J Mol Evol (1999) 49:601–608

JOURNAL OF MOLECULAR EVOLUTION

Protein Tyrosine Kinase cDNAs from Amphioxus, Hagfish, and Lamprey: Isoform Duplications Around the Divergence of Cyclostomes and Gnathostomes

Hiroshi Suga,[1] Daisuke Hoshiyama,[1] Shigehiro Kuraku,[1] Kazutaka Katoh,[1] Kaoru Kubokawa,[2] Takashi Miyata[1]

[1] Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
[2] Ocean Research Institute, The University of Tokyo, Nakano-ku, Tokyo 164-8639, Japan

Received: 28 April 1998 / Accepted: 30 June 1999

Abstract. Animals evolved a variety of gene families involved in cell–cell communication and developmental control by gene duplication and domain shuffling. Each family is made up of several subtypes or subfamilies with distinct structures and functions, which diverged by gene duplications and domain shufflings before the divergence of parazoans and eumetazoans. Since the separation from protostomes, vertebrates expanded the multiplicity of members (isoforms) in the same subfamily by further gene duplications in their early evolution before the fish–tetrapod split. To know the dates of isoform duplications more closely, we have conducted isolation and sequencing cDNAs encoding the fibroblast growth factor receptor, Eph, *src*, and platelet-derived growth factor receptor subtypes belonging to the protein tyrosine kinase family from *Branchiostoma belcheri*, an amphioxus, *Eptatretus burgeri*, a hagfish, and *Lampetra reissneri*, a lamprey. From a phylogenetic tree of each subfamily inferred from a maximum likelihood (ML) method, together with a bootstrap analysis based on the ML method, we have shown that the isoform duplications frequently occurred in the early evolution of vertebrates around or just before the divergence of cyclostomes and gnathostomes by gene duplications and possibly chromosomal duplications.

Key words: Protein tyrosine kinases — isoforms — hagfish — lamprey — amphioxus — gene duplication — phylogenetic tree — evolution

Introduction

Multicellular animals evolved a variety of gene families involved in cell–cell communication and developmental control by gene duplication and domain shuffling. Each family diverged from one or a few ancestral genes, which are shared with plants and fungi, or from an ancestral gene created uniquely in animal lineage (e.g., Iwabe et al. 1996). From phylogenetic analyses of several gene families involved in the signal transduction and developmental control, we recently showed that the pattern of gene diversification is characterized by two active periods in gene duplication interrupted by considerably long periods of silence, instead of proceeding gradually (Suga et al. 1997, 1999; Koyanagi et al. 1998a, 1998b; Hoshiyama et al. 1998; Ono et al. 1999). In the early period before the parazoan–eumetazoan split, animals underwent extensive gene duplications (subtype duplications) that gave rise to different subtypes with diverse functions. Almost complete sets of present-day subtypes had been established within that period. After the divergence

Data deposition: The nucleotide sequence data reported in this paper have been deposited in the DDBJ, EMBL, and GenBank nucleotide sequence databases (accession nos. AB025534–AB025557)
Correspondence to: T. Miyata; *e-mail:* miyata@biophys.kyoto-u.ac.jp

Table 1. Degenerate primers used for the cloning of cDNAs from cyclostomes and amphioxus

	Primer			
Subtype		Name	Sequence	Amino acid sequence
FGFR	sense	S0	5′-GGTAARCCIYTIGGIGARGGIKSNTTYGG-3′	GKPLGEG(C/A)FG
		S1	5′-GGCAAGCAYAARAAYATIRTIAAYYT-3′	GKHKNI(I/V)NL
	antisense	A1	5′-GACGACICCIWRISHCCANAYRTC-3′	D(V/I)W(S/A)(F/Y/L)GVV
Eph	sense	S0	5′-GTIATIGGIGNIGGNGARTTYGG-3′	VIGXGEFG
		S1	5′-GTICAYMGIGAYYTIGCIRCNMGNAA-3′	VHRDLA(A/T)RN
	antisense	A1	5′-GCYTCIGGIGCIGTCCANC-3′	RWTAPEA
		A2	5′-GTCCAICKIAYIGGIATYTT-3′	KIP(I/V)RWT
src	sense	S1	5′-GGAVMNGGIYRITTYGGIGANGTNTGG-3′	GXG(Q/C)FG(E/D)VW
		S2	5′-GGNGARGTNTGGATGGGNACNTGGAAYGG-3′	GEVWMGTWNG
	antisense	A1	5′-GCYTCIGGNGCIGTCCAYTTDATNGG-3′	PIKWTAPEA
		A2	5′-GGYTCNGTNGCNGTRAARTARTCYTCNARRAA-3′	FLEDYFTATEP
PDGFR	sense	S0	5′-GGTMRACCIYTIGGINVIGGIGCNTTYGG-3′	G(K/R)PLGXGAFG
		S1	5′-GGCAAGCAYAARAAYATIRTIAAYYT-3′	GKHKNI(I/V)NL
	antisense	A1	5′-GACGACICCIWRISHCCANAYRTC-3′	D(V/I)W(S/A)(F/Y/L)GVV

