Gene Expansion and Retention Leads to a Diverse Tyrosine Kinase Superfamily in Amphioxus

Salvatore D'Aniello,[1] *Manuel Irimia,*[1] *Ignacio Maeso,*[1] *Juan Pascual-Anaya, Senda Jiménez-Delgado, Stephanie Bertrand, and Jordi Garcia-Fernàndez*

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

Tyrosine kinase (TK) proteins play a central role in cellular behavior and development of animals. The expansion of this superfamily is regarded as a key event in the evolution of the complex signaling pathways and gene networks of metazoans and is a prominent example of how shuffling of protein modules may generate molecular novelties. Using the intron/exon structure within the TK domain (TK intron code) as a complementary tool for the assignment of orthology and paralogy, we identified and studied the 118 TK proteins of the amphioxus *Branchiostoma floridae* genome to elucidate TK gene family evolution in metazoans and chordates in particular. Unlike all characterized metazoans to date, amphioxus has members of all known widespread TK families, with not a single loss. Putting amphioxus TKs in an evolutionary context, including new data from the cnidarian *Nematostella vectensis*, the echinoderm *Strongylocentrotus purpuratus*, and the ascidian *Ciona intestinalis*, we suggest new evolutionary histories for different TK families and draw a new global picture of gene loss/gain in the different phyla. Surprisingly, our survey also detected an unprecedented expansion of a group of closely related TK families, including TIE, FGFR, PDGFR, and RET, due most probably to massive gene duplication and exon shuffling. Based on their highly similar intron/exon structure at the TK domain, we suggest that this group of TK families constitute a superfamily of TK proteins, which we termed EXpanding TK, after their seemingly unique propensity to gene duplication and exon shuffling, not only in amphioxus but also across all metazoan groups. Due to this extreme tendency to both retention and expansion of TK genes, amphioxus harbors the richest and most diverse TK repertoire among all metazoans studied so far, retaining most of the gene complement of its ancestors, but having evolved its own repertoire of genetic novelties.

Introduction

The signaling and regulatory networks involved in metazoan development and cellular behavior have an intrinsic modular structure (Bhattacharyya et al. 2006; Davidson and Erwin 2006), in which proteins with modular domains play a key role in interconnecting the different units (Pawson 1995). Although many of these proteins are unique to metazoans, few of their component domains are so; much of this protein richness has been achieved by modular recombination of preexisting domains in metazoan ancestors (Müller et al. 1999; Patthy 2003; Benito-Gutierrez et al. 2006). A paradigmatic example is the superfamily of tyrosine kinase (TK) proteins. TKs drive phosphorylation events through transfer of phosphate from adenosine triphosphate to tyrosine residues of their target proteins, thereby regulating the target's activity. Although TKs have also been predicted in plants and amoebas (Miranda-Saavedra and Burton 2007), TKs have only expanded and diversified in metazoans and their closest unicellular relatives, the choanoflagellates (King and Carroll 2001). In metazoans, TKs are involved in several aspects of animal development, tissue differentiation, immune responses, and cell death (Geer et al. 1994; Hunter 1998; Hubbard and Till 2000). They are essential components of one of the few signal transduction pathways conserved throughout metazoans (Pires-daSilva and Sommer 2003). Besides, TKs have broad medical importance with mutation or malfunction of TKs responsible for numerous human malignancies and implicated in a wide variety of congenital disorders (Hunter 1998; Chang et al. 2007; Nelson and Grandis 2007).

TK proteins typically consist of a TK domain, responsible for the Tyr phosphorylation of target proteins, and a variable array of other protein motifs, which interact with various components in the signal transduction pathways (Hubbard and Till 2000). Based on sequence similarity and type and number of secondary domains (i.e., those functional protein domains different from the TK domain), TKs have been classified into several protein families grouped into 2 major classes, nonreceptor TKs and receptor TKs (RTKs). For instance, the human genome contains 90 TK genes, 32 nonreceptor TKs grouped in 10 families and 58 RTKs in 19 families (Robinson et al. 2000). However, accurate classification, and thus evolutionary insights, is often hampered by the high degree of similarity of the kinase domain, distinct evolutionary rates, and diversity of protein domain organization in given clades. Here we demonstrate the utility of intron positions within the TK domain (which we termed "intron code"), which greatly simplifies and improves assignment of TK domains to known TK families.

Unlike all previously studied bilaterians that have lost individual TK families, amphioxus contains representatives of all widespread TK families, including all 29 vertebrate families, underscoring the ancestral features of the amphioxus genome (Putnam et al. 2008). On the other hand, we unveiled a massive expansion of some closely related TK families, expanded by means of extensive gene duplication and domain shuffling over millions of years of amphioxus evolution. This retention and expansion have yielded the richest TK protein complement so far known in a single species.

Methods

Search for Previously Described TKs

For each described vertebrate family (Robinson et al. 2000) and the invertebrate-specific SHARK family (Chan et al. 1994), we blasted the whole protein sequence of each

[1] These authors contributed equally to this work.

Key words: amphioxus, tyrosine kinase, genome evolution, gene expansion.

E-mail: jordigarcia@ub.edu.

Mol. Biol. Evol. 25(9):1841–1854. 2008
doi:10.1093/molbev/msn132
Advance Access publication June 11, 2008

representative against the amphioxus genome JGI v1.0, using TBlastN online at the JGI Web page (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html). We analyzed the best hits covering most of the full-length query protein. We then downloaded the corresponding genomic region (for both haplotype scaffolds, if present) and built different gene models (containing information for intron/exon structure and, when possible, start and stop codons) using various software: GenomeScan (Yeh et al. 2001), GeneID (Parra et al. 2000), GeneMarkHMM (Lukashin and Borodovsky 1998), and GeneWise2 (Birney and Durbin 2000). We compared these predictions with expressed sequence tags (ESTs) and the JGI automatic gene prediction, when available. With all this information, we manually built the best possible model considering the orthologous genes as best guides. Finally, to confirm orthology, we used reciprocal Blast and specific intron pattern similarity (Irimia and Roy 2008). All scaffold positions and JGI gene prediction IDs (when available) for "classical" TK genes are provided in supplementary table SM1 (Supplementary Material online).

Full TK Complement

In order to find all potential genes containing TK domains in the genome, we blasted 6 TK domains from different human families (from the genes *ABL1*, *BTK*, *FGFR1*, *INSR*, *MUSK*, and *ROS1*) against the amphioxus genome using TBlastN under highly unrestrictive conditions (*e* value = 100) and then filtered for redundancy. With this approach, we obtained 668 unique hits in 326 scaffolds. We then filtered these hits to eliminate Ser/Thr kinases, using Prosite (Hulo et al. 2006) and National Center for Biotechnology Information (NCBI) Conserved Domain (Marchler-Bauer et al. 2007) Web pages.

For the remaining 415 hits, we built consensus gene models using gene predictions obtained from GenomeScan (Yeh et al. 2001), GeneID (Parra et al. 2000), GeneMarkHMM (Lukashin and Borodovsky 1998), and GeneWise2 (Birney and Durbin 2000) software and comparing with ESTs and the JGI automatic gene prediction when available and using information from both haplotypes when present. Finally, each predicted TK domain was aligned with previously confirmed TK domains; if necessary, gene models were carefully corrected manually to avoid spurious insertions or deletions, by taking advantage of the high sequence similarity and intron/exon structure conservation (Coghlan and Durbin 2007; Siegel et al. 2007).

Classification of these proteins was based on intron patterns within the TK domain and sequence similarity using standard phylogenetic methods. All genes found using the approach described in the previous section were also detected under this global approach. The complete set of TK domain sequences with annotated intron positions is given as supplementary file 1 (Supplementary Material online). Gene models without introns but with sequence similarity to other intron-containing TKs were considered processed pseudogenes (Vanin 1985; D'Errico et al. 2004; Irimia and Roy 2008). It should be noted that due to the draft nature of the amphioxus genome assembly, some TK genes may have not been detected in our survey.

Analysis of Intron/Exon Structures and TK Intron Code Comparisons

Nucleotide coordinates for the start and end of each exon were extracted from gene annotations from different software and databases (NCBI or JGI) using custom Perl scripts. With these coordinates, it is possible to calculate the nucleotide length of each exon and the codon reading frame and therefore calculate the position and phase of each intron (an intron is in phase 0, 1, and 2 if it falls before the first, second, and third bases of a codon, respectively). Once the position and phase of each intron was obtained, we used Perl scripts modified from scripts provided to us by Scott W. Roy to map these positions onto protein-level alignments of the TK domain of all TK genes analyzed. These positions/phases thus define the "TK intron code" of a given TK, which may be then compared across different genes. An example is provided in with intron positions indicated by digits corresponding to the phase of the intron located in between the 2 surrounding amino acids (in phase 0 introns) or after the amino acid that the intron is disrupting (in phase 1 and 2 introns). If 2 introns with the same phase fall in homologous positions of the alignment of 2 different TK domains (in an ungapped and relatively well-conserved region of the alignment), we consider that this intron position is conserved between the 2 TK domains. We can thus compare intron codes of 2 TK genes by comparing all intron positions from the 2 genes in this way (for further information and examples, see supplementary file 2, Supplementary Material online).

