5

Paired-like genes in the cnidarian, *Acropora millepora*

5.1. INTRODUCTION

Some of the key regulatory genes involved in early head organiser activity or nervous system patterning belong to the Paired-type homeobox superclass (Galliot and Miller, 2000). Such genes include members of the *Arx* gene family, many of which are involved in head and brain development during embryogenesis in bilateral animals, and some of which have roles that appear to have been conserved between flies and vertebrates. For example, the *Drosophila* gene, *aristaless (al),* is involved in the ontogeny of specific head segments and tail organs, and the initiation of appendage development (Campbell et al., 1993; Schneitz et al., 1993). Related *Arx* genes in the mouse and zebrafish are expressed in specific regions of the vertebrate diencephalon and floor plate; a second expression zone is also present in the presumptive cerebral cortex of the mouse (Miura et al., 1997).

The *otd*/*Otx* family also includes *paired*-class genes that share functional roles in vertebrates and invertebrates. Both *Otx* orthologs found in the mouse (Boncinelli et al., 1993), and frog (Blitz and Cho, 1995), as well as the three *Otx* genes in zebrafish (Mori et al., 1994) are required for the segmental division of the developing head region. In the fly, *orthodenticle (otd)* is a head gap gene, which functions to define the eye, antenna and parts of the embryonic brain (Finkelstein and Perrimon, 1990; Royet and Finkelstein, 1995).

5.1.1. The Paired-type superclass

The Paired-type (Prd-type) homeodomain superclass is distinct from the Antennapedia superclass and is characterised by six invariant amino acids in the homeodomain (P_{26}, D_{27}, E_{32}, R_{44}, Q_{46} and A_{54}) that distinguish it from other classes (Schneitz et al., 1993). Within the Prd-type superclass, the amino acid residue at homeodomain position 50 is a

key determinant of DNA-binding specificity (Pellizzari et al., 1997; Treisman et al., 1989). The Prd-type homeodomains present in some Pax proteins, in addition to the ubiquitous presence of a PAIRED domain, are characterised by the presence of a serine residue at position 50; S_{50} homeodomains occur only in Pax proteins, and homeodomains of this type bind cooperatively as dimers to P2-sites (TAAT homeodomain half-sites that are separated by two spacer amino acids). Q_{50}-paired type or K_{50} Prd-like homeodomains (those with glutamine or lysine residue at position 50) on the other hand generally prefer P3 sites (three spacer nucleotides). Note that these are generalisations, and differences are apparent in the DNA-binding behaviour of specific homeodomain types. Although the nature of the amino acid at position 50 can influence DNA-binding specificity, it does not appear to be a good phylogenetic character. Previous phylogenetic analyses of homeodomain sequences identified approximately 12 classes within the Prd-type superclass, but the K_{50} and S_{50} classes did not form monophyletic groups, suggesting multiple origins for these groups (Figure 5.1; Galliot et al., 1999). However, Q_{50} homeodomains related to that encoded by the *aristaless* gene consistently formed the basal clade in phylogenetic analyses, implying that this is the ancestral state of the Paired-type superclass.

In addition to the homeodomain, Prd-class genes also encode several structural features that are shared with other homeobox gene classes. The octapeptide sequence, a mediator of transcriptional repression in triploblasts (Mailhos et al., 1998; Smith and Jaynes, 1996), is found N-terminal of the homeodomain but is not specific to the Prd-class nor is it a defining feature (reviewed in Galliot et al., 1999). This structural motif, related to the eh1/GEH domain (Hemmati-Brivanlou et al., 1991) is also found in non-Prd classes including, for example, members of the Msx, Engrailed and NK-2 families.

Many Q_{50} Prd-like homeoproteins also possess an OAR (*otp, aristaless, Rx*) domain (Furukawa et al., 1997; Miura et al., 1997; Simeone et al., 1994), also termed the C-peptide (Mathers et al., 1997), in the carboxyl-terminal region. To date this motif has been found only in Prd-class homeoproteins. The OAR motif has been shown to function as a transcriptional activation domain in the case of Otp (Simeone et al., 1994); its restricted distribution but occurrence in a number of distinct Prd-type homeodomains suggests that it may have been present in the ancestral Prd-type protein, and is consistent with the monophyly of Prd-class homeoproteins.

