analysis, we did not take into account those events of single gene duplications when they have occurred in terminal branches of clades.

However, we could demonstrate that ancestral segmental duplications have also shaped the vertebrate RTK repertoire before the WGDs. This is the case for the *PdgfrA-Kit* and *PdgfrB-Csf1r* gene pairs that are linked to the ParaHox clusters

(Rousset et al. 1995; Suga 1999; Williams et al. 2002; Gu and Gu 2003; Leveugle et al. 2004; Braasch et al. 2006; Siegel et al. 2007; Kottler et al. 2013).

In our analysis, we could reconstruct the evolutionary history of the Vegf and Pdgfr RTK subfamilies (fig. 4). An ancestral head-to-head duplication of a *Vegfr*-like gene occurred before the 1R-WGD, which led to the emergence of the *Vegf* and *Pdgfr* genes. This event most likely occurred after the vertebrate ancestor split off from the non-vertebrate species including amphioxus and sea squirt, since only one *Vegf*-like gene is found in these species. In the meantime, an intron change turning the phase-0 into a phase-2 intron occurred. This was followed by a second ancestral tandem duplication, this time tail-to-head, of the *Pdgfr* ancestor. This cluster of three genes was then duplicated by the 1R/2R-WGDs, followed by several gene loss events. Two possible evolutionary scenarios are proposed, with different paralogy relationships between duplicates (fig. 4). Scenario 1 is supported by the phylogeny of Vegfr and by the presence of other putative 2R paralogs within the postulated 2R paralogons (*Chic1* and *Chic2* for *Kdr-Kit-Pdgfra* vs. *Kdrl*, and *Cdx1* and *Cdx2* for *Flt4-Csf1r-Pdgfra* vs. *Flt1-Flt3*). The more basal phylogenetic position of *Flt3*, which has changed its gene orientation after WGD, might be explained by a higher rate of evolution. Alternatively, 2R paralogons might have been formed by *Flt1-Flt3* versus *Flt1l*, and *Kdr-Kit-Pdgfra* versus *Flt4-Pdgfrb*, this being supported by the phylogenetic relationship between *Kit* and *Csf1r*.

Lineage-Specific Diversity of the RTK Repertoire also Occurred by Gene Loss

Several RTK gene losses were recorded all along the vertebrate evolutionary tree. This is the case for *Axll* that was lost in the tetrapod lineage, and *EphA4l* and *Styk1l*, which were lost in all amniotes. *Ddr2l* and *Kdrl* were not maintained in mammals and eutherians, respectively. Of note, these five genes are those missing in human and mouse. Other lineages have lost some RTK genes like *EphB4*, which is missing in both sauropsids and amphibians. This shows that some genes may have a higher propensity to become useless and may have been lost independently in two lineages. Birds have lost three RTK genes: *Axl*, *Ddr1*, and *Lmtk3* from the Tam, Ddr, and Lmr subfamilies, respectively.

Why could such primordial RTK genes be dispensable in some species? A hint may be found in looking at their function. In mammals, Tam receptors do not play any essential role in embryonic development but are more involved in homeostatic regulations in adult tissues and organ systems. Knockout mice for *Ddr1* present multiple morphological alterations and reproductive defects (Leitinger 2014). In contrast, $Tyro3^{-/-}$, $Axl^{-/-}$, or $Mer^{-/-}$ mice are viable and fertile, as is the triple knockout of these Tam subfamily genes (Lemke, 2013). $Lmtk3^{-/-}$ mice present an increase in locomotor activity and

reduced anxiety-like behavior and decreased depression-like behavior (Inoue et al. 2014). Thus, the lack of two of these three missing genes in birds is not lethal in mouse. This might be consistent with the hypothesis of an adaptive loss in birds. Expression of these three genes (*Axl*, *Ddr1* and *Lmtk3*) is reported in the mammalian brain, and thus may motivate further investigation on the role of these losses in the evolution of the bird lineage.

With the additional 3R-WGD, RTKs rose up to a total number of 84 RTK genes in teleost fish, with at most 81 present in each of the studied species. It has been noted that the human *EphA1* and *Insrr* have no counterpart in zebrafish (Challa and Chatti 2013). Here, we extend this finding and show that these gene losses characterize all fish genomes available to date, including teleosts and the more basal lineages. Among the Teleleostei, Ostariophysii have experienced some distinct losses compared with the Neotelostei (Percomorpha and Gadiformes). *EphA6b*, *ephB4b*, and *pdgfrBb* are present neither in zebrafish (Challa and Chatti 2013) nor in cavefish, extending these features to be common for the Ostariophysi lineage. *EphA4lb*, *tek*, and *ros1* are missing in Neoteleostei. These consecutive losses may indicate that the functions were at some point dispensable in some teleost sublineages. Alternatively, some of the other duplicated genes might have grasped the functions of the deleted genes.