The identical primer is used for the cloning of cDNAs from *Branchiostoma belcheri* (amphioxus), *Eptatretus burgeri* (hagfish), and *Lampetra reissneri* (lamprey), except for the lamprey Eph (see Materials and Methods).

of protostomes and deuterostomes, the multiplicity of members in the same subtype rapidly increased in the first half of chordate evolution before the fish–tetrapod split by further gene duplications (isoform duplications), which gave rise to different isoforms; in most cases, different isoforms in the same subfamily are virtually identical in structure and function, but differ in tissue distribution (for isoforms in the protein tyrosine kinase [PTK] family, see Kraus et al. 1989; Partanen et al. 1991; Shier and Watt 1992; Mustelin and Burn 1993; Takahashi and Shirasawa 1994; Fox et al. 1995). A remarkable consequence suggested by these analyses is that there might be no direct link between the Cambrian explosion and the burst of subtype duplication.

The extensive isoform duplication in the first half of chordate evolution has been identified in many subtypes belonging to various gene families (Miyata et al. 1994; Iwabe et al. 1996; Suga et al. 1997, 1999; Hoshiyama et al. 1998). To know the dates of the active period of isoform duplication more closely, we have conducted isolation and sequencing cDNAs encoding the fibroblast growth factor receptor (FGFR), Eph, *src*, and platelet-derived growth factor receptor (PDGFR) subtypes of the PTK family from an amphioxus, a hagfish, and a lamprey. We report here that the extensive isoform duplication is observed in a period around or immediately before the cyclostome–gnathostome split.

Materials and Methods

Isolation and Sequencing of Cyclostome and Amphioxus cDNAs

Total RNA of *Eptatretus burgeri*, a hagfish, was extracted from each of the liver, brain, and vestigial eye, and those of *Branchiostoma belcheri*, an amphioxus, and larva of *Lampetra reissneri*, a lamprey, were extracted from the whole body using TRIZOL Reagent (Gibco BRL). These total RNAs were reverse-transcribed to cDNAs using oligo(dT) primer with reverse transcriptase (SuperScript II, Gibco BRL) and were used as templates for PCR amplifications with Expand High-Fidelity PCR System (Roche). The sense and antisense degenerate primers were designed from conserved amino acid residues as shown in Table 1. PCR amplifications were carried out under annealing condition of 46°C with the primers S1 and A1 (see Table 1) for all subtypes shown in Table 1, followed by nested PCR with S1 and A2 for the Eph subtype of *L. reissneri*. In *E. burgeri*, cDNAs prepared from the liver, brain, and vestigial eye were used as templates, except for the Eph subtype, for which the cDNA from the brain was used. For the *src* subtype, we carried out another PCR procedure with the primers S1 and A2 under annealing condition of 46°C, followed by nested PCR with S2 and A1 under 50°C.

The PCR-amplified fragments were purified and cloned into the pT7Blue vector (Novagen). More than three independent clones were isolated for each gene and sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit and ABI PRISM 377 DNA Sequencer (Perkin-Elmer).

The 3′ ends of cDNAs were amplified using 3′ RACE System for Rapid Amplification of cDNA Ends (Gibco BRL). The 5′ portions of cDNAs belonging to the FGFR, PDGFR, and Eph subtypes were amplified by PCR using respective subtype-specific sense primers (S0) and gene-specific antisense primers. These amplified fragments were purified, subcloned, and sequenced in the same way as above.

Sequence Alignment and Phylogenetic Tree Inference

Alignments were made by the methods of Needleman and Wunsch (1970) and Berger and Munson (1991), together with manual inspection.

The method of phylogenetic tree inference we adopted here is an approximate method for inferring the maximum likelihood (ML) tree of protein phylogeny (Kishino et al. 1990). This method consists of performing rearrangement of tree topology for a limited number of initial trees by the methods of nearest-neighbor interchange (NNI; Swofford et al. 1996; Adachi and Hasegawa 1996) and subtree pruning and regrafting (SPR; Swofford et al. 1996). The calculation procedure is as follows:

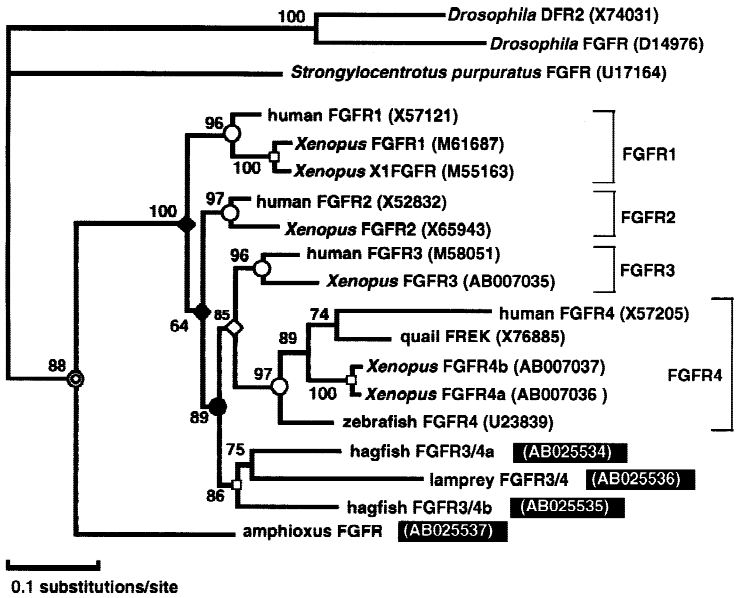

Fig. 1. Maximum likelihood tree of FGFR subfamily. From a comparison of the kinase domain sequences, the tree was inferred by an approximate ML method described in Materials and Methods, using *Drosophila* FGFRs as an outgroup. The number at each branch node represents the local bootstrap probability estimated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994). Open circles, fish–tetrapod split or amphibian–amniote split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split, but antedates the fish–tetrapod split; filled rhombi, gene duplications that antedate the cyclostome–gnathostome split; open boxes, gene duplications on the lineage leading to cyclostomes or *Xenopus*. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

1. Based on the neighbor joining (NJ) method (Saitou and Nei 1987) using the distance matrix estimated by the ML method (Adachi and Hasegawa 1996), 4,000 tree topologies, including that inferred from actual alignment are generated by the bootstrap resampling procedure (Felsenstein 1985).
2. For each of the 4,000 trees, the log-likelihood value is calculated by the ML method of protein phylogeny (Kishino et al. 1990; Adachi and Hasegawa 1996) based on the JTT model (PROTML version 2.3 in Adachi and Hasegawa's program package MOLPHY), and the top 40 trees are selected by log-likelihood criterion.
3. The tree topology with the largest log-likelihood value (approximate ML tree) is searched through repeated local rearrangements (Adachi and Hasegawa 1996) by NNI for all the 40 trees obtained in step 2.
4. To exclude a possibility that the approximate ML tree obtained in step 3 is local optimum, the approximate ML tree is subjected to further rearrangements by SPR. If no tree with log-likelihood value being larger than that of the approximate ML tree in step 3 is generated by SPR, the calculation procedure is completed. Alternatively if trees with larger log-likelihood value are obtained, the procedures in steps 3 and 4 are repeated for these trees.

The local bootstrap probability (LBP) at each node of a tree was calculated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994).

Results and Discussion

Each of the animal gene families involved in cell–cell communication and developmental control diverged from one or a few ancestral genes during animal evolution by gene duplications and domain shufflings. The family tree comprises several independent clusters corresponding to different subtypes or subfamilies that diverged before the divergence of parazoans and eumetazoans, the earliest divergence among extant animal phyla (Koyanagi et al. 1998a, 1998b; Hoshiyama et al. 1998; Siga et al. 1999; Ono et al. 1999). After the separation from protostomes, chordates expanded the multiplicity of members in the same subfamily by further gene duplications (isoform duplications). These isoform duplications have been completed until dates before the fish–tetrapod split (Miyata et al. 1994; Iwabe et al. 1996; Suga et al. 1997, 1999; Hoshiyama et al. 1998).

To determine the divergence times of these isoforms more closely, we have conducted isolation and sequencing cDNAs encoding the FGFR, Eph, *src*, and PDGFR subtypes belonging to the PTK family from *B. belcheri*, *E. burgeri*, and *L. reissneri* by the method described in Materials and Methods. Including these sequences, phylogenetic trees of the four subfamilies have been inferred by the ML analysis described in Materials and Methods.

Phylogenetic Tree of FGFR Subfamily

At least four distinct members (FGFR1–FGFR4) belonging to the FGFR subfamily have already been identified in gnathostomes, and they exhibit different tissue distribution (Partanen et al. 1991). We have isolated two cDNAs from *E. burgeri*, one cDNA from *L. reissneri*, and one cDNA from *B. belcheri*, and their nucleotide sequences have been determined for regions encoding the kinase domain. The amino acid sequences encoded by these cDNAs were aligned with known FGFR sequences for a highly conserved region corresponding to amino acid positions 491–780 in human FGFR1, for which unambiguous alignment is possible (alignment not shown). On the basis of the alignment comprising 270 amino acid sites, excluding gap positions, a phylogentic tree was inferred, using two *Drosophila* FGFRs as an outgroup. The obtained ML tree is shown in Fig. 1.