Based on TK intron code conservation, we defined 3 groups: if >70% (e.g., 5/7 or more) of intron positions of both intron codes are coincident, we consider that the intron codes are similar or analogous, consistent with these TK domains belonging to the same TK family or superfamily; on the other hand, if <30% of intron positions are shared, TK intron codes would be inconsistent with the 2 TK domains being similar; finally, intermediate values may indicate close phylogenetic relationship between different TK families (e.g., NOK and EXTK families or ALK and AATYK). It should be noted that intron position conservation may vary widely across lineages; therefore, these cutoffs are only valid for comparing species with little intron loss/gain.

Importantly, considering that TK domains usually comprise 270 amino acids and that there are 3 possible intron phases, the probability of an intron from 2 different TK domains falling in the homologous position by chance is ~0.001; however, the probability of, for instance, 4 out of 5 introns of 2 TK domains falling in the homologous positions (as in the case of the human SYK protein in fig. 2*B*) is $<10^{-10}$.

Finally, the diagnostic relevance of particular introns is not equivalent as some intron positions are more conserved than others. For instance, the last intron in phase 1 in 10/19 RTKs is likely the same intron position. On the other hand, most introns are unique to specific TK families, and thus, they have higher diagnostic importance.

Phylogenetic Analysis

TK sequences from *Anopheles gambiae* (mosquito), *Ciona intestinalis* (sea squirt), *Drosophila melanogaster*

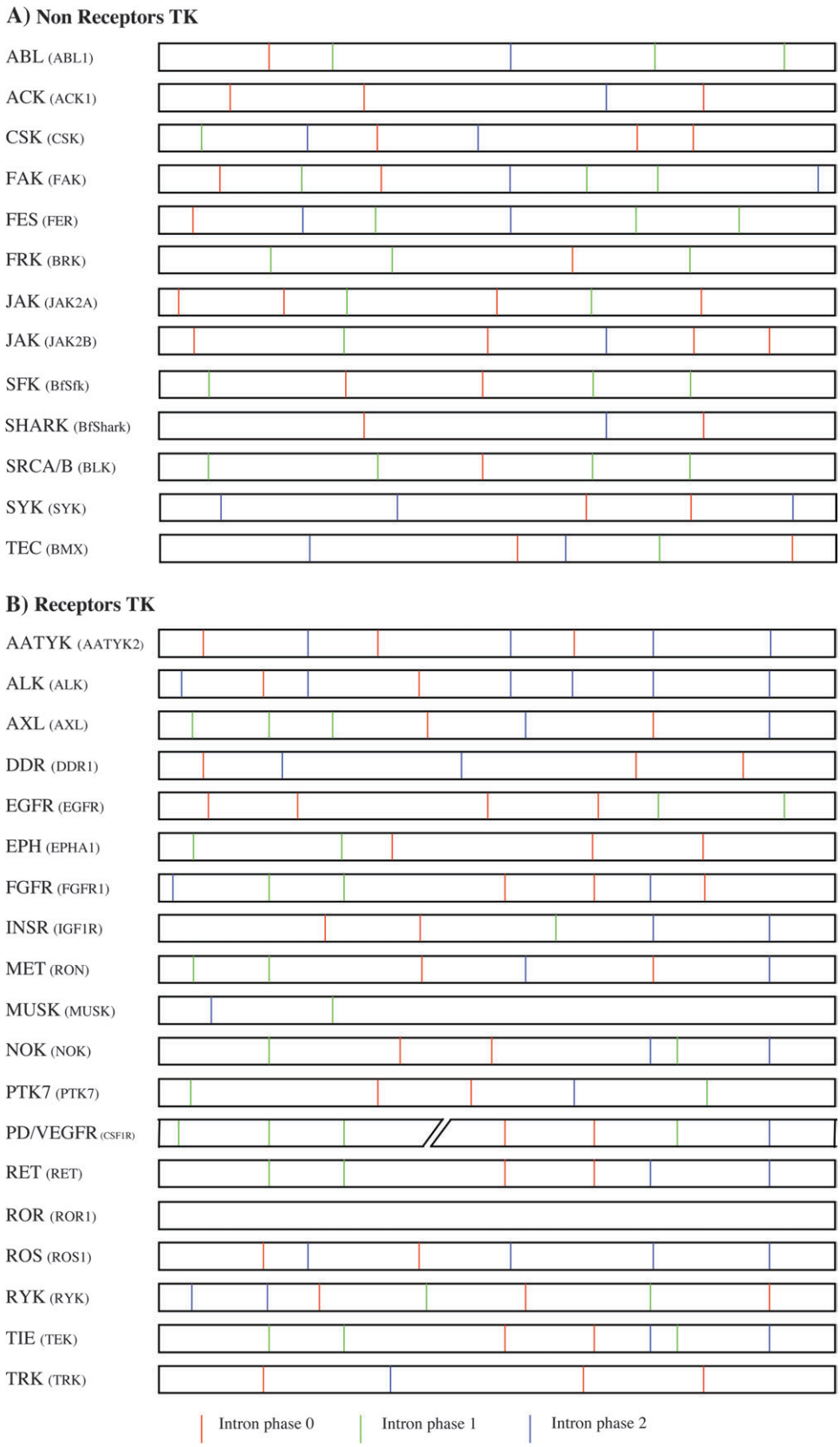
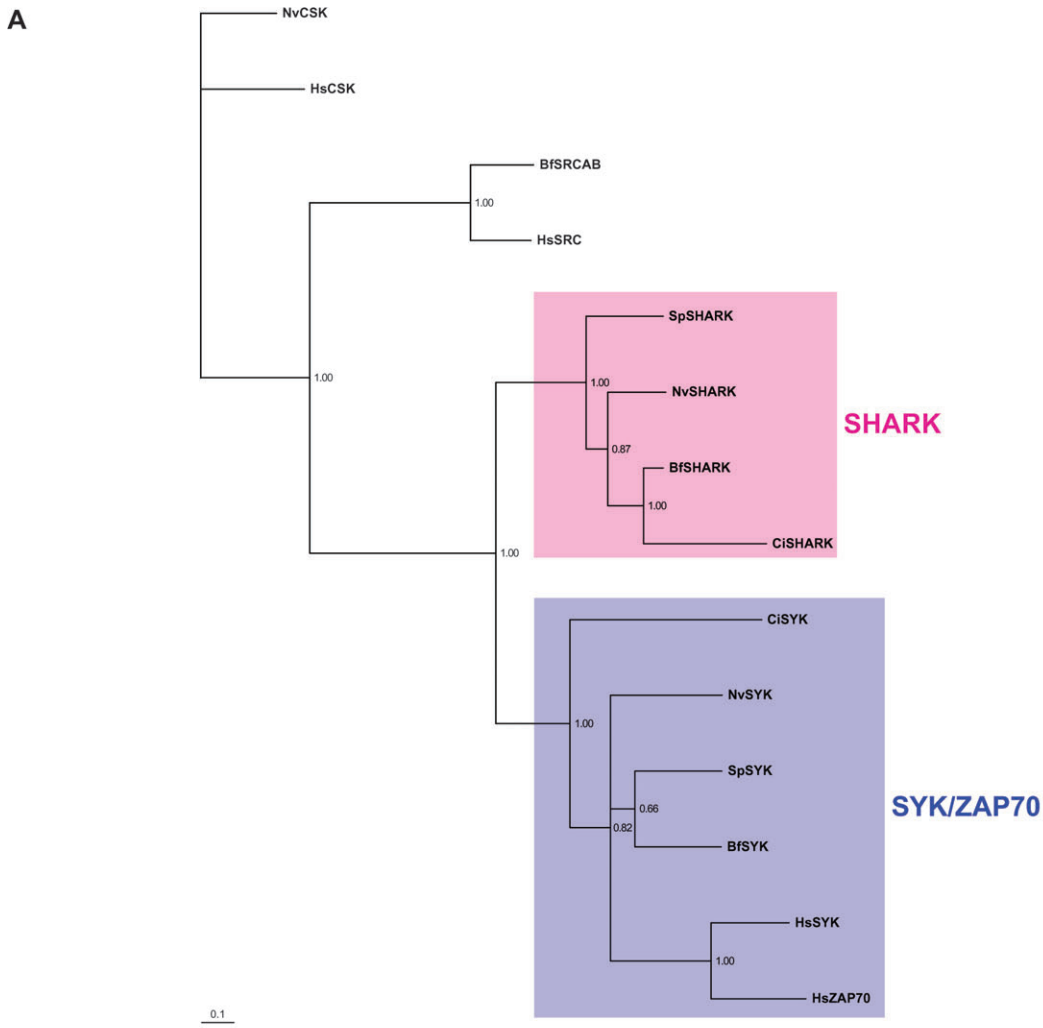


FIG. 1.—TK intron codes schematic representation of intron/exon structures within the TK domains (TK intron codes) of the different widespread TK families in metazoans. Each family is represented by one human member (in parentheses), except in the case of the SHARK and SFK families, for which the single amphioxus orthologs are shown. All the members of a given family show very similar intron codes within and between species. Red bars correspond to positions of phase 0 introns, green to phase 1 introns, and blue bars to phase 2 introns. The hydrophilic stretch of PD/VEGFR is not depicted, and the position in the TK domain is represented by a gap (//). *x* axis corresponds to the relative position of an intron along the alignment of all represented TK domains using ClustalW with default parameters.