The coelacanth harbors the largest RTK repertoire, possibly best reflecting RTK gene content in the last common ancestor of fish and tetrapods. This feature could be associated with the weak genetic diversity observed in this species (Lampert et al. 2012) coupled with the low substitution rate as a whole or among protein sequences (Ameniya et al. 2013; Nikaido et al. 2013). Ameniya et al. (2013) inferred the loss of 55 genes in tetrapods compared with their Coelacanth data. Many of these genes are involved in important developmental processes that imply genes as important as components of signaling pathways (Fgf, Wnt, and Bmp) or transcription factors. These losses could specify some critical morphological transition (Ameniya et al. 2013).

Conclusion

Positive correlations between paralogous gene enrichment and cell or tissue diversification as observed in metazoan species was proposed to explain the emergence of multicellularity (Lynch and Conery 2003). In agreement, our results show that RTK amplifications occurred during the vertebrate WGDs, potentially providing the basis for specific evolutionary innovations. As RTKs are involved in cell proliferation, survival, adhesion, migration, and differentiation and more generally development through regulatory networks, our work provides a baseline for comparative expression and function analysis towards a better understanding of the diversification of cell metabolism and the development complexity in vertebrates.

Supplementary Material

Acknowledgments

Literature Cited

Amemiya CT, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–316.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics 188:799–808.

Ahn S, Tanksley SD. 1993. Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci. 90:7980–7984.

Annenkov A. 2014. Receptor tyrosine kinase (RTK) signalling in the control of neural stem and progenitor cell (NSPC) development. Mol Neurobiol 49:440–471.

Berthelot C, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun. 5:3657.

Bertrand S, et al. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol Biol Evol. 21:1923–1937.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16:1679–1691.

Blomme T, et al. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. 7:R43.

Braasch I, Salzburger W, Meyer A. 2006. Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Mol Biol Evol. 23:1192–1202.

Bridges CB. 1936. The bar "gene" a duplication. Science 83:210–211.

Brunet FG, et al. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol. 23:1808–1816.

Brunet FG, et al. 2015. The evolutionary conservation of the A disintegrin-like and metalloproteinase domain with Thrombospondin-1 motif metzincins across vertebrate species and their expression in teleost zebrafish. BMC Evol Biol. 15:281.

Challa AK, Chatti K. 2013. Conservation and early expression of zebrafish tyrosine kinases support the utility of zebrafish as a model for tyrosine kinase biology. Zebrafish 10:264–274.

Chen F-C, Chen C-J, Li W-H, Chuang T-J. 2010. gene family size conservation is a good indicator of evolutionary rates. Mol Biol Evol. 27:1750–1758.

Christoffels A, et al. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol. 21:1146–1151.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. 2006. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol. 23:121–136.

de Souza SJ, et al. 1998. Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc Natl Acad Sci. 95:5094–5099.

Davis JC, Petrov DA. 2005. Do disparate mechanisms of duplication add similar genes to the genome? Trends Genet. 21:548–551.

Decatur WA, Hall JA, Smith JJ, Li W, Sower SA. 2013. Insight from the lamprey genome: glimpsing early vertebrate development via neuroendocrine-associated genes and shared synteny of gonadotropin-releasing hormone (GnRH). Gen Comp Endocrinol 192:237–245.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3:e314.

Ding X, Jiang QB, Li R, Chen S, Zhang S. 2012. NOK/STYK1 has a strong tendency towards forming aggregates and colocalises with epidermal growth factor receptor in endosomes. Biochem Biophys Res Commun. 421:468–473.

Douard V, et al. 2008. The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? BMC Evol Biol. 8:336.

Ferris S, Whitt GS. 1979. Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol. 12:267–217.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16:805–814.

Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci. 94:6809–6814.

Gómez A, Volff JN, Hornung U, Schartl M, Wellbrock C. 2004. Identification of a second egfr gene in Xiphophorus uncovers an expansion of the epidermal growth factor receptor family in fish. Mol Biol Evol. 21:266–275.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27:221–224.

Grassot J, Gouy M, Perrière G, Mouchiroud G. 2006. Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains. Mol Biol Evol. 23:1232–1241.

Gschwind A, Fischer OM, Ullrich A. 2004. The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat Rev Cancer 4:361–370.

Gu J, Gu X. 2003. Natural history and functional divergence of protein tyrosine kinases. Gene 317:49–57.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.

Hanks SK, Quinn AM, Hunter T. 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42–52.

