

Genomic Survey of Tyrosine Kinases Repertoire in *Electrophorus electricus* With an Emphasis on Evolutionary Conservation and Diversification

Evolutionary Bioinformatics
Volume 16: 1–13
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934320922519



Ling Li¹, Dangyun Liu², Ake Liu³, Jingquan Li¹, Hui Wang¹ and Jingqi Zhou¹

¹School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, P.R. China. ²Department of Central Laboratory, The Affiliated Huaian No.1 People's Hospital, Nanjing Medical University, Huai'an, P.R. China. ³Faculty of Biological Science and Technology, Changzhi University, Changzhi, P.R. China.

ABSTRACT: Tyrosine kinases (TKs) play key roles in the regulation of multicellularity in organisms and involved primarily in cell growth, differentiation, and cell-to-cell communication. Genome-wide characterization of TKs has been conducted in many metazoans; however, systematic information regarding this superfamily in *Electrophorus electricus* (electric eel) is still lacking. In this study, we identified 114 TK genes in the *E. electricus* genome and investigated their evolution, molecular features, and domain architecture using phylogenetic profiling to gain a better understanding of their similarities and specificity. Our results suggested that the electric eel TK (EeTK) repertoire was shaped by whole-genome duplications (WGDs) and tandem duplication events. Compared with other vertebrate TKs, gene members in Jak, Src, and EGFR subfamily duplicated specifically, but with members lost in Eph, Axl, and Ack subfamily in electric eel. We also conducted an exhaustive survey of TK genes in genomic databases, identifying 1674 TK proteins in 31 representative species covering all the main metazoan lineages. Extensive evolutionary analysis indicated that TK repertoire in vertebrates tended to be remarkably conserved, but the gene members in each subfamily were very variable. Comparative expression profile analysis showed that electric organ tissues and muscle shared a similar pattern with specific highly expressed TKs (ie, epha7, musk, jak1, and pdgfra), suggesting that regulation of TKs might play an important role in specifying an electric organ identity from its muscle precursor. We further identified TK genes exhibiting tissue-specific expression patterns, indicating that members in TKs participated in subfunctionalization representing an evolutionary divergence required for the performance of different tissues. This work generates valuable information for further gene function analysis and identifying candidate TK genes reflecting their unique tissue-function specializations in electric eel.

KEYWORDS: *Electrophorus electricus*, tyrosine kinase, phylogenetic analysis, metazoans, expression pattern

RECEIVED: February 19, 2020. **ACCEPTED:** April 7, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is supported by the grant from the National Science Foundation of China (31801049, 81630086), National Key R&D Program of China(2018YFC2000700), Shanghai Jiao Tong University medical-engineering interdisciplinary fund (YG2017QN69), the Fundamental Research Funds for the Central Universities, and the Major Science and Technology Innovation Program of Shanghai Municipal Education Commission (2019-01-07-00-01-E00059).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Hui Wang, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, P.R. China. Email: huiwang@shsmu.edu.cn

Jingqi Zhou, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, P.R. China. Email: jingqizhou@sjtu.edu.cn

Introduction

Tyrosine kinases (TKs) are a large and diverse superfamily of enzymes that catalyze the phosphorylation of select tyrosine residues in target proteins using ATP.¹ TKs are important mediators of signal transduction processes that regulate cell proliferation, differentiation, migration, metabolism, and programmed cell death.^{2,3} According to whether their protein sequences contain transmembrane domains, TKs are further classified into 2 major families, receptor TKs (RTKs) and non-receptor or cytoplasmic TKs (CTKs). The typical structural organization of RTKs typically includes a multidomain extracellular receptor that conveys ligand specificity, a transmembrane hydrophobic helix and a cytoplasmic portion containing a kinase domain (KD).⁴ Most CTKs are associated with phosphotyrosine (pTyr) binding within cells and are likely to transmit the pTyr signals initiated by receptors, while RTKs are involved primarily in responding to extracellular ligands by phosphorylating intracellular target proteins to initiate signal

transduction cascades.^{5,6} At present, TKs have become the subject of an increasing number of studies in physiology and pathology, including inflammation, autoimmunity, neurodegeneration, and infectious diseases.

Similar to many other gene families, the TK family underwent differential expansion in the history of metazoan evolution. After years of research, TKs have been characterized in many species, including many metazoans and a few premetazoans, and these findings have provided numerous vital insights into the structure, function, and regulation of TKs.⁷ Phylogenetic analyses have shown that the RTKs underwent extensive diversification in each of the filasterean, choanoflagellate, and metazoan clades.^{8,9} For instance, the VEGFR and Ephrin receptor subfamilies expanded through single-gene duplications in jawed vertebrates, and specific expansions of the Eph and EGFR subfamilies occurred between humans and zebrafish.¹⁰ Although the highly conserved TK domain is present in all TK proteins, these proteins also contain a diverse



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

arrangement of sequence domains involved in interactions with other molecules that allow for different signal transduction mechanisms.¹ Such divergent architectures are thought to result from gene duplication and domain shuffling events.¹¹ Based on sequence similarity and secondary domain architecture, TKs are further primarily divided into 30 subfamilies that are composed of CTKs and RTKs with specific functions in most metazoans. It has been assumed that 2 major episodes of expansions occurring in TK family. The initial diversification occurred before poriferans and the other metazoans diverged. The other expansion occurred around the diverged time of the cyclostomes and gnathostomes. The diversity of TKs may imply the complexity of the pTyr-based signaling system in metazoans.^{12,13} The evolution of TK activity, its regulation in metazoans, and its involvement in the evolution of multicellularity have been the subject of intriguing studies,^{14,15} and such research is facilitated by the completion of genome sequencing projects for a variety of different animals.

The electric eel (*Electrophorus electricus*) is a freshwater fish native to South America that is best known for its ability to produce high-voltage electric discharges that are used for communication, navigation, and even predation and defense.¹⁶ The taxonomic diversity of fishes that generate electricity is particularly interesting that electric organs have evolved not only in multiple independent fish lineages from myogenic precursors; even within this lineage, there is a tremendous amount of variation in their function.¹⁷ One of the challenges in understanding protein function evolution involves the identification of a tractable model system that allows for an assessment of the core assumptions. The electric fish is the only known species that has evolved 3 distinct electric organs (EO) and provides a unique case study opportunity. Gene family construction is a widely used approach to underlying assumptions that are used to characterize the evolutionary process in both systematics and functional biology.¹⁸ Despite extensive studies of TK genes on many other species, little is known about these supergene families in *E. electricus*. Understanding the regulation of TK genes will be valuable for efforts to induce the differentiation of electrogenic cells in other tissues and organisms and to control the intrinsic electric behaviors of these cells. Recently, the availability of the complete electric fish genome sequence provides an opportunity to perform a genome-wide analysis of TK gene family.^{16,19}

In this study, we analyzed the whole genome sequence of *E. electricus* and systematically identified putative full complement of TK genes (EeTKs). Through comprehensive phylogenetic approaches together with a comparison of the orthology, protein domain organizations, as well as analyzed their expression profiles in different tissues, we have elucidated in detail the evolution of TK expansion and diversity. This article provides the first comprehensive resource of electric eel TKs. The results of this study reveal commonalities and differences for the EeTKs among vertebrates and provide valuable information of EeTK diversity that we hypothesize underlie the functional differences

of specific organs, for example, in the production of voltages. Our findings may also provide further insights into metazoan TK genes evolution and would facilitate addressing the physiological and developmental function of electric eel TKs.

Materials and Methods

Retrieval of TK sequences

To perform genome-wide identification and obtain sequences of the PTK gene family in the electric eel, published genome sequences of *E. electricus* were first downloaded from the Ensembl database (https://useast.ensembl.org/Electrophorus_electricus/Info/Index), and redundant sequences were deleted using an in-house Perl script. The Pfam database (<http://pfam.xfam.org/>)²⁰ was used to screen the genome of *E. electricus*. Proteins with Pkinase_Tyr domains (PF07714) were used to identify the putative TK proteins in the *E. electricus* using the hidden Markov model (HMM) method. For the HMM search (3.1b2),²¹ a coverage of 0.3 was used as the cutoff, and an E-value of 10^{-10} was used for alignments longer than 150 amino acids. The online tools SMART (<http://smart.embl-heidelberg.de/>)²² and CDD (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) were used to verify the TK domains of the predicted proteins. The human and zebrafish TK proteins were retrieved from previous publication data.^{10,23,24} We also selected an additional 30 representative metazoan species to identify their TK members, and the information for these species is listed in Table S1. The TK sequences of these species were obtained as described above for those of electric eel.

