Divergence pattern of animal gene families and relationship with the Cambrian explosion

Takashi Miyata* and Hiroshi Suga

Summary

There are many gene families that are specific to multicellular animals. These have either diverged from ancestral genes that are shared with fungi and/or plants or evolved from an ancestral gene unique to animals. The evolution of gene families involved in cell–cell communication and developmental control has been studied to establish whether the number of member genes increased dramatically immediately prior to or in concert with the Cambrian explosion. A molecular phylogeny-based analysis of several animal-specific gene families has revealed that gene diversification by duplication occurred during two active periods interrupted by a long intervening quiescent period. Intriguingly, the Cambrian explosion is situated in the silent period, indicating that there is no direct link between the first burst of gene diversification and the Cambrian explosion itself. The importance of gene recruitment as a possible molecular mechanism for morphological diversity, and its possible role for the Cambrian explosion, are discussed. *BioEssays* 23: 1018–1027, 2001.

Introduction

In multicellular animals, several different gene families involved in cell–cell communication and developmental control have evolved through gene duplication and gene shuffling, basic mechanisms for generating diverse genes with novel functions.[1] Each of these animal gene families has originated either from ancestral genes that are shared with plants and fungi or from an ancestral gene created uniquely in the animal lineage (e.g., Ref. 2). The major groups of bilateral animals are thought to have diverged explosively at or prior to the Vendian–Cambrian boundary.[3] No direct molecular evidence has been provided to date as to whether the Cambrian explosion was triggered by a dramatic increase in the number of genes involved in cell–cell communication and developmental control either immediately prior to or in concert with the Cambrian explosion.

Department of Biophysics, Kyoto University, Japan.
Funding agency: The Ministry of Education, Science, Sports and Culture of Japan.
*Correspondence to: Takashi Miyata, Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. E-mail: miyata@biophys.kyoto-u.ac.jp

In this article, we will review molecular evidence on the relationship of animal gene families to the Cambrian explosion. Specifically, we will compare the rates of gene duplication in several crucial periods. On the basis of molecular phylogenetic analyses of eight animal-specific gene families, we provide evidence that gene diversification comprised two active periods of gene duplication, interrupted by considerably long periods of silence. In the first period, which took place before the parazoan–eumetazoan split, the earliest divergence among extant animal phyla,[4] animals underwent extensive gene duplications (subfamily-generating duplications) that gave rise to different subfamilies with diverse functions. Almost all the present-day subfamilies were established within this period. The second round of gene diversification in metazoan history occurred in the early evolution of vertebrates around the divergence of cyclostomes and gnathostomes. During this period, the number of members in each subfamily increased by gene duplications (intra-subfamily duplications) giving rise to different members that are expressed tissue specifically (Fig. 1). Remarkably, this analysis suggests that the Cambrian explosion was not immediately preceded or accompanied by active gene diversification, and there may be no direct link between the Cambrian explosion and the burst of gene duplication. Rather, these findings highlight the importance of the origination of a small number of key genes that might have triggered the Cambrian explosion and of gene recruitment of pre-existing genes. These alterations to the genome, together with ecological and environmental changes, are discussed as alternative mechanisms for generating diverse animals at the morphological level.

Classification of family members and their divergence times

By comparing the amino acid sequences of functional domains of transcripts of members of a gene family, it is possible to infer a phylogenetic tree. These results give valuable information on the divergence of family members by gene duplications. In most eukaryotic gene families analyzed to date, the phylogenetic tree of members belonging to a gene family comprises several subfamilies. Members in the same subfamily have virtually identical function in most cases and exhibit sequence similarity over the entire region, but they often differ

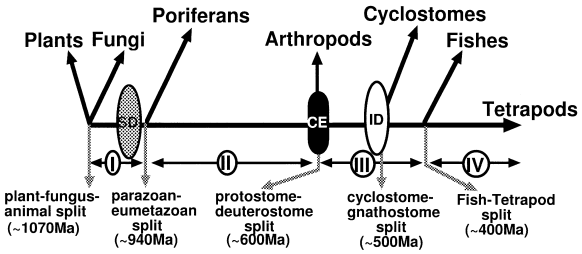

Figure 1. Branching order of the major groups of animals and phylogenetic positions of the bursts of gene duplications. The whole animal lineage leading to the present-day tetrapods since the separation from fungi and plants was divided into four periods, period I, period II, period III, and period IV; the boundaries were tentatively defined by the times of plant–fungus–animal splits, parazoan–eumetazoan split, protostome–deuterostome split, and fish–tetrapod split, which were assumed to have occurred about 1070 million years ago[5] (Ma) , 940 Ma,[5] 600-700 Ma,[4,6,7] 500 Ma,[6] and 400 Ma,[6] respectively. SD and ID, extensive gene duplications that gave rise to subfamilies and members of a particular subfamily, respectively; CE, the Cambrian explosion.

in tissue distribution (e.g., Ref. 2). We refer to members of the same subfamily as members of a subfamily or subfamily members. In contrast, members belonging to different subfamilies differ in structure and/or function of the encoded proteins; they often have structures distinct in domain organization. The presence of such subfamily structure in a family tree suggests that gene duplications (subfamily-generating duplications) that gave rise to different subfamilies antedate gene duplications (intra-subfamily duplications) that gave rise to different members of a particular subfamily, and, as we will discuss, there is no overlap between the dates of subfamily-generating duplications and those of intra-subfamily duplications.

To understand the possible relationships between gene diversification and organismal diversification, the divergence times of family members by gene duplication must be known. A rough estimate for the divergence time is possible by inferring a composite tree of a gene family using members from various species; the branching node of orthologous genes from different species corresponds to the divergence of these species. Molecular clocks do not provide reliable estimates for the divergence times, because the evolutionary rates of family members differ for different periods of animal evolution and different subfamilies. For animal-specific gene families that diverged during animal evolution, many family members from vertebrates and arthropods have already been isolated and sequenced. In contrast, only a few members have been identified in such primitive animals as sponges and diploblasts including hydra. To complete the composite tree of an animal-specific gene family, the comprehensive isolation of family members from these primitive animals is required.