B

```
SpSYK     lnq * gdq-i 1 gkgnfg--------svlkgtc-man * gql * ipvavktlkadfgipnse 0 p * eivreaelmagldhphivrmi 1 gichaaemml * vl
NvSYK     ktq * wsi-l 1 shgnfg--------svlkgtykmpn * ger * ipvavkslksdd-innpk 0 s * eilhearvmmeldhpyivrii 1 gmcqgpsmml * vm
BfSYK     lql * kek-l 1 gagnfg--------svlsgty-qlg * rkt * ievavktlkteq-vpnee 0 p * eilkeanmmkdldhphivrmi * gvchgetiml * vl
HsSYK     ltl * edkel * gsgnfg--------tvkkgyy-qmk 2 kvv * ktvavkilkneandpalk * d * ellaeanvmqqldnpyivrmi * giceaeswml * vm
HsZAP70   lli * adiel * gcgnfg--------svrqgvy-rmr 2 kkq * idvaikvlk-qgtekadt * e * emmreaqimhqldnpyivrli * gvcqaealml * vm
NvSHARK   lel * gse-l * gqgefg--------svlrgvwrdpk * gkk 0 vqvalktlhpek-ivhge * q 0 eflrearvmygldhpcivrli * gvclgpplil 0 vq
BfSHARK   lql * gre-l * gqgefg--------svlmgvwtspd * gre 0 vpvalktlrgeh-iqhge * q * eflrearvmmgldhfcivqli * gvclgppmmm 0 vq
SpSHARK   ier 1 gne-i * gqgefg--------svlegkyke-k * drw * vkvalktlhadh-lqtgq * k * eflreakvmcgldhpcivklm * gvclgppmml 0 vq
CiSHARK   lqe * aaa-r * gfndlvlelykhgadvktrdyegss * alh 0 ipvalktlhgeh-intge * t * efqreaevmmeldhpcivqly * gicrgetlmm 0 vq

SpSYK     elaelgplhkylkkh 2 q-emstrnvlelmyqvaq * gmcylesr * qfvhrdlaarnvllvdetf * akisdfgmskalgldsqyyv 0 aetagkwplk * wyap
NvSYK     elagegplnkylkkn 2 k-gmpllnilvlmlqvae 0 gmqylesl * qfvhrdlaarnvlvvndsf * vkisdfgmsramgagsdyyk 0 agtagrwplk * wyap
BfSYK     elaplgplnkylkkn 2 a-svrperivdlmtqvae * gmaylesk * nfvhrdlaarnillvself * akisdfgmskalglgsdyyk 0 tdtagkwplk * wyap
HsSYK     emaelgplnkylqqn 2 r-hvkdkniielvhqvsm * gmkylees * nfvhrdlaarnvllvtqhy * akisdfglskalradenyyk 0 aqthgkwpvk * wyap
HsZAP70   emagggplhkflvgk 2 reeipvsnvaellhqvsm * gmkyleek * nfvhrdlaarnvllvnrhy * akisdfglskalgaddsyyt 0 arsagkwplk * wyap
NvSHARK   elvtmgalldflidh * qpeisprelklwaaqiaw * gmmylekk * rfvhrdlatrnilmaskqq 0 lkisdfglsravgsgsdyyk * asqggrwpvk 2 wyap
BfSHARK   elvsmgsvldylldy * pqkvslpdlklwsaqias * gmmyleek * rfvhrdlaarnillackdq 0 lkisdfglsravgagsdyyk * asaggrwpvk 2 wyap
SpSHARK   elvsqgalldflqnd * s--itasdlklwatqias * gmmylegk * kfvhrdlaarnillenkgq 0 akisdfglsratganndyyr * sttggrwpvk 2 wyap
CiSHARK   elvamgsaldyildy * pmhtavsdfklwaaqias * gmtyleek 2 gfvhrdlalrnillaskql * vkisdfglsravgagsnyyk * asaggrwpvk 2 wyap

SpSYK     eciykfkfssksdvwsygvtlwealsygkrpya 0 smrgqe * lmqfien-gdrlsqpdrcpddvyslmr * rcwls 2 eakdrpgfgdiesvlsdil--
NvSYK     eciyfrkfsskdvwsygitlweatsygarpyq 0 slsgqa * ilekies-gyrlpapaklpscvyqlmk * dcwqw * e--------------------
BfSYK     eciyyykfsskadvwsygvtlweafsngrkpya 0 gmkgqd * iltwies-drrldrpptcseevyavmr * scwqy 2 kakdrptfaelsatmkdl---
HsSYK     ecinyykfsskdvwsfgvlmweafsygqkpyr 0 gmkgse * vtamlek-germgcpagcpremydlmn * lcwty 2 dvenrpgfaavelrlrn----
HsZAP70   ecinfrkfssrsdvwsygvtmwealsygqkpyk 0 kmkgpe * vmafieq-gkrmecppecppelyalms * dcwiy 2 kwedrpdfltveqrmrac---
NvSHARK   esinygtfshksdvwsygvtlwemysfgqlpyg * emtgge 0 vikmlenegkrldrpdacpeyvyklml * kcwdl * spenrptfnelhnif------
BfSHARK   esinygtfshasdvwsygvtlwemfsygqqpyg * dmtgaq 0 viqfieeegkrlskpdkcpekvyqiml * rcwsy * epsqrptfqmlntifsadpey
SpSHARK   esiyygtfshssdvwgfgvtlwemnsrgaqpyg * ektgae 0 vikqien-ghrlnrpegcpqnvyqimn * kcwsy * kpcnrptfsqlndmfrddpey
CiSHARK   esinygtfssssdvwsygvtlwemfsrgdqpyg * nmtgaq 0 viqyieddkrrlqkpdgcpdkvysims 2 qcway * qanerptfrrlhtkf------
```

(fruit fly), *Homo sapiens* (human), and *Takifugu rubripes* (pufferfish) were collected from the Ensembl (http://www.ensembl.org) and NCBI (http://www.ncbi.nlm.nih.gov) databases following published GenBank accession numbers (Robinson et al. 2000; Shiu and Li 2004). *Strongylocentrotus purpuratus* (purple sea urchin) sequences (Bradham et al. 2006) were downloaded from http://kinase.com. *Monosiga brevicollis* sequences were obtained by blasting the orthologous genes previously described in *Monosiga ovata* (King and Carroll 2001). TK domain amino acid sequences were aligned using ClustalW (Higgins et al. 1996) and manually reviewed. Bayesian inference (BI) trees were inferred using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), with the model recommended by ProtTest 1.4 (Drummond and Strimmer 2001; Guindon and Gascuel 2003; Abascal et al. 2005) under the Akaike information and the Bayesian information criterions (we used WAG + Γ + I model for the SRC, SFK, and SRC families and WAG + Γ model for the PDVEGFR and FGFR families and the SHARK and SYK families). Two independent runs were performed, each with 4 chains. For convention, convergence was reached when the value for the standard deviation of split frequencies stayed below 0.01. Burn-in was determined by plotting parameters across all runs for a given analysis: all trees prior to stationarity and convergence were discarded, and consensus trees were calculated for the remaining trees. In total, we used 2 MrBayes runs of 2,000,000 generations each and 350,000 generation burn-in for the SHARK and SYK families' analysis (1,650,000 postburn-in trees); 2 MrBayes runs of 5,500,000 generations each and 4,165,000 generation burn-in for the PDVEGFR and FGFR families' analysis (1,335,000 postburn-in trees); and 2 MrBayes runs of 8,250,000 generations each and 6,895,000 generation burn-in for the SRC, SFK, and SRC families' analysis (1,355,000 postburn-in trees).

Maximum likelihood (ML) analyses were performed using RAxML version 7.0.3 (Stamatakis 2006) with the model recommended by ProtTest, 1,000 bootstrap replicates and the rapid Bootstrapping algorithm. Phylogenetic trees obtained using ML had topologies consistent with those obtained by BI (data not shown).

Results and Discussion

TK Intron Code as Signature of Orthology and Paralogy

We have studied intron positions and phases within the TK domains (~270 amino acids in length) of the different members of all TK families in mammals. Despite the high similarity at the amino acidic level, the intron/exon structures were strikingly different among most of the different TK families, with generally fewer than 2 intron positions in common (fig. 1), with the exception of few TK families (e.g., ALK–AATYK). Thus, the pattern of intron positions/phases within the TK domain constitutes an intron code that contains valuable information about TK family membership.

Remarkably, the highly divergent TK intron codes observed among the different families allow for clear assignment of a given TK protein to a particular TK family (or small group of highly related TK families). Furthermore, as expected by the low rates of intron loss/gain in orthologous genes from cnidarians to mammals in the deuterostome line (Roy et al. 2003; Sullivan et al. 2006; Coulombe-Huntington and Majewski 2007; Putnam et al. 2007, 2008), orthologous members of the different TK families can be easily identified by the sharing of TK intron codes across wide evolutionary times: whereas intron codes from different TK families in the same species differ in all or nearly all intron positions, TK intron codes from orthologs rarely differ by more than 1 or 2 positions. We applied this strategy as an additional complementary criterion for identification of specific family members in amphioxus and the cnidarian *Nematostella vectensis* and to assess evolutionary relationships between TK families across metazoans.