According to Fig. 1, the three gene duplications that gave rise to the four gnathostome FGFRs obviously antedate the divergence of amphibian and amniotes, as previously shown (Iwabe et al. 1996; Suga et al. 1997). It is also evident that these gene duplications postdate the

cephalochordate–vertebrate split. In the accompanying paper (Kuraku et al. 1999), we have shown that lampreys and hagfishes form a monophyletic group (see also Stock and Whitt 1992; Mallatt and Sullivan 1998). The FGFR tree also supports the cyclostome monophyly; the hagfish FGFR3/4a shows a close association with the lamprey FGFR3/4, although the bootstrap probability is not significantly high. It is therefore likely that the hagfish FGFR3/4a and the lamprey FGFR3/4, together with the hagfish FGFR3/4b that was generated in the cyclostome lineage by gene duplication, are the cyclostome homologs of vertebrate FGFR3 and FGFR4. Thus, Fig. 1 suggests that, of the three gene duplications that generated the four gnathostome FGFRs, two gene duplications antedate the cyclostome–ganthostome split, but the latest gene duplication that gave rise to FGFR3 and FGFR4 occurred after the cyclostome–gnathostome split.

Two more gene duplications were observed on *Xenopus* lineage (Fig. 1). Such gene duplications were also observed in other genes, which might be derived by very recent chromosomal duplications (Suga et al. 1997). The gene duplications on *Xenopus* lineage were also observed in Eph4 (Sek-1/Pag) and src (*src-1/2*) (see below). We do not discuss further these recent gene duplications in *Xenopus.* In addition to the gene duplication in hagfish (FGFR3/4a and 3/4b), two more gene duplications were also observed in Eph subfamily (hagfish EphC*1/2*) and in PGDFR subfamily (lamprey *kit*-like receptor A/B) (see below). Divergence of the hagfish EphC1 and EphC2 may have occurred by a very recent gene duplication, judging from the number (k_s) of synonymous substitution per site (Miyata and Yasunaga 1980); the value is only 0.67, which is comparable to the corresponding value (0.64 ± 0.18 on the average of 574 genes) between human and rodents (unpublished data).

Phylogenetic Tree of Eph Subfamily

A similar analysis was carried out for Ephs, members of the PTK family. At least 13 gnathostome Ephs have been identified to date (Flanagan and Vanderhaeghen 1998); because EphB6 is highly divergent and the structure is unusual, this member was excluded from all the analyses (Matsuoka et al. 1997). We have isolated four cDNAs from *E. burgeri,* two cDNAs from *L. reissneri,* and two cDNAs from *B. belcheri,* and their nucleotide sequences have been determined for regions encoding the kinase domain. The amino acid sequences of the cyclostome and amphioxus Ephs were aligned with those of known Ephs for regions corresponding to amino acid positions 632–973 in human EphB1. On the basis of the alignment of 306 sites in total, excluding gap sites, a phylogenetic tree of Eph subfamily was inferred, using nematode and sponge Ephs as an outgroup (Fig. 2).

The phylogenetic tree of Fig. 2 shows that all gene duplications occurred after the divergence of cephalochordates and vertebrates. The tree also shows that all gene duplications (represented by rhombi in Fig. 2) that gave rise to known Eph isoforms (EphA1–EphA8 and EphB1–EphB5) antedate the amphibian–amniote split or the fish–tetrapod split, as expected (Iwabe et al. 1996; Suga et al. 1997); for four gene duplications that gave rise to EphA3 and EphA5–EphA8, their divergence times are unknown. The hagfish EphA and the hagfish and lamprey EphBs are likely homologs of gnathostome EphA4 and EphB2/5, respectively; note that the cyclostome EphBs support the cyclostome monophyly. Of the 14 gene duplications, excluding gene duplications on the *Xenopus*, zebrafish, hagfish, and amphioxus lineages (represented by open boxes), six gene duplications (filled rhombi) antedate the cyclostome–gnathostome split, and one gene duplication (open rhombus) postdates that split; for the remaining seven gene duplications (half-filled rhombi), the divergence times are unknown, although three antedate the fish–tetrapod split. It is therefore likely that, since the separation from cephalochordates, vertebrates created many Eph isoforms by gene duplications, most of which occurred at dates before the divergence of cyclostomes and gnathostomes.

In addition to gene duplications on *Xenopus* and cyclostome lineages (see above), one gene duplication was observed in each of the amphioxus and zebrafish lineages (Eph1/2 and *rtk5/8*, respectively). The amphioxus Eph1 and Eph2 might diverged recently, judging from the k_s value (0.79) between them. One more example of gene duplication (*src*-like protein A/B) on amphioxus lineage was found in the *src* subfamily (see below).

Phylogenetic Tree of src *Subfamily*

We have isolated three cDNAs from *E. burgeri,* two cDNAs from *L. reissneri,* and two cDNAs from *B. belcheri* and compared these sequences with known sequences belonging to the *src* subfamily. On the basis of the sequence alignment for regions corresponding to the amino acid positions 292–531 in human *src* (238 amino acid sites total, excluding gap sites), a phylogenetic tree of *src* subfamily was inferred, using hydra and sponge *src*-related sequences as an outgroup (Fig. 3). The hagfish *src*-like protein C might be orthologous to the gnathostome *src*s. The hagfish and lamprey *src*-like protein As are possibly paralogous, if the cyclostome monophyly is correct, and the former is the ortholog of human *lck.* There is, however, still a possibility that the two cyclostome sequences form a cluster.