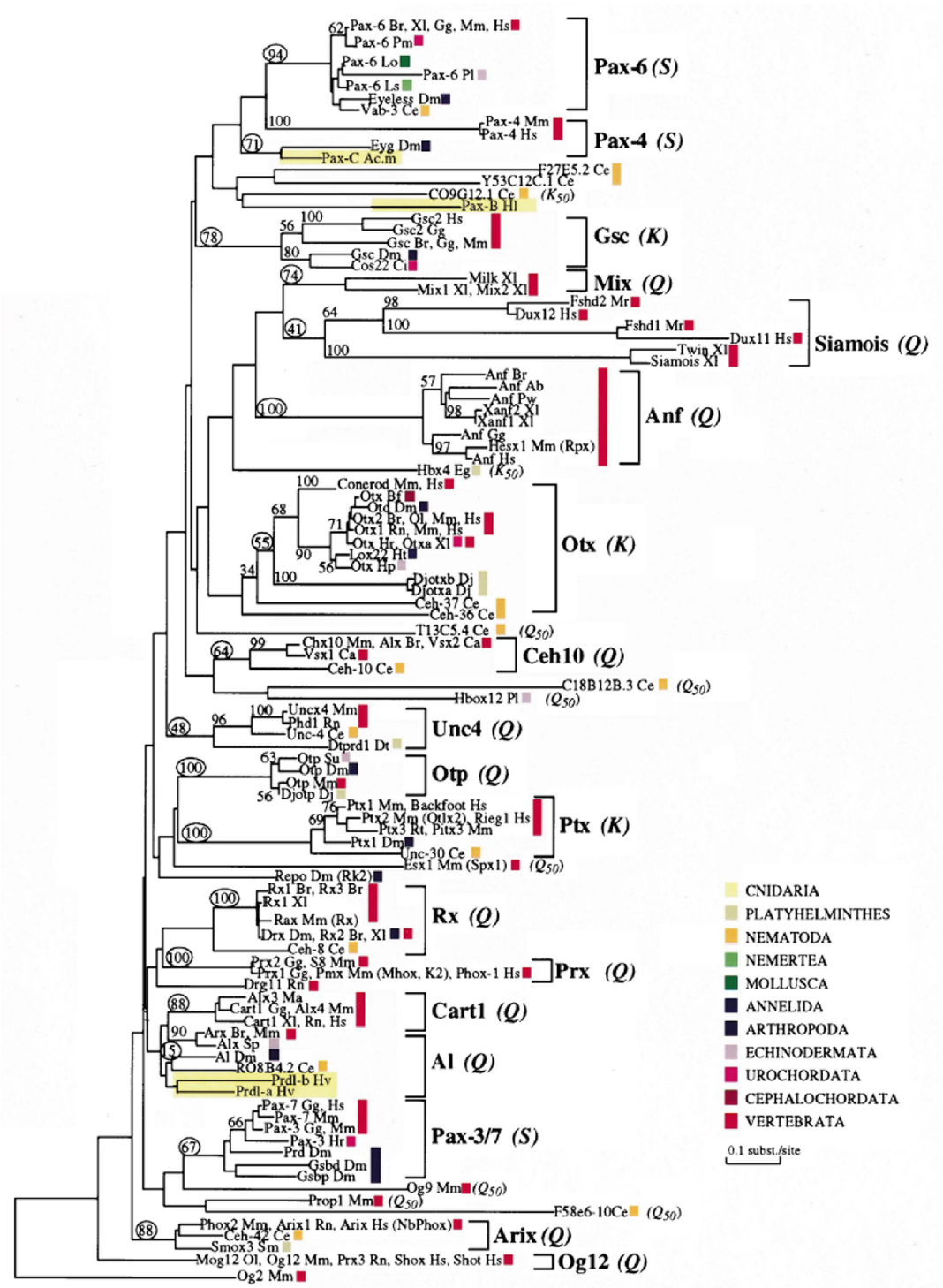


Figure 5.1: Phylogenetic relationships of the homeodomain region between 146 Prd-class genes, using the Neighbour Joining method with a PAM-Dayhoff matrix. The Og-2 sequence was chosen as outgroup. Percentages of 200 bootstrap replicates supporting the topology shown are indicated; those circled correspond to nodes defining families. The residue at position 50 is indicated in brackets (Q, S, K). (Source: Galliot et al., 1999).

5.1.2. Paired-class genes in the Cnidaria

Of the approximately twelve Prd-type classes, representative of at least four classes have been identified in cnidarians (refer to Table 1.1). The two Pax classes are represented by *PaxB*, *PaxC* (Pax4/6-type) and *PaxD* (Pax3/7-types) (Reece-Hoyes, 2001; Miller et al., 2000). Representatives of the Arx class include *HyAlx*, *prdl-a* and *prdl-b* from *Hydra* (Smith et al., 2000; Gauchat et al., 1998), while the cnidarian Otx class includes *CnOtx* and *manacle* in *Hydra* (Smith et al., 1999; Bridge et al., 2000) and the jellyfish *PcOtx* (Muller et al., 1999). Of all these cnidarian genes only *Hydra prdl-a* appears to have a conserved role in anterior patterning. Its expression in the apical region of the *Hydra* head resembles the role of related genes in the developing head and brain (Gauchat et al., 1998). *HyAlx* is expressed in a band of ectodermal epithelial cells at the base of each tentacle where it is specifically involved in tentacle formation, but is not involved in the formation of the head as a whole (Smith et al., 2000a). *Hydra prdl-b* was detected only in nematoblast cells in the body column (Gauchat et al., 1998).