Divergence of gene families involved in cell–cell communication and developmental control in the early evolution of animals

As a typical example of an animal-specific gene family, we first consider the diversification of the protein tyrosine kinase (PTK) family of genes. The protein kinases are a large family of proteins that share similarities in their catalytic (kinase) domain sequences. They are divided into two major families, the protein serine/threonine kinase (PSK) family and the PTK family, on the basis of the respective substrate specificities (reviewed in Refs. 8–10). The phylogenetic tree of the protein kinases revealed an independent cluster of the PTKs, which originated from a group of PSKs.[8] The PTKs are recognized only in multicellular animals, as demonstrated by the lack of PTK homologs in *S. cerevisiae* genome,[11] while the PSKs exist in a wide range of species covering almost all eukaryotes. Functionally, the PTKs are involved in the signal transduction and cell-cell interactions that control cell proliferation and differentiation (reviewed in Refs. 8–10, 12,13). It is therefore likely that the PTKs derived from a precursor PSK of a unicellular protist through rounds of gene duplication and extensive divergence during animal evolution.[8,9]

Figure 2 shows part of the phylogenetic tree of the PTK family inferred from a comparison of the kinase domain by a heuristic maximum likelihood (ML) method[15] (for the full description of the phylogenetic tree, see Ref. 14). This tree includes family members not only from vertebrates and arthropods, but also from sponge, a parazoan. Figure 2 contains five different subfamilies generated by four gene duplications that occurred prior to the divergence of parazoans and eumetazoans. The actual PTK tree comprises many subfamilies with distinct domain structures, as mentioned above.[8,16,17] According to the complete phylogenetic tree,[14] most of the gene duplications that gave rise to different subfamilies antedate the divergence of sponge and eumetazoans, the earliest branching among extant animal phyla; among twenty-nine subfamily-generating duplications, nineteen duplications antedate the parazoan–eumetazoan split and two duplications are found after that split; for eight subfamily-generating duplications, their divergence times are unknown.

According to a statistical analysis based on the bootstrap resamplings,[14] the number of subfamily-generating duplications found in a period (first period) of 130 million years (myrs) between the animal–fungal split and the parazoan–eumetazoan split and that (latter period) of 940 myrs after the parazoan–eumetazoan split are, respectively, 20.8 ± 1.8 and 1.9 ± 1.3. The number of subfamily-generating duplication per 100 myrs in the first period is therefore 16.0 ($= 20.8/1.3$), which is 80 times higher than that ($= 1.9/9.4 = 0.20$) in the latter period. In the *src* subfamily, three gene duplications are found in the first period, suggesting that this group is further

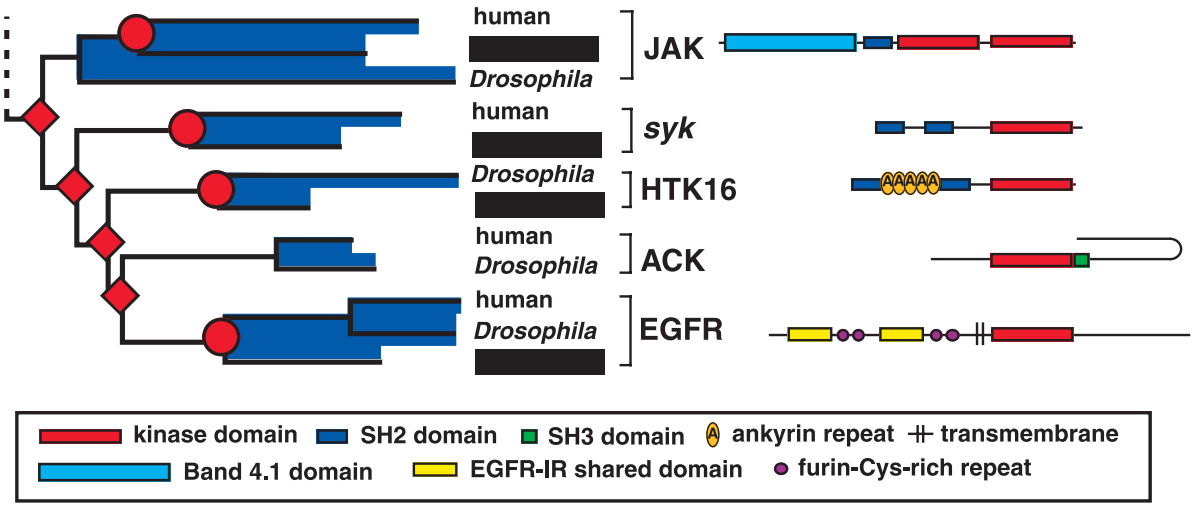

Figure 2. A part of the phylogenetic tree of the protein tyrosine kinase family inferred from the kinase domain. Members in the same subfamily are shaded. Filled circles, parazoan (sponge)–eumetazoan split. Filled rhombi, subfamily-generating duplications. The branch length is proportional to the number of accumulated amino acid substitutions. The domain organization of each subfamily is shown schematically. For the full description of the phylogenetic tree, as well as the method for tree inference, see ref. 14.

subdivided into several subfamilies with distinct functions.[16] Since the PTKs are recognized only in animals.[8,9] and the phylogenetic tree of the protein kinases including the PSKs shows an independent cluster from the PTKs that originated from a group of PSKs,[8] it is highly likely that the extensive subfamily-generating duplications were completed in the early evolution of animals before the parazoan-eumetazoan split.

