Evolution and Functional Diversification of the Paired Box (*Pax*) DNA-Binding Domains

Kristi A. Balczarek,† Zhi-Chun Lai,*† and Sudhir Kumar**

*Institute of Molecular Evolutionary Genetics, Department of Biology, and
†Department of Biochemistry and Molecular Biology, The Pennsylvania State University

The *Pax* gene family consists of tissue-specific transcriptional regulators that always contain a highly conserved DNA-binding domain with six α-helices (paired domain), and, in many cases, a complete or residual homeodomain. Numerous genes of this family have been identified in animals, with the largest number found in vertebrates. Our evolutionary analyses indicate that the vertebrate *Pax* gene family consists of four well-defined and statistically supported groups: group I (*Pax-1, 9*), II (*Pax-2, 5, 8*), III (*Pax-3, 7*), and IV (*Pax-4, 6*). Group I paired domains share a most recent common ancestor with *Drosophila Pox meso,* group II with *Pox neuro,* group III with *paired* and *gooseberry,* and group IV with the *eyeless* gene. Two groups containing complete homeodomains (III and IV) are distantly related, and the intergroup relationships are ((I,III),(II,IV)). These four major groups arose before the divergence of *Drosophila* and vertebrates prior to the Cambrian radiation of triploblastic metazoan body plans. We conducted an analysis of fixed radical amino acid differences between groups in a phylogenetic context. We found that all four fixed radical amino acid differences between groups I and III are located exclusively in the N-terminal α-helices. Similarly, groups II and IV show three fixed radical differences in these α-helices but at positions different from those in groups I and III. Implications of such fixed amino acid differences in potentially generating sequence recognition specificities are discussed in the context of some recent experimental findings.

Introduction

The *Pax* genes constitute a family of transcription factors that play important roles during development as tissue-specific transcriptional regulators. The defining characteristic of this family is the presence of a 128-amino acid long paired domain, which encodes a unique DNA-binding motif (Chalepakis et al. 1991; Treisman, Harris and Desplan 1991). In addition, many of these genes contain a complete or truncated DNA-binding homeodomain (Maulbecker and Gruss 1993). The paired box was first identified in *Drosophila* in three segmentation and two tissue-specific genes, and homologs have been found in a variety of vertebrates and invertebrates (table 1).

Studies of the different *Pax* gene expression patterns in mouse and other model organisms have suggested their involvement in a variety of different developmental functions, especially in the development of the central nervous system. *Pax-2* and *Pax-8* are expressed during the development of the excretory system, including the kidney (Dressler et al. 1990; Plachov et al. 1990), and *Pax-7* contributes to muscular system specification (Jostes, Walther, and Gruss 1991). In mouse and human, *Pax-5* plays a critical role in the development of the immune system, particularly the B-lymphocytes (Adams et al. 1992). Consequently, mutations in the highly conserved regions of the paired domain have been associated with a number of disease phenotypes, e.g., vertebral column malformation (*Pax-1,* Balling, Deutsch, and Gruss 1988) and *Splotch* phenotype in mouse (*Pax-3,* Epstein, Vekemans and Gross 1991), and Waardenburg's syndrome in humans (*Pax-3,* Baldwin et al. 1992; Tassabehji et al. 1992). Aniridia in humans is attributed to mutations in *Pax-6* (Glaser, Walton, and Maas 1992), as is the *Small eye* phenotype in mice (Hill et al. 1991). In addition, oncogenic potentials of *Pax* genes are now well known (Eccles et al. 1992; Maulbecker and Gruss 1993; Poleev et al. 1992; Stapelton et al. 1993).

Of all the *Pax* genes, *Pax-6* has been studied most extensively because of its function in eye development in vertebrates (including human and mouse) and in the specification of sense organs in lower invertebrates (Zhang and Emmons 1995). The *Pax-6* homolog in *Drosophila* (*eyeless*) has been described as a master control gene in eye development. Targeted ectopic expression of the mouse *Pax-6* gene in *Drosophila* leads to ectopic fly eyes (Halder, Callaerts, and Gehring 1995). As the genes responsible for eye morphogenesis in vertebrates and invertebrates are perceived as functional homologs, it is thought that *Pax-6* was responsible for eye morphogenesis in the common ancestor of vertebrates and *Drosophila* (Quiring et al. 1994).

The knowledge of evolutionary relationships among diverse paired domains is critical for understanding the tempo and mode of evolutionary processes responsible for generating the observed diversity of paired domains in animals. A number of authors have attempted to infer the relationships of paired domains by comparing their molecular sequences. For instance, Walther et al. (1991) proposed six major classes based on an analysis of mouse, human, and *Drosophila* paired domains (see also Stapelton et al. 1993), whereas Noll (1993) has suggested four major groups. In Walther et al.'s system, some classes consist of only vertebrate paired domains, whereas others consist exclusively of one or more *Drosophila* paired domains. If this is true, it is unclear when and why these individual families emerged during evolution (Duboule 1994). These classifications are based largely on qualitative amino acid

Key words: *Pax,* paired box, evolution, DNA-binding, gene duplication, eyeless.

Address for correspondence and reprints: Sudhir Kumar, 208 Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802. E-mail: sxk132@psu.edu.

Mol. Biol. Evol. 14(8):829–842. 1997

Table 1
***Pax* Genes Used and Their GenBank/EMBL Accession Numbers**

Species and Genes Used	Accession No.
Mammals	
Homo sapiens (*Hs*)	
Pax-1 (Burri et al. 1989)	P15863
Pax-2 (Eccles et al. 1992)	Q02962
Pax-3 (Burri et al. 1989)	P23760
Pax-5 (Adams et al. 1992)	M96944
Pax-6 (Ton et al. 1991)	M77844
Pax-7 (Schafer et al. 1994)	Z35141
Pax-8 (Kozmik et al. 1993)	L19606
Pax-9 (Stapleton et al. 1993)	S36115
Mus musculus (*Mm*)	
Pax-1 (Chalepakis et al. 1991)	M69222
Pax-2 (Dressler et al. 1990)	P32114
Pax-3 (Goulding et al. 1991)	X59358
Pax-4 (Walther et al. 1991)	P32115
Pax-5 (Adams et al. 1992)	Q02650
Pax-6 (Walther and Gruss 1991)	X63963
Pax-7 (Jostes, Walther, and Gruss 1991)	P47239
Pax-8 (Plachov et al. 1990)	Q00288
Pax-9 (Neubuser, Koseki, and Balling 1995)	P47242
Rattus norvegicus (*Rn*)	
Pax-6 (Matsuo et al. 1993) (homeodomain only)	S74393
Pax-8 Zannini (GenBank)	X94246
Canis familiaris (*Cf*)	
Pax-8 Van Renterghem et al. (GenBank)	X83592
Birds	
Gallus gallus (*Gg*)	
Pax-1 (Peters, Doll, and Niessing 1995)	X82442
Pax-6 (Li et al. 1994) (homeodomain only)	P47237
Pax-9 (Peters, Doll, and Niessing 1995)	X82443
Coturnix coturnix (*Cc*)	
Pax-6 (Martin et al. 1992)	X82151
Fish	
Brachydanio rerio (*Br*)	
Pax-2 (Krauss et al. 1991)	X61389
Pax-6 (Puschel, Gruss, and Westerfield 1992)	X63183
Amphioxus	
Branchiostoma floridae (*Bf*)	
Pax-1 (Holland, Holland, and Kozmik 1995)	U20167
Sea urchin	
Paracentrotus lividus (*Pl*)	
Pax-6 (Czerny and Busslinger 1995)	U14621
Ribbonworm	
Lineus sanguineus (*Ls*)	
Pax-6 (Loosli, Kmita-Cunisse, and Gehring 1996)	X95594
Arthropod	
Drosophila melanogaster (*Dm*)	
eye (Quiring et al. 1994)	X79492
prd (Frigerio et al. 1986)	P06601
gsb-p (Baumgartner et al. 1987)	P09083
gsb-d (Baumgartner et al. 1987)	P09082
meso (Bopp et al. 1989)	P23757
neuro (Dambly-Chaudiere et al. 1992)	P23758
Nematode	
Caenorhabditis elegans (*Ce*)	
Pax-6 (Chisholm and Horvitz 1995)	U31537

sequence comparisons in which the assumption that amino acid changes have occurred at a more or less constant rate is implicitly made. For this reason, it is difficult to evaluate the merit of these classifications without further analysis. In addition, the interrelationships of major groups are not yet defined clearly.

We have conducted rigorous phylogenetic analyses of all available paired domain sequences (table 1). Our analyses included sequences from a primitive vertebrate (amphioxus), *Caenorhaliditis elegans,* and other invertebrates that have not been considered in previous evolutionary analyses. The objectives of our study were to determine the closest invertebrate relatives of the vertebrate paired domains, predict the potential *Pax* gene contents in the genomes of unexplored animal lineages, and construct an evolutionary scenario for the generation of current diversity of vertebrate paired domains. The knowledge of these aspects of evolution of paired domains will facilitate drawing parallels between the functions of closely related homologs in vertebrates and invertebrates (e.g., Neubuser, Koseki, and Balling 1995).

