J Mol Evol (2000) 50:302–311
DOI: 10.1007/s002399910035

JOURNAL OF MOLECULAR EVOLUTION

Protein Tyrosine Phosphatases from Amphioxus, Hagfish, and Ray: Divergence of Tissue-Specific Isoform Genes in the Early Evolution of Vertebrates

Kanako Ono-Koyanagi, Hiroshi Suga, Kazutaka Katoh, Takashi Miyata

Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502 Japan

Received: 12 November 1999 / Accepted: 13 December 1999

Abstract. Since separation from fungi and plants, multicellular animals evolved a variety of gene families involved in cell–cell communication from a limited number of ancestral precursors by gene duplications in two separate periods of animal evolution. In the very early evolution of animals before the separation of parazoans and eumetazoans, animals underwent extensive gene duplications by which different subtypes (subfamilies) with distinct functions diverged. The multiplicity of members (isoforms) in the same subtype increased by further gene duplications (isoform duplications) in the first half of chordate evolution before the fish–tetrapod split; different isoforms are virtually identical in structure and function but differ in tissue distribution. From cloning and phylogenetic analyses of four subfamilies of the protein tyrosine kinase (PTK) family, we recently showed extensive isoform duplications in a limited period around or just before the cyclostome–gnathostome split. To obtain a reliable estimate for the divergence time of vertebrate isoforms, we have conducted isolation of cDNAs encoding the protein tyrosine phosphatases (PTPs) from *Branchiostoma belcheri*, an amphioxus, *Eptatretus burgeri*, a hagfish, and *Potamotrygon motoro*, a ray. We obtained 33 different cDNAs in total, most of which belong to known PTP subfamilies. The phylogenetic analyses of five subfamilies based on the maximum likelihood method revealed frequent isoform duplications in a period around or just before the gnathostome–cyclostome split. An evolutionary implication was discussed in relation to the Cambrian explosion.

Key words: Protein tyrosine phosphatase — Amphioxus — Hagfish — Ray — Tissue-specific isoform — Gene duplication — Phylogenetic tree — Evolution

Introduction

Protein tyrosine phosphatases (PTPs) are involved in various physiological events, including cell proliferation, differentiation, migration, and adhesion in concert with protein tyrosine kinases (PTKs) (Hunter 1995; Tonks and Neel 1996 for reviews). The PTPs comprise a diverse family that shares one or two PTP domains carrying the PTP activity in common (Mauro and Dixon 1994; Brady-Kalnay and Tonks 1995; Streuli 1996; Stoker 1996; Chien 1996 for reviews). Since separation from fungi and plants, animals evolved a variety of PTPs from a common ancestral gene by gene duplication and domain shuffling. The PTP family is classified into at least 17 subtypes or subfamilies (9 receptor-type and 8 nonreceptor-type PTPs), which are characterized by different organization of functional domain and independent cluster in the family tree (Ono et al. 1999). Cloning of PTP cDNAs from a freshwater sponge *Ephydatia fluviatilis*, one of the most primitive multicellular animals, and a

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases with accession numbers AB033560-AB033592.

Correspondence to: Takashi Miyata; *e-mail:* miyata@biophys.kyoto-u.ac.jp

Table 1. Degenerate primers used for the cloning of cDNAs

Subtype		Name	Sequence	Amino acid sequence
Specific primers				
1. PTPR5	sense	S1	5′-TGCCGATG-AAISANTTYSYNAARCA-3′	(K/N)(H/Q/E)F(P/V/A)KH
	sense	S2	5′-CAGGATCG-MANCAYCCNGANAAYAA-3′	(N/Q)HP(D/E)NK
	antisense	A1	5′-GTGAATTC-CKYTGISWICKDATRTG-3′	HIR(S/T)QR
2. PTPR4	antisense	A1	5′-GTGAATTC-ADYTCNGTRTCNCCRTA-3′	YGDTE(L/I)
3. PTPN6	sense	S1	5′-CAGTAGGC-TTYTGGGARGARTTYGA-3′	FWEEFE
	antisense	A1	5′-TGTCGCGG-ATIAYRTCDATNARDAT-3′	ILID(I/V)I
	antisense	A2	5′-CGATGTAC-TGYTGNACNGCCATRTA-3′	YMAVQ(Q/H)
4. PTPR2A	sense	S1	5′-AC-GGNCARCANTTYACNTGG-3′	GQ(Q/H)FTW
5. PTPN3	sense	S1	5′-GTGCATTC-AAYMGNTAYMGNGAYGT-3′	NRYRDV
	antisense	A1	5′-GTGAATTC-GTYTGDATNARNCCCAT-3′	MGLIQT
General primers				
	sense	S1	5′-CAGGATCC-TYTGGMGNATGRTNTGG-3′	FWRM(I/V)W
	sense	S2	5′-CAGGATCC-INGAYTTYTGGMGNATG-3′	XDFWRM
	antisense	A1	5′-GTGAATTC-RYICCIGCNSWRCARTG-3′	HCSAG(V/T/A)
	antisense	A2	5′-GAGAATTC-GTICKNCCNACNCCNGC-3′	AGVGRT

Cloning was performed by using specific primers, general primers, or by the combination of specific primers and general primers for sense and antisense strands (i.e., specific primers for the sense strand and general primers for the antisense strand or vice versa). For example, ray ryPTPR5a belonging to the subtype PTPR5 has been cloned by using the specific primers S1 and A1 of PTPR5 in primary PCR and the general primers S1 and A1 in nested PCR.

phylogenetic analysis of the family members revealed extensive gene duplications (subtype duplications) that gave rise to different subtypes (subfamilies) in the very early evolution of animals: creation of most, if not all, of the subtypes present in triploblast animals have been completed before the parazoan–eumetazoan split, the earliest divergence among extant animal phyla (Ono et al. 1999). Although sponges are thought to be lacking cell cohesiveness and coordination typical of eumetazoans (Margulis and Schwartz 1998), multiple receptor-type PTPs involved in cell adhesion phenomena found in triploblast animals exist in sponges (Ono et al. 1999). The same pattern of divergence was also found in other animal-specific gene families involved in cell–cell communication (Koyanagi et al. 1998a, 1998b; Suga et al. 1999a, 1999b) and developmental control (Hoshiyama et al. 1998).