Other gene families involved in the signal transduction and developmental control examined to date also show the same pattern of subfamily divergence. These include the G protein α subunit (Gα) family,[17] the protein tyrosine phosphatase (PTP) family,[18] the cyclic nucleotide phosphodiesterase (PDE) family,[19] the phospholipase C (PLC) family,[20] the protein kinase C (PKC) family,[20] the transforming growth factor-β receptor (TβR) family,[21] and the Pax family.[22] The number n_b of subfamily-generating duplications in the first period and that n_a in the latter period are summarized in Table 1. As the table shows, most, if not all, of subfamily-generating duplications are observed in the first period, while in the latter period extending about 900 myrs, this type of gene duplication is seldom observed.

In addition to the number of gene duplications, the evolutionary rate of gene family averaged among different subfamilies also differ greatly between the first and latter periods. The average evolutionary rate v_b of the first period is always higher than that v_a of the latter period for the eight gene families examined to date (Table 1). The rapid evolutionary rate in the first period might be related to the frequent subfamily-generating duplications, which resulted in relaxed functional constraints.

These results have several evolutionary implications. (1) Frequent gene duplications that gave rise to different subfamilies with distinct domain organizations and functions occurred in the very early evolution of animals, and these subfamilies were formed before the divergence of parazoans and eumetazoans, the earliest divergence among extant animals. (2) These duplications are characterized by explosive occurrence within a limited period, instead of proceeding gradually. (3) These duplications occurred at approximately the same time (i.e., in a limited period before the parazoan–eumetazoan split) for different gene families examined to date. It is not known, of course, whether these duplications occurred piecemeal or involved genome duplications. (4) The explosive subfamily-generating duplication precedes the Cambrian explosion by about 200–300 myrs. This strongly suggests that there was no direct link between subfamily-generating duplications (i.e., gene diversification) and the Cambrian explosion (i.e., organismal diversification). Instead, these duplications might be related to the acquisition of multicellularity during animal evolution rather than to the Cambrian explosion. A similar phylogenetic analysis of gene families including genes from choanoflagellates, unicellular protists, is interesting in this regard, because the choanoflagellates are thought to be the sister group of metazoans.[4,23,24]

Frequent domain shufflings in the early evolution of animals

A possibility still exists, however, that there was a direct link between the Cambrian explosion and subfamily diversification. Note that the results shown in Table 1 are those obtained

Table 1. The number of subfamily-generating duplications and the average evolutionary rates in the first and the latter periods of animal evolution

Family	n_b	f_b	v_b	n_a	f_a	v_a	f_b/f_a	v_b/v_a
PTK	20.8 ± 1.8	16.0	1.4	1.9 ± 1.3	0.20	0.34	80	4.1
Gα	8.5 ± 0.9	6.5	1.5	1.1 ± 1.2	0.12	0.24	54	6.3
PDE	5.0 ± 0.1	3.8	2.1	0.0 ± 0.0	0.00	0.36	—	5.8
PLC	3.0 ± 0.3	2.3	0.91	0.0 ± 0.1	0.00	0.26	—	3.5
PKC	2.8 ± 0.7	2.2	0.43	0.1 ± 0.4	0.01	0.28	220	1.5
PTP	11.9 ± 1.8	9.2	1.1	0.9 ± 1.0	0.10	0.38	92	2.9
TβR	6.4 ± 1.3	4.9	4.0	1.3 ± 1.3	0.14	0.45	35	8.9
Pax	2.5 ± 0.6	1.9	0.33	0.2 ± 0.6	0.02	0.10	95	3.3

n_b and n_a, the numbers of subfamily-generating duplications in the first and the latter periods, respectively; f_b and f_a, the average numbers of subfamily-generating duplications per 100 myrs in the first and the latter periods, respectively; v_b and v_a, the average evolutionary rates ($\times 10^{-9}$/site/year) in the first and the latter periods, respectively. The first period is tentatively defined as an early period of animal evolution corresponding to a period of 130 myrs between the animal–fungal–plant splits and the parazoan–eumetazoan split, and the latter period corresponds to the remaining periods of 940 myrs after the parazoan–eumetazoan split. The first and the latter periods correspond to period I and periods II–IV in Fig. 1, respectively. Abbreviations of gene family names and data sources: PTK, protein tyrosine kinase family[14]; Gα, G protein α subunit family[17]; PTP, protein tyrosine phosphatase family[18]; PDE, cyclic nucleotide phosphodiesteras family[19]; PLC, phospholipase C family[20]; PKC, protein kinase C family[20]; TβR, transforming growth factor β receptor family[21]; Pax, Pax family[22]. For the data sources of divergence times of animal groups, see legends to Figure 1.

from phylogenetic analyses based on comparisons of amino acid sequences of domains shared among all members of each family. Thus the divergence pattern mentioned above actually represents that of shared domains, but not of the subfamilies. Different subfamilies in animal-specific gene families often have different domain organizations that have been generated by the mechanism of domain shuffling, together with domain duplications. Therefore, there is still a possibility that, although domain duplications antedate the parazoan–eumetazoan split, significant domain shufflings occurred immediately prior or concomitant with the Cambrian explosion. This hypothesis could be tested by comparing the domain organizations of subfamily members from sponges and other eumetazoans.

In Figure 3, the domain organization of sponge PTK is compared with those of human and *Drosophila* PTKs for each of ten subfamilies belonging to the PTK family. For each of these subfamilies, lineages corresponding to human, *Drosophila* and sponge are closely associated with one another in tree to form a cluster, as shown in Figure 2. Thus, in each of ten subfamilies, the PTK sequences from human, *Drosophila* and sponge are possibly orthologous to each other. As Figure 3 shows, the domain organizations of the PTK sequences are virtually identical among human, *Drosophila* and sponge for each of the subfamilies, except for the *ror*/Musk subfamily, in which the domain organizations appear to differ for different species groups.[14] This result strongly suggests that most, if not all, of domain shufflings antedate the divergence of parazoans and eumetazoans. A similar result is also found in the PDE family,[19] which provides further supporting evidence for the ancient domain shufflings before the parazoan–eumetazoan split. Although much more data are needed, these results suggest that the creation of subfamilies in each of the animal-specific gene families by gene duplications and domain shufflings had already been completed prior to the parazoan–eumetazoan split.

