FEBS Letters 453 (1999) 346–350 FEBS 22203

Multiple TGF-β receptor related genes in sponge and ancient gene duplications before the parazoan–eumetazoan split[1]

Hiroshi Suga, Kanako Ono, Takashi Miyata*

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

Received 10 May 1999; received in revised form 26 May 1999

Abstract Members of the transforming growth factor β (TGF-β) family mediate key events in cell growth and development. Various receptors for diverse members of the TGF-β family have recently been isolated and sequenced. These receptors form a family (TβR family) with a Ser/Thr kinase domain in common. To understand the divergence pattern of the TβR family during animal evolution, we have conducted cloning of cDNAs encoding the TβR family members from *Ephydatia fluviatilis*, a freshwater sponge. We obtained seven cDNAs (sALK-1–sALK-7) which are closely related in structure to known family members. Including these sponge sequences, a phylogenetic tree of the family members was inferred by a maximum likelihood method. The phylogenetic tree suggests that the sponge receptors sALK-1–sALK-3, which are closely related to each other, are sponge homologs of vertebrate activin type I receptor (ActR-I). sALK-5 is likely to be a homolog of TGF-β type II receptor. sALK-4 and sALK-6 might be ancestral precursors of type I and type II receptors, respectively, and sALK-7 is possibly an ancestral precursor of both types. The tree revealed that most, if not all, of the gene duplications that gave rise to known subtypes with distinct ligand specificities antedate the divergence of parazoans and eumetazoans, the earliest divergence of extant animal phyla.

Key words: Protein kinase; Transforming growth factor β receptor family; Sponge; Gene duplication; Phylogenetic tree; Evolution

1. Introduction

The transforming growth factor β (TGF-β) family comprises diverse members which mediate key events in cell growth and development through signal transduction (e.g. [1–3] for review). Various receptors for members of the TGF-β family have already been isolated and their structures characterized [1]. These receptors are transmembrane proteins with an amino-terminal signal sequence, an extracellular region with an cysteine-rich segment, a transmembrane segment, and an intracellular region containing a Ser/Thr kinase domain, and they form a protein family (TβR family) with a Ser/Thr kinase domain in common. All known members of the TβR family are classified into two groups, type I and type II receptors, which are distinguished from each other by the presence or absence of a short stretch of tandem Ser/Gly residues (GS box) immediately preceding the kinase domain [1]. The tetrameric form comprising the type I and type II receptors is thought to be essential for signal transduction (e.g. [2,3] for review).

Although recent progress has been made in isolating and characterizing the members of the TβR family, little is known about the evolutionary diversification of the family members. In this paper, we have conducted a molecular phylogenetic analysis of this family. Particularly, to obtain a rough estimate for dates of divergence of the family members by gene duplications, we have carried out cloning and sequencing of cDNAs from *Ephydatia fluviatilis*, a freshwater sponge, and have shown that there exist many members of the TβR family in sponges, the most primitive multicellular animals, which are thought to be lacking the cell cohesiveness and coordination typical of eumetazoans [4]. From a phylogenetic analysis of the family members, we report here that gene duplications that gave rise to distinct subtypes antedate the divergence of the parazoans and eumetazoans, the earliest branching among extant animal phyla. This pattern of gene diversification is quite similar to those observed in other families involved in cell-cell communication and developmental control [5–9].

2. Materials and methods

2.1. Isolation and sequencing of sponge cDNAs

Total RNA of *Ephydatia fluviatilis* (a freshwater sponge) was extracted from the cells hatched from the gemmules using Trizol reagent (Gibco BRL) and reverse transcribed to cDNA using oligo(dT) primer with reverse transcriptase (SuperScript II, Gibco BRL). The degenerate primers for PCR were designed from conserved amino acid residues as follows: (S1), 5′-AAGCCIGCIATIDSICAYMGNGAY-3′, corresponding to the amino acid sequence KPA(I/M)(A/S)HRD for a sense primer; (A1), 5′-TCIGGIGCCATRTAICKIIBIGTNCC-3′, corresponding to GTXRYMAPE, and (A2), 5′-TCCCATAGTAC-TAGTCCIAIIGMVYAIAYRTC-3′, corresponding to D(I/M/V)(Y/W)(A/S)(F/L/M/V)GLVLWE for antisense primers.