Phylogenetic analysis of TK proteins and gene nomenclature

The identified TK proteins were aligned using T-Coffee with the default settings.²⁵ Phylogenetic analyses of all identified TK domains were performed using the neighbor-joining (NJ) and maximum likelihood (ML) methods. The NJ and ML trees were constructed using the program MEGA 6.0²⁶ with the gaps/missing data parameter set to “pairwise deletion” to account for variable amino acid sites and topology support estimated using a 1000-replicate bootstrap test. We used ProtTest 3.4 (<https://github.com/ddarriba/prottest3>) to assess the best amino acid substitution model to infer the phylogeny.²⁷ PhyML 3.1²⁸ was used for ML reconstruction under the LG model (LG model is a method of amino acid replacement matrix, which called after the two authors [Le SQ and Gascuel O]). Invariable sites and the γ -parameter were set to the values generated by ProtTest 3.4. Statistical support for the resulting topology was determined by a 1000-replicate bootstrap analysis. All phylogenetic trees were calculated based on the alignment of common TK domains. Genes that had been previously reported were named accordingly, and others were named according to their homologies to those of zebrafish and human PTK subfamily genes in the phylogenetic tree and were tagged with a species abbreviation.

Protein architecture, conserved motif prediction, and synteny analysis

Protein domain organization was analyzed using HMMER software by searching the SMART (<http://smart.embl-heidelberg.de>) and Pfam (<http://pfam.sanger.ac.uk/>) databases. We confirmed the presence of these predicted domains by manually inspecting the alignments with HMMER. The conserved motifs in the PTK protein sequences were first identified using the online software MEME (Multiple Expectation-Maximization for Motif Elicitation, <http://meme-suite.org>)²⁹ with the following parameters: the optimum motif width was set from 6 to 200, and the maximum number of motifs was 15. The discovered motifs were then annotated using the Pfam program. By checking their physical locations on individual chromosomes, tandem duplications were identified as 2 TK genes when they were separated by no more than 1 intervening gene. For syntenic analysis of the EeTK genes, MCScanX³⁰ with the default settings was used to identify syntenic gene pairs in the *E. electricus* genome.

Gene ontology annotation

The gene ontology (GO) terms associated with the genes were obtained from the *E. electricus* genome databases (<https://efish-genomics.integrativebiology.msu.edu/>), and WEGO was used to perform GO functional classification and identify the distribution of gene functions in the *electric eel* at the macro level. To determine the function of the unigenes, BLASTx alignments with an E-value $\leq 10^{-5}$ were performed with different databases, including Eukaryotic Orthologous Groups (KOG, <http://www.ncbi.nlm.nih.gov/KOG/>), the NCBI nonredundant protein database (nr, <http://www.ncbi.nlm.nih.gov/>), GO (<http://www.geneontology.org/>), and Swiss-Prot (<http://www.expasy.ch/sprot>).

Transcriptomic data of TK genes from different tissues of various animals

Publicly available transcriptomic data from 7 tissues (brain, spinal cord, heart, skeletal muscle, Sachs' electric organ, primary electric organ, and kidney tissues) of *E. electricus* from Gallant et al¹⁶ were used. High-throughput RNA sequencing raw data were downloaded from the SRA database (<http://www.ncbi.nlm.nih.gov/sra>). Quality control of the raw reads was performed with FastQC v0.11.5. After clipping of the Illumina adapter sequences and trimming low-quality bases using fastx-toolkit, short reads were individually mapped to their respective transcriptome assemblies using Bowtie2 (v2.2.8) with default parameters.³¹ To estimate gene expression levels, clean reads were mapped to nonredundant unigenes to calculate the fragments per kilobase transcript length per million fragments mapped (FPKM) value using RNA-Seq by Expectation-Maximization (RSEM).³² These expression values were normalized for sequencing depth and transcript length and then scaled via the trimmed mean of M values normalization under the assumption that most transcripts are not differentially expressed.

Generally, genes with an expression of FPKM < 1 were regarded as not expressed.³³

We obtained the human TK expression values for the abovementioned tissues from the Genotype-Tissue Expression (GTEx) project³⁴ (the electric organ was excluded) and selected the median value as the expression value in a given tissue for each gene. To assess the tissue-specific expression of genes, in this study, we used the parameter tau (τ)³⁵ to determine whether a given gene was expressed in a tissue-specific manner. The τ value was calculated as follows for each gene among the different tissues according to Yanai et al³⁶

$$= \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1}; \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} (x_i)}$$

where x_i is the expression of the gene in tissue i , and n is the number of tissues. The reads per kilo base per million mapped reads (RPKM) values were normalized by performing a \log_2 transformation after adding 1 to avoid negative values. The τ values ranged from 0 to 1, which indicate genes ranging from broadly to specifically expressed.

Results

Identification of the TK repertoire of *E. electricus*

For the genome-wide identification of EeTKs, we initially identified the proteins using HMMER²¹ and SMART²² as described in the Materials and Methods section and ensured the integrity of the data using BLAST searches.³⁶ Our computational analyses led to the identification of 114 distinct TK genes in the *E. electricus* genome, with 70 RTKs organized into 20 subfamilies and 44 CTKs organized into 10 subfamilies. Five of these 30 subfamilies have single members (CCK4, Musk, Sev, Ryk, and Ret subfamilies of RTKs), while the other subfamilies have multiple members, with the Src and Eph subfamilies having the greatest numbers of members (16 and 17, respectively). The detailed information regarding these genes, including the gene name, scaffold location, and the longest transcript ID, is summarized in Table S1.

Phylogenetic position of EeTKs is consistent with the conservation of core motifs

To elucidate the origin of EeTKs, we first conducted a preliminary phylogenetic analysis of the EeTK superfamily. We observed that neither the CTKs nor the RTKs grouped into clearly distinct monophyletic clusters (Figure 1), indicating that CTKs and RTKs were not derived from single genes.³⁷ In principle, conserved protein sites typically correspond to functional significance. To analyze the organization of motifs in EeTK proteins, 8 distinct motifs in the TK domain were identified with high E-values using the MEME tool. A schematic distribution of these motifs is shown in Figure 1. Most TK members within the same clade, especially the most closely related members, typically share common motif compositions