The ML tree revealed that all gene duplications (represented by rhombi in Fig. 3) that gave rise to known isoform genes belonging to the *src* subfamily postdate the cephalochordate–vertebrate split, but antedate the amphibian–amniote split or the fish–tetrapod split. The

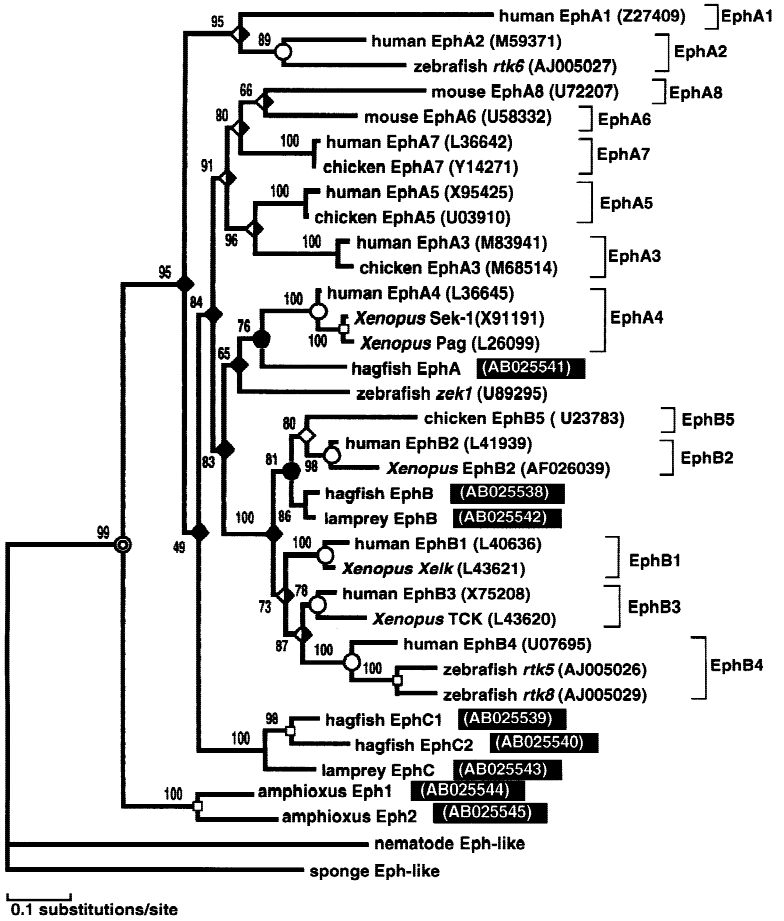

Fig. 2. Maximum likelihood tree of Eph subfamily. From a comparison of the kinase domain sequences, the tree was inferred by an approximate ML method described in Materials and Methods, using nematode and sponge Eph-like sequences as an outgroup. The number at each branch node represents the local bootstrap probability estimated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994). Open circles, fish–tetrapod split or amphibian–amniote split; filled circles, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; filled rhombi, gene duplications that antedate the cyclostome–gnathostome split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split, but antedates the amphibian–amniote split; half-filled rhombi, gene duplications whose divergence times are unknown; open boxes, gene duplications on the lineage leading to cyclostomes, *Xenopus*, fish, or cephalochordates. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

seven gene duplications out of 10 antedate the cyclostome–vertebrate split; for the remaining three, their divergence times are unknown.

Phylogenetic Tree of PDGFR Subfamily

We have isolated two cDNAs from *E. burgeri*, two cDNAs from *L. reissneri*, and one cDNA from *B. belcheri.* These sequences were aligned with those reported to date for regions corresponding to the amino acid positions 606–953 in human PDGFαR (245 amino acid sites total, excluding gap sites). Figure 4 shows the ML tree inferred by using human, purple urchin, and *Drosophila* FGFRs as an outgroup. According to Fig. 4, the hagfish *kit*-like receptor and the lamprey *kit*-like receptor A support the cyclostome monophyly, although the bootstrap probability is not high enough. Unlike the above three subfamilies, the deepest gene duplication antedate the cephalochordate–vertebrate split. Because the group of five isoforms, PDGFα/βR, Flt3, CSF-1R, and *c-kit*, has five immunoglobulin (Ig)-like repeats in the extracellular region, whereas the group of VEGFR, FLT4, and Flk1 has seven Ig-like repeats (van der Geer et al. 1994), it remains possible that they comprise separate subfamilies. In the former group, all the four gene duplications that gave rise to distinct isoforms antedate the cyclostome–gnathostome split.