Cnidarian orthologs of vertebrate *Otx* genes in *Hydra* (*CnOtx*) and the jellyfish *Podocoryne carnea* (*PcOtx*) are highly conserved at the sequence level, but the head-specific role of the Otx gene family appears to have evolved much later. Jellyfish Otx is expressed only during the asexual medusa budding process in the striated muscle, and is not involved in the development of head-like structures in the polyp stage (Müller et al., 1999). Similarly, the Otx ortholog in *Hydra*, *CnOtx*, is expressed in the prechordal mesoderm in developing buds and throughout the body column, but does not appear to have a direct association with anterior patterning (Smith et al., 1999). *Manacle*, another K_{50} *paired*-like gene in *Hydra*, which is expressed in the differentiating basal disk ectoderm in the adult animal, also lacks any association with an anterior patterning role (Bridge et al., 2000). No cnidarian gene of the *Otx* class appears to play a role in anterior development, suggesting the head-specific role of this gene family evolved after the Cnidaria separated from the higher Metazoa.

5.1.3. Statement of Goals

The decision to clone an ortholog of *prdl-a* from *A. millepora* was based on data from *Hydra*, which suggested conservation of function of at least some *aristaless*-related genes might extend to the Cnidaria, and that this gene might play a key role in anterior patterning throughout the Metazoa. The fact that embryonic expression data are not

available for *Hydra prdl-a*, and the derived nature of *Hydra* within the Cnidaria, led to our interest in cloning related *Acropora* genes.

Attempts to clone the *Acropora prdl-a* ortholog led to the isolation of a total of four *paired*-like genes, which were characterised in terms of genomic structure and patterns of expression. The main aim in characterising these *Acropora paired*-like genes was to investigate their role during early embryonic development of the coral. It was hoped that by studying these genes in a member of the basal cnidarian class, the Anthozoa, it might be possible gain insights into likely ancestral functions of specific *paired*-type genes, and perhaps to the superclass overall.