In many eukaryotic gene families, including the PTP family, the multiplicity of members in the same subfamily 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 (e.g., Iwabe et al. 1996). In most cases, different isoforms in the same subfamily are virtually identical in structure and function but differ in tissue distribution. The chromosomal duplications in the early evolution of vertebrates might be responsible for the isoform diversification in part (Ohno 1970; Rousset et al. 1995; Bailey et al. 1997; Amores et al. 1998; Pèbusque et al. 1998). From cloning of the amphioxus, hagfish, and lamprey PTKs and phylogenetic analyses of four subfamilies belonging to the PTK family, Suga et al. (1999c) recently showed frequent isoform duplications in the early evolution of vertebrates around or just before the divergence of cyclostomes. To obtain a reliable estimate for the divergence time of isoforms, it is nevertheless required to accumulate many data from other gene families.

In this paper, we have conducted isolation of cDNAs encoding PTPs from *Branchiostoma belcheri,* an amphioxus, *Eptatretus burgeri,* a hagfish, and *Potamotrygon motoro,* a ray. Phylogenetic analyses of five subfamilies belonging to the PTP family revealed extensive isoform duplications in a period around or just before the gnathostome–cyclostome split. The result is consistent with the previous one obtained from the phylogenetic analyses of PTKs.

Materials and Methods

Isolation and Sequencing of Amphioxus, Hagfish, and Ray cDNAs. Total RNA of *B. belcheri,* an amphioxus, was extracted from the whole body; that of *E. burgeri,* a hagfish, was extracted from each of the brain and liver; and that of *P. motoro,* a ray, was extracted from each of the brain, liver, ovary, spleen, and eye 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) or Ampli Taq Gold (PE Applied Biosystems). 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 45–46°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 (PE Applied Biosystems). The 3′ ends of cDNAs were amplified using 3′ RACE System for Rapid Amplification of cDNA Ends (GIBCO BRL).

Sequence Data. Sequence data used for phylogenetic tree inference were taken from DDBJ release 38.

Sequence Alignment. Multiple alignment of the amino acid sequences of PTPs was carried out by a method developed recently by us

(Katoh et al. in preparation). This method is basically an extended version of the progressive approach of Feng and Doolittle (1987) by improving the calculation procedure of dynamic programming (Needleman and Wunsch 1970), by which the speed of computation has been greatly improved without sacrificing accuracy and efficiency; the computation time required by the new method is only about one-tenth of that by the standard method (for example, CLUSTAL W [Thompson et al. 1994], a widely distributed multiple alignment program).

Phylogenetic Tree Inference. On the basis of the alignment described above, the phylogenetic trees of five subfamilies belonging to the PTP family were inferred by a maximum likelihood (ML) method developed recently by us (Katoh and Miyata 1999): This method is a heuristic approach (ML-TBR) of ML analysis based on tree topology search by the tree bisection and reconnection (TBR) algorithm (Swofford et al. 1996). The tree topology search is repeated until no improvement on the log-likelihood is found for a given initial tree. In the present work, 100 initial trees were generated by bootstrap resamplings and neighbor-joining (NJ) method (Saitou and Nei 1987). The actual computation was performed on a PC cluster composed of 32 Pentium III 500 MHz processors. The bootstrap probability (Felsenstein 1985) was calculated by the method of Adachi and Hasegawa (1996) (local bootstrap probability; LBP).

Each tree inferred by ML-TBR was reexamined by the intact ML method (Kishino et al. 1990), in which all possible tree topologies are examined, by assuming the tree topology of branches within a cluster with a high bootstrap probability at the deepest node. The bootstrap probability in the intact ML analysis was calculated by the RELL method (Kishino et al. 1990; Hasegawa and Kishino 1994). For subfamilies with OTUs fewer than nine, the intact ML analysis was directly performed. The overall tree of the PTP family was inferred by the NJ method; the distance matrix was estimated by the ML method, using the JTT model (Adachi and Hasegawa 1996).

Results and Discussion

Each of animal gene families including the PTP family 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 phylogenetic tree of members of a gene family comprises several independent clusters corresponding to different subtypes or subfamilies that diverged before the divergence of parazoans and eumetazoans (Koyanagi et al. 1998a, 1998b; Hoshiyama et al. 1998; Suga et al. 1999a, 1999b; 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 before the fish–tetrapod split (Miyata et al. 1994; Iwabe et al. 1996; Suga et al. 1997, 1999a, 1999b; Hoshiyama et al. 1998). From cloning and sequencing of the protein tyrosine kinase (PTK) cDNAs from an amphioxus, a hagfish, and a lamprey and phylogenetic analyses of four subfamilies belonging to the PTK family, we recently showed that the isoform duplications occurred at dates around or just before the divergence of gnathostomes and cyclostomes (Suga et al. 1999c). To accumulate more data on the isoform duplications and to determine the divergence times more closely on the basis of a large amount of data, we have addressed isolation of cDNAs encoding PTPs from members of cephalochordates, cyclostomes, and cartilaginous fishes, for which no PTP sequence has been reported to date, except for shark CD45 (Ac. No. U34750).

Amphioxus, Hagfish, and Ray PTPs and Phylogenetic Tree of PTP Family

We have isolated 9 PTP-related cDNAs from *B. belcheri,* 11 cDNAs from *E. burgeri,* and 13 cDNAs from *P. motoro* by the method described in Materials and Methods. These cDNAs encode amino acid sequences of fragment length covering from the C-terminal 2/3 of the first PTP domain to the stop codon; for ray ryPTPN6b and ryPTPN6c, their cDNAs encode the complete PTP domain sequences. These sequences were aligned with known PTP sequences from animals for the PTP domain. According to the alignment, the cloned cDNAs share 27–97% amino acid sequence identities with known animal PTPs. In addition, the amino acid sequences of these cDNAs contain the PTP signature motif HCxxGxxR(S/T) (e.g., Denu et al. 1996) at the precise position. It is therefore highly likely that the isolated cDNAs are the products of PTP genes.