Bootstrap Analysis for Estimating the Number of Isoform Duplications

The above phylogenetic analyses of the four subfamilies provides clear evidence that most if not all isoform duplications occurred in a period between the cephalochordate–vertebrate split and the amphibian–amniote split. It is, however, not obvious whether the majority of isoform duplications predate or postdate the cyclostome–gnathostome split. This may be due to simultaneous isoform duplications around the cyclostome–gnathostome split, judging from the short branch lengths and the low bootstrap probabilities around that split on trees. It is therefore necessary to estimate the number of isoform duplication on a solid statistical basis. For this purpose, we carried out a bootstrap analysis similar to that described previously (Suga et al. 1997). Based on the standard bootstrap procedure (Felsenstein 1985), we generated 100 ML trees with the largest log-likehood value by repeated local rearrangements (Adachi and Hasegawa 1996), using the NJ trees as initial trees. From these set of ML trees, the numbers of isoform duplications that

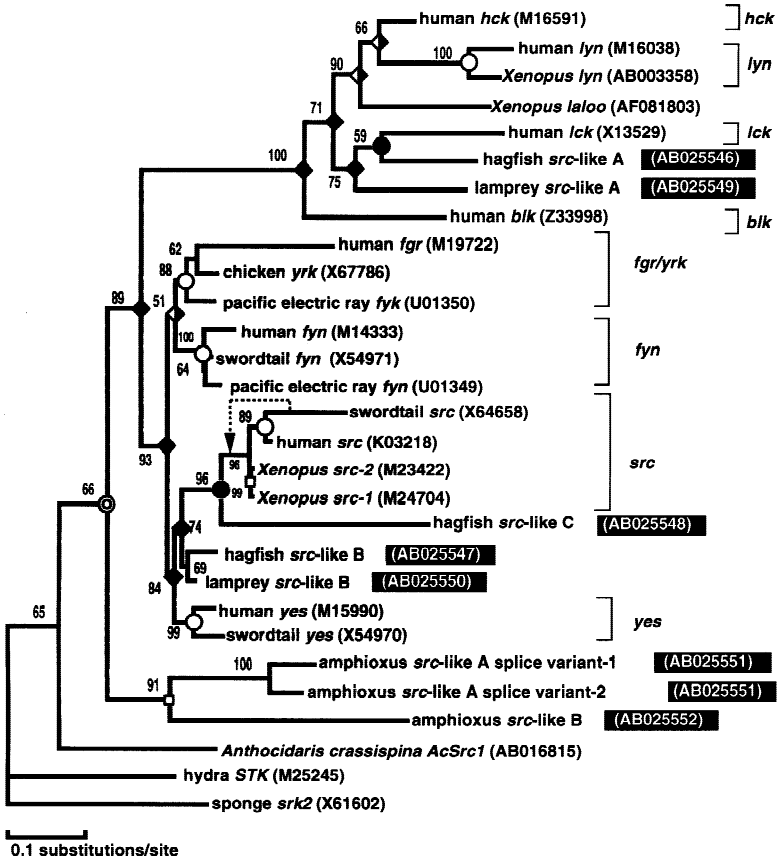

Fig. 3. Maximum likelihood tree of *src* subfamily. From a comparison of the kinase domain sequences, the tree was inferred by an approximate ML method described in Materials and Methods, using hydra and sponge sequences as an outgroup. The number at each branch node represents the local bootstrap probability estimated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994). The dotted arrow means that the swordtail *src* is the outgroup of the human and *Xenopus src*s, when their complete sequences are compared. Open circles, fish–tetrapod split or amphibian–amniote split; filled circles, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; filled rhombi, gene duplications that antedate the cyclostome–gnathostome split; half-filled rhombi, gene duplications whose divergence times are unknown; open boxes, gene duplications on the lineage leading to *Xenopus* or cephalochordates. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