Divergence of members of subfamily in the early evolution of vertebrates

In addition to the subfamily-generating duplications, most gene families increased family members in the same subfamily by further gene duplications (i.e., intra-subfamily duplications) during early vertebrate evolution.[2] Different members belonging to the same subfamily exhibit extensive sequence similarity over their entire regions and, although they are virtually identical to each other in function, they often differ in tissue distribution. The extensive intra-subfamily duplication in the first half of chordate evolution has been identified in many subfamilies belonging to various gene families.[2,16–18,22,25] To estimate the dates of the active period of intra-subfamily duplication more closely, we have isolated and sequenced cDNAs encoding the fibroblast growth factor receptor (FGFR), Eph, *src* and platelet-derived growth factor receptor (PDGFR) subfamilies of the PTK family from amphioxus, hagfish and lamprey. These sequences showed that the extensive intra-subfamily duplication occurred in a period around or immediately before the cyclostome–gnathostome split.[26]

Figure 4 shows a phylogenetic tree of the PDGFR subfamily inferred from a comparison of the kinase domain sequences; using a heuristic approach of maximum likelihood (ML) method.[15] After the separation from arthropods, at least seven intra-subfamily duplications (rhombi and boxes in Fig. 4) were identified on lineages leading to the modern vertebrates,

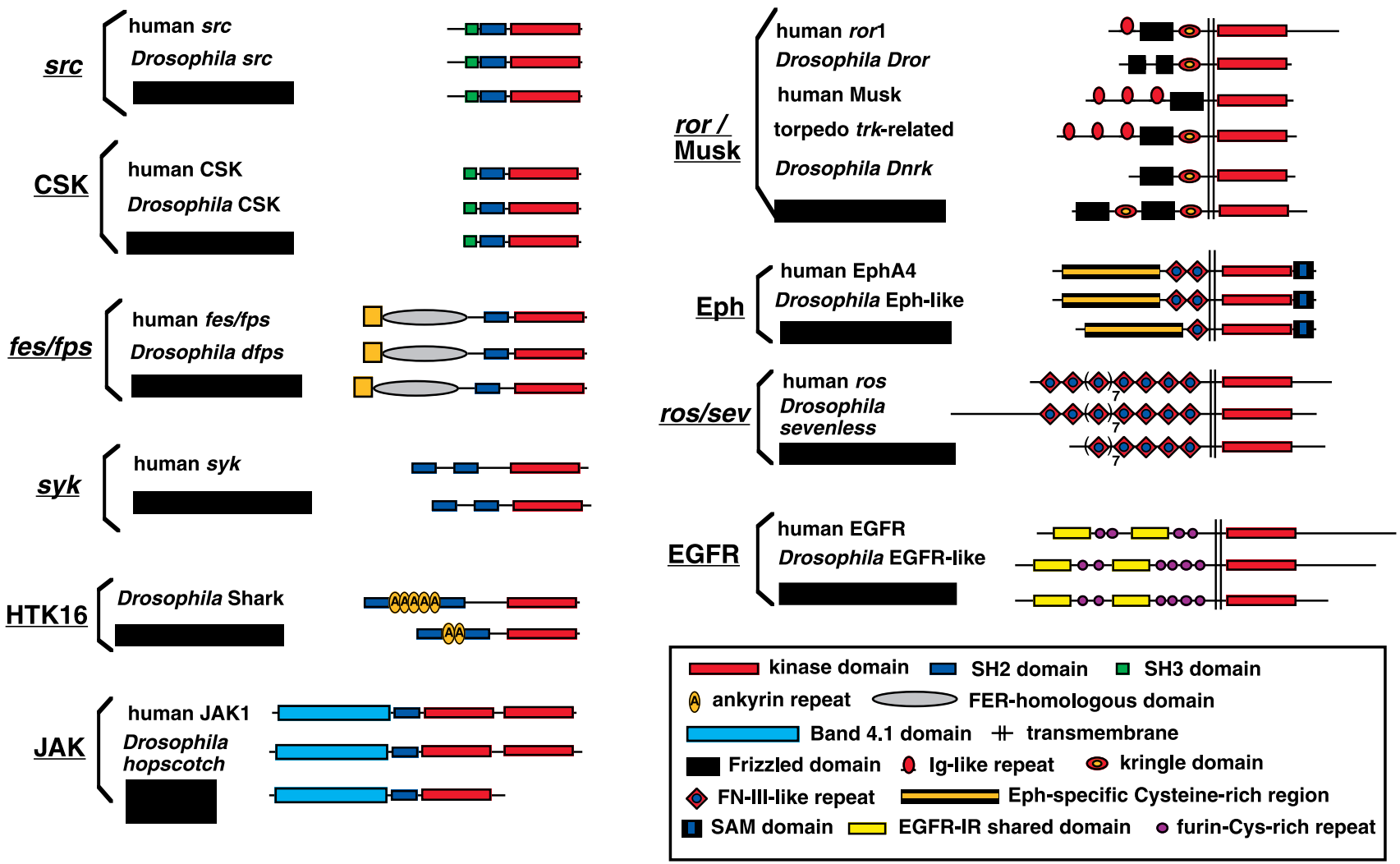

Figure 3. Comparison of domain organization of sponge PTK with those of human and *Drosophila* PTKs in each of ten different subfamilies.