On the basis of the alignment, a phylogenetic tree of the PTP family was inferred by the NJ method (Fig. 1). According to the NJ tree, each of 33 cDNAs isolated here belongs to any one of subfamilies classified recently (Ono et al. 1999) with high bootstrap probability, except for amphioxus amPTPR10. The amPTPR10 has tandemly duplicated PTP domains and thus is likely to be a receptor-type PTP. Judging from the phylogenetic position, the amPTPR10 is likely to be a member of a novel subtype, although there is still a possibility of a member of PTPR2B subtype. The phylogenetic tree of Fig. 1 confirmed our previous result that most subtype duplications antedate the parazoan–eumetazoan split (represented by a filled circle), although the phylogenetic position of two sponge PTPs, sPTPR4 and sPTPR5, differs from that described previously (Ono et al. 1999). The subtype classification of these sponge PTPs should be reexamined based on the overall domain organization.

Phylogenetic Tree of PTPR4 Subfamily

For five subfamilies (PTPR2A, PTPR4, PTPR5, PTPN3, and PTPN6 subfamilies), the amino acid sequences are available for comparison among cephalochordates, cyclostomes, cartilaginous fishes, and tetrapods, and thus the number and the time of occurrence of isoform duplication are possible to estimate from the phylogenetic tree of each subfamily. Figure 2 shows the phylogenetic tree of the PTPR4 subfamily. From a comparison of the amino acid sequences of six PTP cDNAs cloned here