An illustrative example of the utility of TK intron codes is the study of the SYK and SHARK families (fig. 2). The vertebrate SYK family has a similar organization of protein domains to the invertebrate SHARK family (Chan et al. 1994; Ferrante et al. 1995), and the 2 families have been considered to some extent counterparts (Shiu and Li 2004). Using TK intron code comparisons, we easily identified bona fide members of both families in amphioxus, *Nematostella*, sea urchin, and *Ciona* (fig. 2), indicating a very early split of these families at the origin of metazoans (Steele et al. 1999) and reciprocal losses of the SHARK family in vertebrates and of the SYK family in Ecdysozoans. The intron code similarities/divergences between the SYK and SHARK members (fig. 2*B*) are in agreement with classical phylogenetic analysis (fig. 2*A*).

Importantly, the intron code constitutes a qualitative homology identifier. Homology assignment by intron code similarity should be at least as reliable as those by standard phylogenetic methods, based on posterior probabilities. In cases in which the TK domain sequence does not provide enough phylogenetic information, or in cases of differential rates of sequence evolution, TK intron code helps to overcome these problems (see "BRK and SRC Families of Nonreceptor TK" for an example). Moreover, similarities in the TK intron code may indicate close phylogenetic relationship between different families, as in the case of MET and AXL or ALK and AATYK (fig. 1). On the other hand, for highly related groups of TK families that share the same intron code, only standard phylogenetic methods would allow for further

←

FIG. 2.—Example of the use of TK intron codes in phylogenetic classification (*A*) Phylogenetic analyses of the SYK and SHARK families. Bayesian phylogenetic tree of SHARK and SYK/ZAP70 genes from several metazoan species using the TK domain sequence, estimated under the WAG + Γ model (2 MrBayes runs of 2,000,000 generations each; 350,000 generation burn-in; 4 chains per run). ABL kinases were used as outgroups. Ag, *Anopheles gambiae*; Bf, *Branchiostoma floridae*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Nv, *Nematostella vectensis*; and Sp, *Strongylocentrotus purpuratus*. (*B*) Intron codes for the SYK/ZAP70 and SHARK families. Each intron position and phase is indicated by bold numbers. Blue: SYK/ZAP70 specific introns and purple: SHARK specific introns. Bf, *B. floridae*; Ci, *C. intestinalis*; Hs, *H. sapiens*; and Nv, *N. vectensis*. CiZap70 was not included in the alignment because it has lost all the introns within the TK domain.

Table 1
Number of TK Proteins of Each Family Identified in the Amphioxus Genome

Type	Gene Family	Genes in *Homo sapiens*	Genes in *Branchiostoma floridae*
Nonreceptor TKs	ABL	2	1
	ACK	2	3
	BRK	3	1 + 1 Ψ
	CSK	2	1
	FAK	2	1
	FES	2 + 1 Ψ	1
	JAK	4	1
	SFK	—	1 + 3 Ψ
	SHARK	—	1
	SRCA/B	4 + 4 + 1 Ψ	1
	SYK	2	1
	TEC	5	1
Receptor TKs	AATYK	3	1
	ALK	2	1
	AXL[b]	3 + 1 Ψ	1
	DDR	2	1
	EGFR	4	1
	EPHA/B	14	2
	FGFR[c]	4	1
	INSR	3	1
	MET[b]	2	1
	MUSK	1	1
	NOK	1	22
	PD/VEGFR[c]	5 + 3 + 1 Ψ	1
	PTK7	1	1
	RET[c]	1	1 + >100 Ψ
	ROR	2	1
	ROS	1	1
	RYK	1 + 1 Ψ	1
	TIE[c]	2	2 + 5
	TRK	3	1
	Other MARTK[b]	—	8
	Other EXTK[c]	—	47
Total		90 + 5 Ψ	118 + >100 Ψ[d]

[a] Data from Robinson et al. (2000).
[b] The MARTK superfamily includes the MET and AXL families.
[c] The EXTK superfamily includes the TIE, PD/VEGFR, RET, and FGFR families.
[d] The number of pseudogenes in amphioxus is likely an underestimate.

assignment. Therefore, the TK intron code is a useful tool to complement standard phylogenetic analysis.

Nonetheless, the utility of the intron code can go beyond the confirmation of phylogenetic analyses. Intron positions can also be used to improve protein annotations (Irimia and Roy 2008), and intron codes are especially useful when only the TK domains can be confidently predicted from the genome sequence, for instance, if no expression data are available, as is increasingly common with the explosion of genomic sequencing projects.

Amphioxus Has Members of All Widespread TK Families

We identified and annotated (see Methods) members of all widespread receptor and nonreceptor TK families previously described in vertebrates and protostomes in the amphioxus genome (table 1, fig. 3). The domain organization of the amphioxus counterparts always matches the domain organization of the multiple members in vertebrates (fig. 3). As expected from its nonduplicated genome, amphioxus possesses single members of most families, although it shows striking lineage-specific expansions (table 1). Importantly, amphioxus is the only known metazoan that has members of all widespread TK families (fig. 3).

Complex TK Repertoire at the Origin of Metazoans

We have also identified members of most TK families in the genome of the cnidarian *N. vectensis* (fig. 3). This high complexity at the base of the metazoans is in consonance with previous reports for other important developmental genes and networks (Kusserow et al. 2005; Miller et al. 2005; Matus et al. 2006, 2008). However, despite this high complexity, *Nematostella* has some notable absences, such as MUSK, MET, EGFR, AXL, ALK, AATYK, TIE, ROS, and RET. Interestingly, most of these genes are required for the development of complex organic systems, such as the nervous, circulatory, or immune systems (Sato et al. 1995; Alroy and Yarden 1997; Gaozza et al. 1997; Wang et al. 2002; Pulford et al. 2004; Bradham et al. 2006; Lemke 2006; Runeberg-Roos and Saarma 2007; Kim and Burden 2008), thus suggesting that the origin of these families could have played a role in the evolution of organismal complexity through bilaterian evolution.

Slow-Evolving Genomes Clarify the Evolutionary Relationships of Specific TK Families

BRK and SRC Families of Nonreceptor TKs

The BRK family has been tightly linked to breast cancer (Mitchell et al. 1994; Barker et al. 1997) but is also involved in normal development of the pancreas and small intestine (Haegebarth et al. 2006; Akerblom et al. 2007). On the other hand, SRC family members regulate several cellular processes, such as cell division, adhesion, and motility, and have also been associated with different types of cancer (Thomas and Brugge 1997). Both BRK and SRC families have high sequence similarity and share the same protein domain organization (one SH3 and one SH2 domain, in addition to the TK domain), making the classification of members of these families relatively difficult; however, their TK intron codes allow for a clear distinction (fig. 1 and supplementary fig. SM1 [Supplementary Material online]) (Serfas and Tyner 2003). Amphioxus BRK and SRC families consist of 1 and 2 members, respectively (plus 1 BRK and 3 SRC pseudogenic copies). Of the 2 SRC related members, one is the ancestral ortholog of both human SRC families (*BfSRCA/B*), whereas the other amphioxus member (*BfSFK1*) groups with the nonchordate genes in the phylogenetic analysis (fig. 4*A*).

In the cnidarian *N. vectensis*, we found 3 genes in tandem. One of the genes seems to be a member of the BRK family and another one appears basal to all SRC-related genes (the SRCA/B genes and the invertebrate SRC-like genes, red and yellow groups, respectively, in fig. 4*A*); the phylogenetic position of the third gene is not conclusive (fig. 4*A*).

Therefore, only chordates seem to have clear orthologs of the vertebrate family SRCA/B. The nonchordate genes (usually termed Src family kinases [O'Neill et al. 2004;