The paired domain binds DNA at two adjacent major grooves (Czerny, Schnaffer, and Busslinger 1993; Xu et al. 1995). Because the N-terminal domains are more conserved among major groups than are the C-terminal domains, it is sometimes thought that the specificity of sequence recognition between groups may be due to amino acid substitutions in the C-terminal domain (Stuart, Kioussi, and Gruss 1993). We have evaluated this possibility by examining the differences between closely related *Pax* genes as well as between major groups of *Pax* genes.

Materials and Methods

Amino Acid Sequence Data

The published nucleotide sequences of paired box and homeodomains in the *Pax* genes were obtained from GenBank (table 1). Duplicate and partial paired domain sequences were removed from the analysis. The amino acid sequences of the paired domain and the homeodomain were aligned separately, and the nucleotide sequences were adjusted to reflect the amino acid alignments. A representative amino acid sequence alignment is in figure 1.

Evolutionary Analysis

The amino acid and nucleotide sequences of the paired domain and homeodomain were used for the evolutionary analysis. The neighbor-joining (Saitou and Nei 1987) and the maximum parsimony (Swofford 1993) methods were employed for inferring phylogenetic relationships. For estimating the evolutionary distance between two sequences (number of amino acid substitutions per site), the Poisson correction distance was used (Kumar, Tamura, and Nei 1993). Table 2 shows the estimates of Poisson correction distance for the mouse *Pax* genes. For nucleotide sequences, only the 1st and 2nd codon position data were used (3rd codon positions were excluded because the *Pax* genes are distantly related). The Tamura–Nei (1993) model was used to estimate

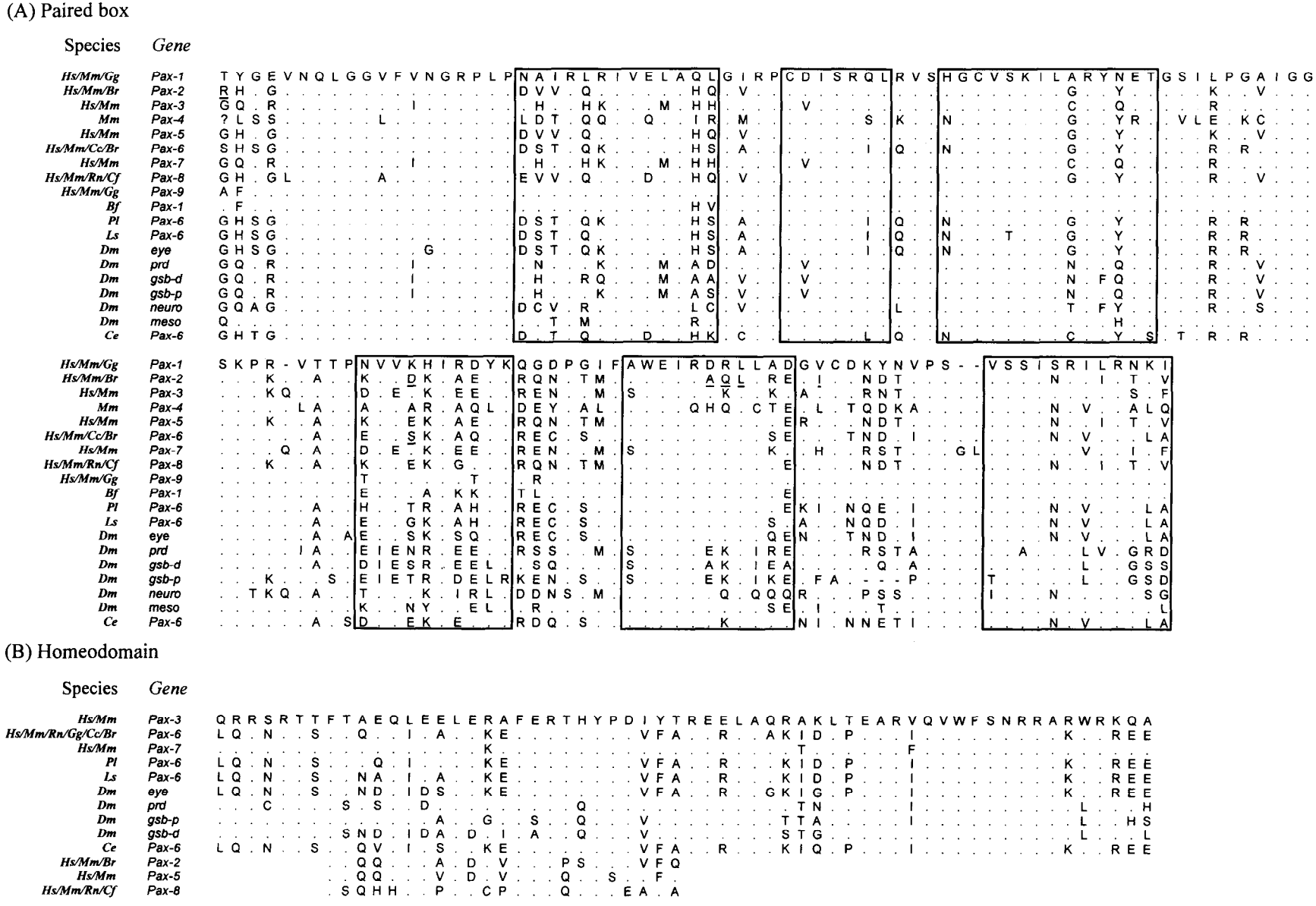

FIG. 1.—Condensed alignments of paired box and homeodomain amino acid sequences (species abbreviations from table 1). Residues identical to the first sequence are indicated by a dot (.), missing data by a question mark (?), and alignment gaps with a dash (–). For paired box domains, one member of each vertebrate homolog is shown (*Mus musculus*). Amino acid residues are underlined whenever amino acid differences are found in different species. In *Pax-6,* the difference occurs at site 83, where the zebrafish sequence has Gly, all other vertebrates have Ser. In *Pax-2,* there are six sites that show differences. At three of these sites, the mouse sequence differs: Ala at site 102, Gln at site 103, and Arg at site 106. At two other sites, the zebrafish sequence differs from the consensus: Glu at site 83 and Val at site 109. At site 1, the human sequence has Gly where the others have Arg. The α-helices are enclosed in rectangular boxes (Xu et al. 1995). The N-terminal domain is from site 1 to 72 and the C-terminal domain is from 78 to 128. In *Pax-2, 5,* and *8* a residual homeodomain is found.

Table 2
The Number of Amino Acid Substitutions per 100 Sites in *Mouse Pax* genes

Gene	*Pax-1*	*Pax-9*	*Pax-3*	*Pax-7*	*Pax-2*	*Pax-5*	*Pax-8*	*Pax-4*
Pax-9.....	4.7							
Pax-3.....	30.3	30.3						
Pax-7.....	30.3	30.3	6.3					
Pax-2.....	36.8	36.8	35.7	35.7				
Pax-5.....	33.5	33.5	32.4	33.5	5.5			
Pax-8.....	34.6	35.7	35.7	35.7	10.5	6.4		
Pax-4.....	55.4	55.4	64.0	58.2	46.2	48.8	50.0	
Pax-6.....	37.9	37.9	43.7	39.0	31.4	29.3	32.4	35.6

NOTE.—Genes have been ordered following the phylogeny in figure 2.

evolutionary distances between nucleotide sequences because it can account for differences in the transitional and transversional rates and the inequality of nucleotide frequencies. However, evolutionary analyses conducted using Kimura's 2-parameter model (Kimura 1980) produced essentially the same results. In these analyses, all sites containing alignment gaps or missing information were excluded (complete-deletion option) and the reliabilities of neighbor-joining trees were determined using the bootstrap test in the MEGA program (1,000 replications: Felsenstein 1985; Kumar, Tamura, and Nei 1993). The outcomes of the bootstrap test are the *P*-values for the interior branches in the tree. For a given branch, the bootstrap *P*-value (BP) is the percentage of bootstrap replicates in which the branch is reconstructed.

For parsimony analysis, a branch-and-bound search was done using PAUP. Unweighted amino acid parsimony was used, and only unique sequences were included in the analysis for computational efficiency.