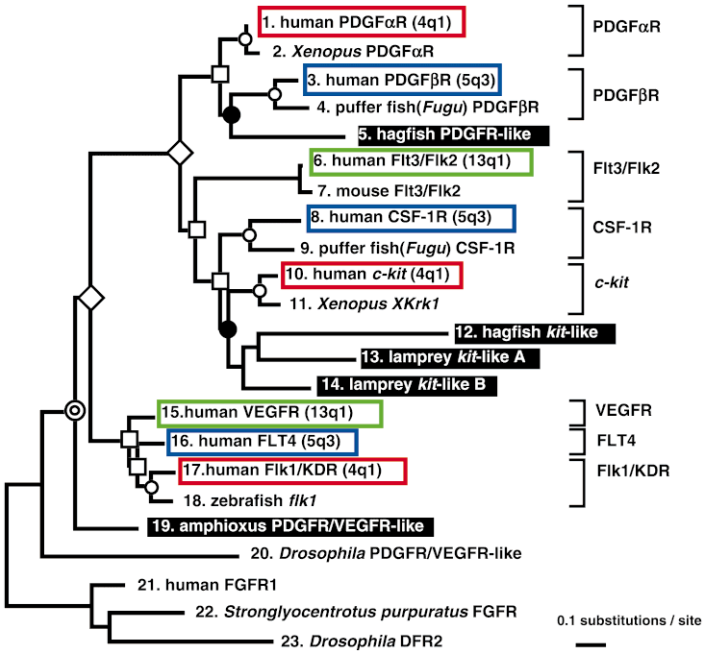

Figure 4. Maximum likelihood tree of PDGFR subfamily. From a comparison of the kinase domain sequences, a phylogenetic tree was inferred by a heuristic approach using the ML method,[15] with human, purple urchin and *Drosophila* FGFR sequences as an outgroup. Heterogeneity of evolutionary rates among sites is modelled by a discrete Gamma distribution[27] with shape parameter α of 0.6. Open circles, fish—tetrapod split or amphibian—amniote split; filled circle, cyclostome—gnathostome split; double circle, cephalochordate—vertebrate split; open rhombi, intra-subfamily duplications; open boxes, intra-subfamily duplications generated possibly by duplications of large chromosomal regions. Data on chromosomal mapping of human genes were taken from van der Geer et al.[12] Accession numbers of sequences used are as follows: 1, human PDGFαR (M21574); 2, *Xenopus* PDGFαR (M80798); 3, human PDGFβR (J03278); 4, *Fugu* PDGFβR (U63926); 5, hagfish PDGFR-like (AB025554); 6, human Flt3/Flk2 (U02687); 7, mouse Flt3/Flk2 (X59398); 8, human CSF-1R (U63963); 9, *Fugu* CSF-1R (U63926); 10, human *c-kit* (U63834); 11, *Xenopus XKrk*1 (Z48770); 12, hagfish *kit*-like (AB025553); 13, lamprey *kit*-like A (AB025555); 14, lamprey *kit*-like B (AB025556); 15, human VEGFR (AF063657); 16, human FLT4 (U43143); 17, human Flk1/KDR (AF035121); 18, zebrafish *flk1* (AAF03237); 19, amphioxus PDGFR/VEGFR-like (AB025557); 20, *Drosophila* CG8222 (AAF52626); 21, human FGFR1 (X57121); 22, *Stronglyocentrotus* FGFR (U17164); 23, *Drosophila* DFR2 (X74031).

(although one intra-subfamily duplication is found on the cyclostome lineages, we do not discuss this duplication further, and only intra-subfamily duplications on gnathostome lineages are considered in detail). All these duplications antedate the divergence of fishes and tetrapods, but follow the divergence of cephalochordates and vertebrates. Furthermore, these seven intra-subfamily duplications probably occurred before the divergence of cyclostomes and gnathostomes, although the differences are not statistically significant. From the bootstrap analysis,[26] the number of intra-subfamily duplications that took place before and after the divergence of cyclostomes and gnathostomes were estimated to be, on average, 3.9 ± 1.1 and 0.8 ± 0.9, respectively. Thus the members of the PDGFR subfamily identified on the vertebrate lineage are likely to have diverged at or just before the divergence of cyclostomes and gnathostomes.

Other subfamilies from the PTK family and those from the PTP family also show the same pattern of divergence.[26,28] The number and timing of intra-subfamily duplications obtained from different subfamilies belonging to the PTK and PTP are summarized in Table 2, together with those of three other protein families. Note that intra-subfamily duplication is not observed either before the cephalochordate-vertebrate split or after the fish-tetrapod split. Table 2 shows that the number of intra-subfamily duplications in the period before the cyclostome-gnathostome split is about 2.5 times greater than that in the period after that split, although the ratio varies greatly for different subfamilies due to the small sample size. It seems therefore likely that most intra-subfamily duplications occurred about the time of or just before the cyclostome–gnathostome split.

Figure 4 also provides information about a possible mechanism for generating diverse members of a particular subfamily in the early evolution of vertebrates. It is possible that intra-subfamily duplications have been derived in part from chromosomal duplications.[1,29–33] From a structural and phylogenetic analysis of the PDGFR subfamily, together with the chromosomal mappings of family members, Rousset et al.[30] suggested that the present-day PDGFR subfamily diverged by both gene duplications and chromosomal duplications. It is generally difficult to infer the chromosomal duplication events on ancient lineages, however, because of frequent translocation and deletion events during evolution. Figure 4 supports the argument by Rousset et al.[30] The PDGFR subfamily is composed of 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 two successive gene duplications (indicated by rhombi). Each group has two more intra-subfamily duplications, which gave rise to three different genes (in PDGFα/βR group one member might be deleted during evolution). Their chromosomal locations in the human

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

	m_b	m_a
a) PTK subfamily		
FGFR	2.6	1.3
Eph	6.5	2.3
src	7.2	1.7
PDGFR	3.9	0.8
b) 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
c) Other proteins		
Aldolase	1.3	1.2
Enolase	2.1	1.1
Complements	2.4	0.1
a) + b) + c)	30.4	12.0

m_b and m_a, the numbers of intra-subfamily duplications that occurred in vertebrate evolution before and after the cyclostome–gnathostome split, respectively. Data were taken from Ref. 26 for (a) and (c) and Ref. 28 for (b). Note that the intra-subfamily duplication is not observed either before the cephalochordate–vertebrate split or after the fish–tetrapod split[26,28]. FGFR, fibroblast growth factor receptor; PDGFR, platelet-derived growth factor receptor; complements, gene group encoding the complement components C3, C4, and C5.

genome differ within the same group, but are coincident between different groups. It is therefore likely that two chromosomal duplications are responsible for the two duplications in each group.