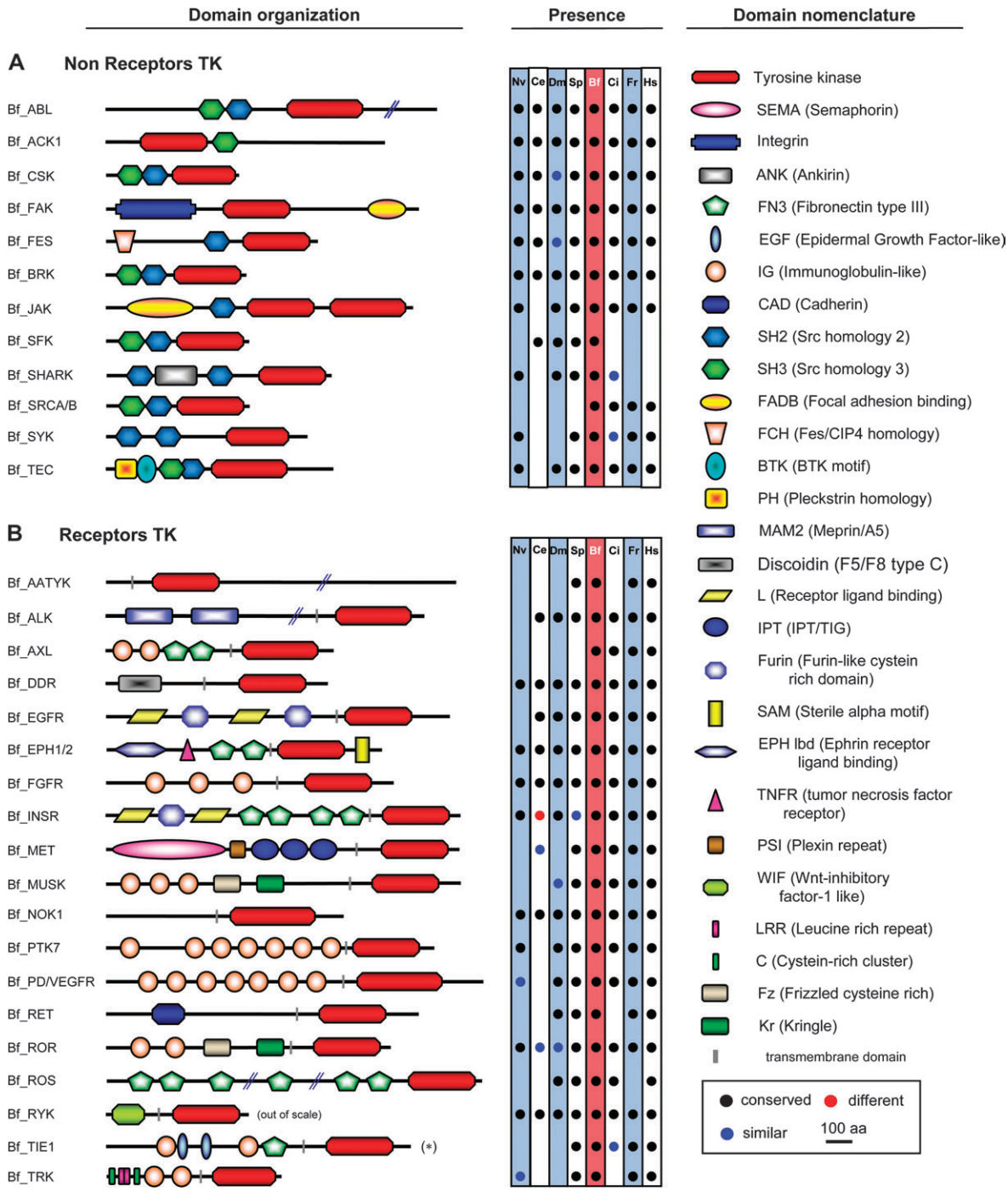


FIG. 3.—Protein domain organization of amphioxus TK proteins and presence/absence in metazoan lineages. (*A*) Nonreceptor TK proteins. (*B*) TK receptors. Protein domains were identified using Prosite and Conserved Domain (NCBI) software. The protein size is shown to scale, except where indicated by bars (//). Bf, *Branchiostoma floridae* and Nv, *Nematostella vectensis*. Data of *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Ciona intestinalis* (Ci), *Fugu rubripes* (Fr), and *Homo sapiens* (Hs) were taken from Shiu and Li (2004) and of *Strongylocentrotus purpuratus* (Sp) from Bradham et al. (2006).

Bradham et al. 2006]), as is also the case of the amphioxus *BfSFK1*, are only distantly related to their vertebrate counterparts (O'Neill et al. 2004; Shiu and Li 2004; Bradham et al. 2006): their intron code is slightly different from the SRCA/B family (supplementary fig. SM1, Supplementary Material online) and they constitute a separate monophyletic group (fig. 4*A*).

We suggest that only an SFK/BRK gene existed in the metazoan ancestor and that this gene was subsequently duplicated in tandem giving rise to an SFK and a BRK gene. Later, in chordates, an SFK gene was duplicated and one of the copies evolved into an ancestral SRC gene, ancestor of *BfSRCA/B*, and the vertebrate *SRCA* and *SRCB* families. Finally, the SFK family was lost both in the ancestor of

vertebrates and ascidians but maintained in amphioxus (fig. 4*B*).

FGFR and PDGFR/VEGFR Families of TK Receptors

FGFR receptors are an evolutionarily conserved and functionally diverse family with a broad range of biological functions in development and adult physiology (Itoh and Ornitz 2004). The PDGFR/VEGFR family is characterized by a long stretch of hydrophilic amino acids in the middle of the TK domain, and its members play different roles in development and organogenesis, especially in endothelium development and angiogenesis and vasculogenesis (Yancopoulos et al. 2000; Alvarez et al. 2006). Both FGFR and PDGFR/VEGFR families of RTKs have characteristic arrays of immunogloblin (Ig) domains at the extracellular portion of the protein (fig. 3). The amphioxus genome contains one canonical member of each FGFR and PDGFR/VEGFR families, the latest with an extracellular organization more similar to the vertebrate VEGFR submembers (fig. 3), as in the case of *Ciona* and sea urchin, which also have members more related to vertebrate VEGFRs by phylogenetic analyses (fig. 5*A*). Intriguingly, phylogenetic analyses place vertebrate PDGFRs at the base of the family (fig. 5*A*). However, a late origin of the PDGFR subfamily in the vertebrate lineage seems the most parsimonious explanation: if a PDGFR gene was already present in early deuterostomes, it would have to have been lost independently at least 3 times (sea urchin, ascidian, and amphioxus lineages). Instead, it is more likely that the PDGFR family has evolved at a higher evolutionary rate after its genesis by tandem duplication at the root of the vertebrate lineage, seeming a basal branch probably due to a long-branch attraction effect in the gene phylogeny (fig. 5*A*).

On the other hand, *Nematostella* genome does not contain any canonical member of the PDGFR/VEGFR family but it does contain 3 members basal to PDGFR/VEGFR. These members lack the typical hydrophilic stretch, suggesting that this stretch was later inserted in the bilaterian ancestors (fig. 5*B*).

Remarkable Lineage-Specific Expansions of Some TK Families in the Amphioxus Genome

MET and AXL Families of TK Receptors

MET proteins are required for liver development (Aoki et al. 1997; Gherardi et al. 2004) and macrophage differentiation (Wang et al. 2002), whereas AXL plays important roles in development of the immune, vascular, and central nervous systems (Bradham et al. 2006). Despite their different functions and extracellular domain organization, MET and AXL TK domains share a very similar intron code (fig. 1), indicating a close evolutionary relationship and a relatively recent split. In amphioxus, we identified a single canonical member of each of the AXL and MET families. However, in addition, we found 8 copies containing an MET-/AXL-related TK domain (similar by sequence and harboring an MET/AXL intron code), which we named Met/Axl-related TKs, MARTKs. Remarkably, the extracellular portion of these extra copies contained a varied combination of protein domains, suggesting that they were probably generated by exon shuffling in the amphioxus lineage.

NOK Family of Oncogenic TK Receptors

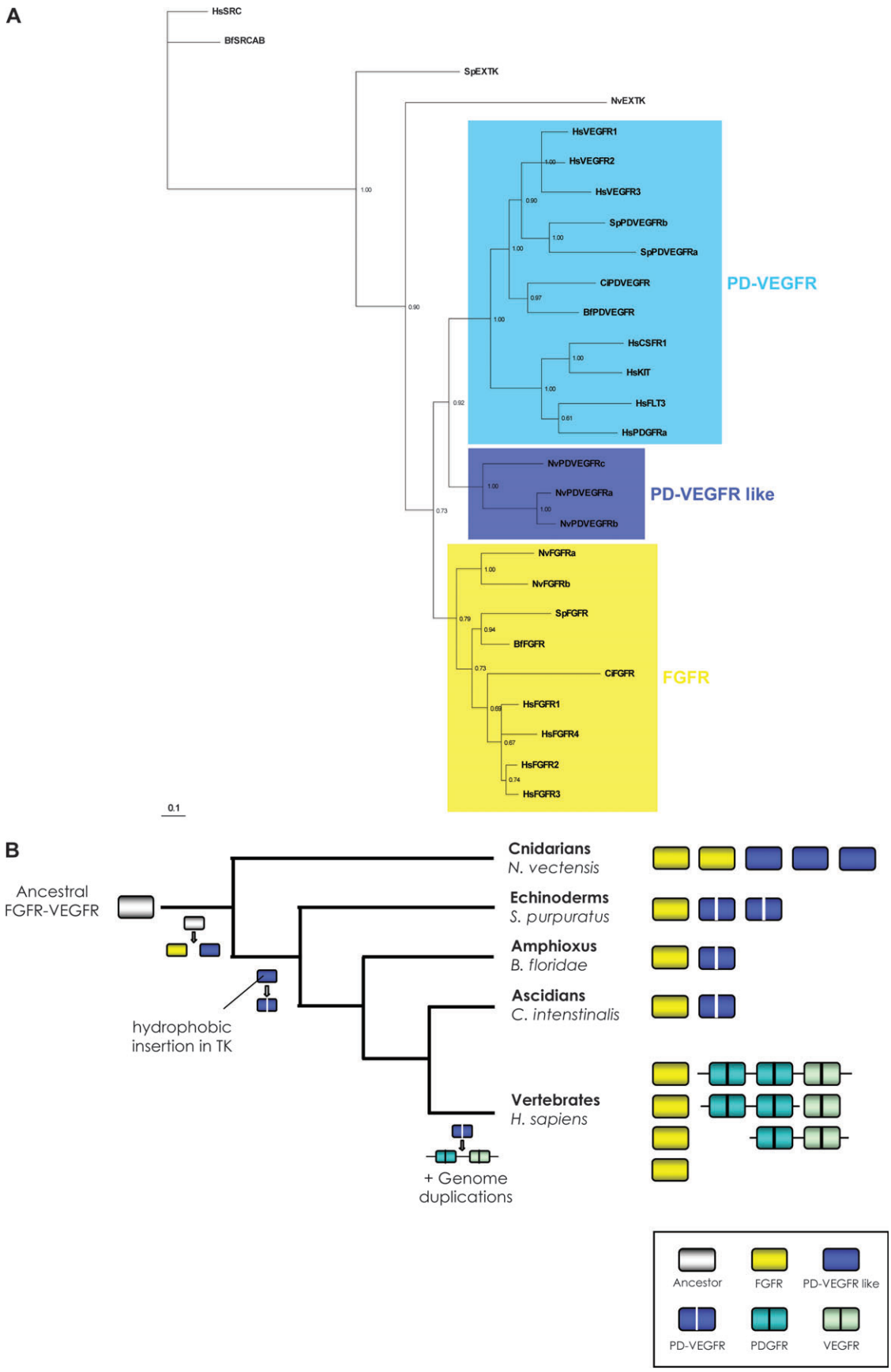
The vertebrate NOK family (after novel oncogene kinase [Liu et al. 2004]) has received little attention in the literature and has been nearly neglected from the evolutionary studies of TK proteins. Its cellular functions remain widely unknown, although it has been implicated with cancer (Liu et al. 2004). We found 22 NOK-related genes in amphioxus, easily recognizable by a distinct TK intron code (fig. 1 and supplementary fig. SM2 [Supplementary Material online]). However, in contrary to the mammalian protein, which does not show any recognizable extracellular domain, most of the amphioxus copies harbor a variety of extracellular domains (fig. 6*A*), again highlighting the propensity of the amphioxus genome for exon shuffling evolution. Remarkably, we also identified orthologs, with the characteristic NOK-like TK domains, in *Nematostella*, sea urchin, and *Ciona* genomes (fig. 1), an indication that this family predated the bilaterian origin and is highly conserved across metazoans.

