Zoological Studies **38**(4): 367-372 (1999)

Zoological Studies

Review Article

What the Cnidaria Tell Us about *Pax* Gene Evolution

David J. Miller

Department of Biochemistry and Molecular Biology, James Cook University, Townsville, Queensland 4811, Australia
Tel: 61-7-47814473. Fax: 61-7-47251394. E-mail: david.miller@jcu.edu.au

CONTENTS

ABSTRACT

David J. Miller (1999) What the Cnidaria tell us about *Pax* gene evolution. *Zoological Studies* **38**(4): 367-372. *Pax* genes are defined by the possession of a motif that was first identified in the *Drosophila* segmentation gene *paired*. They encode transcription factors often containing a homeodomain (or a part of it) as well as the paired domain. Genes in this family play central roles in the development of animals; however, the complexity of the family–9 human and 8 *Drosophila Pax* genes are now known–and the diversity of their functions has effectively obscured the identification of functional homology between different organisms. The availability of data on a number of *Pax* genes in cnidarians should facilitate unravelling the complex evolutionary history of the *Pax* gene family. This review makes use of the cnidarian data and the extensive genomic sequence data now available for the nematode *Caenorhabditis* to propose a model for evolution of the various *Pax* gene types.

Key words: Development, Paired, *Pax*, Evolution, Cnidaria.

INTRODUCTION

Pax genes encode an important class of transcription factors that play central roles in the development of animals. Nine human and 8 *Drosophila Pax* genes (9 if the *eye gone* gene, which does not have a complete paired box, is included) are now known, but this may not be their entire complements. *Pax* gene functions include roles in anterior patterning in *Caenorhabditis* (*vab-3/mab-18*), segmentation in *Drosophila* (*paired, gooseberry*), B-cell specification (*Pax-5*), and development of the thyroid (*Pax-8*), vertebral column (*Pax-1, Pax-9*), limb buds (*Pax-3*), eye (*Pax-6, Pax-2*) and nervous system, in vertebrates. This bewildering diversity of *Pax* gene functions obscures the identification of functional homology between different organisms (see for example, Slack et al. 1993).

While several of the fly and vertebrate genes are clearly orthologous – *sparkling* a *Pax-2/5/8* homolog, *poxmeso* a *Pax-1/9* homolog, and *paired/gooseberry/gooseberry neuro* corresponding to vertebrate *Pax-3/7* – at this stage, only in the case of *eyeless* and *Pax-6* is there clear evidence for conservation of function between *Drosophila* and vertebrates. Furthermore, independent duplications of *Pax* genes

appear to have occurred frequently during evolution. For example, *Drosophila* has 2 *Pax-6* homologs, *eyeless* and *twin of eyeless*, and the zebrafish gene *no isthmus* is a fish-specific duplication of vertebrate *Pax-2* (Pfeffer et al. 1998). Finally, some *Pax* genes appear to be taxon-specific – for example, *Pax-4* is probably unique to vertebrates, and no clear homologs of *Drosophila pox neuro* have yet been identified.

There have been a few attempts to unravel the complex evolutionary history of the *Pax* gene family (Noll et al. 1993, Balczarek et al. 1997). The availability of data on a number of *Pax* genes in cnidarians facilitates this task, because these are the simplest animals at the tissue level of organization, and they can essentially be regarded as the sister group of higher animals. In addition, the extensive genomic sequence data now available for *Caenorhabditis* provide insights into the ancestral gene set of triploblastic animals. This review makes use of these data to propose a model for evolution of the various *Pax* gene types.