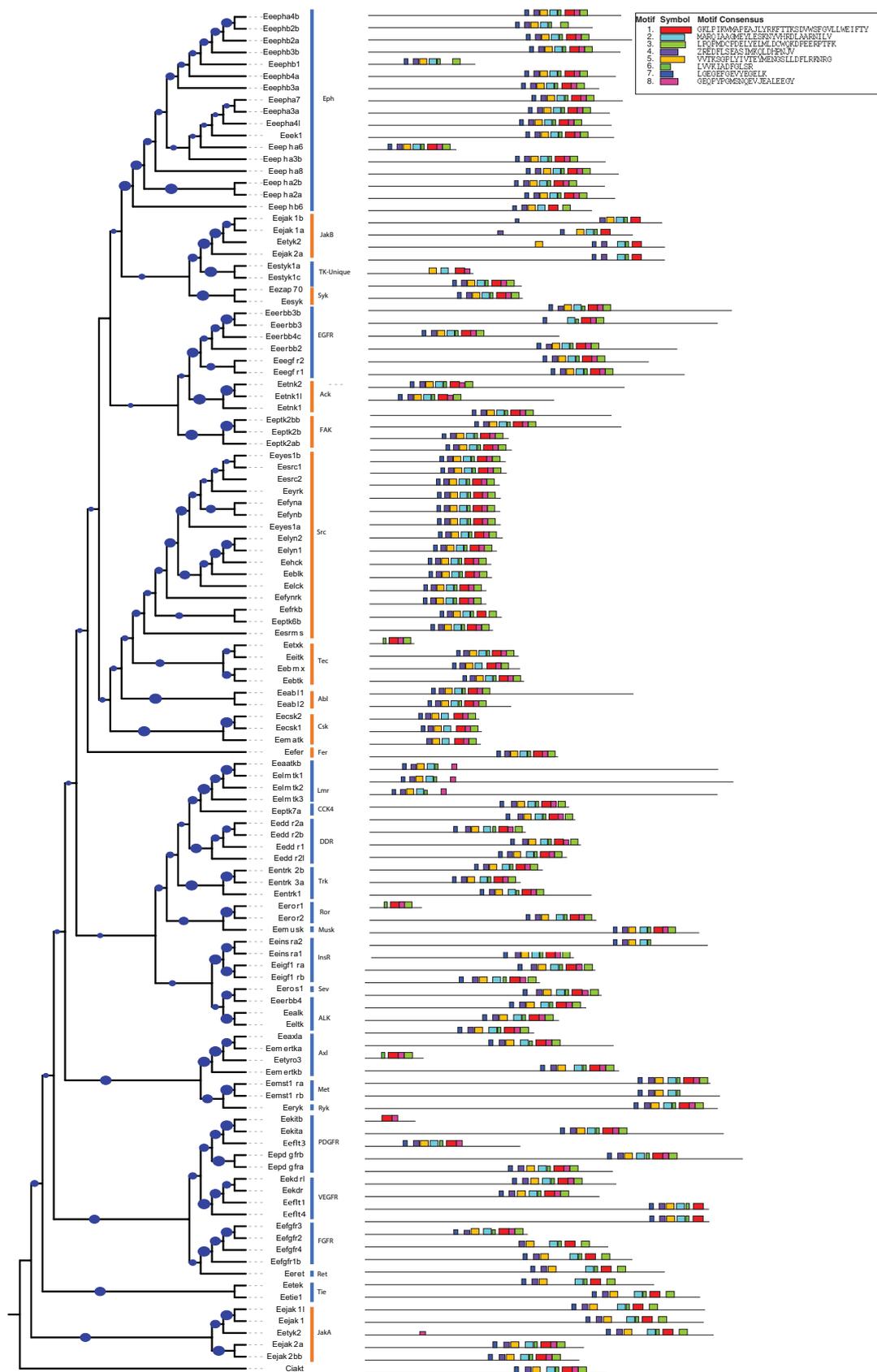


Figure 1. Phylogeny and motif architecture of electric eel TK genes. The left panel shows a maximum likelihood (ML) tree with 1000 bootstraps rooted by sea squirt *atk* gene. For simplicity, the bootstrap support percentages are plotted as circle marks on the branch (only higher than 50% are indicated), and circle size is proportional to the bootstrap values. The right panel shows the motif architecture for each EeTK gene identified using the MEME suite according to the protein sequence. EeTK indicates electric eel TK; MEME, Multiple Expectation-Maximization for Motif Elicitation; ML, maximum likelihood; TK, tyrosine kinases.

(eg, Eph and Src), indicating potential functional similarities among TK proteins. Among them, all groups of the RTK and CTK protein subfamilies contain motifs 1, 7, and 8, in which the highly conserved residues constitute the catalytic loop. In addition, motifs 4, 5, and 7 contain the conserved glycine (G) or glutamic acid (E) residues required for ATP binding, and motifs 2, 5, 7, and 8 contain conserved glycine (G) or proline (P) residues required for activity (Figure S1). It was found that motifs 2, 3, 4, 5, 7, and 8 in Lmr subfamily are conservative, while only motif 3 and 8 are conservative in Axl subfamily. In the JakB (Janus kinase B) subfamily, each member has the same motifs, with some members having additional motifs, and the motif architecture differs among these proteins. These results reveal that the phylogenetic position of EeTKs is consistent with the conservation of core motifs, while the structures of gene members were much more complicated.

Functional conservation and diversification of EeTKs

To better understand the functional diversity of EeTKs, we conducted a domain architecture survey for TK-encoding genes (Figure S2). Of the 114 putative TK-encoding proteins identified, 70 RTKs contain an intracellular KD, a transmembrane (TM) segment, a signal peptide, and known protein domains or motifs in the extracellular region. The other 44 EeTKs lack a signal peptide and a TM and are thus classified as CTKs (Figure 2). The CTK repertoire shared the domain architecture with SH2 or SH3, which can mediate interprotein interactions and promote n-Src catalytic activity, and thus influence signal transduction.³⁸ The Jak subfamily has dual TK domains, it grouped into 2 clusters, designated JakA and JakB, and this result was consistent with the report that 1 domain is inactive and may regulate the catalytic activity and autophosphorylation of the other domain, which is active.³⁹ RTKs show an extensive divergence in their architectures, containing 19 subfamilies with distinct organizations of protein domains. The most common domains are fibronectin type III (FN3) and immunoglobulin (IG) domains, which are extracellular domains involved in interactions with extracellular ligands or other receptors.⁴⁰ Members of the Eph, InsR, Axl, and Tie subfamilies typically have a single KD and 2 or 3 FN3 repeats. The PDGFR, VEGFR, and FGFR subfamilies are very similar in structure, with 5 or 7 IG domains characterizing the extracellular portion of the proteins.³⁸ In addition, the Tie, Axl, Ror, Musk, Trk, and CCK4 subfamilies also possess an IG domain.

“General” kinase genes are associated with many diverse biological processes, suggesting that different subfamily of kinases may be involved in distinct functions. To study the functions of EeTKs, gene ontology (GO) annotations were obtained from the Ensembl genome databases to construct GO graphs. As shown in Figure S3, almost all of the EePKs (97.5% for CTKs, 100% for RTKs) are involved in catalytic activity, binding, cellular process, and metabolic process, consistent with their primary functions. The results also showed

that the majority of CTKs are involved in diverse biological regulatory functions (52.5%) and predominantly participate in responses to stimuli (50%), while RTKs tend to be related to membrane part (64.3%), molecular transducer activity (71.4%), and biological regulation (75.7%). Compared with the wide range of genes with GO annotations, a high proportion of the EeTK genes play roles in catalytic activity, binding, signal transducer activity, metabolic process, localization, signaling, and so on.

Whole-genome and tandem duplications contribute to EeTKs diversity

Gene duplications in genomes could provide important information for gene evolution analysis.^{41,42} Ray-finned fishes are believed to have undergone a third genome duplication, the role of which in specific fish genomes has yet to be understood.⁴³ Thus, we assigned orthology to the EeTKs based on their zebrafish counterparts (Figure 3), and a list of electric eel TKs, their orthologues, and percent identities are shown in Table S2. The zebrafish TKs were identified by Challa and Chatti,¹⁰ with the results highlighting the effects of zebrafish genome duplication events on TK evolution in teleosts. The orthology assignments show a clear clustering of EeTKs and zebrafish TKs into identical families, indicating that the duplication of EeTKs in the *E. electricus* presumably resulted from a teleost-specific genome-wide duplication event (ie, ploidy).^{44,45} The results revealed that each EeTK has a corresponding zebrafish orthologue, except for *bmx*, which does not have zebrafish orthologues. In addition, the zebrafish genes *ptk2ab*, *jak1*, *jak2b*, *src*, *yes1*, *lyn*, and *erbB4* have lineage-specific duplicated copies in *E. electricus*, whereas *met*, *epha10*, *fgfr1b*, *ptk2aa*, *ptk6a*, *axlb*, and *tnk2a* do not have electric eel orthologues. Therefore, the TK composition of the electric eel, as a bony fish, is similar to that of zebrafish but also has its own specific features. We further investigated the number of EeTK orthologues present in the human genome, which encodes 90 TKs, with 32 CTKs and 58 RTKs.²³ Of the 50 human TKs that have single electric eel orthologues, whereas of the 30 human TKs with multiple electric eel orthologues, only 9 are CTKs and 21 are RTKs. The 9 human CTKs are represented by 19 electric eel orthologues, whereas the 21 human RTKs are represented by 44 electric eel orthologues (Figure S4). Such an expansion of TKs is in accordance with genome duplication events that occurred during the teleost radiation. Despite the difference in the actual number of genes, we observed that every human TK subfamily is represented in the teleost, suggesting conserved TK evolution in vertebrates. In addition, based on genome localization and MCScanX analysis results, we found 7 of the 114 EePTKs were tandemly duplicated genes, namely, *erbB4*, *erbB4c*, *kdr*, *pdgfra*, *pdgfrb*, *kita*, and *csf1ra* (Table S3, Figure S5), which are clustered into 3 tandem duplication event regions on the *E. electricus* genome scaffolds. Furthermore, 3 pairs of segmental duplication events were detected in our analysis, namely, *erbB4* and *erbB4c*, *pdgfra* and