TIE Family of TK Receptors

TIE members have been so far identified in vertebrates, sea urchin (Bradham et al. 2006) and *Ciona* (Shiu and Li 2004). Despite the fact that the different members do not have identical extracellular organization in all deuterostomes, TIE-like proteins are characterized by combinations of 3 different protein domains: FN3 (fibronectin type III), Ig (immunoglobulin), and EGFs (epidermal growth factor–like domains). In amphioxus, we found 2 distinguishable TIE receptors, plus at least 5 extra copies with distinct

←

FIG. 4.—Proposed scenario for the evolution of the BRK and SRC gene families in metazoans. (*A*) Phylogenetic analysis of the SRC/SFK and BRK TK families. Bayesian phylogenetic tree of SRC, SFK, and BRK genes from several metazoan species using the TK domain sequences, estimated under WAG + I + Γ model (2 MrBayes runs of 8,250,000 generations each; 6,895,000 generation burn-in; 4 chains per run). CSK and ABL kinases were used as outgroups. Ag, *Anopheles gambiae*; Am, *Asterina miniata*; Bf, *Branchiostoma floridae*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Nv, *Nematostella vectensis*; and Sp, *Strongylocentrotus purpuratus*. MbSFK like, *Monosiga brevicollis*; CiBRK1, CiBRK2, and CiBRK3 were previously published as SRC-Ci (*Ciona* specific). (*B*) An ancestral SFK/BRK gene (gray block) was duplicated in tandem at the root of the metazoan clade, giving rise to a BRK (green) and an SFK (yellow) gene. The ancestral organization in tandem (plus an extra lineage-specific duplication [olive green]) is still present in cnidarians. Along the echinoderm lineage, the SFK gene suffered an extensive expansion (up to 7 copies in sea urchin and at least 3 in starfish [Bradham et al. 2006]). In the ancestral chordate, the SFK gene duplicated resulting in an SFK gene and a vertebrate-like SRCA/B gene (red, which lost one ancestral SFK intron). The SFK gene was lost in the olfactorian clade. At the root of the vertebrates, prior to the whole-genome duplications, the SRCA/B gained its vertebrate-specific intron and duplicated in tandem (Gu J and Gu X 2003), subsequently generating the SRCA (pink) and SRCB (brown) families.



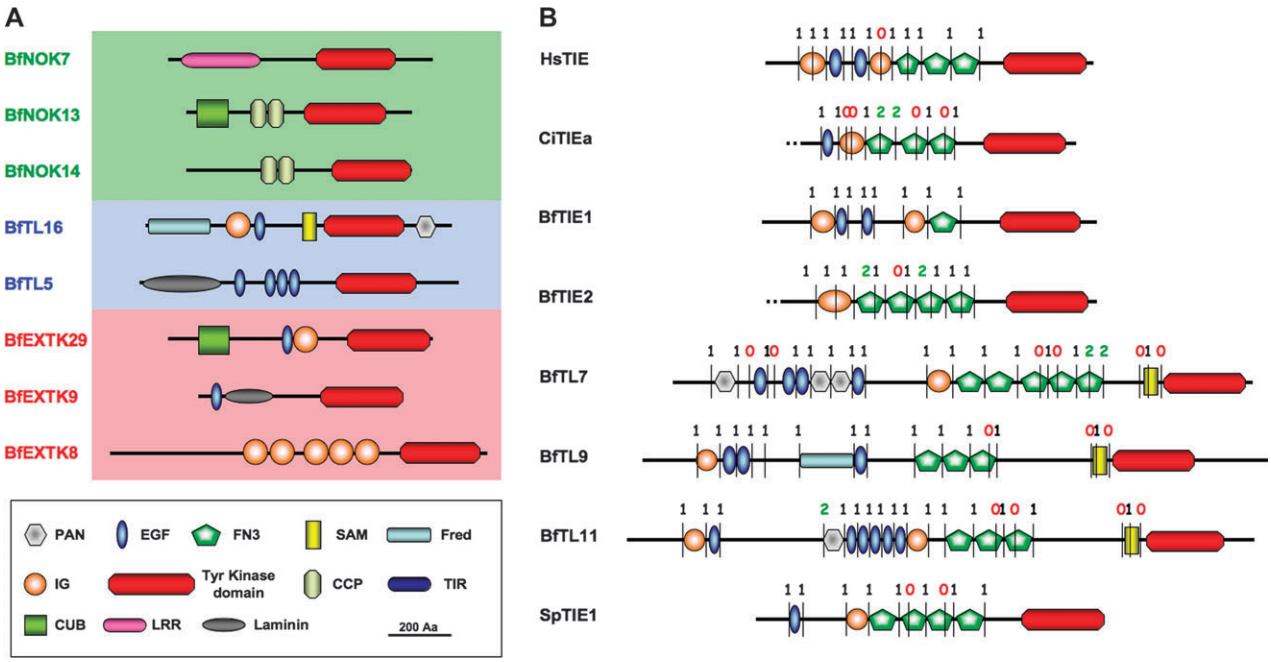


FIG. 6.—Exon shuffling and examples of new TK combinations in the amphioxus genome. (*A*) Examples of new domain combinations in specific RTKs of amphioxus. Green background, NOK-like proteins; blue background, TIE like proteins; and red background, EXTK proteins. (*B*) Intron phases and domain combinations in TIE and TIE-like in deuterostomes. Only introns surrounding the extracellular domains are shown: red, phase 0, black, phase 1; and green, phase 2. Hs, *Homo sapiens*; Ci, *Ciona intestinalis*; Bf, *Branchiostoma floridae*; and Sp, *Strongylocentrotus purpuratus*.

combinations of the characteristic extracellular domains of TIE proteins (fig. 6*B*). On the other hand, we have not been able to identify any TIE-like protein in the genome of *N. vectensis*, suggesting that the origin of the TIE family predates the origin of deuterostomes but is posterior to the split of cnidarians and the ancestor of deuterostomes.

Interestingly, TIE proteins are primarily involved in endothelium development (Sato et al. 1995), a tissue which is specific to vertebrates. Thus, the early deuterostome origin of this family suggests that these proteins played other functions, perhaps in primitive deuterostome circulatory systems without real endothelium, being later recruited in endothelium evolution in vertebrates.

EXTK: A New Superfamily of Related TK Proteins with Widespread Tendency to Gene Duplication and Exon Shuffling across Metazoans

Our characterization of the TK domain intron code in 5 metazoan genomes trigger us to suggest that RET, PDGFR/VEGFR, FGFR, and TIE families are very closely related and originated early in metazoan evolution from a single gene that harbor a unique 7-intron code in the TK domain (supplementary fig. SM2, Supplementary Material online). Further duplications and exon shuffling followed by specific TK intron losses in the lineage to deuterostomes accounted for the great diversification in these TK families.

Strikingly, the amphioxus genome contains more than 50 genes with this distinct TK domain, the largest expansion of TKs described so far (for comparison, humans have 15 members of these superfamily, after 2 rounds of whole-genome duplication, table 1). Appealingly, independent expansions of these families have also been reported in all studied metazoan clades (although in numbers not comparable to amphioxus), generally referred to as FGFR-like expansions (Manning et al. 2002; Shiu and Li 2004; Bradham et al. 2006). We thus propose a new superfamily of TK proteins, related by early gene duplication in metazoans, which we name EXTK (from EXpanding TKs).