PAX GENES, PAIRED DOMAINS AND *PAX* HOMEODOMAINS

Pax genes are defined by the possession of a *pa*ired-bo*x*, first identified in the *Drosophila* segmentation gene *paired* (*prd*). They encode transcription factors often containing a complete or partial homeodomain as well as the paired domain. The paired domain is a large (128 AA residue) DNA-binding domain containing 2 distinct helix-loop-helix motifs (Xu et al. 1995). The N-terminal PAI subdomain is exclusively responsible for DNA-binding in the case of the paired protein (Xu et al. 1995); however, in the case of several other Pax proteins, the RED subdomain also makes DNA contacts (Czerny et al. 1993, Epstein et al. 1994a,b) and appears to be involved in modulating the binding specificity of the PAI subdomain (Vogan and Gros 1997). Residues 42, 44, and 47 (PAI subdomain) are critical in determining the specificity of DNA-sequence recognition by Pax-6 and Pax-5 (Czerny and Busslinger 1995). In addition to this mode of binding, Pax (and paired-like) proteins can bind as dimers via their homeodomains to a palindromic motif with the consensus sequence TAAT(N_{2-3})ATTA (P2 or P3 sites respectively; Wilson et al. 1995), and other targets are likely to exist for the RED subdomain (Epstein et al. 1994b). Hence some Pax proteins have at least 3 distinct modes of DNA-binding.

THE KNOWN CNIDARIAN *PAX* GENES

Recently, the sequences of 2 genes from each of 3 cnidarians, *Hydra littoralis* (Hydrozoa), the jellyfish *Chrysaora quinquecirrha* (Scyphozoa), and the coral *Acropora millepora* (Anthozoa) have been published (Sun et al. 1997, Catmull et al. 1998). It is not clear at this stage whether the full complement of cnidarian *Pax* genes has yet been identified, however these 3 organisms are likely to represent the range of diversity within the phylum Cnidaria. The Anthozoa are the basal class (Bridge et al. 1992, Odorico and Miller 1997); thus the coral is more likely

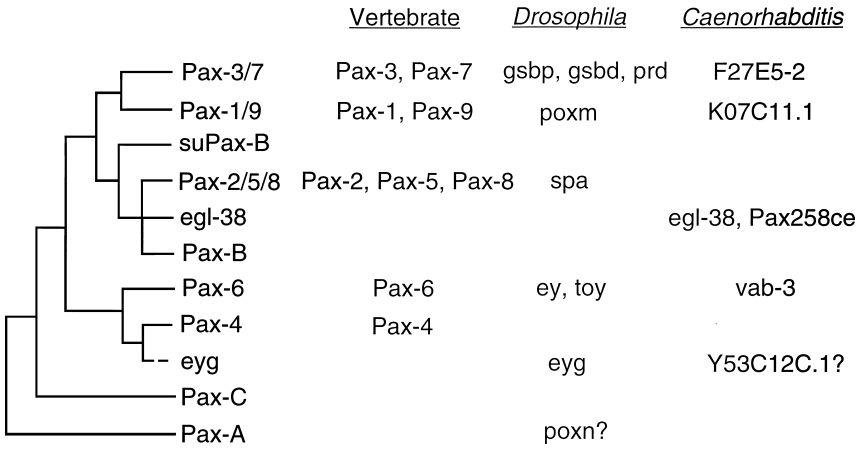

Fig.1. Phylogenetic analysis of paired domain sequences from a representative range of animals. The tree summarizes analyses conducted using both maximum parsimony and distance methods on a dataset consisting of 39 paired domain sequences encompassing the diversity of known Pax proteins. The position of the eye gone sequence (which contains only a partial paired domain) was determined in a more restricted set of analyses of the Pax-6/Pax-4 group.

to represent ancestral character states than are the more derived hydra and jellyfish.

As might be expected, relating the cnidarian genes to specific *Pax* gene classes identified from vertebrates and *Drosophila* is not straightforward. The paired domains encoded by all of the cnidarian *Pax* genes are most closely related to those in the vertebrate Pax-2/5/8 class and, in some cases, to that in *Drosophila pox neuro*. One of these genes, *Pax-A*, in each of the 3 cnidarians is clearly orthologous. The only recognizable motif in the corresponding proteins is the paired domain – they contain neither a homeodomain (or a homeodomain relic) nor an octapeptide; the paired domains are approximately 87.5% identical among the 3 organisms.

