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.

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.

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 (intrasubfamily 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 preexisting 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...
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, respectively; CE, the Cambrian explosion.

**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, 940 Ma, 600–700 Ma, and 400 Ma, respectively. SD and ID, extensive gene duplications that gave rise to subfamilies and members of a particular subfamily, respectively; CE, the Cambrian explosion.

**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. 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, 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, the number of subfamily-generating duplications found in a period (first period) of 130 million years (myrs) between the animal–fungus 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 (± 20.8/1.8), 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
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 \(\alpha\) subunit (G\(\alpha\)) 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-\(\beta\) receptor (T\(\beta\)R) 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...
compared with those of human and Drosophila eumetazoans. Organizations of subfamily members from sponges and other lower eumetazoans indicate 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 the PTK sequences from human, Drosophila and sponge for each of the subfamilies, except for the src 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. 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. 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. 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.

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

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

\( n_b \) and \( n_a \), the numbers of subfamily-generating duplications per 100 myrs 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 \((> 10^{-9} \text{site/year})\) 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.

**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. 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. 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.
Figure 3. Comparison of domain organization of sponge PTK with those of human and Drosophila PTKs in each of ten different subfamilies.
(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, the number of intra-subfamily duplications that took place before and after the divergence of cyclostomes and gnathostomes were estimated to be, on average, $3.9 \pm 1.1$ and $0.8 \pm 0.9$, 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. 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. From a structural and phylogenetic analysis of the PDGFR subfamily, together with the chromosomal mappings of family members, Rousset et al. 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. The PDGFR subfamily is composed of three separate groups, PDGF$/$$b$R group, CSF-1R group (Flt3/Flk2, CSF-1R and c-kit) and VEGFR group (VEGFR, FLT4 and Flk1/KDR), which were generated by two successive gene duplications (indicated by rhombi). Each group has two more intra-subfamily duplications, which gave rise to three different genes (in PDGF$/$$b$R 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

<table>
<thead>
<tr>
<th>Subfamily</th>
<th>mb</th>
<th>ma</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) PTK subfamily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR</td>
<td>2.6</td>
<td>1.3</td>
</tr>
<tr>
<td>Eph</td>
<td>6.5</td>
<td>2.3</td>
</tr>
<tr>
<td>src</td>
<td>7.2</td>
<td>1.7</td>
</tr>
<tr>
<td>PDGFR</td>
<td>3.9</td>
<td>0.8</td>
</tr>
<tr>
<td>b) PTP family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTPR4</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PTPN3</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>PTPR5</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>PTPN6</td>
<td>1.7</td>
<td>0.7</td>
</tr>
<tr>
<td>PTPR2A</td>
<td>0.5</td>
<td>1.6</td>
</tr>
<tr>
<td>c) Other proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldolase</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Enolase</td>
<td>2.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Complements</td>
<td>2.4</td>
<td>0.1</td>
</tr>
<tr>
<td>a) + b) + c)</td>
<td>30.4</td>
<td>12.0</td>
</tr>
</tbody>
</table>

mb and ma, 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. 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. Instead, 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, 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. Other, 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 triggered 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.

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.\(^{34}\)
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. (46) 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 proceeded 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