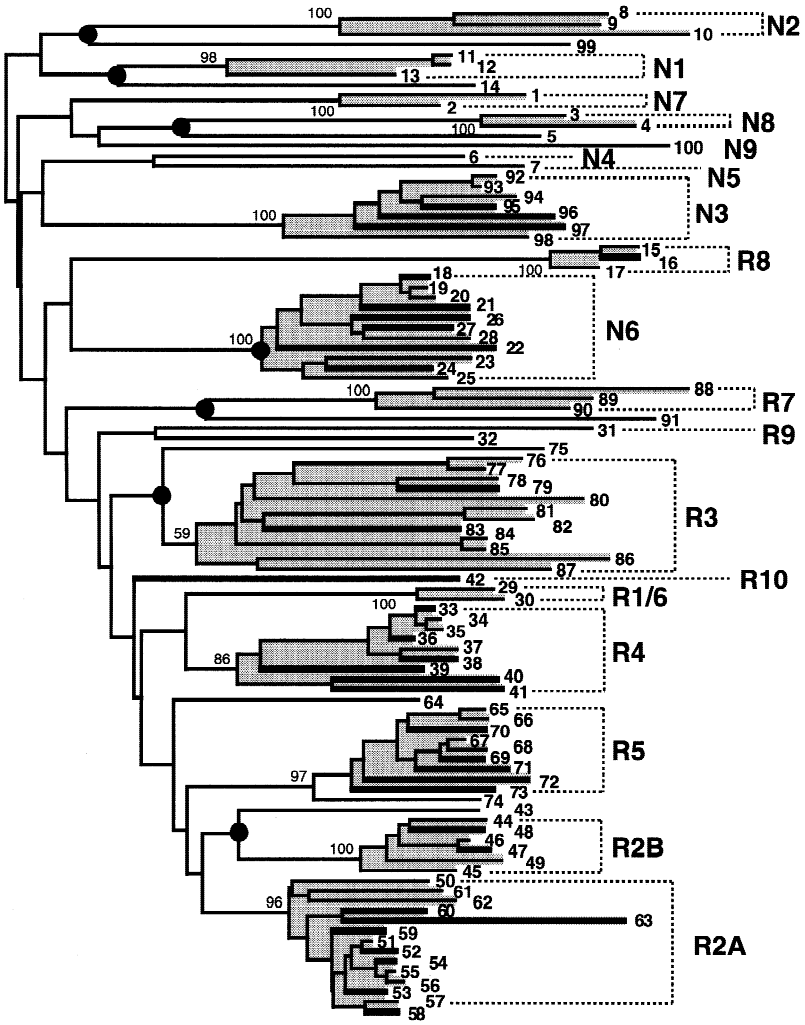


Fig. 1. Unrooted tree of PTP family. From a comparison of the PTP domain sequences of fragment length (141 amino acids), the tree was inferred by the NJ method (Saitou and Nei 1987); gap sites were excluded from the analysis. Filled circles correspond to parazoan–eumetazoan split. Clusters corresponding to PTP subfamilies are shaded. The bootstrap probability at the deepest node of each cluster is shown. The classification of subfamilies and their names were followed by Ono et al. (1999). The lineages of PTPs whose sequences have been determined in the present work are shown by heavy lines. The sequence names and accession numbers are as follows: 1, human PTP (M64572); 2, human PTP (M68941); 3, human pez (X82676); 4, human PTP D1 (X79510); 5, sponge sPTPN8 (AB019133); 6, human PTP-BAS type 1 (D21209); 7, human DKFZp566K0524 (AL050040); 8, human 70zpep (AF077031); 9, human PTP G1 (D13380); 10, human BDP1 (X79568); 11, *Xenopus* PTPX1 (L33098); 12, *Xenopus* PTPX10 (L33099); 13, human PTP MEG2 (M83738); 14, sponge sPTPN1 (AB019129); 15, human IA-2/PTP (L18983); 16, ray ryPTPR8 (AB033585, present work); 17, human IAR (AF007555); 18, ray ryPTPN6b (AB033591, present work); 19, human PTP 1D (X70766); 20, *Xenopus* SH-PTP2 (U15287); 21, hagfish hgPTPN6a (AB033576, present work); 22, hagfish hgPTPN6b (AB033578, present work); 23, *Drosophila* csw Y1229 (U19909); 24, amphioxus amPTPN6 (AB033567, present work); 25, sponge sPTPN6 (AB019132); 26, ray ryPTPN6c (AB033592, present work); 27, ray ryPTPN6a (AB033580, present work); 28, human hPTP1C (U15528); 29, human LCA (Y00638); 30, shark CD45 (U34750); 31, *Drosophila* Ptp69D (M27699); 32, sponge sPTPR5 (AB019127); 33, ray ryPTPR4b (AB033589, present work); 34, human PTPα (M34668); 35, *Xenopus* PTPα (U09135); 36, hagfish hgPTPR4 (AB033572, present work); 37, human HPTPε (X54134); 38, ray ryPTPR4a (AB033583, present work); 39, amphioxus amPTPR4c (AB033564, present work); 40, amphioxus amPTPR4a (AB033562, present work); 41, amphioxus amPTPR4b (AB033563, present work); 42, amphioxus amPTPR10 (AB033561, present work); 43, sponge sPTPR2B (AB019126); 44, human hR-PTPμ (X58288); 45, human RPTP-ρ (AF043644); 46, human h-PTPκ (Z70660); 47, ray ryPTPR2B (AB033587, present work); 48, hagfish hgPTPR2B (AB033574, present work); 49, human hPTP-J (U73727); 50, *Drosophila* Lar (M27700); 51, human PTPδ (L38929); 52, ray ryPTPR2Ab (AB033584, present work); 53, amphioxus amPTPR2A (AB033566, present work); 54, ray ryPTPR2Ac (AB033586, present work); 55, chicken CRYPα1 (L32780); 56, human PTPσ (U35234); 57, human LAR (Y00815); 58, ray ryPTPR2Aa (AB033581, present work); 59, hagfish hgPTPR2Aa (AB033569, present work); 60, hagfish hgPTPR2Ab (AB033571, present work); 61, leech HmLAR1 (AF017084); 62, leech HmLAR2 (AF017083); 63, hagfish hgPTPR2Ac (AB033575, present work); 64, sponge sPTPR4 (AB019125); 65, human PTPRζ (M93426); 66, chicken CPTPζ (L27625); 67, human PTPRγ (L09247); 68, chicken PTPγ (U38349); 69, ray ryPTPR5b (AB033590, present work); 70, ray ryPTPR5a (AB033588, present work); 71, hagfish hgPTPR5a (AB033570, present work); 72, hagfish hgPTPR5b (AB033577, present work); 73, amphioxus amPTPR5 (AB033565, present work); 74, *Drosophila* Ptp99A (M80539); 75, sponge sPTPR3 (AB019128); 76, human DEP-1 (U10886); 77, chicken PTP (AJ238216); 78, human HPTPβ (X54131); 79, hagfish hgPTPR3 (AB033573, present work); 80, rat PTPRQ (AF063249); 81, *Drosophila* Ptp10D (M80465); 82, *Drosophila* Ptp4E (L20894); 83, amphioxus amPTPR3 (AB033560, present work); 84, human PTP-U2 (Z48541); 85, chicken CRYP-2 (U65891); 86, mouse Esp (U36488); 87, human SAP-1 (D15049); 88, human STEP (U27831); 89, human HPCPTP1 (D64053); 90, human LC-PTP (D11327); 91, sponge sPTPR7 (AB019131); 92, human PTPN1 (M33689); 93, chicken PTP1B (U86410); 94, human PTPRF (M25393); 95, ray ryPTPN3 (AB033582, present work); 96, amphioxus amPTPN3 (AB033568, present work); 97, hagfish hgPTPN3 (AB033579, present work); 98, *Drosophila* Ptp61F (L11253); 99, sponge sPTPN2 (AB019130); 100, human DKFZp564F0923 (AL110210).

with those of known PTPs belonging to the same subfamily for a region of 416 amino acids, including the C-terminal 2/3 of the first PTP domain and the complete second PTP domain, the tree was inferred by a heuristic approach of the ML method (Katoh and Miyata 1999), using human, chicken, and shark PTPs belonging to the PTPR1/6 subfamilies as an outgroup.

According to Fig. 2, the gnathostome PTPs are classified into two different isoforms (designated as PTPα and PTPε), which were generated by gene duplication (isoform duplication). In the PTPα isoform, the branching order of PTPs coincides exactly with species order, suggesting that the ray ryPTPR4b and hagfish hgPTPR4 are orthologs of human/chicken/*Xenopus* PTPα. The am-

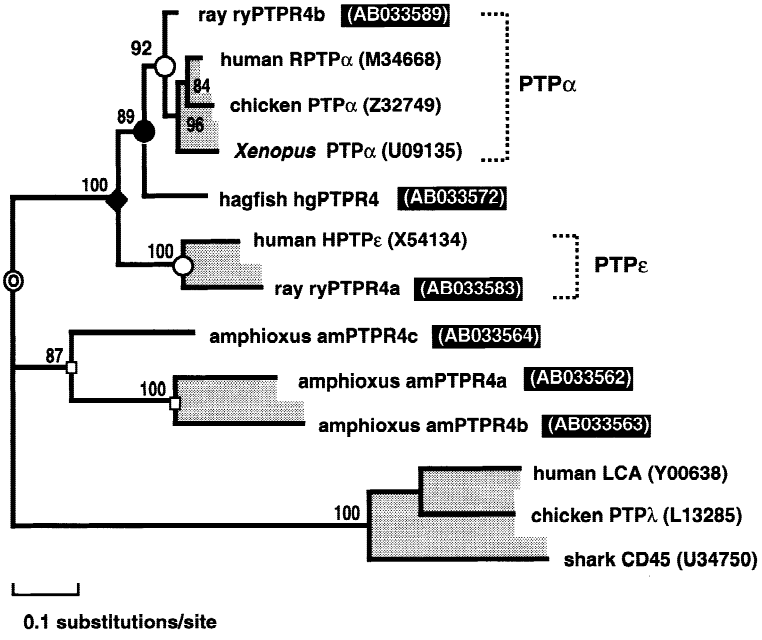


Fig. 2. Maximum likelihood tree of PTPR4 subfamily. From a comparison of the N- and C-terminal PTP domain sequences of fragment length (416 amino acids), the tree was inferred by a heuristic approach of ML method (Katoh and Miyata 1999), using human, chicken, and shark PTPs belonging to the PTPR1/6 subfamily as an outgroup; gap sites were excluded from the comparison. The number at each branch node represents the local bootstrap probability estimated by the LBP method (Adachi and Hasegawa 1996). For shaded clusters, monophyletic grouping was assumed in the intact ML analysis (see text). Isoforms that are considered to be present in gnathostomes in common are shown on the right hand side of the tree. Open circle, ray–tetrapod split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; filled rhombus, gene duplication that antedates the cyclostome–gnathostome split; open box, gene duplication in amphioxus lineage. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

phioxus amPTPR4a, amPTPR4b, and amPTPR4c form a cluster in the tree, which was generated by gene duplications on cephalochordate lineage, independently from isoform duplications in vertebrates. Judging from the phylogenetic position of the hagfish hgPTPR4, the isoform duplication that gave rise to the gnathostome isoforms PTPα and PTPε is likely to predate the divergence of gnathostomes and cyclostomes, but postdates the cephalochordate–vertebrate split. To confirm the ML tree inferred from the heuristic approach, an ML analysis based on the intact ML method of protein phylogeny (Kishino et al. 1990; Adachi and Hasegawa 1996) has been performed, assuming the tree topology within each cluster (shaded in Fig. 2) with a high bootstrap probability (LBP > 95%) at the deepest node. The intact ML analysis supported the tree of Fig. 2 at the significance level of ± 1 SE.