5.2. RESULTS

5.2.1. Isolation of the *paired*-like genes

The search for an *Acropora* ortholog of *prdl-a* first began using redundant primer pairs designed to amplify the homeobox region of *paired*-like genes. Using *A. millepora* genomic DNA as a template, these primers amplified two different size PCR products of 700 bp and 850 bp. Sequence from these PCR products indicated they each contained distinct *paired*-like homeoboxes, interrupted by introns of different sizes. Each amplicon was then used as a probe to screen 50,000 plaques of an *A. millepora* genomic library, and pure plaque populations corresponding to each probe were isolated; λPC2.3 and λPC1.1. The 'primer-walking' strategy was employed to sequence each genomic clone insert, starting at the PCR-amplified region and walking outwards towards the vector.

Sequence from genomic clone λPC1.1 revealed that this homeobox most closely resembled that of the *Arx* gene family and a corresponding cDNA, *arx-Am*, was later isolated from a pre-settlement stage cDNA library. During the same cDNA library screen, cross-hybridisation resulted in the isolation of another *Arx*-related homeobox gene, *hbn-Am*, which closely resembled the *Drosophila homeobrain* gene.

Sequence from the λPC2.3 genomic clone indicated this clone contained not one, but two different homeobox genes in close proximity, which both closely resembled

members of the *Otx* gene family. Two incomplete cDNAs corresponding to each of the genes were isolated from a pre-settlement stage cDNA library.

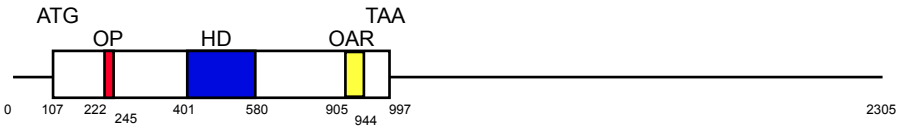
Thus, in total, four homeobox genes of the *paired*-class were isolated from the coral, *A. millepora*. The isolation and preliminary characterisation of these four genes is described below.

5.2.2. *The hbn-Am* cDNA

The *hbn-Am* cDNA clone was fortuitously isolated from the λZAP-II pre-settlement stage cDNA *A. millepora* library by Ms. Dalma Seboek, an Honours student under my supervision, during a screen for the *arx-Am* cDNA clone (see section 5.2.3). A 350 bp PCR product was amplified from the λPC1.1 genomic clone and corresponded to the 5'-end of the *arx-Am* coding region, including 138 bp of the homeobox. This PCR amplicon was used as a probe to screen 50,000 plaques under conditions of low stringency. As well as the *arx-am* cDNA, an additional two identical clones were also isolated and sequence comparisons with the BlastX database (Altschul et al., 1990) revealed that they most closely resembled the *Drosophila homeobrain* gene. One clone was selected and is herein referred to as *hbn-Am*.

The *hbn-Am* cDNA is 2305 bp, comprising 888 bp of protein coding sequence, 106 bp of 5' UTR and 1311 bp of 3' UTR. Figure 5.2 shows the nucleotide sequence and conceptual protein translation of the *hbn-Am* cDNA. The predicted protein sequence indicates this gene encodes a protein of 297 amino acids with a molecular weight of approximately 33 kDa. From a conceptual protein translation of the *hbn-Am* cDNA, the initiating codon was predicted to be located at base pairs 107-109. A homeodomain is located at amino acids 100 to 159 in the N-terminal half of the protein. An octapeptide (Smith and Jaynes, 1996) is also present in the Hbn-Am protein, beginning 40 residues from the aminoterminus and separated from the homeodomain by 52 residues. An OAR domain (Simeone et al., 1994), a motif that is restricted to the *Prd*-class, is located 109 amino acids C-terminal of the homeodomain, and differs in only four positions from the consensus (see Figure 5.3C).