PCR was performed with primers S1 and A1, or S1 and A2 using the Expand High Fidelity PCR system (Roche). PCR conditions were 40 cycles of denaturation at 94°C for 20 s, annealing at 46°C for 1 min, and extension at 72°C for 1 min. 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 the BigDye Terminator Cycle Sequencing Ready Reaction kit and ABI Prism 377 DNA Sequencer (Perkin-Elmer). The full-length sponge TβR related sequences were obtained by the 3′ and 5′ RACE system for rapid amplification of cDNA ends (Gibco BRL).

2.2. Sequence alignment and phylogenetic tree inference

Alignments were made by the methods of Needleman and Wunsch [10] and Berger and Munson [11], together with manual inspection.

The phylogenetic tree of the TβR family was inferred by the method described previously [12]. This method is an approximate method

*Corresponding author. Fax: (81) (75) 753-4223.
E-mail: miyata@biophys.kyoto-u.ac.jp

[1] The nucleotide sequence data reported in this paper have been deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases (accession numbers AB026824–AB026830).

Abbreviations: TGF-β, transforming growth factor β; TβR, TGF-β receptor; ActR, activin receptor; BMPR, bone morphogenetic protein receptor; AMHR, anti-Müllerian hormone receptor

PII: S0014-5793(99)00749-8

for inferring the maximum likelihood (ML) tree of protein phylogeny [13] and consists of performing rearrangement of tree topology for a limited number of initial trees by the methods of nearest-neighbor interchange (NNI) [14,15] and subtree pruning and regrafting (SPR) [14]. The local bootstrap probability (LBP) at each node of a tree was calculated by the RELL method [13,16].

3. Results and discussion

3.1. Evolutionary relationship of the TβR family with other families belonging to the protein kinase superfamily

Over 4000 different protein kinases (PKs) have already been identified and sequenced from a wide range of eukaryotic species, and they form an extremely large superfamily comprising a variety of families including the TβR family (the Protein Kinase Resource, http://www.sdsc.edu/Kinases/pk_home.html). To understand the phylogenetic relationship of the TβR family with other families belonging to the PK superfamily, a phylogenetic tree was inferred for 21 major families, based on the comparison of the shared kinase domain of 88 amino acids in aligned length (Fig. 1). According to Fig. 1, members in the same family are clustered in the tree to form an independent family [18,19], and all these families diverged before the divergence of animals–fungi–plants. The TβR family is most closely related to the *raf* and protein tyrosine kinase (PTK) families [18,19]. The bootstrap probability that members in the TβR family have diverged from a common ancestral gene is very high, being 97%. Because no member of the TβR family has been identified in eukaryotic species other than animals, it is highly likely that members of the TβR family diverged from a common ancestral precursor during animal evolution by gene duplication.

From phylogenetic analyses of many gene families, we previously showed that, in a gene family whose members diverged specifically in animal lineages from a common ancestral gene, the family members are classified into several subtypes or subfamilies with distinct structures and functions, and gene duplications (subtype duplications) that gave rise to different subtypes antedate at least the divergence of vertebrates and arthropods. After the separation from arthropods, vertebrates underwent further gene duplications (isoform duplications) that gave rise to multiple members in the same subfamily which are virtually identical in structure and function. Most of these isoform duplications have been completed at least before the divergence of fishes and tetrapods [20,21]. The correspondence between the clustering pattern of subtypes in family tree and structural and/or functional difference may be useful for classifying family members whose functions are not yet identified [20,21].

3.2. Cloning and sequencing of sponge cDNAs belonging to the TβR family

Various members of the TβR family have already been identified in vertebrates and *Drosophila*. To know whether or not the related receptors exist in sponges, the most primitive animals among extant metazoans [4], we have carried out cloning and sequencing cDNAs of the TβR family members from *Ephydatia fluviatilis*, a freshwater sponge, by the method described in Section 2. We have obtained seven cDNAs (sALK-1–sALK-7), all of which are transmembrane proteins with an amino-terminal signal sequence, an extracellular region with a cysteine-rich segment, a transmembrane segment, and an intracellular region containing a Ser/Thr kinase domain. The structures of the sponge receptors were compared with those of known family members (Fig. 2).