Phylogenetic Tree of PTPN3 Subfamily

The phylogenetic tree of PTPN3 subfamily was inferred by the intact ML method (Kishino et al. 1990; Adachi and Hasegawa 1996) by comparing sequences of a region of 180 amino acids including the C-terminal 2/3 of the PTP domain and using the *Drosophila* Ptp61F belonging to the same subfamily as an outgroup (Fig. 3). The ML tree of Fig. 3 was supported at the significance level of ± 1 SE. According to Fig. 3, gnathostomes have two different isoforms, PTPN1 and PTPRF, which were generated by one isoform duplication. In addition, the isoform duplication occurred on vertebrate lineage after the separation from cephalochordates, as in the case of PTPR4 subfamily, but the date of occurrence clearly differs: the isoform duplication postdates the divergence of gnathostomes and cyclostomes.

Phylogenetic Tree of PTPR5 Subfamily

Figure 4 shows the phylogenetic tree of PTPR5 subfamily inferred from the intact ML method (Kishino et al. 1990; Adachi and Hasegawa 1996); the ML tree was inferred from a comparison of sequences for a region of 435 amino acids, including the C-terminal 2/3 of the first PTP domain and the complete second PTP domain, using the *Drosophila* Ptp99A belonging to the same subfamily as an outgroup (Ono et al. 1999). According to the tree of PTPR5 subfamily, at least two different isoforms, PTPγ and PTPζ, exist in gnathostomes. The cloned ryPTPR5a and ryPTPR5b from a cartilaginous fish are likely to be orthologous to the human/chicken PTPζ and PTPγ isoforms, respectively. The hagfish hgPTPR5a is likely to be orthologous to the human/chicken PTPγ. The amphioxus amPTPR5 is possibly an ancestral precursor of the two isoforms. Judging from the phylogenetic position of hgPTPR5a, the isoform duplication that gave rise to PTPγ and PTPζ might have occurred on an ancestral lineage of vertebrates before the separation of gnathostomes and cyclostomes, but after the separation from cephalochordates.

Although the tree shown in Fig. 4 has the largest log-likelihood value (Lmax) and the largest bootstrap probability p (= 31.1) among all possible trees, alternative trees (tree i, i = 1, 3) are also possible (Fig. 4 insert): they have the difference ΔLi (= Lmax − Li) of log-likelihood values from Lmax being smaller than ± 1 SE. All the four trees, including the ML tree, evidently show the occurrence of isoform duplications in a period between the cephalochordate–vertebrate split and the cartilaginous fish–tetrapod split. However, the number of isoform duplication (Nb) before and (Na) after the cyclostome–vertebrate split varies depending on tree con-

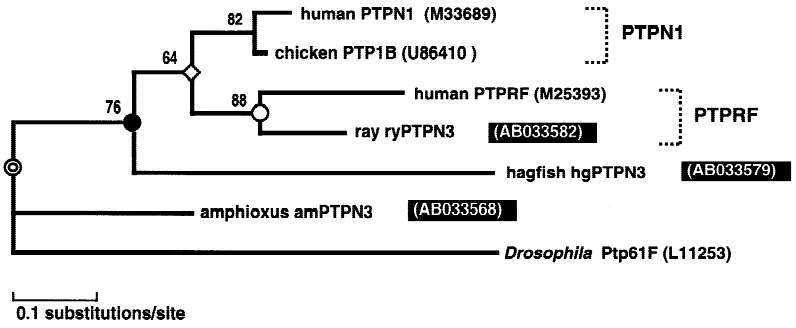
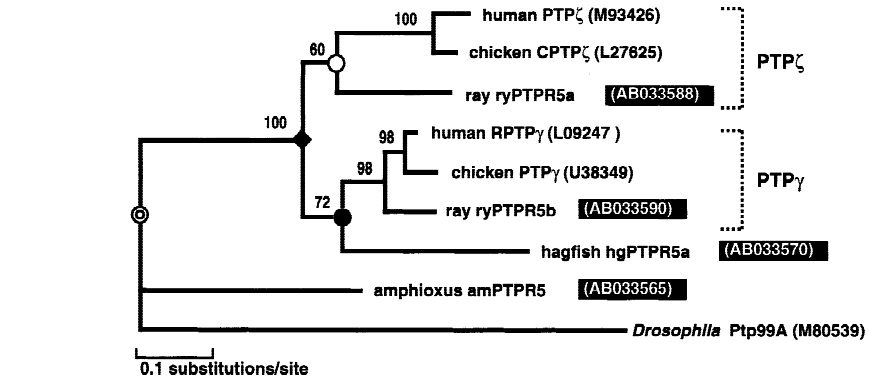


Fig. 3. Maximum likelihood tree of PTPN3 subfamily. From a comparison of the PTP domain sequences of fragment length (180 amino acids), the tree was inferred by the intact ML method (Kishino et al. 1990; Hasegawa and Kishino 1994), using *Drosophila* PTPs belonging to the same subfamily as an outgroup; gap sites were excluded from the comparison. The number at each branch node represents the bootstrap probability estimated by the RELL method (Hasegawa and Kishino 1994). Isoforms that are considered to be present in gnathostomes in common are shown on the right-hand side of the tree. Open circle, cartilaginous fish–tetrapod split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split, but antedates the cartilaginous fish–tetrapod split. Accession numbers of sequences are shown in parentheses; reverse letters, present work.