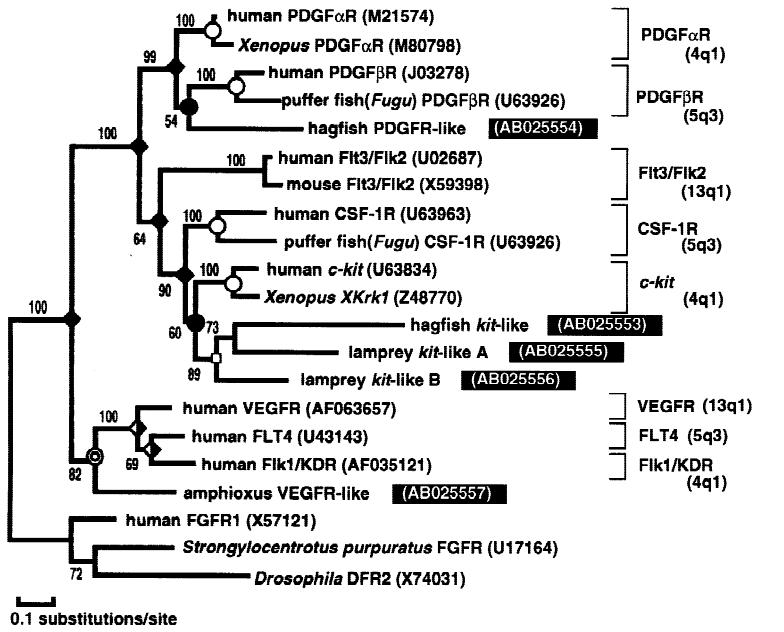

Fig. 4. Maximum likelihood tree of PDGFR subfamily. From a comparison of the kinase domain sequences, the tree was inferred by an approximate ML method described in Materials and Methods, using human, purple urchin, and *Drosophila* FGFR sequences as an outgroup. The number at each branch node represents the local bootstrap probability estimated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994). Open circles, fish–tetrapod split or amphibian–amniote split; filled circles, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; filled rhombi, gene duplications that antedate the cyclostome–gnathostome split; half-filled rhombi, gene duplications whose divergence times are unknown. Data on chromosomal mapping of human genes were taken from van der Geer et al. (1994). Accession numbers of sequences are shown in parentheses; reverse letters, present work.

took place at dates before and after the cyclostome–gnathostome split were counted (designated as N_b and N_a, respectively).

Table 2 shows the results of the four subfamilies, together with those of the other three protein groups. In the seven gene groups examined here, N_b is always larger than N_a; the former is three times larger in value than the latter, on the average. Since the value of N_a is not negligibly small, it may be reasonable to conclude that extensive isoform duplications occurred in a limited period around or just before the divergence of cyclostomes and gnathostomes.

There is a possibility that these isoform duplications were derived from chromosomal duplications in part

Table 2. The numbers of isoform duplications in evolutionary periods before and after the divergence of cyclostomes and gnathostomes

	N_b	N_a	N_b/N_a
PTK subfamily			
FGFR	2.6 ± 1.2	1.3 ± 1.0	2.0
Eph	6.5 ± 1.4	2.3 ± 1.8	2.8
src	7.2 ± 1.4	1.7 ± 1.5	4.2
PDGFR	3.9 ± 1.1	0.8 ± 0.9	4.9
Other proteins			
Aldolase	1.3 ± 0.9	1.2 ± 0.9	1.1
Enolase	2.1 ± 0.9	1.1 ± 1.0	1.9
Complements	2.4 ± 0.5	0.1 ± 0.3	24
Total	26.0	8.5	3.1

N_b and N_a, the numbers of isoform duplications that occurred at dates before and after the cyclostome–gnathostome split, respectively. Isoform duplications whose divergence times are unknown were excluded from the calculations of N_b and N_a. For the calculation procedures, see text. Complements, gene groups encoding the complement components C3, C4, and C5. For the phylogenetic trees of three gene groups in other proteins, see the accompanying paper (Kuraku et al. 1999).

(Ohno 1970; Lundin 1993; Rousset et al. 1995; Bailey et al. 1997; Pébusque et al. 1998; Amores et al. 1998). From a structural and phylogenetic analysis of PDGFR subfamily, together with the chromosomal mappings of the family members, Rousset et al. (1995) suggested that the present-day PDGFR subfamily diverged by gene duplications and chromosomal duplications, although it is generally difficult to infer the chromosomal duplication events on ancient lineages because of frequent translocation and deletion events during evolution. Figure 4 supports the argument by Rousset et al. (1995). According to Fig. 4, PDGFR subfamily comprises three separate groups, PDGFα/βR group, CSF-1R group (Flt3/Flk2, CSF-1R, and *c-kit*), and VEGFR group (VEGFR, FLT4, and Flk1/KDR), which were generated by the first and second gene duplications. Each group has two more isoform duplications, which gave rise to three different isoforms (in PDGFα/βR group one member might be deleted during evolution). Their chromosomal locations on the human genome differ in the same group, but coincide to each other between different groups. It is therefore likely that two chromosomal duplications are responsible for the two isoform duplications in each group.

In conclusion, after the separation from cephalochordates, vertebrates evolved a variety of tissue-specific genes with virtually identical structure and function in each subfamily in a limited period around or just before the divergence of cyclostomes and gnathostomes by gene duplications and possibly chromosomal duplications. Cephalochordates are likely to have multiple isoforms generated by gene duplications on cephalochordate lineage, independently from vertebrates, although less frequent. Cyclostomes also increased the multiplicity of family members by such lineage-specific gene duplications. From phylogenetic analyses of several gene families involved in the signal transduction and developmental control, we recently showed that most subtype duplications that gave rise to different subfamilies with diverse functions had been completed before the parazoan–eumetazoan split, the earliest divergence of extant animal phyla (Suga et al. 1997, 1999; Koyanagi et al. 1998a, 1998b; Hoshiyama et al. 1998; Ono et al. 1999). These results suggest that the Cambrian explosion, the burst of diversification of the major group of animal phyla at the Cambrian/Vendian boundary (Conway Morris 1993), has been accomplished without creating further new genes. Thus, the molecular mechanism of the Cambrian explosion should be understood based on mechanisms that could generate organismal diversity by utilizing or recruiting preexisting genes, but not by creating new genes with novel functions.