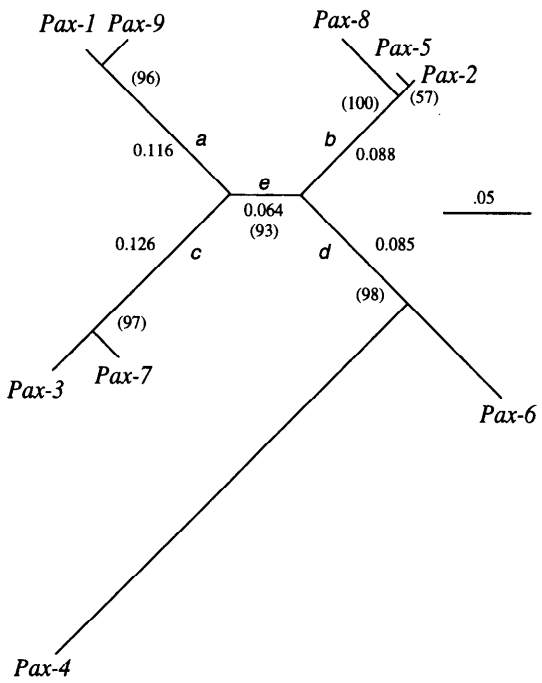

FIG. 2.—Evolutionary relationships of the vertebrate paired box domains. The scale bar has the unit of the number of amino acid substitutions per site. Branch lengths are drawn to scale and bootstrap *P*-values (BP) are shown in parentheses, determined using 128 amino acid sites from the paired domain.

The amino acid sequences of *Pax* genes were compared in an evolutionary context to study radical and conservative differences in the α-helical and intervening regions in N- and C-terminal regions of the paired domains. We compared fixed amino acid differences between closely related *Pax* genes (e.g., *Pax-1* and *Pax-9* in fig. 3). In these comparisons, a site is said to be fixed if every homologous sequence has the same residue. For instance, in the comparison of *Pax-1* and *9* sequences, site 1 is fixed in *Pax-1* (Thr) as well as in *Pax-9* (Ala) (see fig. 6). For comparisons between groups, a site is said to be fixed if all the *Pax* genes within a group have the same amino acid residue (e.g., site 4 in group I; fig. 6). For the analysis of radical and conservative differences, three properties of the amino acid residue's side chain were considered: nonpolar, uncharged polar, and charged polar (Voet and Voet 1990). For instance, site 1 shows a fixed difference between *Pax-1* and *9*, which is a radical difference in polarity. The size of the side chain of each amino acid was also considered using two criteria: the length of the side chain (considered radical if their lengths differed by two or more links) and the presence of a linear versus a ring structure.

Results

Evolutionary Relationships of Vertebrate *Pax* Genes

Figure 2 shows a schematic of the evolutionary relationships of the vertebrate *Pax* genes as inferred from a distance-based phylogenetic analysis of the paired domain amino acid sequences. Identical evolutionary relationships among vertebrate *Pax* homologs were obtained in the analyses of 1st + 2nd codon position data as well as in the maximum parsimony analyses. Thus, the vertebrate *Pax* gene family can be arranged into four groups, based on figures 2 and 3: I (*Pax-1, 9*), II (*Pax-2, 5, 8*), III (*Pax-3, 7*), and IV (*Pax-4, 6*). In this phylogeny, *Pax* genes without a homeodomain comprise one group (group I) and those with a residual homeodomain fall into a single group (group II). By contrast, *Pax* genes containing full homeodomains are divided into two distantly related groups (III and IV). The bootstrap test shows high statistical support for the inclusion of respective *Pax* genes within groups (BP > 95%, fig. 2). Somewhat different groupings were suggested by other authors without statistical analysis (Walther et al. 1991; Noll 1993). An analysis of homeodomain amino acid sequences of the full and residual homeodomain-con-

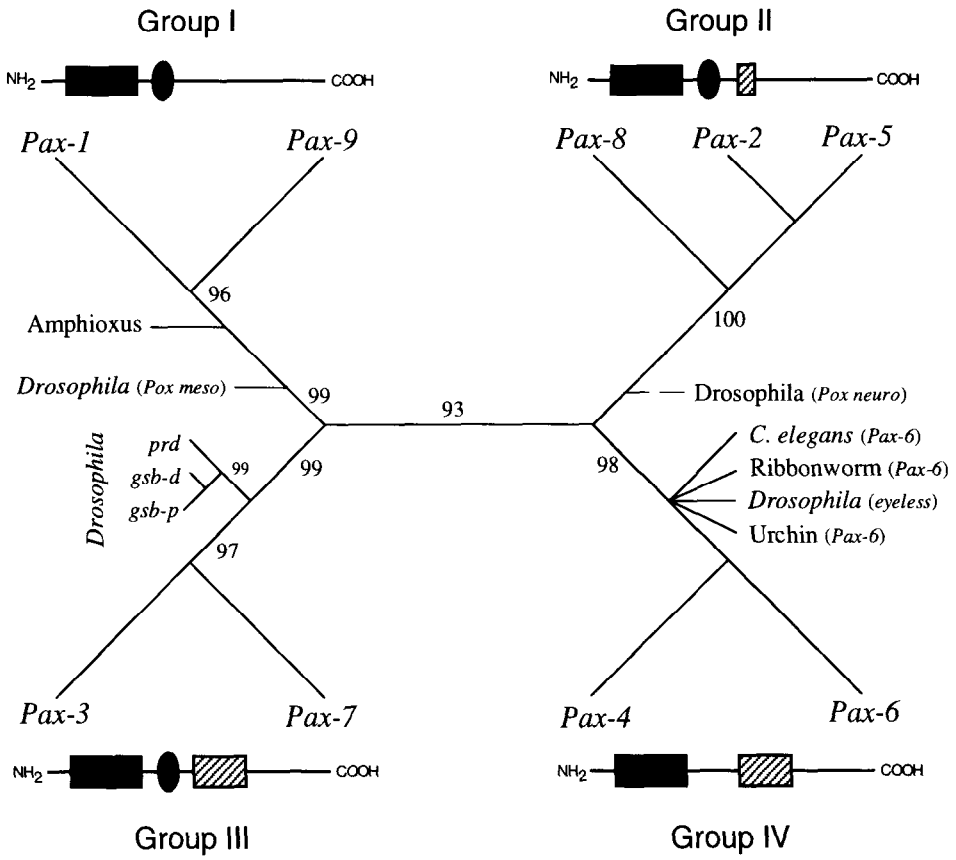

FIG. 3.—A schematic showing the phylogenetic relationship of all available paired domain sequences. Based on amino acid (aa) sequence analysis (128-aa paired domain, 23-aa homeodomain) for among-group relationships and nucleotide sequence analysis for within-groups relationships. Bootstrap *P*-values >90% are shown on the corresponding branches. Gene structure in terms of the paired domain (black rectangle), octapeptide (black ellipse), and homeodomain (hatched rectangle) are shown. The dotted line indicates the uncertainty in the phylogenetic placement. The tree is not drawn to scale.

taining *Pax* genes provides additional support for these group constitutions.

Groups I, III, and IV consist of two *Pax* genes each, and group II contains three genes (*Pax-2 ,5,* and *8*). In the analysis of the paired domain sequences, *Pax-2* and *Pax-5* were closer to each other than either is to *Pax-8* or any other *Pax* gene (figs. 2 and 3). However, this inference was weakly supported in the bootstrap tests (BP = 57%; fig. 2). Analysis of the partial homeodomain sequences also groups *Pax-2* and *5* together to the exclusion of *Pax-8.*

Cephalochordates are the closest relatives of vertebrates, and one member of the *Pax* gene family has been reported from amphioxus (*BfPax-1*; Holland, Holland, and Kozmik 1995). Our analysis indicated that this gene is a member of group I (BP > 95%) and that the vertebrate *Pax-1* and *Pax-9* are closer to each other than either is to *BfPax-1* (BP > 95%; fig. 3).

Evolutionary Relatives of Vertebrate *Pax* Genes

The evolutionary relationships of vertebrate and invertebrate paired domains were inferred by using the amino acid as well as the nucleotide sequences (fig. 3). Phylogenetic analyses suggest that each of the vertebrate *Pax* groups contains at least one *Drosophila* paired box gene, and that these *Drosophila* genes are distantly related to all other chordate *Pax* homologs in their respective groups (fig. 3). With the exception of group II, the evolutionary relationships of *Drosophila* and vertebrate *Pax* genes were strongly supported in the bootstrap tests (BP > 95%) as well as in the analysis of 1st + 2nd codon position data. These inferences were also congruent with the results of the phylogenetic analyses of the common homeodomain sequences. These results show that the paired box containing genes in all triploblastic animals fall into the following four major groups.

Group I consists of vertebrate *Pax-1* and *9,* amphioxus *Pax-1,* and *Drosophila Pox meso,* which do not contain the homeodomain. This composition is statistically supported at 99% bootstrap probability. Both *Pax-1* and *Pax-9* are known to play a role in the development of the vertebral column and function in the development of endoderm and mesoderm. In contrast, *Pox meso* is only active in mesodermal structures (e.g., Holland, Holland, and Kozmik 1995). In this group, the vertebrate *Pax-1* and *9* are more closely related to each other than either is to amphioxus *Pax-1.* Thus, the duplication event that resulted in *Pax-1* and *9* appears to have occurred after the amphioxus–vertebrate splitting, and the amphioxus genome is likely to contain fewer *Pax* genes than those in vertebrates.