←

FIG. 5.—Proposed scenario for the evolution of the PDGFR, VEGFR, and FGFR gene families in metazoans. (*A*) Phylogenetic analyses of the FGFR and PDGFR/VEGFR TK families. Bayesian phylogenetic tree of FGFR and PDGFR/VEGFR genes from several metazoan species using the TK domain sequence, estimated under WAG + I + Γ model (2 MrBayes runs of 5,500,000 generations each; 4,165,000 generation burn-in; 4 chains per run). SRCA/B members from amphioxus and human were used as outgroups. NvEXTK and SpEXTK were previously considered as fast-evolving FGFR members. (We suggest that they are indeed members of the EXTK superfamily, not necessarily more related to FGFR than the other members of the superfamily.) CiPD-VEGFR, SpPD-VEGFRa, SpPD-VEGFRb, NvPD-VEGFRa, NvPD-VEGFRb, NvPD-VEGFRc, SpFGFR, SpEXTK, and NvEXTK were previously published as CiVEGFR, SpVEGFR7, SpVEGFR10, NvVEGFRa, NvVEGFRb, NvVEGFR16, SpFGFR1, SpFGFR2, and NvFGFRc, respectively. Bf, *Branchiostoma floridae*; Ci, *Ciona intestinalis*; Hs, *Homo sapiens*; Nv, *Nematostella vectensis*; and Sp, *Strongylocentrotus purpuratus*. (*B*) The FGFR (yellow) and PDGFR/VEGFR (dark blue) families had a common ancestor early in the evolution of metazoans (gray). Before the split of cnidarians and bilaterians, a gene duplication event generated FGFR-like and PDGFR/VEGFR-like genes. The PDGFR-/VEGFR-like gene did not have yet the distinct hydrophilic insertion in the TK domain, which was acquired later in early bilaterians. In the vertebrate lineage, before the whole-genome duplications, this PDGFR-/VEGFR-like gene duplicated in tandem giving rise to a PDGFR gene (turquoise) that followed a faster rate of sequence evolution.

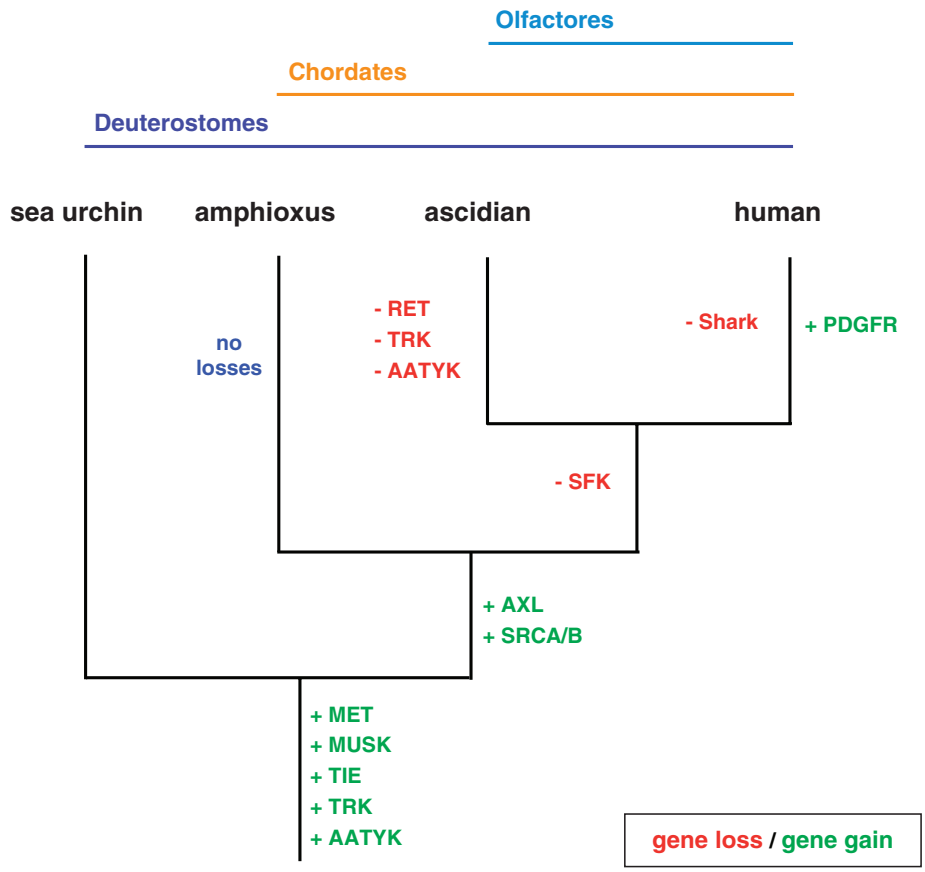


FIG. 7.—TK gene loss and gain within the deuterostomian clade. Gene families gained (green) and lost (red) are indicated at the relevant cross nodes leading to the groups analyzed: sea urchin (*Strongylocentrotus purpuratus*), amphioxus (*Branchiostoma floridae*), ascidian (*Ciona intestinalis*), and human (*Homo sapiens*). Gene gains at the basis of deuterostomes indicate those families not present in the genome of the cnidarian *Nematostella vectensis*.

More intriguingly, MET, AXL, and NOK families are also related to EXTK members, both phylogenetically and by intron code (supplementary fig. SM2, Supplementary Material online), and have also been expanded in amphioxus and in some other metazoans (Shiu and Li 2004). Hence, we hypothesize that, for uncertain reasons, the EXTK and related groups are more prone to undergo gene duplication and exon shuffling than other TK families, thus providing a major substrate for evolutionary innovation.

Expansion of RET Processed Pseudogenes

Finally, in addition to a single canonical RET receptor, we identified in the amphioxus genome more than 100 processed pseudogenes (i.e., sequences with high similarity and analogous domain organization to the canonical RET gene but lacking introns, as a result of their origin by retrotranscription of an mRNA [Vanin 1985; D'Errico et al. 2004; Irimia and Roy 2008]). We compared the adjacent regions of each copy and found that sequence conservation is limited to the coding region (data not shown), further supporting the origin by retroinsertion. Intriguingly, few of the copies include stop codons, the cadherin and TK domains are more conserved in sequence than the rest of the protein, and the average *Ka*/*Ks* ratio is ~0.5; these 3 data do not prove but strongly suggest that a fraction of these copies may be under negative purifying selection.

To our knowledge, this is the first report of such a massive expansion of any single processed pseudogene in metazoans, with a number of copies comparable to those of non-LTR transposable elements in the same species (Permanyer et al. 2006).

The TK Family in Amphioxus: Prototypical and Unique

In summary, our survey of TKs reveals 2 remarkable aspects of the amphioxus genome. First, it is the only genome where all the TK families are represented. It did not lose any of the genes present in the common ancestor of protostomes and deuterostomes, in contrast to vertebrates (fig. 7). These results underscore that amphioxus has retained most of the components of a prototypical chordate structure in its genome as well as in its body plan (Holland et al. 2008). The TK gene superfamily adds further arguments to the use of amphioxus genes in comparative studies as the reference clade for the origin of chordates and as a simple model system for vertebrates.

However, a second and perhaps more surprising and challenging feature of the amphioxus genome is its high degree of gene creation and expansion. The

unprecedentedly large expansion of the EXTK and related families in amphioxus compared with all other studied metazoans (57 EXTKs and 22 NOKs, compared with, for instance, 15 EXTKs and 1 NOK in humans; table 1) by gene duplication and exon shuffling, and of the RET receptor by retrointegration, may give future insights into the mechanisms of genome plasticity.

Due to these 2 features, the extreme tendency to gene retention and expansion, amphioxus harbors the richest TK repertoire among all metazoans studied so far. Amphioxus, widely considered an evolutionarily static organism, a living fossil, has not only retained most of the gene complement of its ancestors but has dramatically evolved its own repertoire of genetic novelties.

Supplementary Material

Supplementary table SM1, files 1 and 2, and figures SM1 and SM2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Scott W. Roy for critical reading of the manuscript and helpful discussions, Èlia Benito-Gutiérrez and Jon Permanyer for helpful comments on the work, and Eva Lázaro, Marta Riutort, Marta Álvarez-Presas, and Jordi Paps for assistance with phylogenetic analysis. JGF thanks Laura for unsolicited help and support. We thank the Joint Genome Institute for the amphioxus genome sequence resources. This work was funded by grant BFU2005-00252 from the Ministerio de Educación y Ciencia (MEC), Spain. S.A. holds a Juan de la Cierva postdoctoral contract from MEC and S.B. is an EMBO postdoctoral fellow. M.I. holds FPI and I.M. FPU fellowships (MEC) and J.P.A. an FI fellowship (Generalitat de Catalunya).

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 21: 2104–2105.

Akerblom B, Annerén C, Welsh M. 2007. A role of FRK in regulation of embryonal pancreatic beta cell formation. Mol Cell Endocrinol. 270:73–78.

Alroy I, Yarden Y. 1997. The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. FEBS Lett. 410: 83–86.

Alvarez RH, Kantarjian HM, Cortes JE. 2006. Biology of platelet-derived growth factor and its involvement in disease. Mayo Clin Proc. 81:1241–1257.

Aoki S, Takahashi K, Matsumoto K, Nakamura T. 1997. Activation of Met tyrosine kinase by hepatocyte growth factor is essential for internal organogenesis in Xenopus embryo. Biochem Biophys Res Commun. 234:8–14.