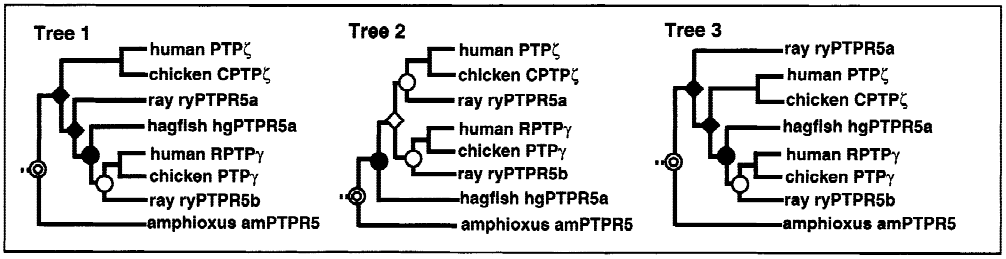


Fig. 4. Maximum likelihood tree of PTPR5 subfamily. From a comparison of the N- and C-terminal PTP domain sequences of fragment length (435 amino acids), the tree was inferred by the intact ML method (Kishino et al. 1990; Hasegawa and Kishino 1994), using *Drosophila* and sponge PTPs belonging to the same subfamily as an outgroup; gap sites were excluded from the comparison. The number at each branch node represents the bootstrap probability estimated by the RELL method (Hasegawa and Kishino 1994). Isoforms that are considered to be present in gnathostomes in common are shown on the right hand side of the tree. **Insert,** three alternative trees with $\Delta L < \pm 1$ SE (branch lengths are arbitrary). Open circle, cartilaginous fish–tetrapod split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split, but antedates the cartilaginous fish–tetrapod split; filled rhombus, gene duplication that antedates the cyclostome–gnathostome split. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

sidered (i.e., Nb = 1 and Na = 0 in ML tree, Nb = 2 and Na = 0 in tree 1, Nb = 0 and Na = 1 in tree 2, and Nb = 2 and Na = 0 in tree 3). Taking a weighted mean of the number of isoform duplications among the four trees by using renormalized bootstrap probabilities as weight factors, we have Nb = 1.18 and Na = 0.24 on the average.

Phylogenetic Tree of PTPN6 Subfamily

Figure 5 shows the ML tree of the PTPN6 subfamily. From a comparison of sequences for a region of 182 amino acids, including the C-terminal 2/3 of the PTP domain, the tree was inferred by the heuristic approach

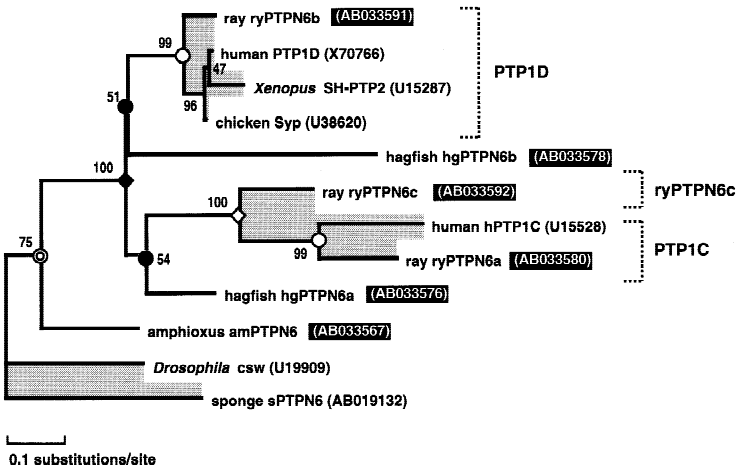


Fig. 5. Maximum likelihood tree of PTPN6 subfamily. From a comparison of the PTP domain sequences of fragment length comprising 182 amino acids, the tree was inferred by a heuristic approach of the ML method (Katoh and Miyata 1999) using *Drosophila* and sponge PTPs belonging to the same subfamily as an outgroup; gap sites were excluded from the comparison. The number at each branch node represents the local bootstrap probability estimated by the LBP method (Adachi and Hasegawa 1996). For shaded clusters, monophyletic grouping was assumed in the intact ML analysis (see text). Isoforms that are considered to be present in gnathostomes in common are shown on the right-hand side of the tree. Open circle, cartilaginous fish–tetrapod split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; filled rhombus, gene duplication that antedates the cyclostome–gnathostome split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split, but antedates the cartilaginous fish–tetrapod split. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

of ML method using *Drosophila* and sponge PTPs belonging to the same subfamily as an outgroup. According to the tree, gnathostomes have at least three isoforms (PTP1C, PTP1D, and ryPTPN6c) belonging to the PTPN6 subfamily, and they diverged by isoform duplications in a period between the cephalochordate–vertebrate split and the cartilaginous fish-tetrapod split. Judging from the bootstrap probability, it is highly likely that one isoform duplication that gave rise to PTP1C and ryPTPN6c postdates the gnathostome–cyclostome split. However, the branching order of four groups leading to PTP1D, PTP1C/ryPTPN6c, hagfish hgPTPN6a, and hagfish hgPTPN6b is statistically unreliable.

These four groups, together with amphioxus amPTPN6, *Drosophila* csw, and sponge sPTPN6, were subjected to the intact ML analysis, assuming the tree topology of human hPTP1C, ray ryPTPN6a, and ryPTPN6c (in PTP1D group, the branching order [ray, (*Xenopus*, [human chicken])] was assumed). We also assumed that the divergence of protostomes and deuterostomes antedates the divergence of cephalochordates and vertebrates (e.g., Margulis and Schwartz 1988), although aldolase and triose phosphate isomerase clocks suggest their close divergence times (Nikoh et al. 1997). The intact ML analysis confirmed that the tree shown in Fig. 5 is the ML tree, but is not significantly supported, as expected. There are nine other trees whose ΔLs are less than ± 1 SE. We have estimated the average numbers of isoform duplications before and after the gnathostome–cyclostome split by the same calculation procedure as described in the case of the PTPR5 subfamily and have obtained to be 1.65 and 0.67, respectively.