Evolutionary implications

The overall divergence patterns of animal-specific gene families involved in the signal transduction and developmental control can be summarized as follows. 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 (Fig. 1). In the early period before the parazoan–eumetazoan split, animals underwent extensive gene duplications (subfamily-generating duplications) and domain shufflings that gave rise to different subfamilies with diverse functions, and almost complete sets of present-day subfamilies had been established within this period. After the divergence of protostomes and deuterostomes, a further increase of family members occurred in vertebrate lineages: The multiplicity of members in the same subfamily rapidly increased in a limited period around the divergence of cyclostomes and gnathostomes by gene duplications, together with chromosomal duplications. Note that this type of duplication is identified neither before the cephalochordate–vertebrate split nor after the fish–tetrapod split. Different mechanisms might have operated in the two active periods. In the early period before the parazoan–eumetazoan split, shufflings of different functional domains might have played an important role for generating diverse subfamilies with distinct functions. In contrast, newly created genes in the latter period are exclusively types of genes that are virtually identical to each other in structure and function, but differ only in tissue distribution. Chromosome-wide duplications or even genome duplications might be a possible mechanism for generating diverse members of each subfamily in the latter period.[1,26,28–33] Different tissue distributions among members may be responsible for the preservation of divergence of functionally redundant genes originated at very ancient times approximately 500 million years ago.[34]

A remarkable consequence suggested by the scenario of gene diversification during animal evolution is that neither the burst of subfamily-generating duplication nor that of intra-subfamily duplication is directly linked with the Cambrian explosion. This suggests the importance of the origination of a small number of key genes that might have triggered the Cambrian explosion and that of gene recruitment of pre-existing genes, together with that of ecological and environmental changes as alternative mechanisms for generating diverse animals at the morphological level.

The *Hox* gene cluster may be a candidate for one of the key genes. As Finnerty and Martindale[35] noted, however, the nearly all protostomes possess remarkably similar *Hox* clusters consisting of three different subsets, 3′, central, and 5′ genes, despite large variation in their body plans. Thus, there is no simple relationship between the origination of *Hox* genes and the morphological diversification of triploblast animals. Rather, the variation of the body plans might be explained by different expression patterns of *Hox* genes.[36,37] It remains possible that, since cnidarians, diploblast animals, have no central genes in the *Hox* clusters,[35] the origination of the central genes in the triploblast animals trigerred the Cambrian explosion. It remains also possible that the conserved Hox cluster consisting of the 3′, central, and 5′ genes originated in the common ancestor of diploblasts and triploblasts, and the central genes were deleted in the ancestral lineage of diploblasts. The *Hox* cluster of *C. elegans*, a member of ecdysozoans, may be a typical example of the gene deletion, in which it now appears that four of ten genes have been lost.[38] Extensive surveys of the *Hox* clusters in the lower animals are needed.

Recent results showing that the divergence pattern of the Pax gene family encoding transcription factors is similar to those of gene families involved in the signal transduction[22] suggests the importance of gene recruitment as one of possible factors for explaining the Cambrian explosion.[17,18,22] By recruiting already existing genes for other purposes in different developmental stages, a variety of animals with diverse body plans might have been able to

evolve without the requirement of new genes with novel functions. The mechanism of gene recruitment was first identified in vertebrate genes encoding taxon-specific components of eye lens crystallins, which have recruited from various enzymes and stress proteins (see Ref. 39 for review). These proteins are multifunctional, functioning as lens crystallins in addition to their original functions as enzymes. Several other examples of the multifunctional proteins have already been reported[40–43] and gene recruitment is now believed to be a general mechanism for generating diverse functions of proteins.

The *Pax6* subfamily provides a typical example for gene recruitment. Different subfamilies of the *Pax* gene family are expressed in various restricted territories in the neural tube.[44,45] One of these subfamilies, *Pax6*, has been shown to be expressed repeatedly for different roles in different developmental stages and adult tissues.[44–47] Note that the *Pax6* subfamily comprises only one member. Furthermore the *Pax2/5/8* subfamily belonging to the *Pax* family exists in a sponge[22] that is thought to lack cell cohesiveness and coordination typical of eumetazoans.[4] In addition, a phylogenetic analysis suggests that most of subfamily-generating duplications antedate the divergence of parazoans and eumetazoans, which is at least 200 myrs earlier than the Cambrian explosion. Thus, there is no direct link between the divergence of *Pax* gene family and the Cambrian explosion.[22] These lines of evidence support the hypothesis that gene recruitment is a possible molecular mechanism involved in the Cambrian explosion. The cross-talk of molecules involved in signal transduction may also be a possible mechanism, by which complex networks of the signal transduction are generated without gene duplication.[48]

Gene recruitment and reduced evolutionary rate

Comparisons of evolutionary rates between different animals would provide insights into the repeated expressions for multiple purposes by gene recruitment. According to the neutral theory of molecular evolution, most evolutionary changes at the DNA sequence level are caused by random drift of selectively neutral mutants, with the rate of change dependent on mutation rate and the degree of functional constraint against sequence variations.[49] The degree of functional constraint depends largely on the numbers of such functionally important sites such as those that (1) are involved in the active centers of molecules, (2) interact directly with other molecules, and (3) are critical for maintaining protein tertiary structures.[49,50] In addition to the intrinsic functional constraints characteristic of individual molecules, there are others that derive from the interactions with surrounding molecules.[51] Due to those functional constraints, the evolutionary rate of a molecule can vary for different tissues or organs in which the molecule is expressed.[51–53] Different expression patterns of a gene at different stages of development, as a consequence of gene recruitment, might result in a considerable increase in the functional constraints against sequence variations on the encoded molecule.