The 2nd gene in the hydra and jellyfish is known as *Pax-B*, and these 2 are again clear orthologs. However *Pax-C*, the 2nd coral gene, is only distantly related to hydra/jellyfish *Pax-B*. Both of these genes encode a homeodomain in addition to the paired domain, and the *Pax-B* genes also encode an unambiguous octapeptide motif (Fig. 1), whereas *Pax-C* does not. It has been suggested that jellyfish/hydra *Pax-B* corresponds to vertebrate *Pax-6* (Sun et al. 1997). However, the homeodomains of hydra Pax-B and human PAX-6 are only 55% identical, and, in contrast to Pax-B, all Pax-6 class proteins lack an octapeptide. In fact, Pax-C more closely resembles PAX-6 than does Pax-B; it lacks an octapeptide, and has 72% identity in the homeodomain. However, it has only 3 of the 10 “Pax-6 diagnostic” amino acid residues in the homeodomain (Callaerts et al. 1997), and it shares a number of key residues with the Pax-3/7 class (Catmull et al. 1998). The Pax-C paired domain shares 4 unique substitutions (two of which are shared with pox neuro) with the Pax-A proteins, suggesting a common origin via a duplication event (see Fig. 2).

PHYLOGENY OF PAIRED DOMAINS – WIDELY-DISTRIBUTED AND LINEAGE-SPECIFIC CLASSES

A number of novel Pax sequences reported recently (including the cnidarian data) have not previously been included in phylogenetic studies. Figure 1 summarizes the results of a number of phylogenetic analyses of a representative range of paired domain sequences. Not all of the available sequences were included; when nearly identical sequences were available from a number of organisms, so as to expedite the analyses, a single representative was selected (for example, Pax-6 sequences are available from a large number of vertebrates, and are all either identical or very similar).