Group II includes vertebrate *Pax-2, 5,* and *8,* and *Drosophila Pox neuro* genes. Vertebrate genes of this group contain a residual homeodomain; however, no homeodomain has been found in *Drosophila Pox neuro.* Members of this group are known to function in the excretory systems (kidney) and in the liver at the onset of B lymphopoiesis. As mentioned above, the phylogenetic affinity of *Drosophila Pox neuro* is not very strong. In amino acid sequence analyses, *Pox neuro* showed weak association to vertebrate *Pax* genes of this group, but it appeared to be more closely related to *Drosophila* segmentation genes (*prd* and *gsb*) in the nucleotide sequence (1st + 2nd codon positions) analyses, which appears to be due to similarity of nucleotide frequencies of these genes. The complete absence of a homeodomain in *Pox neuro* does not support its inclusion in either group II or III. The paired domain of *Pox neuro* was previously thought to be a homolog of vertebrate *Pax-2* (Bopp et al. 1989). Based on the results of the amino acid sequence analysis, we have placed *Pox neuro* tentatively in group II (BP = 55%; fig. 3).

Group III contains two vertebrate *Pax* genes (*3* and *7*) and three *Drosophila* paired box genes (*prd, gsb-d,* and *gsb-p*), all of which contain a complete homeodomain. The three *Drosophila* genes form a monophyletic group. They have been generated by gene duplication in a short time after the protostome–deuterostome split (see fig. 5). These three *Drosophila* genes are pair-rule genes involved in segment specification, and functional studies have suggested that these three genes are able to substitute for one another (Li and Noll 1994). Our phylogenetic placement of these *Drosophila* genes is in disagreement with that by Walther et al. (1991) who assigned these genes in a separate class on the basis of overall amino acid similarity in the paired domains sequences. The use of qualitative strategy in evolutionary analysis is problematic if all *Pax* genes have not evolved with the same evolutionary rate. This is indeed the case for the *prd, gsb-p,* and *gsb-d* genes that have evolved at approximately twice the rate of other genes of this group at the amino acid sequence level. This is indicated by the length of the branch from the common ancestor to these *Drosophila* genes, which was twice the length of the branch leading to the vertebrate homologs in group III (results not shown).

Group IV consists of vertebrate *Pax-4* and *6* and *Drosophila eyeless,* all of which are missing the octapeptide and contain a full homeodomain. With the exception of group I, *Pax-6* is the only gene for which potential orthologous sequences have been reported from a variety of deuterostome and protostome animals (e.g., *C. elegans,* ribbonworm, and urchin). Results of our phylogenetic analysis of amino acid and nucleotide sequences show that *C. elegans* formed the basal lineage in group IV and that its sister lineage consisted of all other *Pax-6* homologs and the mouse *Pax-4* (fig. 4A). The clustering of vertebrate *Pax-4* and *6, Drosophila eyeless,* and all invertebrate *Pax-6* genes is statistically supported.

Root of the *Pax* Gene Tree

The evolutionary relationships among the four *Pax* groups are drawn in the form of an unrooted tree in figure 3. This is because, at present, no *Pax* homologs have been reported from diploblasts, plants, fungi, or other distant relatives of animals. We conducted a BLAST search of the *Saccharomyces cerevisiae* genome using the paired domain as a probe and did not find any significant matches. Therefore, it is unclear if these genes are present outside the animal kingdom. *Pax* genes may be found in diploblasts, in which case, we need to know the number of genes as well as their sequences to establish reliably the root for this tree.

On the other hand, the close relatives of homeodomain sequences of the *Pax* genes are known (*Mix.1* and *odt*; Duboule 1994, p. 36). We utilized these outgroups to root the *Pax* gene homeodomain tree. We found that group II and IV homeodomains were closer to each other than either was to group III, i.e., ((*Mix.1, odt*) (III (II, IV))). Because the paired and homeodomains have evolved as a unit (see *Discussion* later) in *Pax* genes, this result suggests that the root is likely to be on the branch leading to group I, III, or in the middle (branch *a, c,* or *e,* respectively in fig. 2).

In some previous studies on the basis of paired domain sequence similarity, it was suggested that *Pax-4* should be put in a class of its own as an outgroup to all other *Pax* genes (e.g., Burri et al. 1989; Walther et al. 1991; Stapelton et al. 1993). This was done to accommodate the observation that the *Pax-4* sequences showed lowest amino acid sequence similarity to all other *Pax* genes. However, in our analysis of the 1st + 2nd codon positions of vertebrate *Pax-6,* mouse *Pax-4, Drosophila eyeless,* and *C. elegans Pax-6* sequences, with *Pax-2* as an outgroup, we find that mouse *Pax-4* shares a most recent common ancestor with vertebrate *Pax-6* genes (fig. 4A). This means that the gene duplication occurred after the divergence of protostomes and deuterostomes. This is in agreement with the amino acid sequence analysis using the maximum parsimony method, in which *Pax-4* and *Pax-6* genes of vertebrates cluster together to the exclusion of invertebrate *Pax-6* genes (fig. 4B). The grouping is also supported by the lack of an octapeptide in both *Pax-4* and *Pax-6,* while all other *Pax* genes contain the octapeptide.

It appears that the differences in the placement of *Pax-4* in our study and some previous classifications is due to the increased evolutionary rate in *Pax-4* (fig. 2 and Table 2), which may have misled previous authors who used amino acid similarity for organizing *Pax* genes into groups. *Pax-4* is peculiar in that the C content in the first codon position is about 50% higher than that in *Pax-6* because of a rather high frequency of the amino acid leucine. Inclusion of sequences that show large differences in amino acid or nucleotide base composition is likely to make it difficult for tree-making methods to recover phylogenetic affinities with high confidence. Additionally, the results from an analysis of homeodomain sequences of genes in figure 4 suggest that *Pax-4* and *6* share a more recent common ancestor with each

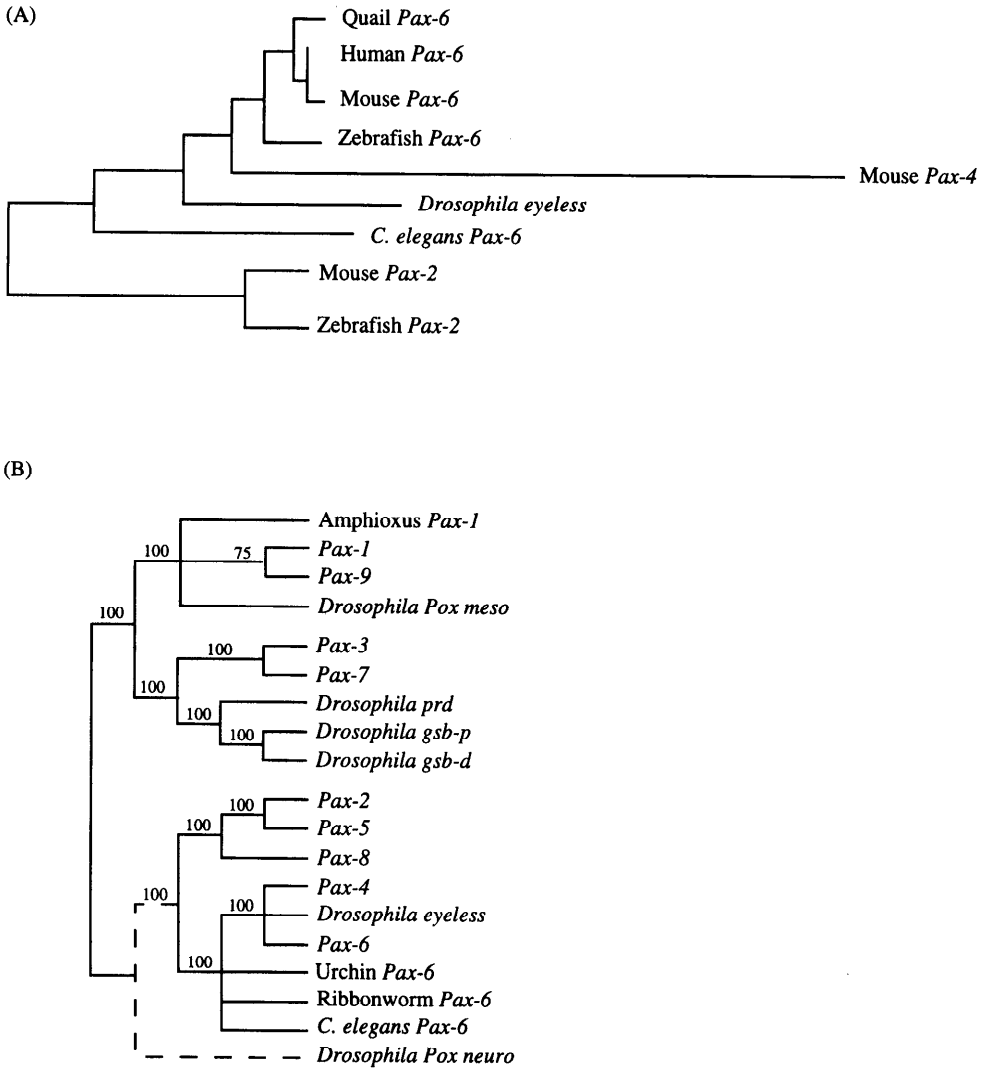

FIG. 4.—Evolutionary relationships of vertebrate and invertebrate *Pax-6* and mouse *Pax-4* sequences using (*A*) the neighbor-joining method and (*B*) the maximum parsimony method (50% majority rule consensus tree from 48 most parsimonious trees). The values above branches are the branch frequencies (in percent) in 48 MP trees. Both (*A*) and (*B*) were based on 128 amino acids in the paired domain. See text for details.

other than either does to any of the invertebrate *Pax-6* genes. Therefore, the root of the *Pax* gene tree does not lie on the lineage leading to *Pax-4*.