We can symbolize the estimated rate of average sequence change by v and compare v at different periods. For example, v_{II} would refer to the rate of sequence change of a gene during period II (Fig. 1), the interval between the parazoan–eumetazoan split and the Cambrian explosion. When we make such comparisons, some significant differences in the rates of sequence evolution become apparent. Thus , the evolutionary rate v_{III} of the Pax-6 protein in the early period of vertebrate evolution (corresponding to period III in Fig. 1) significantly differs from that v_{IV} of the latter period (period IV) .[22] The evolutionary rate v_{III} is almost comparable to those of invertebrates, but in the lineages of higher vertebrates, the evolutionary rate v_{IV} is only about 1/5–1/2 of v_{III}. In contrast, no such difference in evolutionary rate between period III and IV is observed in house-keeping enzymes aldolases and triose phosphate isomerase. It is highly likely that it is increased functional constraint on the Pax6 protein, rather than reduced mutation rate that is responsible for the marked reduction of evolutionary rate in higher vertebrates. As mentioned above, Pax6 plays many different roles in different developmental stages and adult tissues in higher vertebrates.[42–47] Ultimately, the diversity of those roles and requirements reflecting a history of gene recruitments might be responsible for the strong reduction of molecular evolutionary rate of the Pax6 protein.[22]

Conclusion

A molecular phylogenetic study of multicellular animal-specific gene families involved in cell–cell communication and developmental control should help to provide insights into understanding the possible relationship between gene diversification and the Cambrian explosion. Each of these animal-specific gene families has diverged specifically in the animal kingdom from one or a few ancestral genes. Phylogenetic trees of gene families, each of which includes genes from a variety of animals from sponges to mammals, provide a rough estimate for the time of divergence of family members by gene duplications. Gene amplification proceded intermittently, not gradually: The pattern of gene diversification is characterized by two active periods in gene duplication interrupted by a long period of silence. In the early period (period I in Fig. 1) before the parazoan–eumetazoan split, extensive subfamily-generating duplications occurred, and almost complete sets of present-day subfamilies with diverse domain organizations and functions were established within the period. In the early evolution of vertebrates around the divergence of cyclostomes and gnathostomes, the multiplicity of members in the same subfamily explosively increased by intra-subfamily duplications. Different gene families show almost the same pattern

and timing of gene diversification, suggesting chromosome-wide duplications as a possible mechanism. It remains possible that the subfamily-generating duplications, together with domain shufflings, are related to the evolution of multicellularity. Interesting evidence on the origin and divergence of genes involved in cell–cell communication may be obtained from the phylogenetic trees of gene families including members from choanoflagellates, a unicellular protist thought to be the closest relatives to animals.[23,24]

A remarkable consequence suggested by the punctuated pattern of divergence is that the Cambrian explosion occupies a place in the relatively quiet period in gene diversification, and therefore might have no direct relationship with the burst of gene duplications and domain shufflings. These findings suggest the importance of the origination of a small number of key genes, and of gene recruitment of pre-existing genes, together with ecological and environmental changes as possible triggers of the Cambrian explosion. The importance of gene recruitment is emphasized as a possible molecular mechanism for generating diverse animals at the morphological level. It is particularly interesting that diversification of basic (subfamily) genes involved in cell–cell communication and developmental control by gene duplications and domain shufflings was completed in the very early evolution of animals before the divergence of parazoans and eumetazoans. During the remaining eumetazoan evolution of about 900 myrs, new subfamilies have seldom been created. In addition, evidence suggests that some of the genes, which are recognized in such primitive animals as sponges, have been deleted during specialization of animal groups.

Acknowledgments

We thank Drs. N. Iwabe, K. Kuma and D. Hoshiyama for valuable suggestions and discussion.