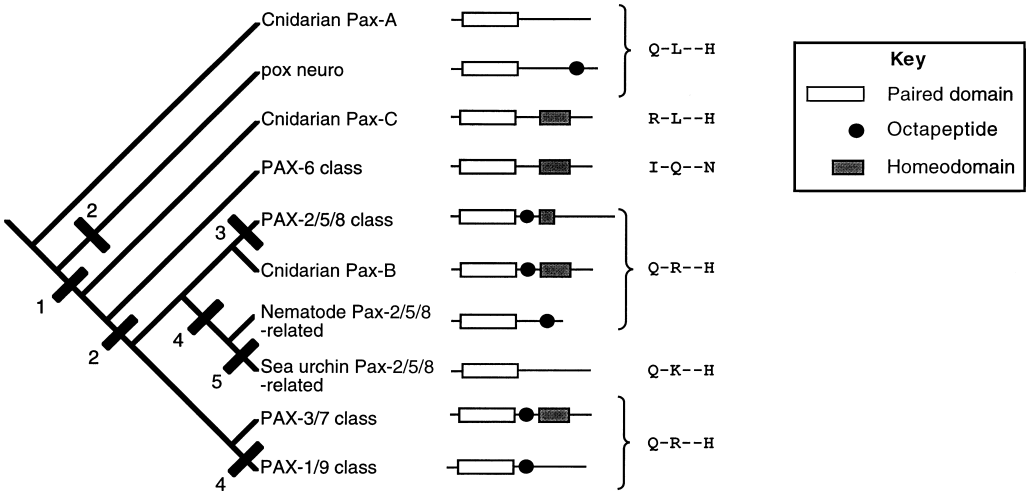

Fig. 2. Scheme representing evolution of Pax proteins. We propose that the ancestor of all vertebrate Pax classes arose by fusion of a *Pax-A*-like gene and a *paired-like* homeobox gene. Pax-C represents the precursor of the vertebrate proteins after the fusion event. Numbered synapomorphies on the cladogram are as follows: (1) acquisition of the homeodomain; (2) acquisition of the octapeptide (n.b. [1] and [2] may have occurred simultaneously, as some paired-like proteins also contain octapeptide motifs); (3) partial loss of the homeodomain; (4) complete loss of the homeodomain; and (5) loss of octapeptide. The central part of the figure shows the schematised structure of the various Pax proteins; the open box represents the paired domain, the filled box the homeodomain, and the filled circle the octapeptide. Note that the homeodomain of Pax-2/5/8 is truncated, hence the filled box in this case is shown as half the standard size. The letters on the right of the figure are the amino acid residues at positions 42-47 in the paired domain; positions #42, 44, and 47 have been shown to be critical in determining the specificity of paired domain binding (Czerny and Busslinger 1995).

Because no *Pax* genes have yet been identified outside of the Metazoa, there is no obvious appropriate outgroup with which to root the phylogenetic tree. Sun et al. (1997) have argued that the vertebrate *Pax-6* and *Pax-2/5/8* groups evolved from an ancestral gene similar to cnidarian *Pax-B*. This suggests that the tree should be rooted using Pax-B. However, on the basis of overall structure, we favor the alternative of rooting the tree using the least complex of the cnidarian Pax proteins, the Pax-A clade (see Fig. 1).

The 4 major clades identified in our analyses correspond to classes I-IV (Pax-1/9, Pax-2/5/8, Pax-3/7, and Pax-6 groups, respectively) in Balczarek et al. (1997) and a number of other studies, and the sister group relationship between the Pax-1/9 and Pax-3/7 clades is also consistent with previous analyses. Clearly the 4 major *Pax* gene classes were distinct before the protostome/deuterostome divergence; not only is *Drosophila* represented in each of the 4 classes, but so is *Caenorhabditis*. Inclusion of the cnidarian data adds some new perspectives to the issue of the evolution of paired domains. Most striking is the high degree of identity between the paired domains of the cnidarian Pax-B and Pax-2/5/8 classes (Fig. 1). The divergent nature of cnidarian Pax-A and Pax-B paired domains is also quite clear from the analyses, as is the much closer relationship between the Pax-A clade and Pax-C.

The inclusion of the cnidarian data also highlights the atypical nature of some of the Pax-2/5/8-related sequences from *Caenorhabditis elegans* and sea urchins. *C. elegans egl-38* (Chamberlin et al. 1997) and *cePax258B* (Czerny et al. 1997) represent an independent (nematode-specific) duplication of a *Pax-2/5/8*-like gene. Although these genes encode proteins whose paired domains are most closely related to the Pax-2/5/8 class, they are only distantly related to other ("true") Pax-2/5/8 homologs. In addition, egl-38 and cePax258B lack any vestige of (and possibly never have possessed) a homeodomain or the C-terminal transactivation and inhibition domains that are well conserved throughout the Pax-2/5/8 class. For these reasons, egl-38 and cePax258B should probably not be considered true Pax-2/5/8 homologs, but placed in a class of their own.

There is a clear *Pax-2/5/8* homolog in the sea urchin (suPax258; Czerny et al. 1997), although again atypical *Pax-2/5/8*-related genes are also present. These authors report the presence of 3 distinct atypical *Pax-2/5/8*-related genes in sea urchins, but for only one of these (sea urchin *Pax-B*) are full sequences available (*suPax-B* data were reported for 2 species, *Strongylocentrotus purpuratus* and *Paracentrotus lividus*). For suPax-A and suPax-C, only partial paired-domain data are available, and these were not included in the phylogenetic analyses. suPax-B is characterized by a highly diverged Pax-2/5/8-related paired-domain sequence, absence of homeodomain and octapeptide (C-terminal regulatory domains almost certainly also not present), and Q-K--H at the positions determining specificity, rather than the Q-R--H which characterizes all other Pax-2/5/8 proteins (see Fig. 2). As in the case of egl-38/cePax258B, sea urchin Pax-B differs sufficiently from true Pax-2/5/8 homologs to merit a class of its own.