A



B

```
1    ATTCGGCACGAGGGACGGAGCAACTTCTAGATCTAAGTTGGTGAACATCAGCTTTGTCTCTTTTTTTTCTTGTTGTTTTCCTTACCAACCCAGAGGATTT

          M  M  E  F  K  A  R  S  K  M  L  M  S  E  P  R  E  K  G  E  R  I  R  G  D  H  S  I  E  S  A      31
101  GGTTTGATGATGGAATTCAAAGCACGTTCAAAGATGTTGATGTCGGAGCCAAGAGAAAAAGGCGAGAGAATTAGAGGGGACCACTCCATAGAATCCGCGT

       L  I  Q  A  T  S  K  G  Y  T  I  D  M  I  L  G  N  H  D  R  K  P  R  Q  G  R  S  L  E  Q  D  E  K  T     65
201  TGATTCAAGCGACGAGTAAAGGTTACACAATCGATATGATTCTTGGAAATCACGATCGAAAGCCAAGGCAAGGACGATCACTGGAACAAGATGAGAAAAC

          S  G  N  L  S  D  N  E  D  E  L  S  V  G  D  E  R  G  S  K  D  C  D  S  I  D  E  N  D  G  K  P  R     98
301  TTCTGGAAATCTCAGTGACAACGAAGACGAGTTAAGCGTGGGCGATGAGAGGGGTAGTAAAGACTGTGACAGCATTGACGAAAACGATGGCAAACCGAGA

       K  M  R  R  S  R  T  T  F  T  T  Y  Q  L  H  Q  L  E  Q  T  F  E  K  T  Q  Y  P  D  V  F  T  R  E    131
401  AAAATGCGACGAAGTAGAACGACCTTCACCACGTACCAGCTTCACCAACTGGAGCAAACCTTTGAGAAGACTCAATATCCGGATGTGTTCACAAGGGAGG
                                                  hbn.F
     E  L  A  L  R  L  D  L  S  E  A  R  V  Q  V  W  F  Q  N  R  R  A  K  W  R  K  R  D  K  S  M  G  R  E    165
501  AACTTGCCTTGAGACTAGATCTCAGTGAAGCGAGAGTTCAGGTATGGTTTCAAAACAGAAGAGCCAAATGGAGGAAGCGTGACAAGTCAATGGGGCGGGA
                                                  hbn.R
          S  P  T  Y  L  Y  G  Q  S  N  D  Y  R  A  L  S  A  D  L  P  H  A  H  S  A  T  F  A  A  G  M  E  R    198
601  AAGCCCTACATACCTGTATGGCCAATCAAATGACTACAGAGCGCTCTCCGCAGATCTGCCGCATGCGCACTCCGCAACATTTGCTGCAGGAATGGAAAGA

       W  H  P  H  V  P  A  L  S  L  L  S  P  Y  A  T  A  P  S  P  Y  S  P  I  G  G  H  L  A  G  I  S  P    231
701  TGGCACCCTCACGTGCCAGCCCTTAGCCTTTTGTCGCCATACGCAACAGCACCGAGTCCATATAGTCCCATCGGGGGCCATCTCGCAGGGATCTCTCCGG

       A  A  M  Y  T  H  H  L  Y  R  Q  Y  A  R  P  T  M  A  P  F  L  K  G  P  L  S  P  N  D  P  Q  E  F  R    265
801  CAGCAATGTACACGCACCATTTGTATAGGCAGTACGCAAGACCGACCATGGCACCCTTTTTGAAGGGCCCTCTTAGTCCCAACGACCCACAGGAATTCCG

          R  S  S  I  E  N  L  R  Y  K  A  K  Q  H  S  S  S  S  S  L  A  G  D  G  S  K  T  A  E  L  I  *      297
901  TAGGTCGAGTATTGAAAATCTGAGATACAAAGCTAAGCAGCACAGCTCGTCAAGCTCACTGGCCGGTGATGGAAGCAAAACTGCTGAATTAATTTAACGA

1001 ACTTGAACGTAGGAAAGCATCGGGTATTCTCACTACAAAGATATTCACACTAAGTAGCCATGTTCTTTAACTGACACGAAGAATGCTTTTCGCATAAATA
1101 GAGATCACTTCCTGAAGGATAACATGGTCTTTTTCACTTTTATGGACGTTTTGATGCCTATGAAAATAGTCTATTAAATTGTAGTTAATCAACTCTCGAA
1201 AAACATCCATTCTGTGCCTTGTTGAAGAAAGGTTTTGCATCAGGAACTGAGTATAAGAGGGGAAAACTTGACTGTCTCAAGTGAATTAGGGAAAAGAAGA
1301 CGAAAAAGCATCACTTAGCTTTCCACAGAGGAGACTTTTTACACGCCTTATACTATTGAAATAGCATTTTGGAAAACCTCTTTTTAAGAGATGAGAAGGT
1401 ACGCTGCAGCTTATTGAACTCGATTGAAATTAGCATACTGTGAACTAGTTTCCTAACTAAAATTACGATGCAAATGTCAAAATATACGTATGCTACGATT
1501 GCGCACGATACCAACACAATAATCTTCTTTCCAGAAACAATTGTATAAGATAGTTTATTTTTTCTAAAGAAATTGCTAAATCTGCAACGCAATGTGCCAT
1601 AAAAGAGCTTCATCAAATCTAAAATACAGAAATATCCATTACTGCGCAATGCTAATTAATGATCAATAGGTATTCATGACATATTTTTGTACGCCGAGAA
1701 GTAAACTCTTCGACCTTCAAGGAATATATAGCATTTTAGTTAGCTTATTCAGTCCTAAAACGGTGTACATAGTTGTTAACTTCGACAAACGTGGCCCATG
1801 GATTTCGGGCAAGCATGTAATTTATTATTGAAAAGTGTGCCTTTTACATCAATAAGGATCTGAAAATATATTTAATAGAAATTATTCTTATTTATAAGCG
1901 TGGCTGTACGCTGTGTGCACCCAACGCGGCACTATAGGCTAGTGTAAAAAAGGACAAGTGCTTCGATATACTTGCGTAACTGCCGTTCGAAGGCGGAGAA
2001 CACGTGTTCTGGCTATTAGAGTGACAAGAAAACAGTAGGTTAGTAGAGCATGCGCACATTGTCCAAAGTTATTCAAATCGATGCAGTTGTCGCAAAGACG
2101 TTCGTATTCAAATGAGCTCCTACCGAGAATTAGAAACATTTGCCGAAGCAGCTATAGTTTTCACTTGTGTTGAGTTAGTAGATGATAACATTTTGGTGCT
2201 TTGTTGATCATGAAGTTACCTTCCATAGCTTATAGTAAGCTGAGCCAAGGTAAATAAAGTATAAATATTAAAAAAAAAAAGCCTGCAAAAAAAAAAAAAA
2301 AAAAA
```