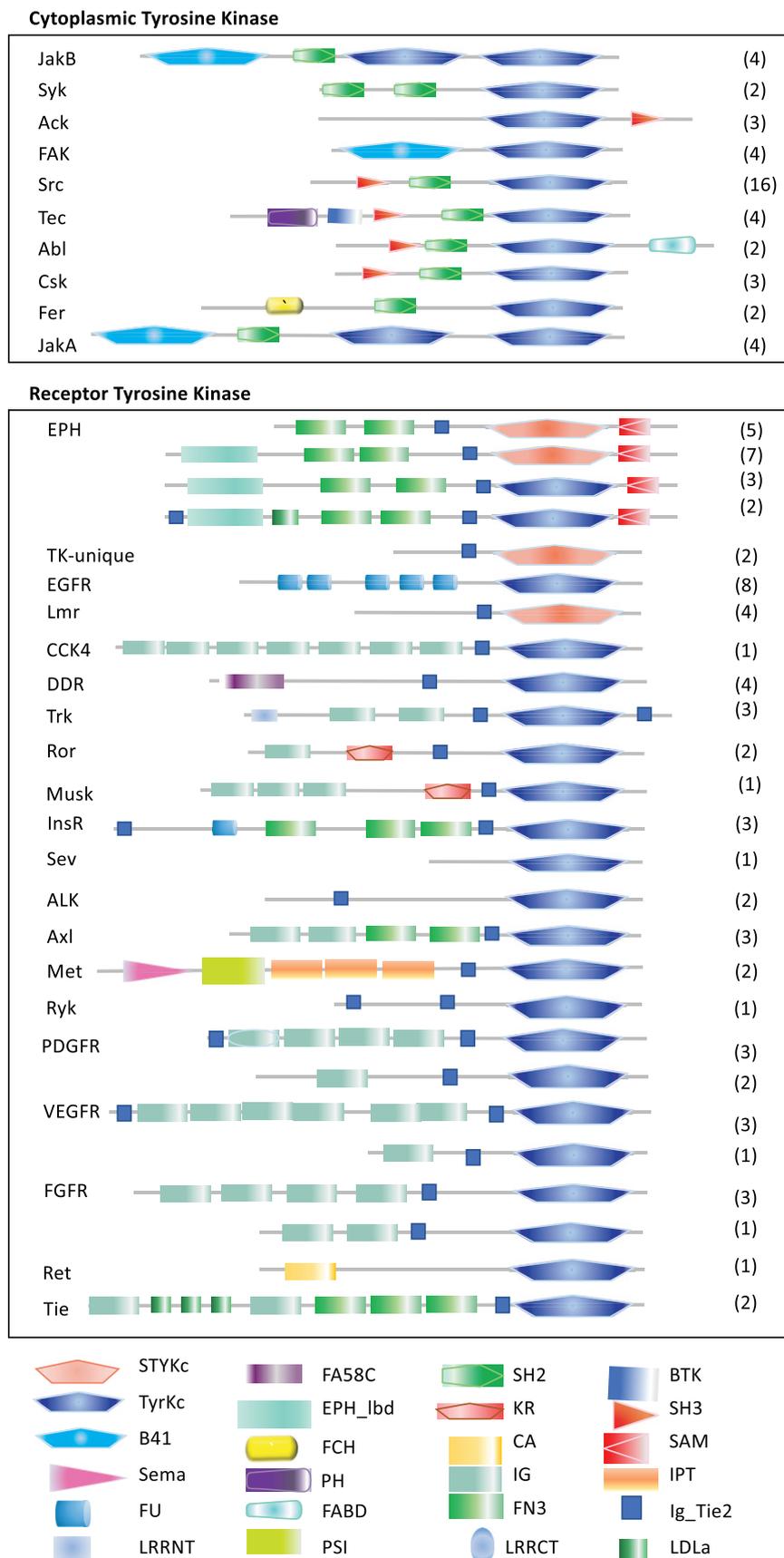


Figure 2. Classification of EeTKs. One hundred fourteen EeTKs were divided into 10 cytoplasmic tyrosine kinase (CTK) and 20 receptor tyrosine kinase (RTK) subfamilies, respectively, by domain architecture. A typical domain organization is schematically displayed for each subfamily, with the number of the genes that belong to each domain architecture shown in parentheses. The Pfam or SMART domain names are shown on the bottom. The domain organizations of all the proteins are shown in Figure S1. CTK indicates cytoplasmic tyrosine kinase; EeTK, electric eel TK; RTK, receptor tyrosine kinase.

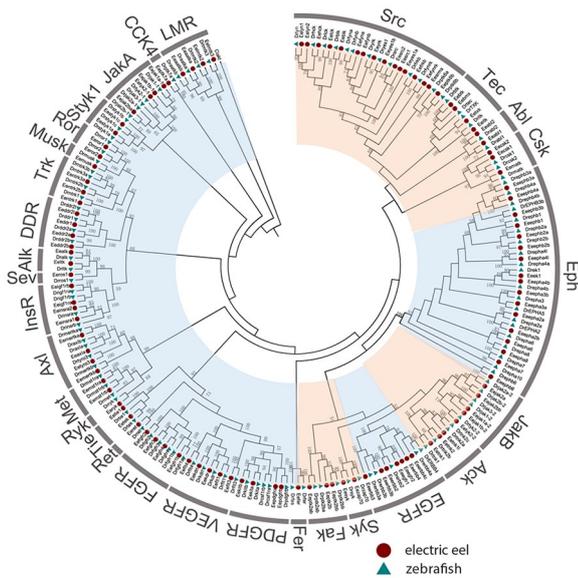


Figure 3. Dendrogram representing orthologous relationships between electric eel and zebrafish TK proteins. Phylogenetic analysis was conducted using the maximum likelihood (ML) method based on an alignment of TK domains from the electric eel and zebrafish, with a TK gene of *Ciona intestinalis* used as the outgroup. Only the subfamily names are shown. The electric eel genes are prefixed with “Ee,” and the zebrafish genes are prefixed with “Dr.” ML indicates maximum likelihood; TK, tyrosine kinases.

pdgfrb, and kita and kdr. The ErbB kinases have ever been reported as specified electric organ proteins⁴⁶; here, we also identified the other 2 closely tandem arranged EeTK genes, which provided potential target genes for unique tissue-function specialization in electric eel. The diversified and specified gene functions in each gene family may be the result of gene expansion from ancient paralogs or multiple origins of gene ancestry. Our results demonstrate that both whole-genome and tandem duplication events have been fundamental in shaping the diversity of TK repertoire in *E. electricus* genome.

Lineage-specific expansion but conservation of TKs in vertebrates

To better understand the phylogenetic relationships in this multigene family during vertebrates evolution, we used the NJ and ML algorithms to analyze EeTKs using sequences from 5 other representative chordates, including amphioxus (*Branchiostoma lanceolatum*), sea squirt (*Ciona intestinalis*), sea lamprey (*Petromyzon marinus*), elephant shark (*Callorhynchus milii*), and zebrafish (*Danio rerio*) (Figure 4). To minimize phylogenetic artifacts, we inferred a simplified ML tree that excluded sequences with identical structural domain compositions in the same species, which were considered to be recent closely related paralogs. The resulting tree topology generally agreed with the original tree that used the complete data set (Figure 1). Due to genome duplication during the teleost radiation, all the EeTK genes clustered with duplicated paralogous genes from vertebrates to form monophyletic clades. All the

TKs were analyzed in the amphioxus and elephant shark genomes, and at least 1 TK member was identified among all of the 30 major subfamilies. This finding strengthened the notion that the basic repertoire of the CTK and RTK families had already been established before the divergence of urochordates and vertebrates. After the divergence of protostomes and deuterostomes, the multiplicity of members in the same subtype rapidly increased by further gene duplication during the first half of chordate evolution before the fish-tetrapod split, giving rise to gene family expansion. Phylogenetic analyses revealed that TK repertoire expanded and was maintained in at least 1 major group of teleosts. We found gene members of Jak, Src, and EGFR subfamilies expanded specifically in electric eel, whereas Eph, FGFR, and Axl have more members in zebrafish. Teleosts have a considerably larger repertoire of TKs than tetrapods, which is in accordance with most recent large-scale changes being limited to simple duplications, including whole-genome duplications within vertebrates. Our results supported that the TK repertoire showed a conserved but expanded pattern that is markedly consistent with that observed in the electric eel and other species.