The analyses shown in figure 2 also highlight the uniqueness of *Drosophila* pox neuro. While pox neuro is often lumped with the Pax-2/5/8 class (e.g., Noll 1993), this is likely to simply reflect the fact that the paired domains in the Pax-2/5/8 class most resemble those in the ancestral Pax protein. Hence, paired sequences that are not obviously assignable to other classes tend to look most like Pax-2/5/8 and therefore have been grouped with them.

THE ANCESTRAL *PAX* GENE – STRUCTURE AND SEQUENCE

Analyses of paired domain sequences (Fig. 1), together with those of paired-related homeodomains (Galliot et al. 1999), provide some novel insights into the evolutionary history of the Pax family. We have previously proposed a model for the evolution of the various Pax classes (Catmull et al. 1998), and the present analyses permit an expansion on this, as shown in figure 2. Briefly, we suggest that the precursor of all vertebrate classes arose through fusion of a *Pax-A*-like gene with a *paired*-like homeobox gene; this ancestral gene probably most closely resembled *Pax-6*, and its closest extant relative is likely to be *Pax-C*. It is unclear whether the homeodomain and octapeptide were acquired consecutively or simultaneously, as the 2 motifs are found in both some Pax and some paired-like proteins. The presence of an unambiguous octapeptide in hydra Pax-B indicates that acquisition of both homeodomain and octapeptide occurred early in animal evolution.

As discussed above, the cnidarian data imply that the ancestral *Pax* gene encoded a paired domain most like those in the Pax-2/5/8 class. Why should Pax-2/5/8 most closely reflect the ancestral paired-domain sequence? The answer may be that the loss of a substantial part of the homeodomain resulted in restriction of the DNA-binding options of

the protein, effectively constraining the divergence of the paired domain. In the Pax-2/5/8 class, homeodomain dimerization cannot occur, because this requires the 3rd helix (Wilson et al. 1995). As the residual homeodomain features only helix 1 and the extreme N-terminal end of helix 2, it is unlikely to function in DNA-binding but presumably interacts with other transcription factors. There are precedents for this – for example, in the case of the paired-like homeodomain protein Phox1, residues on the surface of helix 1 interact with other transcription factors when it binds to the SRE (Simon et al. 1997). The reduction of DNA-binding options may have led to the "fixing" of the Pax-2/5/8 paired domain – its divergence may have been much more constrained than the paired domains in those Pax classes in which a full homeodomain was maintained.

The paired domain in cnidarian Pax-B is closely related to those of the vertebrate Pax-2/5/8 class. However, in this case a complete homeodomain is present–why in this case should the paired-domain have become fixed? One possible explanation is that in Pax-B the homeodomain has lost the ability to dimerize–effectively, *Pax-B* may be on its way to becoming a *Pax-2/5/8*-like gene. The recognition helix (helix 3) of Pax-B contains the conserved residues, E42 and R44, that play critical roles in recognition of the P3 site and dimer formation. However, the presence of a proline residue at position 43 (in place of the alanine present in almost all other Pax homeodomains) is likely to distort the helix, possibily sufficiently to prevent recognition of the P2 site, and may also effectively prevent dimerization, as this position is known to contribute to the dimer interface (Wilson et al. 1995). This has important implications in relation to the suggestion that Pax-B is the cnidarian Pax-6 ortholog (Sun et al. 1997), as it is thought that the interaction between Pax-6 and the rhodopsin promoter (and genes for other visual pigments) represents its most evolutionarily ancient function (Sheng et al. 1997). This occurs through the homeodomain dimerization on P3 sites, an interaction known to occur for most paired-like and several Pax proteins, but unlikely to occur with Pax-B.