Evolutionary Rates of Change in *Pax* Genes

The sequence alignment in figure 1 clearly shows that the paired domains for *Pax-6* are identical in sequence between birds and mammals, which diverged about 310 million years ago (MYA) (Benton 1990, 1993; Hedges et al. 1996). This is true for other *Pax* genes as well (fig. 1). This indicates that vertebrate paired box-containing genes are evolving considerably slower than other developmentally important genes. For instance, the highly conserved N-terminal domain of the *Sonic hedgehog* gene has experienced, on average, 1.16 substitutions per 100 amino acids between birds and mammals (Kumar, Balczarek, and Lai 1996). By contrast, no amino acid differences are observed in the comparison of available sequences of bird and mammal *Pax* homologs. In fact, the evolutionary rates in *Pax* genes are on the order of those shown by some of the slowest-evolving histones (see Nei 1987, p. 50).

Even though *Pax* genes show very little change at the amino acid sequence level, the DNA sequences of orthologous *Pax* genes have frequently undergone changes that keep the amino acid residues unaltered (synonymous substitutions). For instance, in *Pax-6*, there are no amino acid substitutions between human and mouse, but the synonymous distance is 0.1505 substitutions per site (Nei and Gojobori 1986). Therefore, ever since their origin in the ancestors of all vertebrates, *Pax* domains have evolved under extremely strong purifying selection at the amino acid sequence level, where even conservative amino acid changes are forbidden (see also Halder, Callaerts, and Gehring 1995).

Radical Changes in Vertebrate Paired Domains

The importance of invariant amino acid sites (sites containing the same residue in homologous sequences) are routinely examined in the laboratory by studying the

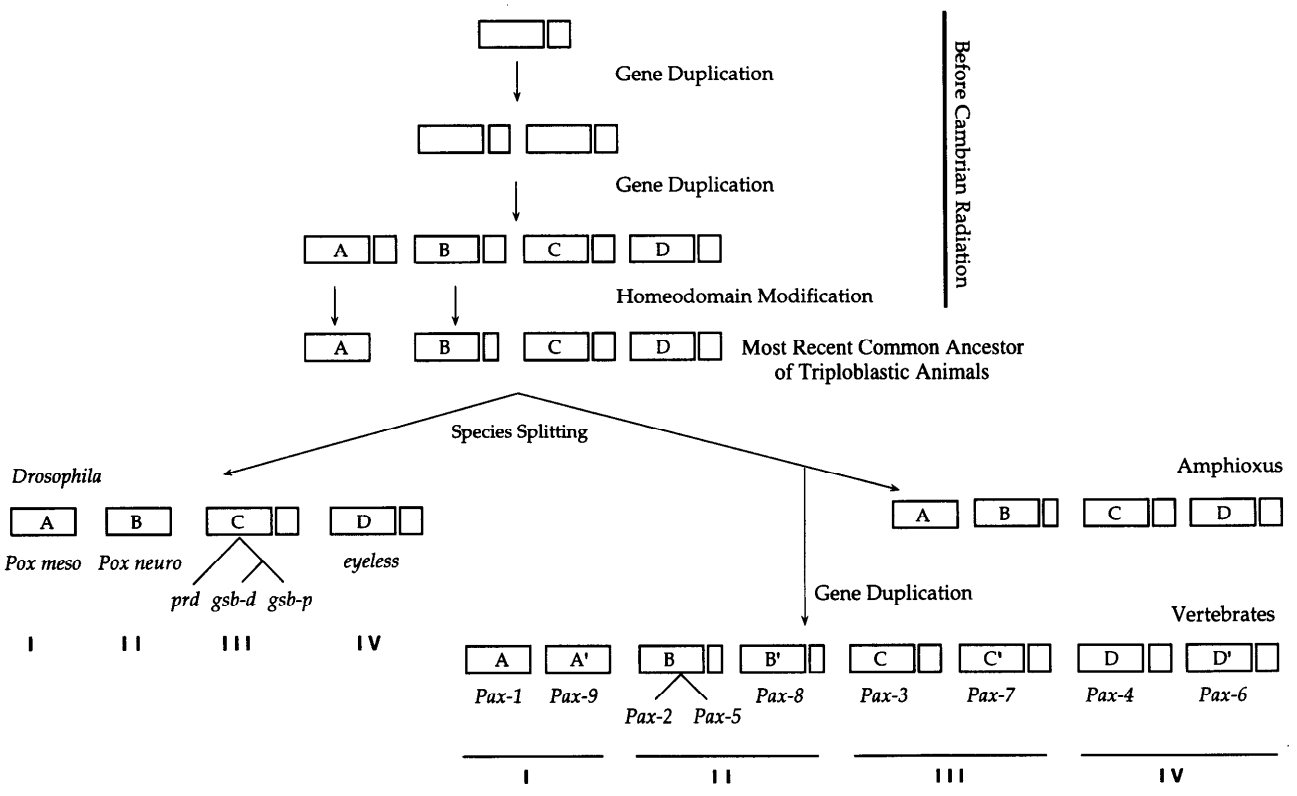

FIG. 5.—One possible evolutionary scenario for the generation of the *Pax* genes by gene duplications.

correspondence between primary sequence lesions and the resulting phenotypes. Similarly, the study of amino acid sites that have been conserved in orthologous sequences (e.g., *Pax-1* genes in different vertebrates) but diverged among paralogous sequences (e.g., *Pax-1* and *9*) may provide valuable insight into the primary precursors of functional diversification (e.g., Wilson, Carlson, and White 1977). This may be particularly true for DNA-binding domains in which the presence of an amino acid residue with a different property may change the DNA-binding affinity and/or specificity. However, the existence of primary amino acid sequence differences is not a prerequisite for evolution of a new function or modification of a previously existing function because, for instance, paralogous genes may have the same function but differ in temporal and tissue-specific expression patterns (e.g., *gsb-p* and *gsb-d* in *Drosophila*; Baumgartner et al. 1987).

We constructed an alignment of vertebrate *Pax* genes in which the sequences were arranged according to their phylogenetic relationships (fig. 6) and classified all fixed amino acid differences as radical or conservative changes (see *Materials and Methods*). All sites showing radical fixed amino acid differences within groups and between groups are shown in figures 6 and 7, and a summary of these changes is given in table 3. Only the paired domain sequences were studied because homeobox sequences are not found in group I, and are truncated in group II.

Analyses within Groups

Within group I and within group III, we find that the radical as well as the conservative fixed differences are found only in the intervening region between α-helices. In contrast, group II genes have undergone mainly conservative changes, and these changes are found in both the helices and the intervening regions. Comparisons between *Pax-4* and *6* (group IV) were not conducted because the *Pax-4* sequence is highly divergent. However, in general, *Pax-4* and *Pax-6* are quite similar at potentially important sites, such as the sites where residues are conserved in all groups except group IV (e.g., in the N-terminal [sites 42, 44, and 47]), so their functions may also be similar.

Analyses between Closely Related Groups

The comparison of 23 fixed differences between groups I and III shows an entirely different pattern from that found in within group comparisons. In this case, all of the radical fixed differences are found in the α-helices in the N-terminal domain (fig. 7). Three of them involve His residues and result in a hydrophilicity change (nonpolar in group I and polar in group III). In contrast to the radical differences, the conservative changes are uniformly distributed between α-helical and intervening regions and are found in both the N- and C-terminal domains. These observations suggest that the radical differences in the N-terminal domain may have led to different binding specificities and strengths of group I and III genes (see Xu et al. 1995).