Diversity and commonality of metazoan TK repertoires

Both the timing and the rapid expansion in the number of TK genes provide a glimpse into the mechanisms involved in promoting the coordinated emergence and increasing sophistication of signal transduction during eukaryotic evolution.⁴⁷ We further traced the origins and diversification of TK subfamilies in 31 genomes covering the major clades of metazoans. As shown in Figure 5, all of the TK members were placed into corresponding clusters, including 1674 TK genes among the 31 subfamilies (Table S4). Among the bilaterians, the number of subfamilies ranges from 13 to 30. The protostomes possess 13~23 subfamilies, and the deuterostomes possess 21~30 subfamilies. A similar number of represented subfamilies (ranging from 26 to 30) are present across the vertebrates, which indicates that the TK repertoire tends to be evolutionarily conserved. In addition, the total number of TKs across different lineages exhibited considerable variation (ranging from 32 in fruit fly to 110 in electric eel), indicating that clade-specific gene duplication and domain shuffling can increase the number of metazoan TK subfamilies with a premetazoan origin. We observed that 11 of the 30 TK subfamilies generally possess a single representative gene among protostomes, and a few subfamilies, including VEGFR, Met, and InsR, have undergone species-specific expansions. The Src subfamily was established before the emergence of protostomes, and this subfamily contains some members (src, yes1, fyn, and lyn) that expanded in the electric eel and may be associated with the electric organs reported in previous studies.^{48,49} For the Tec, Tie, Ddr, and Eph subfamilies, lineage-specific duplications were detected

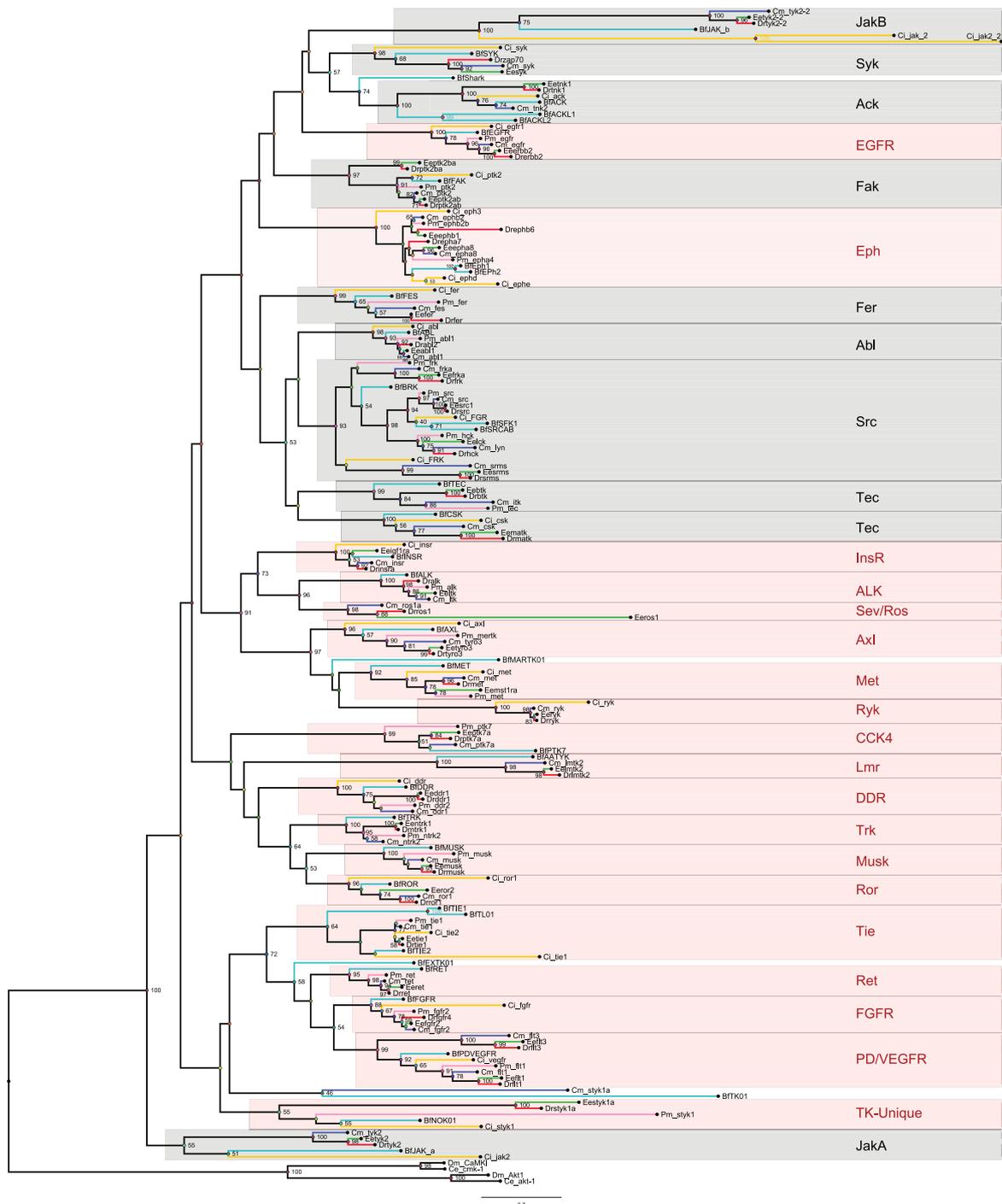


Figure 4. Phylogenetic relationships of major tyrosine kinase subfamilies in 5 representative chordata. The ML tree was based on multiple sequence alignments of TK domains, with bootstrap values shown on the branches. Abbreviated species names are as follows: Bf, *Branchiostoma lanceolatum*; Ci, *Ciona intestinalis*; Pm, *Petromyzon marinus*; Cm, *Callorhynchus milii*; Ee, *Electrophorus electricus*; and Dr, *Danio rerio*. The RTK subfamily is marked in blue, and the CTK subfamily is marked in red. CTK indicates cytoplasmic tyrosine kinase; ML, maximum likelihood; RTK, receptor tyrosine kinase; TK, tyrosine kinases.

in the nonvertebrate deuterostome genomes (in the sea squirt and/or the amphioxus). Subfamily genes *ddr21* and *Kdr1*, which have distinct orthologues in the chicken and stickleback genomes, were not maintained in mammals and eutherians. The RTK genes *epha1* and *insrr* or the CTK genes *fgr*

and *srm* are present in mammals but not the electric eel. These results show that some TKs may have a higher propensity to become useless and may have been independently lost in different lineages. For the electric eel, the exceptional and independent diversification of the TK repertoire through