Phylogenetic Tree of PTPR2A Subfamily

Figure 6 shows the phylogenetic tree of the PTPR2A subfamily inferred by the heuristic approach of ML method. A region of 458 amino acids, including the C-terminal 2/3 of the first PTP domain and the complete second PTP domain were compared, and *Drosophila*, mosquito, and leech PTPs belonging to the same subfamily were used as an outgroup for tree inference. The hagfish hgPTPR2Ac was excluded from the present analysis because the evolutionary rate is extremely high. Judging from the bootstrap probabilities, the three lineages, human PTPσ, chicken CRYPα1, and ray ryPTPR2Ac, represent species relationship and are clustered as an independent group (PTPσ), but the branching order of human PTPδ, ray ryPTPR2Ab, and PTPσ is uncertain. Also, each of two groups, human LAR/ray ryPTPR2Aa (designated as LAR) and hagfish duplicated genes hgPTPR2Aa/hgPTPR2Ab forms an independent cluster, but the phylogenetic relationship among the three groups—PTPδ/σ/ryPTPR2Ab, LAR, and hgPTPR2Aa/b—is unreliable, although their divergence might have occurred around the time of divergence of gnathostomes and cyclostomes.

The intact ML analysis was performed for seven groups, PTPδ, ryPTPR2Ab, PTPσ, LAR, hagfish hgPTPR2Aa/b, amphioxus amPTPR2A, and outgroup. The intact ML analysis showed that the tree topology shown in Fig. 6 has the largest log-likelihood value (i.e., ML tree), but is not statistically significant: the difference ΔLs of nine different trees, including the ML tree, are less than ± 1 SE. Following the calculation procedure described above, the average numbers of isoform dupli-

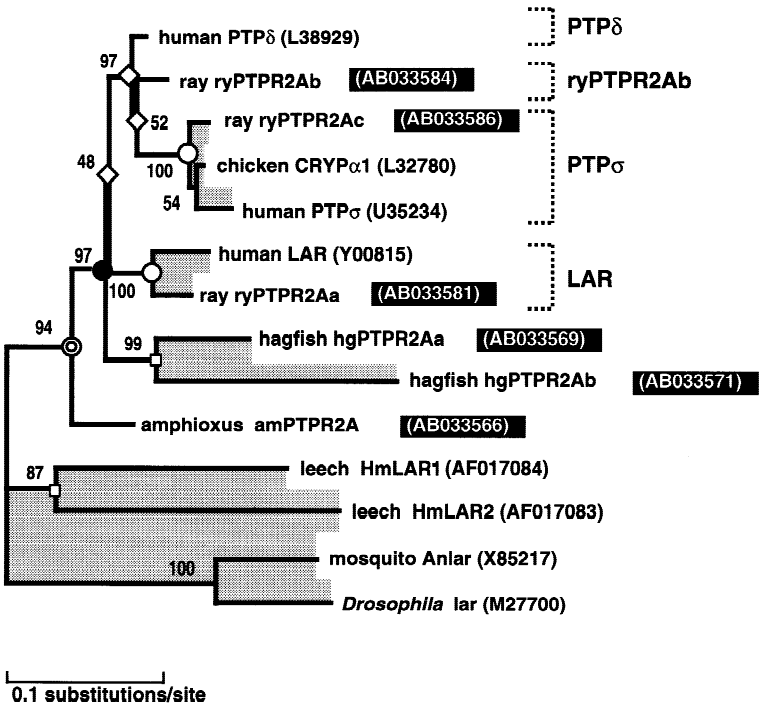


Fig. 6. Maximum likelihood tree of PTPR2A subfamily. From a comparison of the N- and C-terminal PTP domain sequences of fragment length (458 amino acids), the tree was inferred by a heuristic approach of the ML method (Katoh and Miyata 1999) using *Drosophila,* mosquito, and leech PTPs belonging to the same subfamily as an outgroup; gap sites were excluded from the comparison. The number at each branch node represents the local bootstrap probability estimated by the LBP method (Adachi and Hasegawa 1996). For shaded clusters, monophyletic grouping was assumed in the intact ML analysis (see text). Isoforms that are considered to be present in gnathostomes in common are shown on the right hand side of the tree; there is a possibility that PTPδ and ryPTPR2Ab belong to the same isoform. Open circle, cartilaginous fish–tetrapod split; filled circle, cyclostome–gnathostome split; double circle, cephalochordate–vertebrate split; open rhombus, gene duplication that postdates the cyclostome–gnathostome split but antedates the cartilaginous fish–tetrapod split; open box, gene duplication in hagfish or leech lineages. Accession numbers of sequences are shown in parentheses; reverse letters, present work.

cations before and after the gnathostome–cyclostome split were estimated to be 0.52 and 1.60, respectively.

Isoform Duplications in a Period Around or Just Before the Cyclostome–Gnathostome Split

The results on the number and timing of isoform duplications obtained from the five different subfamilies are summarized in Table 2, together with those obtained from previous analyses on the four PTK subfamilies and three other proteins families (Suga et al. 1999c). In Table 2, the number Nb (Na) of isoform duplications that antedate (postdate) the gnathostome–cyclostome split is shown for each subfamily; the method for estimating the Na and Nb used in the previous work slightly differs from that in the present work. Note that isoform duplication was observed neither before the cephalochordate–vertebrate split nor after the fish–tetrapod split. Table 2 shows that the isoform duplication occurred about 2.5 times more frequently in the period before the cyclostome–gnathostome split than in the period after that split, although the ratio Nb/Na varies greatly for different subfamilies due to small sample size. Thus we conclude that isoform duplications have occurred in a period around or just before the cyclostome–gnathostome split.