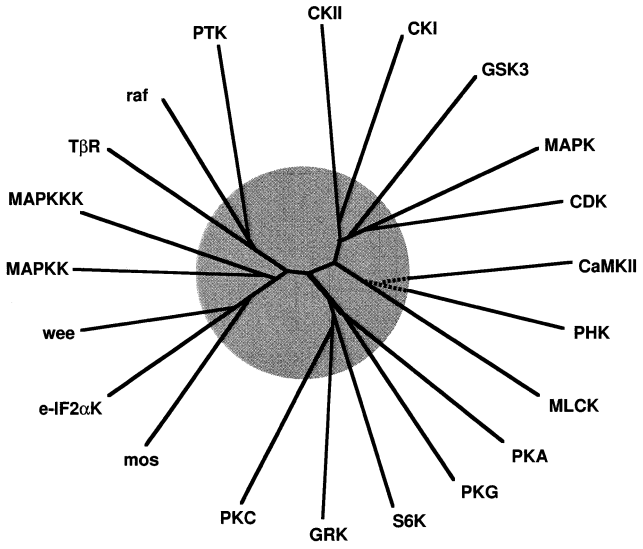

Fig. 1. Phylogenetic relationships among the major groups of the protein kinase superfamily. The tree was inferred as follows. On the basis of the alignment of kinase domains of 662 PK sequences from a wide range of eukaryotic species, a phylogenetic tree was inferred by the neighbor-joining (NJ) method [17]; the alignment used for tree inference consists of 88 amino acids in length, excluding gaps. From the inferred tree, 21 families with large numbers of members were selected and their phylogenetic relationships are schematically shown (branch lengths are arbitrary). Note that members in the same family are clustered in the NJ tree, and gene duplications that gave rise to different families antedate the animal–fungal–plant splits (indicated by a circle); the divergence times of CaMKII, PHK and MLCK are unknown (indicated by dotted lines). Abbreviations of family names: CK, casein kinase; GSK, glycogen synthase kinase; MAPK, mitogen activated protein kinase; CDK, cyclin dependent kinase; CaMK, calcium-calmodulin dependent kinase; PHK, phosphorylase kinase; MLCK, myosin light chain kinase; PKA, cAMP dependent kinase; PKG, cGMP dependent kinase; e-IF2αK, eukaryotic initiation factor-2α kinase; GRK, G protein coupled receptor kinase; TβR, transforming growth factor β receptor; S6K, ribosomal protein S6 kinase; PKC, protein kinase C; MAPKK, mitogen activated protein kinase kinase; MAPKKK, mitogen activated protein kinase kinase kinase; PTK, protein tyrosine kinase. Sequence data were taken from GenBank release 97 and Protein Identification Resource release 49.

sALK-1–sALK-3 are closely related to each other in sequence and have a short series of tandem Ser/Gly residues (GS box), a hallmark of the type I receptor, immediately preceding the kinase domain. sALK-4 has also the GS box. Thus sALK-1–sALK-4 are likely to be type I receptors. In sALK-5 and sALK-6, however, no clear GS box-like sequence is found. From the lack of a GS box, together with phylogenetic positions (see below), sALK-5 and sALK-6 are likely to be type II receptors. In addition to the GS box, there appears to exist another hallmark consisting of a short stretch of cysteine residue (Cys box) in the C-terminal end of the cysteine-rich segment: the amino acid sequence of the Cys box is $CCX_{4-5}CN$ in the type I receptors, whereas in the type II receptors, the corresponding sequence is $CXCX_4CN$ (Fig. 2). Furthermore, in the type I receptors, the homology region of the cysteine-rich segment is more extensive than that in the type II receptors (Fig. 2). sALK-7 is similar to the type I receptors in structure rather than to the type II receptors, although amino acid alternations exist in the GS box. This