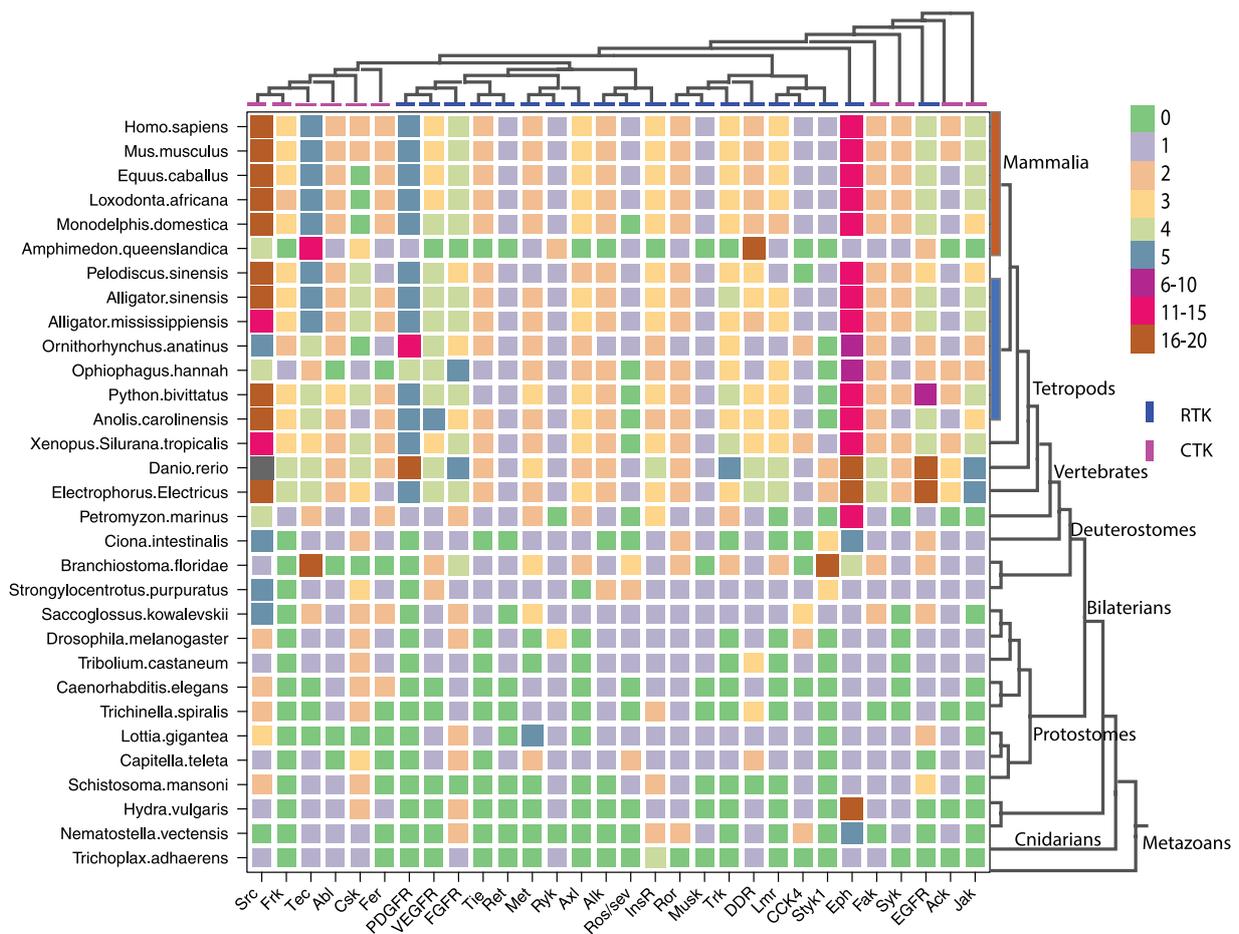


Figure 5. Schematic representation of the occurrence of TKs in 31 representative metazoan species. The square in different colors represents that the number of genes indicated at the bottom was observed in a species. The subfamily metazoan on the left shows the ML tree according to Figure S1. CTK indicates cytoplasmic tyrosine kinase; ML, maximum likelihood; RTK, receptor tyrosine kinase; TK, tyrosine kinases.

the creation of TKs with diverse architectures may be associated with an increase in organismal diversity and tissue-function specializations.

Comparative gene expression profiling of TKs across different tissues

The expression patterns of genes are typically related to their functions, and the results of RNA-Seq experiments have provided preliminary information regarding EeTK gene expression profiles. Therefore, we investigated the expression patterns of EeTK genes in different tissues via RNA-Seq analysis and retrieved expression data for corresponding human tissues from GTEx. Our results showed high variance in the expression levels of human and electric eel TK genes among different tissues (Figure 6). Considering the available RNA-Seq data for the electric eel, the brain showed the highest number of expressed TK genes (87 genes) among the tested tissues, while the kidney and electric organ presented the fewest (only 17, 28, and 25 genes for the kidney, primary electric organ [EO], and Sachs EO, respectively; Figure 6A). In humans, the lung presented the highest

number of expressed TK genes (78 genes), while the skeletal muscle showed the fewest (30 genes). Only a few genes from humans and electric eels, including *ddr1*, *txk*, *epha4b*, *ddr2b*, *fgfr4*, *axl*, and *fynb*, presented no expression in a few or even all tissues, with most genes, especially *fynb*, *syk*, *hck*, and *pdgfrb*, exhibiting relatively high expression in all tissues (Figure 6A). These similar expression patterns may indicate that these genes play fundamental roles in regulating organism growth and tissue development. We also found a few genes presenting distinct expression patterns. For instance, *epha8* showed low expression across almost all the tested tissues in the electric eel but high expression in human tissues. *Ptk2bb* was observed to be expressed at low levels in all of the electric eel tissues except in the brain, while this gene was highly expressed in all human tissues except muscle. In addition, we observed that electric organ tissues and muscle shared a similar pattern with specific highly expressed TKs (ie, *epha7*, *musk*, *jak1*, and *pdgfra*), suggesting regulation of TKs might play an important role in specifying an electric organ identity from its muscle precursor. The *epha4* and *epha6* were shown to be predominantly high expressed in electric organ and lower expression in muscle, but other

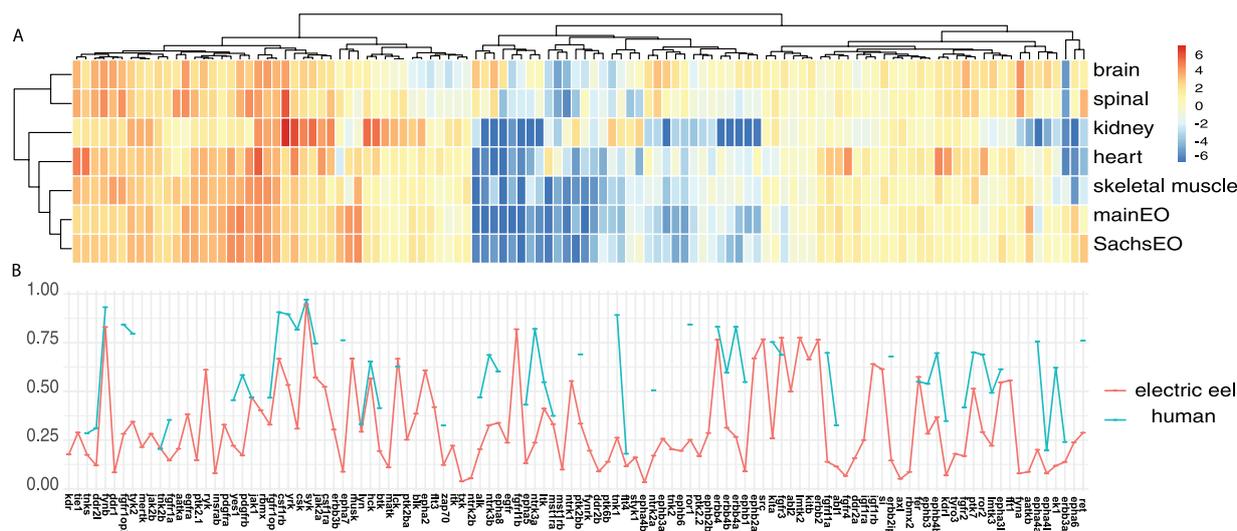


Figure 6. The expression profiles of EeTK genes in different tissues. (A) The relative expression levels are shown corresponding to \log_2 -transformed FPKM values after adding a pseudocount of 0.1. The scaled colors vary from blue showing high expression level to red showing low expression for genes according to FPKM values. (B) The comparison of tau values of TK genes between the electric eel and humans. EeTK indicates electric eel TK; FPKM, fragments per kilobase transcript length per million fragments mapped; TK, tyrosine kinases.

members of Epha members expressed lowly in both electric organ and muscle. These results indicate that members in TKs subfamily participated in subfunctionalization representing an evolutionary divergence required for the performance of electric organs and muscle.