The above result on isoform diversification, together with those on subtype diversification obtained from previous analyses, suggests the pattern of gene diversification characteristic of multicellular animal–specific genes, which is characterized by two active periods in gene duplication interrupted by a considerably long period of silence, instead of proceeding gradually: in the early period before the parazoan–eumetazoan split about 900 million years ago (Nikoh et al. 1997), animals underwent extensive subtype duplications that gave rise to different subtypes with diverse functions and almost complete sets of present-day subtypes had been established within this period. Since the separation from cephalochordates, vertebrates increased the multiplicity of isoforms in the

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
(a) PTP family			
PTPR4	1.0	0.0	
PTPN3	0.0	1.0	
PTPR5	1.2	0.2	
PTPN6	1.7	0.7	
PTPR2A	0.5	1.6	
Total	4.4	3.5	1.3
(b) PTK family*	20.2	6.1	3.3
(c) Other proteins*	5.8	2.4	2.4
(a) + (b) + (c)	30.4	12.0	2.5

N_b and N_a, the numbers of isoform duplications that occurred at dates before and after the cyclostome–gnathostome split, respectively. Note that isoform duplication was observed neither before the cephalochrodate–vertebrate split nor after the fish-tetrapod split.
*Data from Suga et al. (1999c). The numbers of isoform duplications in (b) and (c) are a total of four subfamilies (fibroblast growth factor receptor, Eph, *src*, and platelet-derived growth factor receptor subfamilies) and that of three protein families (aldolases, enolases, and gene group encoding the complement components C3, C4, and C5), respectively.

same subfamily in a period around or just before the cyclostome–gnathostome split about 500 million years ago (Dickerson 1971) by further gene duplications (isoform duplications). Different mechanisms might have operated in the two active periods. In the early period before the parazoan–eumetazoan split, shufflings of different functional domains may play an important role for generating distinct subtypes with diverse functions. In contrast, newly created genes in the latter period are exclusively isoform types that are virtually identical to each other in structure and function, differing only in tissue distribution. Chromosomal duplications might be a possible mechanism for generating diverse isoforms in the latter period (Ohno 1970; Rousset et al. 1995; Bailey et al. 1997; Amores et al. 1998; Pèbusque et al. 1998; Suga et al. 1999c).

A remarkable consequence suggested by the scenario of gene diversification during animal evolution is that the Cambrian explosion, the burst of diversification of the major group of animal phyla at the Cambrian/Vendian boundary about 600 million years ago (Conway Morris 1993) might directly link with neither the burst of subtype duplication nor that of isoform duplication. It still remains possible that the Cambrian explosion directly links to the creation of a certain type of genes encoding the transcription factors involved in the body plan. Although many data should be accumulated before final conclusion, we prefer a hypothesis of gene recruitment as a possible molecular mechanism for explaining the Cambrian explosion (Hoshiyama et al. 1998; Suga et al. 1999a, 1999c; Ono et al. 1999): By recruiting already existing genes for other purposes in different developmental stages, a variety of animals with diverse body plans might have been possible to evolve without creating new genes with novel functions. The *Pax-6* would provide a typical example for gene recruitment. Members of the *Pax* gene family, transcription factors involved in developmental control, express in various restricted territories in the neural tube (Strachan and Read 1994; Mansouri et al. 1996). One of the subtype, *Pax-6*, is shown to express repeatedly in different developmental stages and adult tissues for different roles (Strachan and Read 1994; Mansouri et al. 1996; Callaerts et al. 1997; St-Onge et al. 1997). Furthermore, Hoshiyama et al. (1998) isolated cDNA encoding one of the subtype *Pax-2/5/8* from sponges that are thought to lack cell cohesiveness and coordination typical of eumetazoans (Margulis and Schwartz 1998). In addition, a phylogenetic analysis suggests that most of subtype duplications antedate the divergence of parazoans and eumetazoans, as in the cases of gene families involved in the signal transduction (Hoshiyama et al. 1998). These lines of evidence support the hypothesis of gene recruitment for a possible molecular mechanism of the Cambrian explosion.

Acknowledgments. We thank 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

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

Brady-Kalnay SM, Tonks NK (1995) Protein tyrosine phosphatases as adhesion receptors. Curr Opin Cell Biol 7:650–657

Callaerts P, Halder G, Gehring WJ (1997) *Pax*-6 in development and evolution. Annu Rev Neurosci 20:483–532

Chien CB (1996) PY in the fly receptor-like tyrosine phosphatases in axonal pathfinding. Neuron 16:1065–1068

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

Denu JM, Stuckey JA, Saper MA, Dixon JE (1996) Form and function in protein dephosphorylation. Cell 87:361–364

Dickerson RE (1971) The structures of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

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

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Hasegawa M and 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 *Pax* family. J Mol Evol 47:640–648

Hunter T (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. Cell 80:225–236

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

Katoh K, Miyata T (1999) A heuristic approach of maximum likelihood method for inferring phylogenetic tree and an application to the mammalian SOX-3 origin of the testis-determining gene SRY. FEBS Lett 463:129–132

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

Mansouri A, Hallonet M, Gruss P (1996) *Pax* genes and their roles in cell differentiation and development. Curr Opin Genet Dev 8:851–857

Margulis L, Schwartz KV (1998) Five kingdoms, 3rd ed. W.H. Freeman and Company, New York, NY

Mauro LJ, Dixon JE (1994) "Zip codes" direct intracellular protein tyrosine phosphatases to the correct cellular "address." Trends Biochem Sci 19:151–155

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

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

Nikoh N, Iwabe N, Kuma K, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Shicui Z, Hori K, Shimura Y, Miyata T (1997) An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. J Mol Evol 45:97–106

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 sponges and extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split. J Mol Evol 48:654–662

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

Stoker AW (1996) Axon guidance: motor-way madness. Curr Biol 6:794–797

St-Onge L, Sora-Pineda B, Chowdhury K, Mansouri A, Gruss P (1997) *Pax6* is required for differentiation of glucagon-producing α-cells in mouse pancreas. Nature 387:406–409

Strachan T, Read AP (1994) PAX genes. Curr Opin Genet Dev 4:427–438

Streuli M (1996) Protein tyrosine phosphatases in signaling. Curr Opin Cell Biol 8:182–188

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 (1999a) Extensive gene duplication in the early evolution of animals before parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. J Mol Evol 48:646–653

Suga H, Ono K, Miyata T (1999b) Multiple TGF-β receptor related genes in sponge and ancient gene duplications before the parazoan-eumetazoan split. FEBS Lett 453:346–350

Suga H, Hoshiyama D, Kuraku S, Katoh K, Kubokawa K, Miyata T (1999c) Protein tyrosine kinase cDNAs from amphioxus, hagfish and lamprey: Isoform duplications around the divergence of cyclostomes and gnathostomes J Mol Evol 49:601–608

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tonks NK, Neel BG (1996) From form to function: signaling by protein tyrosine phosphatases. Cell 87:365–368