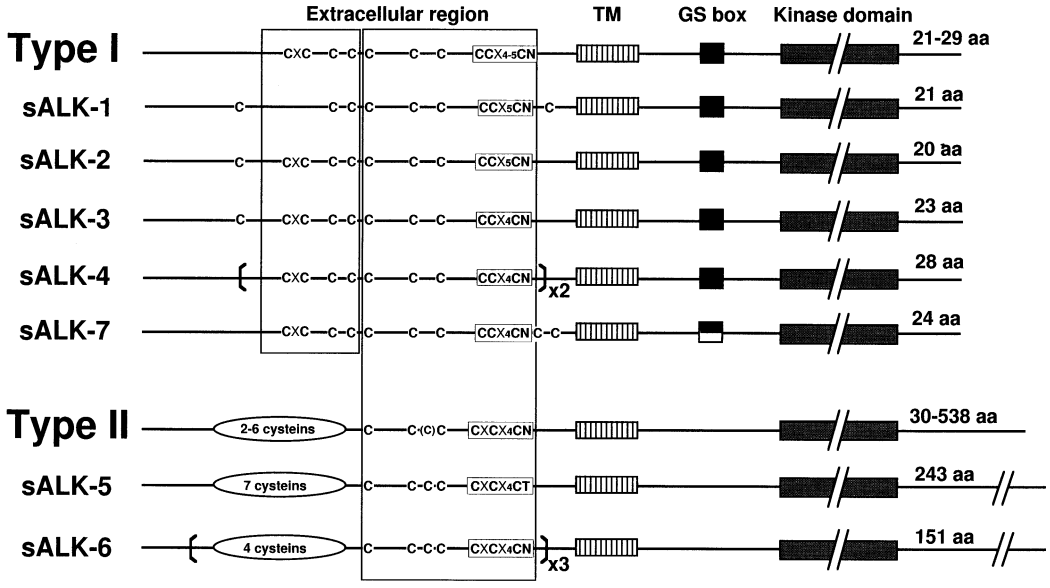

Fig. 2. The domain structures of TβR family members from human and sponge. Type I and Type II, the consensus sequences of known type I and type II receptors, respectively; sALK-1–sALK-7, sponge receptors; TM, transmembrane segment; GS box, a short stretch of tandem Ser/Gly residues. Segments in which sequence similarities are observed between different sequences are boxed. A cysteine residue enclosed by parentheses in Type II is absent in the TGF-β type II receptor. In the extracellular region of the sponge sALK-4 and sALK-6, the cysteine-rich segments enclosed by brackets are tandemly duplicated and triplicated, respectively; only the N-terminal repeat is shown. The sALK-7 has amino acid alternations in GS box (represented by a half-filled box). The number of amino acids from immediately after the conserved arginine in the subdomain XI [18,19] of the kinase domain to C-terminal is shown for each receptor.

receptor may be a common ancestral precursor of the type I and type II receptors (see below). Interestingly, sALK-4 has tandemly duplicated cysteine-rich segments in the extracellular region. Also in the sALK-6, the cysteine-rich segment is triplicated.

3.3. *Phylogenetic tree of the TβR family and subtype duplications*

The amino acid sequences of the sponge cDNAs (sALK-1–sALK-7) were aligned with those of the known members of the TβR family from vertebrates and *Drosophila* for a highly conserved region (amino acid sites 233–524 in human bone morphogenetic protein type IA receptor (BMPR-IA)) of the kinase domain. On the basis of the alignment, a phylogenetic tree was inferred by the method described previously [12] (see also Section 2), using the *raf* sequences as an outgroup (Fig. 3).

As Fig. 3 shows, the inferred tree of the TβR family revealed seven separate clusters (subfamilies), consisting of BMPR-I, activin type I receptor (ActR-I), a group comprising TGF-β type I receptor (TβR-I) and activin type I receptor B (ActR-IB), TGF-β type II receptor (TβR-II), activin type II receptor (ActR-II), BMPR-II and anti-Müllerian hormone receptor (AMHR). In addition, gene duplications that gave rise to different clusters predate the vertebrate–arthropod (*Drosophila*) split. Furthermore, the clustering pattern of members in the tree coincides well with the structural and functional (ligand specificity) differences, except for TβR-I/ActR-IB group. Because the same subfamily-function relationship is observed in many gene families examined to date [20,21], it might be reasonable to consider the above seven groups as separate subtypes (phylogenetically, a subtype corresponds to a subfamily).