Barker KT, Jackson LE, Crompton MR. 1997. BRK tyrosine kinase expression in a high proportion of human breast carcinomas. Oncogene. 15:799–805.

Benito-Gutierrez E, Garcia-Fernandez J, Comella JX. 2006. Origin and evolution of the Trk family of neurotrophic receptors. Mol Cell Neurosci. 31:179–192.

Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA. 2006. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annu Rev Biochem. 75:655–680.

Birney E, Durbin R. 2000. Using GeneWise in the Drosophila annotation experiment. Genome Res. 10:547–548.

Bradham CA, Foltz KR, Beane WS, et al. (21 co-authors). 2006. The sea urchin kinome: a first look. Dev Biol. 300: 180–193.

Coghlan A, Durbin R. 2007. Genomix: a method for combining gene-finders' predictions, which uses evolutionary conservation of sequence and intron-exon structure. Bioinformatics. 23:1468–1475.

Coulombe-Huntington J, Majewski J. 2007. Characterization of intron loss events in mammals. Genome Res. 17:23–32.

Chan TA, Chu CA, Rauen KA, Kroiher M, Tatarewicz SM, Steele RE. 1994. Identification of a gene encoding a novel protein-tyrosine kinase containing SH2 domains and ankyrin-like repeats. Oncogene. 9:1253–1259.

Chang Y-M, Kung H-J, Evans CP. 2007. Nonreceptor tyrosine kinases in prostate cancer. Neoplasia. 9:90–100.

Davidson EH, Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. Science. 311:796–800.

D'Errico I, Gadaleta G, Saccone C. 2004. Pseudogenes in metazoa: origin and features. Brief Funct Genomic Proteomic. 3:157–167.

Drummond A, Strimmer K. 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics. 17:662–663.

Ferrante A Jr, Reinke R, Stanley E. 1995. Shark, a Src homology 2, ankyrin repeat, tyrosine kinase, is expressed on the apical surfaces of ectordermal epithelia. Proc Natl Acad Sci USA. 92:1911–1915.

Gaozza E, Baker SJ, Vora RK, Reddy EP. 1997. AATYK: a novel tyrosine kinase induced during growth arrest and apoptosis of myeloid cells. Oncogene. 15:3127–3135.

Geer PV, Hunter T, Lindberg RA. 1994. Receptor protein-tyrosine kinases and their signal transduction pathways. Annu Rev Cell Biol. 10:251–337.

Gherardi E, Love CA, Esnouf RM, Jones EY. 2004. The sema domain. Curr Opin Struct Biol. 14:669–678.

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

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Haegebarth A, Bie W, Yang R, Crawford SE, Vasioukhin V, Fuchs E, Tyner AL. 2006. Protein tyrosine kinase 6 negatively regulates growth and promotes enterocyte differentiation in the small intestine. Mol Cell Biol. 26:4949–4957.

Higgins DG, Thompson JD, Gibson TJ. 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266: 383–402.

Holland LZ, Satoh N, Azumi K, et al. (62 co-authors). 2008. Primitive and derived characters in the amphioxus genome. Genome Res. doi 10.1101/gr.073676.107.

Hubbard SR, Till JH. 2000. Protein tyrosine kinase structure and function. Annu Rev Biochem. 69:373–398.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA. 2006. The PROSITE database. Nucleic Acids Res. 34:D227–D230.

Hunter T. 1998. The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. Philos Trans R Soc Lond B Biol Sci. 353:583–605.

Irimia M, Roy SW. 2008. Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res. 36:1703–1712.
Itoh N, Ornitz DM. 2004. Evolution of the Fgf and Fgfr gene families. Trends Genet. 20:563–569.
Kim N, Burden SJ. 2008. MuSK controls where motor axons grow and form synapses. Nat Neurosci. 11:19–27.
King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proc Natl Acad Sci USA. 98:15032–15037.
Kusserow A, Pang K, Sturm C, et al. (11 co-authors). 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature. 433:156–160.
Lemke G. 2006. Neuregulin-1 and Myelination. Science's STKE: signal transduction knowledge environment 2006(325):pe11.
Liu L, Yu X-Z, Li T-S, et al. (13 co-authors). 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.
Lukashin A, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107–1115.
Manning G, Plowman GD, Hunter T, Sudarsanam S. 2002. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci. 27:514–520.
Marchler-Bauer A, Anderson JB, Derbyshire MK, et al. (25 co-authors). 2007. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. D237–D240.
Matus DQ, Magie CR, Pang K, Martindale MQ, Thomsen GH. 2008. The Hedgehog gene family of the cnidarian, *Nematostella vectensis*, and implications for understanding metazoan Hedgehog pathway evolution. Dev Biol. 313:501–518.
Matus DQ, Pang K, Marlow H, Dunn CW, Thomsen GH, Martindale MQ. 2006. Molecular evidence for deep evolutionary roots of bilaterality in animal development. Proc Natl Acad Sci USA. 103:11195–11200.
Miller DJ, Ball EE, Technau U. 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. Trends Genet. 21: 536–539.
Miranda-Saavedra D, Barton GJ. 2007. Classification and functional annotation of eukaryotic protein kinases. Proteins. 68:893–914.
Mitchell PJ, Barker KT, Martindale JE, Kamalati T, Lowe PN, Page MJ, Gusterson BA, Crompton MR. 1994. Cloning and characterisation of cDNAs encoding a novel non-receptor tyrosine kinase, brk, expressed in human breast tumours. Oncogene. 9:2383–2390.
Müller WE, Kruse M, Blumbach B, Skorokhod A, Müller IM. 1999. Gene structure and function of tyrosine kinases in the marine sponge *Geodia cydonium*: autapomorphic characters of Metazoa. Gene. 238:179–193.
Nelson EG, Grandis JR. 2007. Aberrant kinase signaling: lessons from head and neck cancer. Future Oncol. 3:353–361.
O'Neill FJ, Gillett J, Foltz KR. 2004. Distinct roles for multiple Src family kinases at fertilization. J Cell Sci. 117:6227–6238.
Parra G, Blanco E, Guigo R. 2000. GeneID in Drosophila. Genome Res. 10:511–515.
Patthy L. 2003. Modular assembly of genes and the evolution of new functions. Genetica. 118:217–231.
Pawson T. 1995. Protein modules and signalling networks. Nature. 373:573–580.
Permanyer J, Albalat R, Gonzàlez-Duarte R. 2006. Getting closer to a pre-vertebrate genome: the non-LTR retrotransposons of *Branchiostoma floridae*. Int J Biol Sci. 2:48–53.
Pires-daSilva A, Sommer RJ. 2003. The evolution of signalling pathways in animal development. Nat Rev Genet. 4:39–49.
Pulford K, Lamant L, Espinos E, Jiang Q, Xue L, Turturro F, Delsol G, Morris SW. 2004. Oncogenic protein tyrosine kinases. Cell Mol Life Sci. 61:2939–2953.
Putnam N, Butts T, Ferrier DEK, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature. 453:1064–1071.
Putnam NH, Srivastava M, Hellsten U, et al. (19 co-authors). 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 317:86–94.
Robinson DR, Wu Y-M, Lin S-F. 2000. The protein tyrosine kinase family of the human genome. Oncogene. 19:5548–5557.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Roy S, Fedorov A, Gilbert W. 2003. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci USA. 100:7158–7162.
Runeberg-Roos P, Saarma M. 2007. Neurotrophic factor receptor RET: structure, cell biology, and inherited diseases. Ann Med. 39:572–580.
Sato TN, Tozawa Y, Deutsch U, Wolburg-Buchholz K, Fujiwara Y, Gendron-Maguire M, Gridley T, Wolburg H, Risau W, Qin Y. 1995. Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. Nature. 376:70–74.
Serfas MS, Tyner AL. 2003. Brk, Srm, Frk, and Src42A form a distinct family of intracellular Src-like tyrosine kinases. Oncol Res. 13:409–419.
Shiu S-H, Li W-H. 2004. Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. Mol Biol Evol. 21:828–840.
Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A. 2007. Free full text 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.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Steele RE, Stover NA, Sakaguchi M. 1999. Appearance and disappearance of Syk family protein-tyrosine kinase genes during metazoan evolution. Gene. 239:91–97.
Sullivan JC, Reitzel AM, Finnerty JR. 2006. A high percentage of introns in human genes were present early in animal evolution: evidence from the basal metazoan *Nematostella vectensis*. Genome Inform. 17:219–229.
Thomas SM, Brugge JS. 1997. Cellular functions regulated by Src family kinases. Annu Rev Cell Dev Biol. 13:513–609.
Vanin EF. 1985. Processed pseudogenes: chracteristics and evolution. Annu Rev Genet. 19:253–272.
Wang MH, Zhou YQ, Chen YQ. 2002. Macrophage-stimulating protein and RON receptor tyrosine kinase: potential regulators of macrophage inflammatory activities. Scand J Immunol. 56:545–553.
Yancopoulos GD, Davis S, Gale NW, Rudge JS, Wiegand SJ, Holash J. 2000. Vascular-specific growth factors and blood vessel formation. Nature. 407:242–248.
Yeh R-F, Lim LP, Burge CB. 2001. Computational inference of homologous gene structures in the human genome. Genome Res. 11:803–816.

Barbara Holland, Associate Editor

Accepted June 3, 2008