There are 25 amino acid differences between groups II and IV, which are evenly distributed between the N- and C-terminal domains. Only three radical differences are found in the α-helices, all of which are in the N-terminal domain. However, these sites are differ-

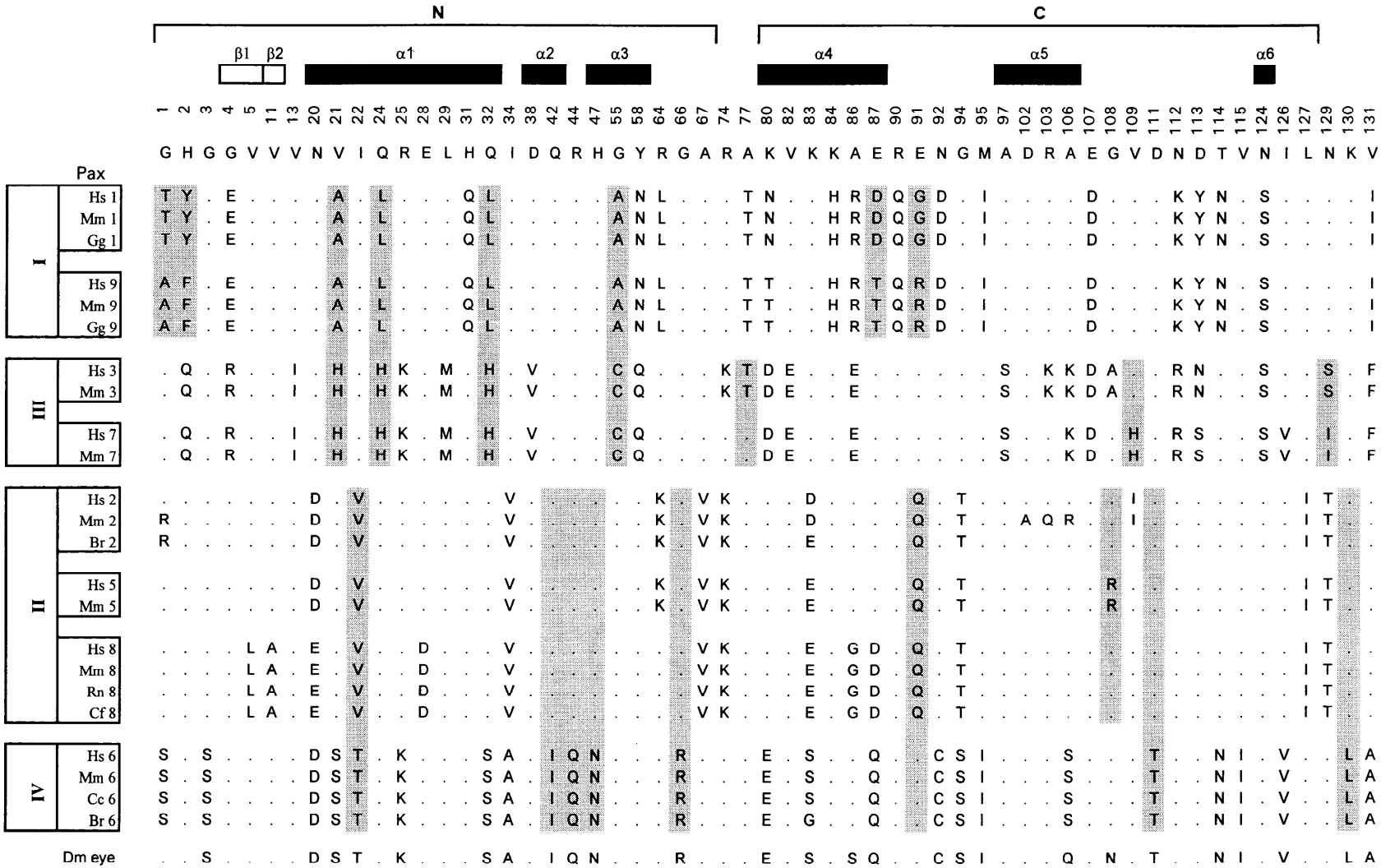

FIG. 6.—Radical and conservative fixed amino acid differences within and between vertebrate *Pax* gene groups. Fixed radical differences are shown with gray background.

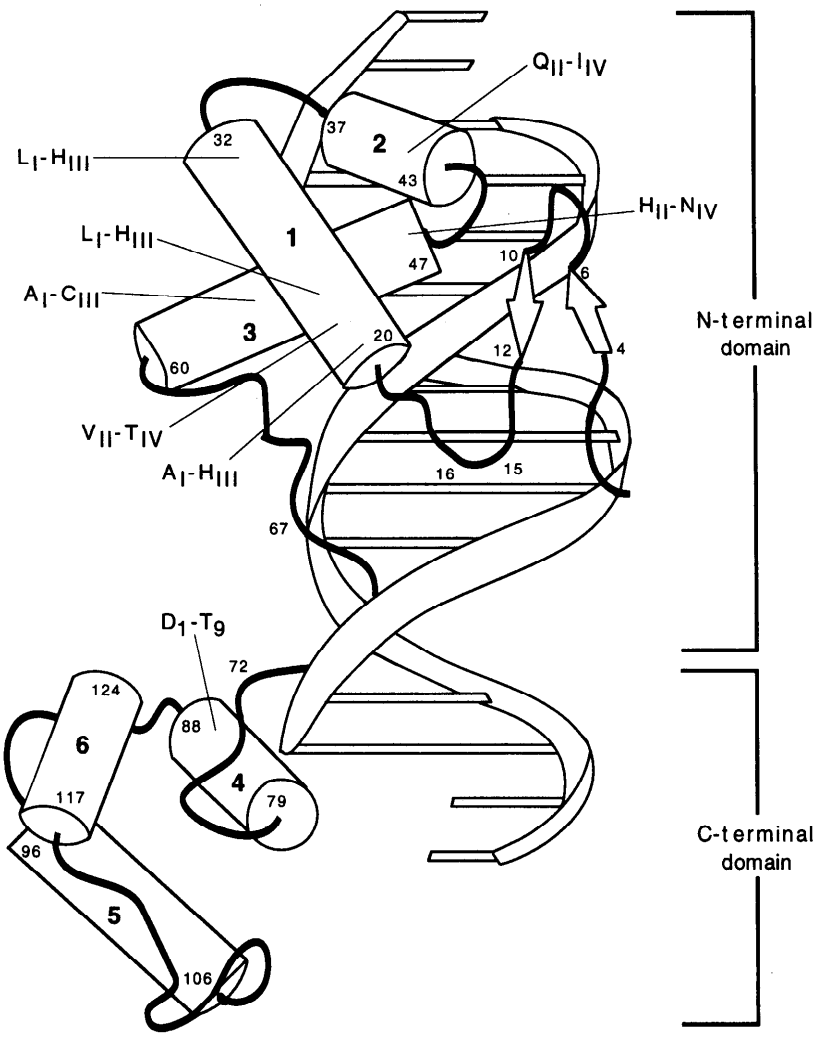

FIG. 7.—The crystal structure of the *prd* protein (Xu et al. 1995). The radical amino acid changes located in α-helices are shown.

ent from those sites showing differences between groups I and III. Again, the C-terminal domain in groups II and IV shows large numbers of conservative changes, most of which are found in sites not participating in α-helical structures. It is worth noting that the radical amino acid changes in hydrophobicity, as well as size, found in the N-terminal α-helices are rarely observed as indicated by Dayhoff's (1978) and Jones–Taylor–Thornton substitution (1992) matrices.

In a comparison of *Drosophila* sequences within each major group to their vertebrate homologs, we find that the amino acid residues identified to be important in the vertebrate groups are conserved in the *Drosophila* homologs of the respective vertebrate groups (e.g., see *Drosophila eyeless* and *Pax-6* in fig. 6). Such conservation of primary structure suggests conservation of function of protostome and deuterostome paired domains. Recent experimental data indicate such function-

Table 3
Characterization and Localization of Fixed Amino Acid Differences in Vertebrate *Pax* Genes

	RADICAL				CONSERVATIVE			
	α-helical		Intervening		α-helical		Intervening	
GROUP	N	C	N	C	N	C	N	C
Within I	0	1	2	1	0	0	0	0
Within III	0	0	0	3	0	1	0	4
Between I and III	4	0	0	0	5	5	3	6
Within II	0	0	0	1	2	2	3	0
Between II and IV	3	0	2	3	3	1	3	10

NOTE.—Amino acids found between the N-terminal and C-terminal domains were tabulated as C-terminal intervening regions.

al conservation in *Pax-6* and *eyeless* genes (e.g., Halder, Callaerts, and Gehring 1995).

Discussion

A Mechanism for Evolution of Paired Box Domains

Because every *Pax* gene group contains a *Drosophila Pax* gene that is distantly related to the vertebrate homologs, it is clear that the members of all four groups were present in the genomes of the most recent common ancestor of vertebrates and *Drosophila.* The gene duplications generating these four genes would then have occurred in progenitors of the ancestors of *Drosophila* and vertebrates (fig. 5). Thus, these four groups may have originated before the Cambrian radiation of triploblastic metazoan body plans, because the *C. elegans Pax-6* gene is a member of group IV and diverged prior to the splitting of *Pax-4* and *Pax-6.*

To determine the number of duplication events that led to the four major *Pax* groups, a root for the tree is necessary. As mentioned earlier, the root of the *Pax* tree is likely to be on branch *a, c,* or *e* in figure 2. It is interesting to note that the lengths of branches *a* and *c,* and *b* and *d* are quite similar (fig. 2). Branch lengths *a* and *c* represent the amount of evolutionary change leading to the two closely related groups I and III. Similarly, *b* and *d* represent the amount of evolutionary change leading to groups II and IV. In the absence of other information, it is possible to construct an evolutionary scenario in which the root lies on branch *e* by the midpoint rooting strategy (excluding *Pax-4,* which is evolving considerably faster). We have employed this possibility to propose an evolutionary scenario for the generation of observed diversity of paired box genes in *Drosophila* and vertebrates (fig. 5). It is worth noting that if the root lies on one of the other branches, the number of duplications will simply increase by 1. In this sense, the scenario we propose is more parsimonious if chromosomal (or genome) duplications were to occur before the Cambrian radiation of animal phyla.