According to Fig. 3, the sponge sALK-1–sALK-3 are clustered with a bootstrap probability of 100% and they belong to the activin-RI/*sax* subfamily, suggesting that sALK-1–sALK-3 are homologs of vertebrate ActR-I. The sponge sALK-5 is closely associated with human TβR-II, suggesting that they are cognate genes, although the bootstrap probability is not high enough (62%). A detailed phylogenetic analysis of the type II receptors based on the ML method of protein phylogeny [13] suggests that there is still a possibility that sALK-5 is clustered with the ancestral lineage of BMPR-II/AMHR (data not shown). The sponge sALK-6 is probably a type II receptor, judging from the high bootstrap probability (= 99%) and the lack of a GS box. sALK-6 might be an ancestral precursor of the type II receptors. Fig. 3 also suggests that the sponge sALK-4, sALK-6 and sALK-7 are new subtypes, although detailed analyses might be necessary before a final conclusion can be drawn.

According to the phylogenetic tree of Fig. 3, the subtype duplications (indicated by filled rhombi) that gave rise to all known subtypes, as well as the newly identified sALK-4, sALK-6 and sALK-7, always predate the divergence of parazoan (sponge) and eumetazoans, the earliest divergence among extant animal phyla; for a subtype duplication (half-filled rhombus) giving rise to BMPR-II and AMHR, the divergence time is unknown. To estimate the numbers of subtype duplications before and after the divergence of parazoans and eumetazoans in a statistically solid manner, we have carried out a bootstrap analysis by the method described previously [21]: on the basis of the standard bootstrap procedure [22], we have generated 100 ML trees with the largest log-likelihood value by repeated local rearrangements [15], using the NJ trees [17] as initial trees. From the set of ML trees, the numbers of isoform duplications that took place at dates before and after the parazoan–eumetazoan split were counted (designated as N_b and N_a, respectively). From the above boot-