Hanks SK, Hunter T. 1995. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. Faseb J. 9:576–596.

Hanks SK. 2003. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biol. 4:111.

He W, et al. 2014. Phylogenetic analysis and positive-selection site detecting of vascular endothelial growth factor family in vertebrates. Gene 535:345–352.

Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol. 59:190–203.

Howe K, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503.

Holland LZ, et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 8:1100–1111.

Hubbard SR, Miller WT. 2007. Receptor tyrosine kinases: mechanisms of activation and signaling. Curr Opin Cell Biol. 19:117–123.

Hufton AL, et al. 2008. Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res. 18:1582–1591.

Hughes AL. 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol. 48:565–576.

Hughes MK, Hughes AL. 1993. Evolution of duplicate genes in a tetraploid animal, Xenopus laevis. Mol Biol Evol. 10:1360–1369.

Hughes T, Liberles DA. 2008. Whole-genome duplications in the ancestral vertebrate are detectable in the distribution of gene family sizes of tetrapod species. J Mol Evol. 67:343–357.

Huminiecki L, Heldin CH. 2010. 2R and remodeling of vertebrate signal transduction engine. BMC Biol. 8:146.

Hurley IA, et al. 2007. A new time-scale for ray-finned fish evolution. Proc Biol Sci. 274:489–498.

Inoue J, Yukuto S, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after while-genome duplication in teleost fish suggested by mathematical modeling. Proc Natl Acad Sci. 112:14918–14923.

Inoue T, et al. 2014. LMTK3 deficiency causes pronounced locomotor hyperactivity and impairs endocytic trafficking. J Neurosci 34:5927–5937.

Iwabe N, Kuma K, Miyata T. 1996. Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. Mol Biol Evol. 13:483–493.

Jaillon O, et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431:946–957.

Jaillon O, Aury JM, Wincker P. 2009. "Changing by doubling", the impact of whole genome duplications in the evolution of eukaryotes. C R Biol. 332:241–253.

Kasahara M. 2007. The 2R hypothesis: an update. Curr Opin Immunol 19:547–552.

Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA. 2009. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res. 19:1404–1418.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. Genome Biol. 3:RESEARCH0008.

Kottler VA, Fadeev A, Weigel D, Dreyer C. 2013. Pigment pattern formation in the Guppy, Poecilia reticulata, involves the Kita and Csf1ra receptor tyrosine kinases. Genetics 194:631–646.

Lajoie M, Bertrand D, El-Mabrouk N, Gascuel O. 2007. Duplication and inversion history of a tandemly repeated genes family. J Comput Biol. 14:462–478.

Lampert KP, et al. 2012. Population divergence in East African coelacanths. Curr Biol. 22:R439–R440.

Lehti-Shiu MD, Shiu S-H. 2012. Diversity, classification and function of the plant protein kinase superfamily. Philos Trans R Soc Lond B: Biol Sci. 367:2619–2639.

Leitinger B. 2014. Discoidin domain receptor functions in physiological and pathological conditions. Int Rev Cell Mol Biol. 310:39–87.

Lemke G. 2013. Biology of the TAM receptors. Cold Spring Harb Perspect Biol. 5:a009076. doi: 10.1101/cshperspect.a009076.

Lemmon MA, Schlessinger J. 2010. Cell signaling by receptor tyrosine kinases. Cell 141:1117–1134.

Leveugle M, Prat K, Popovici C, Birnbaum D, Coulier F. 2004. Phylogenetic analysis of *Ciona intestinalis* gene superfamilies supports the hypothesis of successive gene expansions. J Mol Evol. 58:168–181.

Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. Nature 409:847–849.

Liu L, et al. 2004. A novel protein tyrosine kinase NOK that shares homology with platelet-derived growth factor/fibroblast growth factor receptors induces tumorigenesis and metastasis in nude mice. Cancer Res. 64:3491–3499.

Liu Y, et al. 2013. Natural selection and functional diversification of the epidermal growth factor receptor EGFR family in vertebrates. Genomics 101:318–325.

Louis A, Muffato M, Roest Crollius H. 2012. Genomicus: five genome browsers for comparative genomics in Eukaryota. Nucleic Acids Res. 41:D700–D705.

Lynch M, Conery JS. 2003. The origins of genome complexity. Science 30:1401–1404.

Makino T, McLysaght A. 2010. Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci. 107:9270–9274.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. Science 298:1912–1934.

McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome Res. 24:1665–1675.

Maere S, et al. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci. 102:5454–5459.

McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. Nat Genet. 31:200–204.

Mellgren EM, Johnson SL. 2005. Kitb, a second zebrafish ortholog of mouse Kit. Dev Genes Evol. 215:470–477.

Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 11:699–704.

Naruse K, et al. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res. 14:820–828.

Nikaido M, et al. 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res. 23: 1740–1748.

Ogino Y, Katoh H, Kuraku S, Yamada G. 2009. Evolutionary history and functional characterization of androgen receptor genes in jawed vertebrates. Endocrinology 150:5415–5427.

Ohno S. 1970. Evolution of gene duplication. New York: Springer-Verlag Inc.

Ohno S. 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. Semin Cell Dev Biol. 10:517–522.

Paterson AH, et al. 2006. Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. Trends Genet. 22:597–602.

Postlethwait JH, et al. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. 10:1890–1902.

Putnam NH, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071.

Robinson DR, Wu YM, Lin SF. 2000. The protein tyrosine kinase family of the human genome. Oncogene 19:5548–5557.

Rohner N, et al. 2009. Duplication of fgfr1 permits Fgf signaling to serve as a target for selection during domestication. Curr Biol. 19:1642–1647.

Rousset D, Agne?s F, Lachaume P, André C, Galibert F. 1995. Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulin like domains. J Mol Evol. 41:421–429.

Santini F, Harmon LJ, Carnevale G, Alfaro ME. 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol. 9:194.

Satoh N, Rokhsar D, Nishikawa T. 2014. Chordate evolution and the three-phylum system. Proc R Soc B. 281:20141729.

Schartl M, Volff J-N, Brunet F. 2015. Evolution of receptor tyrosine kinases. In: Wheeler DL, Yarden Y, editors. Receptor tyrosine kinases: structure, functions and role in human disease. New York: Humana Press, Springer. p. 17–36.

Schlessinger J. 2014. Receptor tyrosine kinases: legacy of the first two decades. Cold Spring Harb Perspect Biol. 6:a008912.

Sémon M, Wolfe KH. 2007. Consequences of genome duplication. Curr Opin Genet Dev. 17:505–512.

Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. 20:461–464.

Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. Curr Opin Microbiol. 2:548–554.

Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A. 2007. Comparative genomics of ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. BMC Genomics 8:312.

Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol. 1079:105–116.

Singh PP, et al. 2012. On the expansion of "Dangerous" gene repertoires by whole-genome duplications in early vertebrates. Cell Rep 2:1387–1398.

Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 25:1081–1090.

Smith JJ, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet. 45:415–421.

Srivastava M, et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466:720–726.

Steinke D, Salzburger W, Braasch I, Meyer A. 2006. Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics 7:20.

Stephens SG. 1951. Possible significance of duplication in evolution. Adv Genet. 4:247–265.

Suga H, et al. 1997. Intermittent divergence of the protein tyrosine kinase family during animal evolution. FEBS Lett. 412:540–546.

Suga H, et al. 1999. Protein tyrosine kinase cDNAs from amphioxus, hagfish, and lamprey: isoform duplications around the divergence of cyclostomes and gnathostomes. J Mol Evol. 49:601–608.

Suga H, et al. 2008. Ancient divergence of animal protein tyrosine kinase genes demonstrated by a gene family tree including choanoflagellate genes. FEBS Lett. 582:815–818.

Suga H, Torruella G, Burger G, Brown MW, Ruiz-Trillo I. 2014. earliest holozoan expansion of phosphotyrosine signaling. Mol Biol Evol. 31:517–528.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. 13:382–390.

Turunen O, Seelke R, Macosko J. 2009. In silico evidence for functional specialization after genome duplication in yeast. FEMS Yeast Res. 9:16–31.

Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179.

Volff JN, et al. 2003. The Xmrk oncogene can escape nonfunctionalization in a highly unstable subtelomeric region of the genome of the fish *Xiphophorus*. Genomics 82:470–479.

Volff JN, Schartl M. 2003. Evolution of signal transduction by gene and genome duplication in fish. J Struct Funct Genomics 3:139–150.

Williams H, Brenner S, Venkatesh B. 2002. Identification and analysis of additional copies of the platelet-derived growth factor receptor and colony stimulating factor 1 receptor genes in fugu. Gene 295:255–264.

Wolfe K. 2000. Robustness – it's not where you think it is. Nat Genet. 25:3–4.

Woods IG, et al. 2000. A comparative map of the zebrafish genome. Genome Res. 10:1903–1914.

Woods IG, et al. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res. 15:1307–1314.

Yarden Y, Pines G. 2012. The ERBB network: at last, cancer therapy meets systems biology. Nat Rev Cancer 12:553–563.

Zwick E, Bange J, Ullrich A. 2001. Receptor tyrosine kinase signalling as a target for cancer intervention strategies. Endocr Relat Cancer 8:161–173.

Associate editor: Ross Hardison