In our scheme, there was one gene in the ancestors of all triploblastic animals (protostomes and deuterostomes) that contained a paired domain, a homeodomain, and an octapeptide. The first gene duplication resulted in two copies of this gene. The second gene duplication event involved these two genes and produced four genes (A–D; fig. 5) that were the precursors of the four *Pax* gene groups observed. Prior to the divergence of major triploblastic animal lineages, the homeodomain in precursor A was lost and that in B was modified. However, the exact point of modification of the homeodomain in B is not clear; it may have occurred before the species splitting or in the lineage leading to amphioxus. In either case, the homeodomain was lost independently in *Drosophila Pox neuro.* Clearly, the octapeptide was lost uniquely in the lineage leading to group IV.

In the lineage leading to vertebrates (after the vertebrate–invertebrate split), a "collective" gene duplication of the four genes (A–D) doubled the number of *Pax* genes in each vertebrate *Pax* group. An additional gene duplication in group II resulted in the three vertebrate *Pax* genes in this group. After the split of protostomes and deuterostomes, gene C underwent two gene duplications in the lineage leading to *Drosophila*: the first duplication resulted in the *prd* and the ancestor of *gsb* genes, and the second produced the *gsb-p* and *gsb-d* genes.

In this scenario, we have considered the evolution of the paired box and homeodomain as a single unit. This is for two reasons. First, there is evidence from functional studies that the paired domain and the homeodomain are both required functionally (Jun and Desplan 1996), even though in some cases they may function independently. Second, their evolution as a single unit is more parsimonious than the alternative evolutionary scenarios that would require insertion of octapeptides and homeodomains multiple times in the same relative positions.

The above scenario suggests that the amphioxus genome will contain at most four *Pax* genes (unless some have been lost by chance or independent gene duplications have occurred). This is consistent with the fact that the amphioxus genome contains fewer members of many multigene families as compared to the vertebrates genomes (e.g., Carroll 1995; Holland et al. 1994). In addition, other close relatives of vertebrates should contain four major types of *Pax* genes.

Radical Changes and Binding Specificities

Xu et al. (1995) have elucidated the crystal structure of the *Drosophila prd* gene paired domain. In this gene, only the N-terminal domain comes in close contact with DNA. Because we have found the fixed (Nei and Gojobori 1986) radical changes in the N-terminal α-helices, mapping these changes on the crystal structure (fig. 7) may provide insight into the possible role of the radical changes.

It is interesting that the residues that contact the DNA in the *prd* paired domain are conserved in all *Pax* genes, with one exception. At position 47, which is the first residue to bind DNA in the third α-helix (Xu et al. 1995), *Pax-6* has the amino acid Asn that recognizes the nucleotide T. All other *Pax* genes carry His at this position, which shows higher affinity toward the nucleotide G. This change, which was identified as a fixed radical difference in our analysis, therefore plays a large role in DNA binding specificity (reviewed in Jun and Desplan 1996).

The fact that all other *prd* DNA-contacting sites are conserved in the N-terminal points toward an overall similarity of crystal structure of different *Pax* genes. However, paired domains in other groups may differ from the *prd* paired domain N-terminal DNA-binding structure at some other sites. Sites with fixed radical amino acid differences shown in figure 6 are some such candidate positions. If these sites do come into contact with the DNA in other *Pax* genes, they are likely to have different target DNA sequences and/or different DNA binding affinities.

In comparing groups I and III, the radical difference in hydrophobicity, charge, and the side-chain

length at positions 21, 24, and 32 in the first α-helix of the N-terminal domain stand out (fig. 7). In group I, these positions are occupied by hydrophobic (uncharged) residues, whereas group III paired domains contain the positively charged hydrophilic residue His in each case. Two of these His residues (21, 32) are at the ends of the first α-helix in close proximity to the negatively charged DNA backbone, which may allow this α-helix to be in close contact with DNA. This is in contrast to the situation at the same positions in group I paired domains. Therefore, the differences in binding specificities and/or strengths of group I and III domains may reside in such uniquely fixed sites in the N-terminal domain.

In the case of groups II and IV, the only radical difference in helix 1 is a hydrophobicity difference in a residue that is close to the DNA backbone (hydrophobic in group II, hydrophilic in group IV). Similarly, the only radical difference in the second α-helix in these groups is at position 42 (Gln in group II, Ile in group IV). This difference, however, places the hydrophilic amino acid residue in group II rather than group IV. At position 47, the change from His to Asn in group IV affects the DNA binding both physically and electronically.

Regions outside the α-helices in the N-terminal domain may also play a role in the DNA-binding specificity of the various *Pax* genes. For instance, the paired domain has two β units that are found 5′ of the α-helices. The β1 unit contacts the minor groove in the *prd* gene. Interestingly, a charged residue is present in groups I (Glu) and III (Arg), whereas an uncharged residue (Gly) is found in members of groups II and IV. This may affect the contact to the DNA. Similarly, a β-turn unit between β2 and α1 contacts the minor groove. An Ile at position 13 in group III may help to distinguish its binding from all other groups, which contain Val at this site.

The region between β1 and β2 is highly conserved in all *Pax* proteins, required for an intramolecular interaction with residues 40, 44, and 45. While residues 40 and 45 are invariant, residue 44 is Gln in group IV and Arg in all other groups. This may represent an important mechanism for group IV *Pax* proteins to carry out this intramolecular interaction differently from other groups.

Residues at the tail end of the N-terminal domain may also play a role in recognition, as sites here also contact the minor groove of the DNA. In this region, residues 64 and 66 seem to be important. In group IV, both sites are charged (Arg), while in group I, both sites are uncharged. In both groups II and III, only one of these sites is charged (Arg or Lys at site 64). This difference in the number of charged residues may allow this tail region to contact the minor groove with different affinity.

In DNA recognition in group II and IV genes, experimental evidence indicates a role for C-terminal domain as well. It has been shown that both *Pax-5* and *Pax-6* recognize longer stretches of DNA as compared to *prd* gene, suggesting that both the N- and C-terminal domains are active in DNA binding. Jun and Desplan (1996) found that *Pax-6* in particular had a strict requirement for DNA sequences 3′ to those recognized by the *prd* gene. It is also known that the C-terminal domain of *Pax-5* plays an important role in DNA binding (Czerny, Schnaffer, and Busslinger 1993). However, Jun and Desplan (1996) determined that the consensus binding sequence for this gene is virtually identical to that of *prd*.

The above fixed-site analyses have allowed us to infer amino acid changes (and their locale) in the paired domains that may be responsible for altering their DNA binding affinity and specificity between groups. These differences reside exclusively in the α-helices of the N-terminal domain. In these analyses we used the crystal structure of the *Drosophila* prd protein. With the knowledge of the structures of other paired domains, it should be possible to conduct even more precise analyses. This information may then be useful in studying the influence of specific amino acid sites in functional diversification of *Pax* domains and in further understanding of the evolution of developmental mechanisms.

Acknowledgments

We thank Koichiro Tamura, Esther Siegfried, Mitch Price, and George Zhang for helpful suggestions and comments. This work was supported by NSF and NIH grants to Masatoshi Nei and Z.-C.L.

LITERATURE CITED

ADAMS, B., P. DORFLER, A. AGUZZI, Z. KOZMIK, and P. URBANEK. 1992. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes & Dev. **6**:1589–1607.

BALDWIN, C. T., C. F. HOTH, J. A. AMOS, E. O. DA-SILVA, and A. MILUNSKY. 1992. An exonic mutation in the HuP2 paired domian gene causes Waardenburg's syndrome. Nature **355**:637–638.

BALLING, R., U. DEUTSCH, and P. GRUSS. 1988. Undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax-1. Cell **55**: 531–535.

BAUMGARTNER, S., D. BOPP, M. BURRI, and M. NOLL. 1987. Structure of two genes at the gooseberry locus related to the paired gene and their spatial expression during *Drosophila* embryogenesis. Genes & Dev. **1**:1247–1267.

BENTON, M. J. 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. J. Mol. Evol. **30**:409–424.

———. 1993. The fossil record 2. Chapman and Hall, New York.

BOPP, D., E. JAMET, S. BAUMGARTNER, M. BURRI, and M. NOLL. 1989. Isolation of two tissue-specific *Drosophila* paired box genes, Pox meso and Pox neuro. EMBO J. **8**: 3447–3457.

BURRI, M., Y. TROMVOUKIS, D. BOPP, G. FRIGERIO, and M. NOLL. 1989. Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO J. **8**:1183–1190.

CARROLL, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. Nature **376**:479–485.