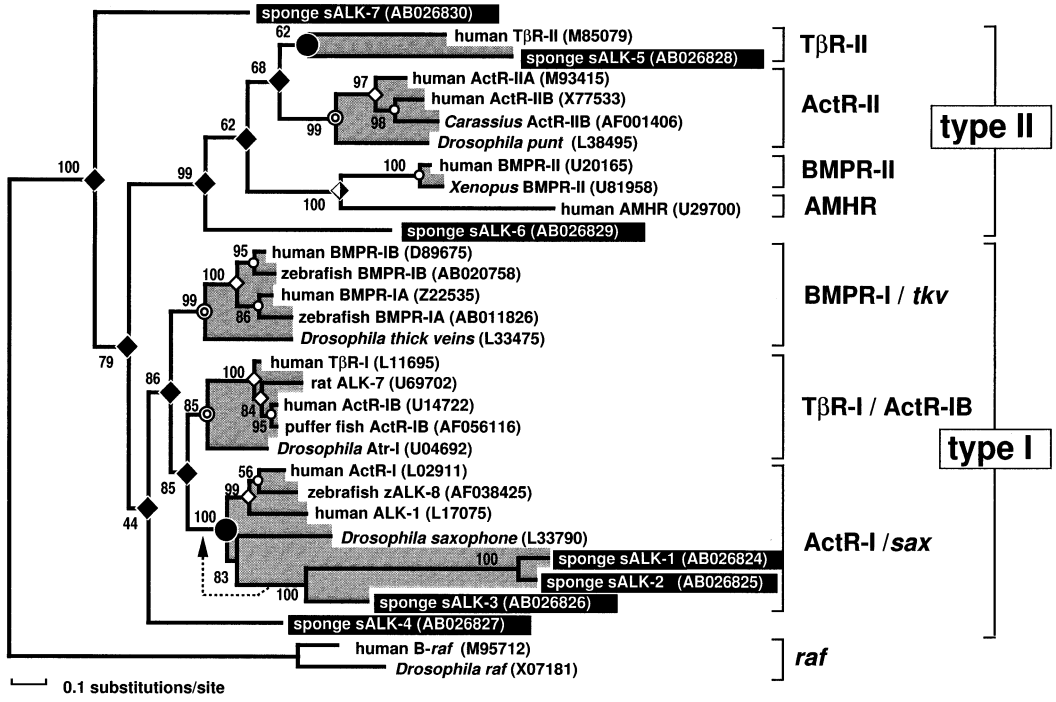

Fig. 3. Phylogenetic tree of the TβR family. From a comparison of the kinase domain sequences, the tree was inferred by a ML method described previously [12], using the *raf* sequences as an outgroup. The number at each branch node represents the local bootstrap probability estimated by the RELL method [13,16]. Open circles, fish–tetrapod split or amphibian–amniote split; double circles, vertebrate–arthropod (*Drosophila*) split; filled circle, parazoan (sponge)–eumetazoan split; open rhombi, gene duplications that gave rise to isoforms in the same subtype during vertebrate evolution. Note that all the open rhombi antedate the open circles; filled and half-filled rhombi, gene duplications that gave rise to different subtypes. Note that all the filled rhombi antedate the filled circles; for the half-filled rhombus, the divergence time is unknown. According to a detailed phylogenetic analysis among the type I receptors, it is equally likely that the cluster of the sponge sALK-1–sALK-3 is an outgroup of the vertebrate ActR-1 and *Drosophila sax*, as indicated by a dotted arrow (data not shown). Accession numbers of sequences are shown in parentheses; reverse letters, present work (sponge).

strap analysis, we have $N_b = 6.4 \pm 1.3$ and $N_a = 1.3 \pm 1.3$ (Fig. 4). Thus we conclude that most, if not all, of the subtype duplications occurred in the very early evolution of animals and an almost complete set of subtypes was established before the divergence of parazoans and eumetazoans. Similar divergence patterns were observed in other families involved in signal transduction and developmental control, including the protein tyrosine kinase family [5], G protein α subunit family [5], phosphodiesterase family [6], phospholipase C family [7], protein kinase C family [7], protein tyrosine phosphatase family [8] and *Pax* family [9].

The ancient divergence of the TβR family members has evolutionary implications. First, from the molecular clocks of aldolase and triose phosphate isomerase, the divergence time of parazoans and eumetazoans and that of animal and fungi (or plants) were estimated to be 940 and 1070 million years (Myrs) ago, respectively [23]. If these divergence times are really correct, it seems reasonable to consider that the set of subtypes of the TβR family was created rapidly within a short time period of 100 Myrs or so. Second, ligand molecules of the TGF-β family are thought to be involved in axis formation in diverse animals during early development and regulate other diverse processes in later development of triploblasts (e.g. [1,3] for review). It is interesting that an almost complete set of basic receptors (subtypes) for the TGF-β family members existed in ancient primitive animals like sponges that are thought to be lacking the cell cohesiveness and coordination typical of eumetazoans [4]. It seems conceivable that in the ancient animals and sponges, these signalling molecules and their receptors might function for roles other than those found in triploblasts. Third, the subtype duplications predate the earliest divergence among extant animal phyla, as shown by the phylogenetic analysis of the TβR family, together with those of other families. In addition, as we have recently shown, isoform diversifications in the same subfamily are likely to have occurred around the divergence of cyclostomes

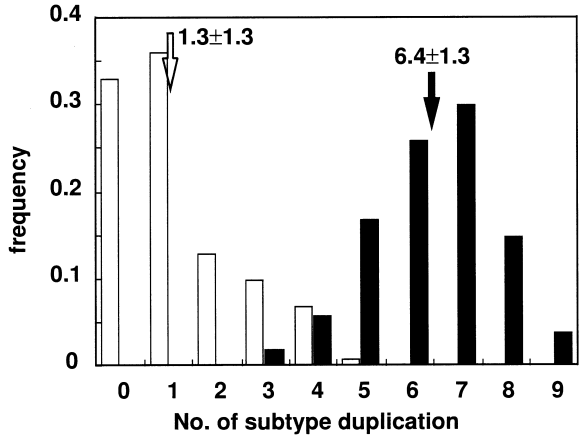

Fig. 4. Distribution of the numbers of subtype duplications before and after the divergence of parazoans and eumetazoans. Histograms of the numbers of subtype duplications before and after the parazoan–eumetazoan split are represented by filled and open bars, respectively. Arrowheads represent the average numbers of subtype duplications. For the method of calculation, see text.