Since the τ parameter has been recently proposed as a regular method to measure expression specificity,³³ we used this value to determine whether a given gene is expressed in a tissue-specific manner. We refer to those genes with τ values greater than 0.8 as being genes exhibiting tissue-specific expression, such as *ephb8*, *alk*, and *fgfr11b* (Figure 6B), and they were observed to be predominantly expressed in the electric eel brain and spinal tissues at high levels. Most genes (~71%) with τ values less than 0.4 were regarded as broadly expressed genes, including *rbmx* and *jak2b* (Figure 6B). *Mst1rb*, *ntrk3b* and *ephb3a* were shown to be expressed at low and similar levels in the electric organ. *Epha8*, *alk*, and *musk* exhibited a similar pattern between humans and electric eels. Expression-specific pattern may reflect their unique tissue-function specializations and provide insight into evolutionary conservation and divergence between duplicated gene expression, maintenance, and regulation.

Discussion

This study was conducted to create an unambiguous resource providing detailed information on EeTKs and their relationship to multiple metazoan TKs. The information we report is crucial for gaining a better understanding of the TK repertoire in an emerging model electric fish and for studies of TK biology in various model organisms. To our knowledge, no such study has previously been reported, and because the annotation of the electric eel genome is ongoing (Figure S6), our study generated systematic and useful annotation data for the EeTK genes.

Evolutionary relationships within the vertebrate TK repertoire

Comparing the EeTK genes with the TKs of human and zebrafish, the majority of their features appear to be largely stable, including subfamily numbers and sequence characteristics.^{50,51} Most metazoan genomes encode 30 TK subfamilies, although the total number of TKs representing each subfamily of orthologues varies among different lineages (Figure 4, 90 in humans, 122 in zebrafish and 114 in the electric eel). Because the conservation of gene family size across multiple species may reflect specific functional constraints,^{52,53} this observation is consistent with our TK identification results in the electric eel. The difference in the total number of TKs is due to the specific expansion of a few subfamilies, such as the Eph and Src subfamilies. Similarly, among the bilaterians, the differences in the total number of TK genes may also result from the specific expansions of specific subfamilies. In the present study, our genome-wide analysis of *E. electricus* identified 114 TK-encoding genes belonging to 30 subfamilies, and sequence identity matrix analysis results suggested that TKs in vertebrates tend to be remarkably conserved and stable.³⁷

Sequence architecture reveals functional conservation and diversification

TKs have significant roles in cell growth, apoptosis, and development. A simple way to assess functional use is to assess the extent to which a consistent order of domain architectures is maintained.⁵⁴ A comparison of the number of distinct domain combinations may more accurately reflect the diversity of functional usage of protein families. The evolutionary origin of the canonical/functional domain organization of each TK subfamily is indicated where a conserved

function has been demonstrated. A highly conserved subfamily of TKs may be responsible for highly conserved signaling pathways regulating target gene expression, and the diversity of domain architecture within a subfamily (such as Eph and Src) can indicate the linkage of historically disparate functional domains into a novel molecule during the evolution of complex multicellular life.⁵⁵ Of the 10 common metazoan CTK subfamilies, at least 4 (Src, Csk, Tec, and Abl) are designated as Src-related CTKs for their shared SH3-SH2-TyrKc domain architecture, whereas in the RTK subfamilies, almost all the subfamilies evolved extracellular domains, which are probably used for sensing extracellular signals. The distantly related domain patterns of RTKs and CTKs reflect that the generation of these 2 types of TKs may have been due to different evolutionary conservation, environmental variability, and biological functions during the unicellular-multicellular transition.⁵⁶ The lineage-specific TK repertoire, which diversified independently of the common set of TKs in metazoans, may reflect the constant exposure of their tissues to the environment.⁵⁷ The more stable evolution of CTKs would then reflect a relatively stable intracellular environment, with CTKs generally acting downstream of RTKs or other receptors to transmit extracellular signals. The far fewer changes in domain architecture in more evolutionarily recent times may coincide with a more stable multicellular environment.

Distinct evolution of the vertebrate CTK and RTK repertoire after gene duplication events

Phylogenetic analysis suggested that expansion and sequence variation events have occurred in the TK family in vertebrates. The TK repertoire appears to be largely stable after the initial expansion, with a unique set of vertebrate TKs retained after the occurrence of whole-genome duplication.⁵⁸ The duplications of genes and entire genomes are believed to be important mechanisms underlying morphological variation and functional innovation in the evolution of life and especially for the wide diversity observed in the speciation of fishes.^{44,59} The results of our analysis further show that gene duplication, which is known to be common in teleost due to the whole-genome duplications in the teleost lineage, occurs more in RTK genes than CTK genes (RTK vs CTK for multiple orthologs is 21 vs 9 and 44 vs 19 in human and electric eel, respectively; Figure S4). This finding is consistent with the theory that following genome duplication, gene duplicates that acquire novel functions and contribute to diversity are retained more frequently to contributors to the evolution of organisms. RTKs are longer and contain more domains with greater variation than CTKs, suggesting a greater probability of the evolution of novel combinations and functions that results in greater duplicate gene retention for RTKs. This explanation is also consistent with the recent discovery of a higher divergence

among RTKs during metazoan evolution, which may have facilitated cell-to-cell communication and allowed for responses to a variety of extracellular cues during the evolution of multicellularity.⁸ The same rationale would help explain the retention of a larger RTK than CTK repertoire in the electric eel. Our studies assessing these features in electric eel provide a unifying and consistent explanation with those observed in zebrafish, which aids in explaining the observed retention of a larger RTK than CTK repertoire in vertebrate lineages.

Expression pattern of TK genes in the electric eel

Expression profiling analyses revealed tissue-specific or sex-dimorphic expression patterns of EeTK genes. Our findings complement the systematic information available on the TK family in the electric eel and increase our understanding of metazoan TKs. The results showed that 90 TKs were expressed in at least 1 tested tissue, and there was high variance in the expression levels among different tissues. The expression pattern analysis of EeTK genes among 7 different tissues showed that 9 EeTK genes exhibited tissue-specific expression, while 12 EeTK genes were broadly expressed. The former group includes *alk* and *ephb2* (Figure 6A), while the latter group includes *ryk* and *jak1* (Figure 6B). Our results suggest that the variable expression patterns of genes exhibiting tissue-specific expression may indicate that they have different roles between species, and the similar patterns of broadly expressed genes may suggest that they have fundamental roles in organism development.^{16,60} We observed that the *tyk1* gene is highly expressed in the muscle and EO, which may indicate that this TK is involved in bioelectricity generation. It has been reported that when this gene is overexpressed, phosphatidylinositol 3'-kinase (PI3K) pathways can be activated to induce cell invasion and metastasis to distant organs.⁶¹ PI3K acts through distinct signaling targets to regulate cell size, cell proliferation, and protein synthesis and degradation. These results are consistent with the previous finding that electrocytes, the electrical cells in EOs, are much larger than muscle fibers, which may be due to changes in insulin-like growth factor (IGF) signaling pathway genes.⁸ Overall, the above results show variable expression of major TKs in different tissues and may indicate that each gene plays different roles during the organogenesis process. Considering that the electric eel has electric organs, making it unusual among teleost fish and that TKs are one of the components of the P-Tyr cell signaling pathway in metazoans, we considered that TKs may be related to the specialization of electric organ discharge. Although the functions of most TK genes in the electric eel remain to be examined, our phylogenetic and expression analyses provide a solid foundation for future research.^{10,62} Follow-up functional studies are required for a better understanding of the roles of TKs in the regulation of key growth and developmental processes.

Conclusions

In this study, we systematically identified putative full complement of TK genes by analyzing the *E electricus* genome sequences and characterized their sequences by phylogenetic analysis among representative metazoans, as well as analyzed their expression profiles in different tissues. Understanding the evolution of and regulation of TK activity across vertebrates and their relationships to the evolution of multicellular organisms has been a crucial subject in recent studies. It should be noted that the currently annotated version of the *E electricus* genome sequence contains unmapped and partial scaffolds. As the genome project continues, the quality of the sequence information will improve, and the gene annotations will be more informative. While we believe that our study includes all of the EeTKs, new and refined sequence information may result in modifications to our computational findings with functionally relevant annotations.