CHALEPAKIS, G., R. FRITSCH, H. FICKENSCHER, U. DEUTSCH, M. GOULDING, and P. GRUSS. 1991. The molecular basis of the undulated/Pax-1 mutation. Cell **66**:873–884.

CHISHOLM, A. D., and H. R. HORVITZ. 1995. Patterning of the Caenorhabditis elegans head regeneration by the Pax-6 family member vab-3. Nature **376**:52–55.

CZERNY, T., and M. BUSSLINGER. 1995. DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). Mol. Cell. Biol. **15**:2858–2871.

CZERNY, T., G. SCHNAFFER, and M. BUSSLINGER. 1993. DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. Genes & Dev. **7**: 2048–2061.

DAMBLY-CHAUDIERE, C., E. JAMET, M. BURRI, D. BOPP, K. BASLER, E. HAFEN, N. DUMONT, P. SPIELMANN, A. GHYSEN, and M. NOLL. 1992. The paired box gene *pox neuro*: a determinant of poly-innervated sense organs in *Drosophila*. Cell **69**:159–172.

DAYHOFF, M. O. 1978. Survey of new data and computer methods of analysis. P. 29 *in* M. O. DAYHOFF, ed. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Springs, Md.

DRESSLER, G. R., U. DEUTSCH, K. CHOWDHURY, H. O. NORNES, and P. GRUSS. 1990. Pax-2, a new murine paired-box-containing gene and its expression in the developing excretory system. Development **109**:787–795.

DUBOULE, D. 1994. Guidebook to the homeobox genes. Oxford University Press, New York.

ECCLES, M. R., L. J. WALLIS, A. E. FIDLER, N. K. SPURR, P. J. GOODFELLOW, and A. E. REEVE. 1992. Expression of the PAX2 gene in human fetal kidney and Wilms' tumor. Cell Growth Differ. **3**:279–289.

EPSTEIN, D., M. VEKEMANS, and P. GROSS. 1991. Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. Cell **67**:767–774.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FRIGERIO, G., M. BURRI, D. BOPP, S. BAUMGARTNER, and M. NOLL. 1986. Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. Cell **47**:735–746.

GLASER, T., D. S. WALTON, and R. L. MAAS. 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. Nature Genet. **2**:232–239.

GOULDING, M. D., G. CHALEPAKIS, U. DEUTSCH, J. R. ERSELIUS, and P. GRUSS. 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J. **10**:1135–1147.

HALDER, G., P. CALLAERTS, and W. J. GEHRING. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. Science **267**:1788–1792.

HEDGES, S. B., P. H. PARKER, C. G. SIBLEY, and S. KUMAR. 1996. Continental breakup and ordinal diversification of birds and mammals. Nature **381**:226–229.

HILL, R. E., J. FAVOR, B. L. M. HOGAN, C. C. T. TON, G. F. SAUDERS, I. M. HANSON, J. PROSSER, T. JORDAN, N. D. HASTIE, and V. VAN HEYNINGEN. 1991. Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature **354**:522–525.

HOLLAND, N. D., L. Z. HOLLAND, and Z. KOZMIK. 1995. An amphioxus Pax gene, AmphiPax1 expressed in embryonic endoderm, but not in mesoderm: implications for the evolution of class I paired box genes. Mol. Mar. Biol. Biotechnol. **4**:206–214.

HOLLAND, P. W. H., J. GARCIA-FERNANDEZ, N. A. WILLIAMS, and A. SIDOW. 1994. Gene duplications and the origins of vertebrate development. Development (Suppl.):125–133.

JONES, D. T., W. R. TAYLOR, and J. M. THORNTON. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–282.

JOSTES, B., C. WALTHER, and P. GRUSS. 1991. The murine paired box gene, Pax-7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev. **33**:27–38.

JUN, S., and C. DESPLAN. 1996. Cooperative interactions between paired domain and homeodomain. Development **122**: 2639–2650.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KOZMIK, Z., R. KURZBAUER, P. DORFLER, and M. BUSSLINGER. 1993. Alternative splicing of Pax-8 gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. Mol. Cell. Biol. **13**:6024–6035.

KRAUSS, S., T. JOHANSEN, V. KORZH, and A. FJOSE. 1991. Expression of the zebrafish paired box gene pax[zf-b] during early neurogenesis. Development **113**:1193–1206.

KUMAR, S., K. A. BALCZAREK, and Z.-C. LAI. 1996. Evolution of the *hedgehog* gene family. Genetics **142**:965–972.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetic analysis. Pennsylvania State University, University Park.

LI, H.-S., J.-M. YANG, R. D. JACOBSON, D. PASKO, and O. SUNDIN. 1994. Pax-6 is first expressed in a region of ectoderm anterior to the early neural plate: implications for stepwise determination of the lens. Dev. Biol. **162**:181–194.

LI, X., and M. NOLL. 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. Nature **367**:83–86.

LOOSLI, F., M. KMITA-CUNISSE, and W. J. GEHRING. 1996. Isolation of a Pax-6 homolog from the ribbonworm *Lineus sanguineus*. Proc. Natl. Acad. Sci. USA **93**:2658–2663.

MARTIN, P., C. CARRIERE, C. DOZIER, B. QUATANNENS, M.-A. MIRABEL, B. VANDENBUNDER, D. STEHELIN, and S. SAULE. 1992. Characterization of a paired box- and homeobox-containing quail gene (Pax–QNR) expressed in the neuroretina. Oncogene **7**:1721–1728.

MATSUO, T., N. OSUMI-YAMASHITA, S. NOJI, H. OHUCHI, and E. KOYAMA. 1993. A mutation in the Pax-6 gene in rat small eye is associated with impared migration of midbrain crest cells. Nature Genet. **3**: 299–304.

MAULBECKER, C. C., and P. GRUSS. 1993. The oncogenic potential of Pax genes. EMBO J. **12**:2361–2367.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

NEUBUSER, A., H. KOSEKI, and R. BALLING. 1995. Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1. Dev. Biol. **170**:701–716.

NOLL, M. 1993. Evolution and role of Pax genes. Curr. Opin. Genet. Dev. **3**:595–605.

PETERS, H., U. DOLL, and J. NIESSING. 1995. Differential expression of the chicken Pax-1 and Pax-9 gene: in situ hybridization and immunohistochemical analysis. Dev. Dynam. **203**:1–16.

PLACHOV, D., K. CHOWDHURY, D. WALTHER, D. SIMON, J.-L. GUENET, and P. GRUSS. 1990. Pax-8, a murine paired box gene expressed in the developing excretory system and thyroid gland. Development **110**:643–651.

POLEEV, A., H. FICKENSCHER, S. MUNDLOS, A. WINTERPACHT, B. ZABEL, A. FIDLER, P. GRUSS, and D. PLACHOV. 1992. Pax-8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilm's tumor. Development **116**:611–623.

PUSCHEL, A. W., P. GRUSS, and M. WESTERFIELD. 1992. Sequence and expression pattern of pax-6 are highly conserved between zebrafish and mice. Development **114**: 643–651.

QUIRING, R., U. WALLDORF, U. KLOTER, and W. J. GEHRING. 1994. Homology of the eyeless gene of *Drosophila* to the Small eye gene in mice and Aniridia in humans. Science **265**:785–789.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCHAFER, B. W., T. CZERNY, M. BERNASCONI, M. GENINI, and M. BUSSLINGER. 1994. Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. Nucleic Acids Res. **22**:4574–4582.

STAPELTON, P., A. WEITH, P. URBANEK, Z. KOZMIK, and M. BUSSLINGER. 1993. Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. Nature Genet. **3**:292–298.

STUART, E. T., C. KIOUSSI, and P. GRUSS. 1993. Mammalian Pax genes. Annu. Rev. Genet. **27**:219–236.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign, Ill.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

TASSABEHJI, M., A. P. READ, V. E. NEWTON, R. HARRIS, R. BALLING, P. GRUSS, and T. STRACHAN. 1992. Waardenburg's Syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. Nature **355**:635–636.

TON, C. C. T., H. HIRVONEN, H. MIWA et al. (12 co-authors). 1991. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. Cell **67**:1059–1074.

TREISMAN, J., E. HARRIS, and C. DESPLAN. 1991. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. Genes & Dev. **5**:594–604.

VOET, D., and J. G. VOET. 1990. Biochemistry. John Wiley & Sons, New York.

WALTHER, C., and P. GRUSS. 1991. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development **113**:1435–1449.

WALTHER, C., J.-L. GUENET, D. SIMON, U. DEUTSCH, B. JOSTES, M. D. GOULDING, D. PLACHOV, R. BALLING, and P. GRUSS. 1991. Pax: a murine multigene family of paired box-containing genes. Genomics **11**:424–434.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.

XU, W., M. ROULD, S. JUN, C. DESPLAN, and C. PABO. 1995. Crystal structure of a paired domain–DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. Cell **80**:639–650.

ZHANG, Y., and S. W. EMMONS. 1995. Specification of sense-organ identity by a *Caenorhabditis elegans* Pax-6 homologue. Nature **377**:55–59.

SHOZO YOKOYAMA, reviewing editor

Accepted May 1, 1997