and gnathostomes [12]. It is therefore likely that the Cambrian explosion, the explosive diversification of the major group of animal phyla at the Cambrian/Vendian boundary [24], was accomplished without creating new genes. Thus the molecular mechanism of the Cambrian explosion should be understood based on mechanisms which could generate organismal diversity by utilizing or recruiting preexisting genes, but not by creating new genes with novel functions.

Acknowledgements: 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

[1] Kingsley, D.M. (1994) Genes Dev. 8, 133–146.
[2] Massagué, J. (1996) Cell 85, 947–950.
[3] Hoodless, P.A. and Wrana, J.L. (1998) Curr. Top. Microbiol. Immunol. 228, 235–272.
[4] Margulis, L. and Schwartz, K.V. (1998) Five Kingdoms, 3rd edn., W.H. Freeman and Company, New York.
[5] Suga, H., Koyanagi, M., Hoshiyama, D., Ono, K., Iwabe, N., Kuma, K. and Miyata, T. (1999) J. Mol. Evol. (in press).
[6] Koyanagi, M., Suga, H., Hoshiyama, D., Ono, K., Iwabe, N., Kuma, K. and Miyata, T. (1998) FEBS Lett. 436, 323–328.
[7] Koyanagi, M., Ono, K., Suga, H., Iwabe, N. and Miyata, T. (1998) FEBS Lett. 439, 66–70.
[8] Ono, K., Suga, H., Iwabe, N., Kuma, K. and Miyata, T. (1999) J. Mol. Evol. (in press).
[9] Hoshiyama, D., Suga, H., Iwabe, N., Koyanagi, M., Nikoh, N., Kuma, K., Matsuda, F., Honjo, T. and Miyata, T. (1998) J. Mol. Evol. 47, 640–648.
[10] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[11] Berger, M.P. and Munson, P.J. (1991) CABIOS 7, 479–484.
[12] Suga, H., Hoshiyama, D., Kuraku, S., Katoh, K., Kubokawa, K. and Miyata, T. (1999) J. Mol. Evol. (in press).
[13] Kishino, H., Miyata, T. and Hasegawa, M. (1990) J. Mol. Evol. 30, 151–160.
[14] Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M. (1996) In: Molecular Systematics, 2nd ed. (Hillis, D.M., Moritz, C. and Mable, B.K., Eds.), pp. 407–514, Sinauer Associates, Sunderland, MA.
[15] Adachi, J. and Hasegawa, M. (1996) Computer Science Monographs, No. 28, MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. The Institute of Statistical Mathematics, Tokyo.
[16] Hasegawa, M. and Kishino, H. (1994) Mol. Biol. Evol. 11, 142–145.
[17] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[18] Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) Science 241, 42–52.
[19] Hanks, S.K. and Hunter, T. (1995) FASEB J. 9, 576–596.
[20] Iwabe, N., Kuma, K. and Miyata, T. (1996) Mol. Biol. Evol. 13, 483–493.
[21] Suga, H., Kuma, K., Iwabe, N., Nikoh, N., Ono, K., Koyanagi, M., Hoshiyama, D. and Miyata, T. (1997) FEBS Lett. 412, 540–546.
[22] Felsenstein, J. (1985) Evolution 39, 783–791.
[23] Nikoh, N., Iwabe, N., Kuma, K., Ohno, M., Sugiyama, T., Watanabe, Y., Yasui, K., Shicui, Z., Hori, K., Shimura, Y. and Miyata, T. (1997) J. Mol. Evol. 45, 97–106.
[24] Conway Morris, S. (1993) Nature 361, 219–225.