SYNTHESIS

In the absence of expression data, attempts to homologize the cnidarian *Pax* genes with those from vertebrates or *Drosophila* are premature. Structurally, the cnidarian genes differ significantly; the *Pax-A* genes differ from Pax-2/5/8 class members in encoding neither a partial homeodomain nor an octapeptide, and the homeodomains in both Pax-B and Pax-C are intermediate between those of the Pax-6 and Pax-3/7 classes. The data presently available imply that the diversification of *Pax* genes to the major classes identified in higher animals probably occurred after the divergence of the Cnidaria; *Pax-C* and *Pax-B* are very different but intermediate types, and *Pax-A* is not a 'true' *Pax-2/5/8* gene.

Nevertheless, the cnidarian data provide valuable insights into Pax evolution, and several testable predictions can be made about the evolution of Pax function based on the scenario outlined above. The Cnidaria allow us to see which *Pax* and *Paired*-like genes in higher animals have changed least since very simple animals diverged. The level of identity in the paired domains of the cnidarian Pax-A proteins and in *Drosophila* pox neuro suggests that *pox neuro* may most closely reflect the ancestral function of *Pax* genes in higher animals.

Nervous system expression is nearly universal for *Pax* genes, implying that this is where these genes originally functioned. *Pox neuro* is a neuroblast cell-fate gene (Dambly-Chaudiere et al. 1992, Nottebohm et al. 1992), and *no isthmus* is central to CNS patterning in zebrafish (Macdonald et al. 1997, Pfeffer et al. 1998). Furthermore, Pax proteins that lack a homeodomain appear to function in cell-fate decisions–this is true of egl-38, despite its function being outside of the nervous system. This leads us to predict that cnidarian *Pax-A* will have cell-fate specification activity, probably in the nervous system. The fusion of *paired*-like and *Pax-A*-like genes, the event that gave rise to the ancestor of the vertebrate (and most invertebrate) *Pax* genes, probably permitted a transition in function from roles in cell-fate specification to roles in head-end patterning. If this is the case, we expect both *Pax-C* and cnidarian *Pax-B* to show head-specific (or at least, axis-specific) expression.

Acknowledgments: The author would like to thank Brigitte Galliot, Eldon Ball, David Hayward, and members of the Miller laboratory for discussions and for commenting on the manuscript, and Colomban de Vargas for assistance with the analyses. This work was supported by grants from the Australian Research Council and James Cook Univ.

REFERENCES

Balczarek KA, ZC Lai, S Kumar. 1997. Evolution and functional diversification of the Paired-box (Pax) DNA-binding proteins. Mol. Biol. Evol. **14:** 829-842.

Bridge D, CW Cunningham, B Schierwater, R Desalle, LW Buss.