Acknowledgments. We thank Prof. M. Hasegawa, Dr. N. Iwabe, and Dr. K. Kuma for discussion. This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan.

References

Adachi J, Hasegawa M (1996) Computer science monographs, no. 28, MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood. Institute of Statistical Mathematics, Tokyo

Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish *hox* clusters and vertebrate genome evolution. Science 282:1711–1714

Balley WJ, Kim J, Wagner GP, Ruddle FH (1997) Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol Biol Evol 14:843–853

Berger MP, Munson PJ (1991) A novel randomized iterative strategy for aligning multiple protein sequences. CABIOS 7:479–484

Conway Morris S (1993) The fossil record and the early evolution of the Metazoa. Nature 361:219–225

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Flanagan JG, Vanderhaeghen P (1998) The ephrins and Eph receptors in neural development. Ann Rev Neurosci 21:309–345

Fox GM, Holst PL, Chute HT, Lindberg RA, Janssen AM, Basu R, Welcher AA (1995) cDNA cloning and tissue distribution of five human EPH-like receptor protein-tyrosine kinases. Oncogene 10: 897–905

Hasegawa M, Kishino H (1994) Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. Mol Biol Evol 11:142–145

Hoshiyama D, Suga H, Iwabe N, Koyanagi M, Nikoh N, Kuma K, Matsuda F, Honjo T, Miyata T (1998) Sponge *Pax* cDNA related to *Pax-2/5/8* and ancient gene duplications in the *Pax* family. J Mol Evol 47:640–648

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

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 30:151–160

Koyanagi M, Suga H, Hoshiyama D, Ono K, Iwabe N, Kuma K, Miyata T (1998a) Ancient gene duplication and domain shuffling in

animal cyclic nucleotide phosphodiesterase family. FEBS Lett 436: 323–328

Koyanagi M, Ono K, Suga H, Iwabe N, Miyata T (1998b) Phospholipase C cDNAs from sponge and hydra: antiquity of genes involved in the inositol phospholipid signaling pathway. FEBS Lett 439:66–70

Kraus MH, Issing W, Miki T, Popescu NC, Aaronson SA (1989) Isolation and characterization of *ERBB3*, a third member of the *ERBB* /epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. Proc Natl Acad Sci USA 86:9193–9197

Kuraku S, Hoshiyama D, Katoh K, Suga H, Miyata T (1999) Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. J Mol Evol (in press)

Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–19

Mallatt J, Sullivan J (1998) 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol Biol Evol 15:1706–1718

Matsuoka H, Iwata N, Ito M, Shimoyama M, Nagata A, Chihara K, Takai S, Matsui T (1997) Expression of a kinase-defective Eph-like receptor in the normal human brain. Biochem Biophys Res Comm 235:487–492

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36

Miyata T, Kuma K, Iwabe N, Nikoh N (1994) A possible link between molecular evolution and tissue evolution demonstrated by tissue specific genes. Jpn J Genet 69:473–480

Mustelin T, Burn P (1993) Regulation of *src* family tyrosine kinases in lymphocytes. Trends Biochem Sci 18:215–220

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Ohno S (1970) Evolution by gene duplication. Springer Verlag, New York, NY

Ono K, Suga H, Iwabe N, Kuma K, Miyata T (1999) Multiple protein tyrosine phosphatases in sponge and explosive gene duplication in the early evolution of animals before the parazoan-eumetazoan split. J Mol Evol 48:654–662

Partanen J, Mäkelä TP, Eerola E, Korhonen J, Hirvonen H, Claesson-Welsh L, Alitalo K (1991) FGFR-4, a novel acidic fibroblast growth factor receptor with a distinct expression pattern. EMBO J 10:1347–1354

Pébusque M-J, Coulier F, Birnbaum D, Pontarotti P (1998) Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 15:1145–1159

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

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Shler P, Watt VM (1992) Tissue-specific expression of the rat insulin receptor-related receptor gene. Mol Endocrinol 6:723–729

Stock DW, Whitt GS (1992) Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. Science 257:787–789

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics, 2d edition. Sinauer Associates, Sunderland, MA, pp 407–514

Suga H, Kuma K, Iwabe N, Nikoh N, Ono K, Koyanagi M, Hoshiyama D, Miyata T (1997) Intermittent divergence of the protein tyrosine kinase family during animal evolution. FEBS Lett 412:540–546

Suga H, Koyanagi M, Hoshiyama D, Ono K, Iwabe N, Kuma K, Miyata T (1999) Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. J Mol Evol 48:646–653

Takahashi T, Shirasawa T (1994) Molecular cloning of rat JAK3, a novel member of the JAK family of protein tyrosine kinases. FEBS Lett 342:124–128

van der Geer P, Hunter T, Lindberg, RA (1994) Receptor protein-tyrosine kinases and their signal transduction pathways. Ann Rev Cell Biol 10:251–337