Author Contributions

J.Z. and H.W. conceived of the study. J.Z. performed the study. J.Z. and L.L. analyzed the data. D.L. collected the data. J.Z. wrote the paper. L.L., A.L., and J.L. edited and revised the manuscript.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Hubbard SR, Till JH. Protein tyrosine kinase structure and function. *Annu Rev Biochem.* 2000;69:373-398.
- Miller WT. Tyrosine kinase signaling and the emergence of multicellularity. *Biochim Biophys Acta.* 2012;1823:1053-1057.
- Purcell AL, Carew TJ. Tyrosine kinases, synaptic plasticity and memory: insights from vertebrates and invertebrates. *Trends Neurosci.* 2003;26:625-630.
- Pincus D, Letunic I, Bork P, Lim WA. Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc Natl Acad Sci USA.* 2008;105:9680-9684.
- Annenkov A. Receptor tyrosine kinase (RTK) signalling in the control of neural stem and progenitor cell (NSPC) development. *Mol Neurobiol.* 2014;49:440-471.
- Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell.* 2010;141:1117-1134.
- Schaap P, Barrantes I, Minx P, et al. The physarum polycephalum genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. *Genome Biol Evol.* 2016;8:109-125.
- Suga H, Dacre M, de Mendoza A, Shalchian-Tabrizi K, Manning G, Ruiz-Trillo I. Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci Signal.* 2012;5:ra35.
- Shiu SH, Li WH. Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. *Mol Biol Evol.* 2004;21:828-840.
- Challa AK, Chatti K. Conservation and early expression of zebrafish tyrosine kinases support the utility of zebrafish as a model for tyrosine kinase biology. *Zebrafish.* 2013;10:264-274.
- Suga H, Kuma K, Iwabe N, et al. Intermittent divergence of the protein tyrosine kinase family during animal evolution. *FEBS Lett.* 1997;412:540-546.
- Hubbard SR, Mohammadi M, Schlessinger J. Autoregulatory mechanisms in protein-tyrosine kinases. *J Biol Chem.* 1998;273:11987-11990.
- Valentinis B, Romano G, Peruzzi F, et al. Growth and differentiation signals by the insulin-like growth factor 1 receptor in hemopoietic cells are mediated through different pathways. *J Biol Chem.* 1999;274:12423-12430.
- Li WQ, Young SL, King N, Miller WT. Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. *J Biol Chem.* 2008;283:15491-15501.
- Manning G, Young SL, Miller WT, Zhai YF. The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc Natl Acad Sci USA.* 2008;105:9674-9679.
- Gallant JR, Traeger LL, Volkening JD, et al. Nonhuman genetics genomic basis for the convergent evolution of electric organs. *Science.* 2014;344:1522-1525.
- Traeger LL, Sabat G, Barrett-Wilt GA, Wells GB, Sussman MR. A tail of two voltages: proteomic comparison of the three electric organs of the electric eel. *Sci Adv.* 2017;3:e1700523.
- Liberles DA, Dittmar K. Characterizing gene family evolution. *Biol Proced Online.* 2008;10:66-73.
- Traeger LL, Volkening JD, Moffett H, et al. Unique patterns of transcript and miRNA expression in the South American strong voltage electric eel (*Electrophorus electricus*). *BMC Genomics.* 2015;16:243.
- Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44:D279-D285.
- Eddy SR. Profile hidden Markov models. *Bioinformatics.* 1998;14:755-763.
- Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 2015;43:D257-D260.
- Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. *Oncogene.* 2000;19:5548-5557.
- D'Aniello S, Irimia M, Maeso I, et al. Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Mol Biol Evol.* 2008;25:1841-1854.
- Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000;302:205-217.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725-2729.
- Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164-1165.
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003;52:696-704.
- Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202-W208.
- Wang YP, Tang HB, Debarry JD, et al. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-359.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
- Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform.* 2017;18:205-214.
- Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013;45:580-585.
- Yanai I, Benjamin H, Shmooish M, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics.* 2005;21:650-659.
- Camacho C, Coulouris G, Avagyan V, et al. BLAST plus: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
- Brunet FG, Volf JN, Schartl M. Whole genome duplications shaped the receptor tyrosine kinase repertoire of jawed vertebrates. *Genome Biol Evol.* 2016;8:1600-1613.
- Groveman BR, Xue S, Marin V, et al. Roles of the SH2 and SH3 domains in the regulation of neuronal Src kinase functions. *FEBS J.* 2011;278:643-653.
- Chen M, Cheng A, Candotti F, et al. Complex effects of naturally occurring mutations in the JAK3 pseudokinase domain: evidence for interactions between the kinase and pseudokinase domains. *Mol Cell Biol.* 2000;20:947-956.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002;298:1912-1934.
- Opazo JC, Butts GT, Nery MF, Storz JF, Hoffmann FG. Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Mol Biol Evol.* 2013;30:140-153.
- Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays.* 2005;27:937-945.
- Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics.* 2014;289:1045-1060.
- Jaillon O, Aury JM, Brunet F, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature.* 2004;431:946-957.
- Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol.* 2006;23:121-136.
- Mate SE, Brown KJ, Hoffman EP. Integrated genomics and proteomics of the *Torpedo californica* electric organ: concordance with the mammalian neuromuscular junction. *Skelet Muscle.* 2011;1:20.

47. Lim WA, Pawson T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell*. 2010;142:661-667.
48. Shores CG, Cox ME, Maness PF. A tyrosine kinase related to Pp60c-Src is associated with membranes of electrophorus-electricus electric organ. *J Biol Chem*. 1987;262:9477-9485.
49. Brickell PM. The p60c-src family of protein-tyrosine kinases: structure, regulation, and function. *Crit Rev Oncog*. 1992;3:401-446.
50. He SZ, Tong XL, Han MJ, Bai YM, Dai FY. Genome-wide identification and characterization of tyrosine kinases in the silkworm, *Bombyx mori*. *Int J Mol Sci*. 2018;19:934.
51. Liu A, He FA, Gu X. Identification and characterization of tyrosine kinases in anole lizard indicate the conserved tyrosine kinase repertoire in vertebrates. *Mol Genet Genomics*. 2017;292:1405-1418.
52. Chen FC, Chen CJ, Li WH, Chuang TJ. Gene family size conservation is a good indicator of evolutionary rates. *Mol Biol Evol*. 2010;27:1750-1758.
53. Zhou JQ, Liu DY, Sa ZN, Huang W, Zou YY, Gu X. Effective estimation of the minimum number of amino acid residues required for functional divergence between duplicate genes. *Mol Phylogenet Evol*. 2017;113:126-138.
54. Knight JDR, Qian B, Baker D, Kothary R. Conservation, variability and the modeling of active protein kinases. *PLoS ONE*. 2007;2:e982.
55. Babonis LS, Martindale MQ. Phylogenetic evidence for the modular evolution of metazoan signalling pathways. *Philos Trans R Soc Lond B Biol Sci*. 2017;372:20150477.
56. Bhattacharyya T, Sowdhamini R. Genome-wide search for tyrosine phosphatases in the human genome through computational approaches leads to the discovery of few new domain architectures. *Evol Bioinform Online*. 2019;15:1176934319840289.
57. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci*. 2002;27:514-520.
58. Arnegard ME, Zwickl DJ, Lu Y, Zakon HH. Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proc Natl Acad Sci USA*. 2010;107:22172-22177.
59. Iwabe N, Kuma K, Miyata T. Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. *Mol Biol Evol*. 1996;13:483-493.
60. Lamanna F, Kirschbaum F, Waurick I, Dieterich C, Tiedemann R. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genomics*. 2015;16:668.
61. Dayde-Cazals B, Fauvel B, Singer M, et al. Rational design, synthesis, and biological evaluation of 7-azaindole derivatives as potent focused multi-targeted kinase inhibitors. *J Med Chem*. 2016;59:3886-3905.
62. Attwood MM, Krishnan A, Almen MS, Schioth HB. Highly diversified expansions shaped the evolution of membrane bound proteins in metazoans. *Sci Rep*. 2017;7:12387.