1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc. Natl. Acad. Sci. USA **89:** 8750-8753.
Callaerts P, G Halder, WJ Gehring. 1997. Pax-6 in development and evolution. Ann. Rev. Neurosci. **220:** 483-532.
Catmull J, DC Hayward, NE Macintyre, JS Reece-Hoyes, R Mastro, P Callaerts, EE Ball, DJ Miller. 1998. *Pax-6* origins – implications from the structure of two coral *Pax* genes. Dev. Genes Evol. **208:** 352-356.
Chamberlin HM, RE Palmer, AP Newman, PW Sternberg, DL Baillie, JH Thomas. 1997. The *PAX* gene *egl-38* mediates developmental patterning in *Caenorhabditis elegans*. Development **124:** 3919-3928.
Czerny T, M Bouchard, Z Kozmik, M Busslinger. 1997. The characterisation of novel *Pax* genes of the sea urchin and *Drosophila* reveal an ancient evolutionary origin of the *Pax-2/5/8* subfamily. Mech. Dev. **67:** 179-192.
Czerny T, 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 Schaffner, 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, M Dumont, P Spielmann A, Ghysen, M Noll. 1992. The paired box gene *pox neuro*: a determinant of poly-innervated sense organs in *Drosophila*. Cell **69:** 159-172.
Epstein JA, J Cai, T Glaser, L Jepeal, RL Maas. 1994a. Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. J. Biol. Chem. **269:** 8355-8361.
Epstein JA, T Glaser, J Cai, L Jepeal, DS Walton, RL Maas. 1994b. Two independent and interactive DNA-binding subdomains of the Pax-6 paired domain are regulated by alternative splicing. Genes Dev. **17:** 2022-2034.
Galliot B, C de Vargas, DJ Miller. 1999. Evolution of homeobox genes: Q50 *Paired*-like genes founded the Paired class. Dev. Genes Evol. **209:** 186-197.
Macdonald R, J Scholes, U Strahle, C Brennan, N Holder, M Brand, SW Wilson. 1997. The Pax protein Noi is required for commissural axon pathway formation in the rostral forebrain. Development **124:** 2397-2408.
Noll M. 1993. Evolution and role of *Pax* genes. Curr. Opinion Genet. Dev. **3:** 595-605.
Nottebohm E, C Dambly-Chaudiere, A Ghysen. 1992. Connectivity of chemosensory neurons is controlled by the gene *poxn* in *Drosophila*. Nature **359:** 829-832.
Odorico DM, DJ Miller. 1997. Internal and external relationships of the Cnidaria: implications of primary and predicted secondary structure of the 5'-end of the 23S-like rDNA. Proc. Roy. Soc. Lond. B **264:** 77-82.
Pfeffer PL, T Gerster, K Lun, M Brand, M Busslinger. 1998. Characterisation of three novel members of the zebrafish pax2/5/8 family: dependency of *Pax5* and *Pax8* expression on the *Pax2.1* (*noi*) function. Development **125:** 3063-3074.
Sheng G, E Thouvenot, D Schmucker, DS Wilson, C Desplan. 1997. Direct regulation of *rhodopsin*1 by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. Genes Dev. **11:** 1122-1131.
Simon KJ, DA Grueneberg, M Gilman. 1997. Protein and DNA contact surfaces that mediate the selective action of the Phox1 homeodomain at the *c-fos* serum response element. Mol. Cell. Biol. **17:** 6653-6662.
Slack JMW, PWH Holland, CF Graham. 1993. The zootype and the phylotypic stage. Nature **361:** 490-492.
Sun H, A Rodin, Y Zhou, DP Dickinson, DE Harper, D Hewett-Emmett, WH Li. 1997. Evolution of paired domains: isolation and sequencing of jellyfish and hydra *Pax* genes related to *Pax-5* and *Pax-6*. Proc. Natl. Acad. Sci., USA **94:** 5156-5161.
Vogan KJ, P Gros. 1997. The C-terminal subdomain makes an important contribution to the DNA binding activity of the Pax-3 paired domain. J. Biol. Chem. **272:** 28289-28295.
Wilson DS, B Guenther, C Desplan, J Kuriyan. 1995. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. Cell **82:** 709-719.
Xu W, MA Rould, S Jun, C Desplan, CO Pabo. 1995. Crystal structure of a paired domain-DNA complex at 2.5A resolution reveals structural basis for Pax developmental mutations. Cell **80:** 639-650.

刺絲胞動物可以告訴我們多少有關 *Pax* 基因之演化

David J. Miller[1]

Pax 基因是一個由果蠅成對體節基因中首先被確認的基因區，它們通常包含了一個或部分的同源區以及成對區的轉錄因子遺傳密碼。這一族的基因在動物發育上扮演著主要的角色；然而，由於 *Pax* 基因族的複雜性和其功能的多樣性 — 已知人類有九個，而果蠅有八個 *Pax* 基因，混淆了我們對於不同生物間此基因族功能同源性的確認。最近獲得的幾個刺絲胞動物 *Pax* 基因資料，應有助於闡明 *Pax* 基因家族複雜的演化歷史。這篇回顧文章利用了刺絲胞動物的資料以及現今大量的線蟲基因組資料，提出一個解釋各種 *Pax* 基因演化的模型。

關鍵詞：發育，成對，成對箱基因，演化，刺絲胞動物。

[1]Department of Biochemistry and Molecular Biology, James Cook University, Townsville, Queensland 4811, Australia