References

1. Ohno S. Evolution by gene duplication. New York: Springer Verlag. 1970.
2. Iwabe N, Kuma K, Miyata T. Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. Mol Biol Evol 1996;13:483–493.
3. Conway Morris S. The fossil record and the early evolution of the Metazoa. Nature 1993;361:219–225.
4. Margulis L, Schwartz KV. Five kingdoms, 3rd edition. New York: W. H. Freeman and Company. 1998.
5. Nikoh N, Iwabe N, Kuma K, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Shicui Z, Hori K, Shimura Y, Miyata T. 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 1997;45:97–106.
6. Dayhoff MO. Survey of new data and computer methods of analysis. In: Dayhoff MO, editor. Atlas of protein sequence and structure Vol 5, Suppl 3.Washington DC: National Biochemical Research Foundation. 1978. p 1–8.
7. Doolittle RF, Feng DF, Tsang S, Cho G, Little E. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 1996;271:470–477.
8. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 1988;241:42–52.
9. Hanks SK, Hunter T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 1995;9:576–596.
10. Hunter T. Protein kinase classification. Methods Enzymol 1991;200:3–37.
11. Mewes HW, Albermann K, Bahr M, Frishman D, Gleissner A, Hani J, Heumann K, Kleine K, Maierl A, Oliver SG, Pfeiffer F, Zollner A. Overview of the yeast genome. Nature 1997;387(6632 Suppl):7–65.
12. van der Geer P, Hunter T, Lindberg RA. Receptor protein-tyrosine kinases and their signal transduction pathways. Annu Rev Cell Biol 1994;10:251–337.
13. Fantl WJ, Johnson DE, Williams LT. Signalling by receptor tyrosine kinases. Annu Rev Biochem 1993;62:453–481.
14. Suga H, Katoh K, Miyata T. Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan–eumetazoan split. Gene in press.
15. Katoh K, Kuma K, Miyata T. Genetic algorithm-based maximum likelihood analysis for inferring molecular phylogeny. J Mol Evol in press.
16. Suga H, Kuma K, Iwabe N, Nikoh N, Ono K, Koyanagi M, Hoshiyama D, Miyata T. Intermittent divergence of the protein tyrosine kinase family during animal evolution. FEBS Lett 1997;412:540–546.
17. Suga H, Koyanagi M, Hoshiyama D, Ono K, Iwabe N, Kuma K, Miyata T. 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 1999;48:646–653.
18. Ono K, Suga H, Iwabe N, Kuma K, Miyata T. Multiple protein tyrosine phosphatases in sponges and extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split. J Mol Evol 1999;48:654–662.
19. Koyanagi M, Suga H, Hoshiyama D, Ono K, Iwabe N, Kuma K, Miyata T. Ancient gene duplication and domain shuffling in animal cyclic nucleotide phosphodiesterase family. FEBS Lett 1998;436:323–328.
20. Koyanagi M, Ono K, Suga H, Iwabe N, Miyata T. Phospholipase C cDNAs from sponge and hydra: Antiquity of genes involved in the inositol phospholipid signaling pathway. FEBS Lett 1998;439:66–70.
21. Suga H, Ono K, Miyata T. Multiple TGF-β receptor related genes in sponge and ancient gene duplications before the parazoan–eumetazoan split. FEBS Lett 1999;453:346–350.
22. Hoshiyama D, Suga H, Iwabe N, Koyanagi M, Nikoh N, Kuma K, Matsuda F, Honjo T, Miyata T. Sponge *Pax* cDNA related to *Pax*-2/5/8 and ancient gene duplications in *Pax* family. J Mol Evol 1998;47:640–648.
23. Wainright PO, Hinkle G, Sogin ML, Stickel SK. Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 1993;260:340–342.
24. Ragan MA, Goggin CL, Cawthorn RJ, Cerenius L, Jameson AVC, Plourde SM, Rand TG, Soderhall K, Gutel RR. A novel clade of protistan parasites near the animal-fungal divergence. Proc Natl Acad Sci USA 1996;93:11907–11912.
25. Miyata T, Kuma K, Iwabe N, Nikoh N. A possible link between molecular evolution and tissue evolution demonstrated by tissue specific genes. Jpn J Genet 1994;69:473–480.
26. Suga H, Hoshiyama D, Kuraku S, Katoh K, Kubokawa K, Miyata T. Protein tyrosine kinase cDNAs from amphioxus, hagfish and lamprey: Isoform duplications around the divergence of cyclostomes and gnathostomes. J Mol Evol 1999;49:601–608.
27. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 1994;39:306–314.
28. Ono-Koyanagi K, Suga H, Katoh K, Miyata T. Protein tyrosine phosphatases from amphioxus, hagfish, and ray: divergence of tissue-specific isoform genes in the early evolution of vertebrates. J Mol Evol 2000;50:302–311.
29. Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 1993;16:1–19.
30. Rousset D, Agnès F, Lachaume P, André C, Galibert F. Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulin like domains. J Mol Evol 1995;41:421–429.
31. Bailey WJ, Kim J, Wagner GP, Ruddle FH. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol Biol Evol 1997;14:843–853.

32. Pébusque M-J, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: Phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 1998;15:1145–1159.
33. 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. Zebrafish *hox* clusters and vertebrate genome evolution. Science 1998;282:1711–1714.
34. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 1999;151:1531–1545.
35. Finnerty JR, Martindale MQ. The evolution of the Hox cluster: insight from outgroups. Curr Opin Genet Dev 1998;8:681–687.
36. Averof M, Akam M. *Hox* genes and the diversification of insect and crustacean body plans. Nature 1995;376:420–423.
37. Grenier JK, Garber TL, Warren R, Whitington PM, Carroll S. Evolution of the entire arthropod *Hox* gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr Biol 1997;7:547–553.
38. Van Auken K, Weaver DC, Edgar LG, Wood WB. *Caenorhabditis elegans* embryonic axial patterning requires two recently discovered posterior-group Hox genes. Proc Nat Acad Sci USA 2000;97:4499–4503.
39. Wistow G. Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem Sci 1993;18:301–306.
40. Yang F, Demma M, Warren V, Dharmawardhane S, Condeelis J. Identification of an actin-binding protein from *Dictyostelium* as elongation factor 1. Nature 1990;347:494–496.
41. Numata O, Takemasa T, Takagi I, Hirono M, Hirono H, Chiba J, Watanabe Y. *Tetrahymena* 14-nm filament-forming protein has citrate synthase activity. Biochem Biophys Res Commun 1991;174:1028–1034.
42. Kurasawa Y, Numata O, Katoh M, Hirano H, Chiba J, Watanabe Y. Identification of *Tetrahymena* 14-nm filament-associated protein as elongation factor 1. Exp Cell Res 1992;203:251–258.
43. Singh R, Green MR. Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. Science 1993;259:265–268.
44. Strachan T, Read AP. PAX genes. Curr Opin Genet Dev 1994;4:427–438.
45. Mansouri A, Hallonet M, Gruss P. *Pax* genes and their roles in cell differentiation and development. Curr Opin Cell Bio 1996;8:851–857.
46. Callaerts P, Halder G, Gehring WJ. *Pax-6* in development and evolution. Annu Rev Neurosci 1997;20:483–532.
47. St-Onge L, Sora-Pineda B, Chowdhury K, Mansouri A, Gruss P. *Pax6* is required for differentiation of glucagon-producing a-cell in mouse pancreas. Nature 1997;387:406–409.
48. Birnbaumer L, Abramowitz J, Brown AM. Receptor-effector coupling by G proteins. Biochim Biophys Acta 1990;1031:163–224.
49. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University press. 1983.
50. Dickerson RE. The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1971;1:26–45.
51. Kuma K, Iwabe N, Miyata T. Functional constraints against variations on molecules from tissue level: Slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. Mol Biol Evol 1995;12:123–130.
52. Hastings KEM. Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. J Mol Evol 1996;42:631–640.
53. Duret L, Mouchiroud D. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. Mol Biol Evol 2000;17:68–74.