Figure 5.2: The *hbn-Am* cDNA. (A) A schematic representation of the *hbn-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region and the homeodomain region is indicated by the blue box. The octapeptide motif is shown by a red box and the OAR domain is indicated by a yellow box. The start (ATG) and stop (TAA) codons are shown at the top. The numbers beneath the box represent the nucleotide sequence. (B) The nucleotide sequence of the *hbn-Am* cDNA and predicted amino acid sequence of the protein. The homeodomain residues are highlighted in blue, the octapeptide is highlighted in red and the OAR domain is highlighted in yellow. Numbers on the left side represent the nucleotide sequence; numbers on the right side represent the amino acid sequence. Primers (hbn.F and hbn.R) used in the real-time quantitative PCR are shown by arrows beneath the nucleotide sequence.

The Hbn-Am homeodomain is a member of the Paired-like gene class (see Galliot et al., 1999) for review) as it possesses the six invariant amino acid residues characteristic of this class. It most closely resembles the homeodomain of Homeobrain in *Drosophila* (Walldorf et al., 2000) and HyAlx of *Hydra vulgaris* (Smith et al., 2000a) (> 88% identity; see Figure 5.3A). Unlike the Prd-type homeodomains, which possess a serine at position 50, *hbn-Am* and its related genes encode homeodomains that possess a glutamine at the corresponding position. The Hbn-Am octapeptide/GEH domain is most similar to that in *Drosophila* Gsbp (75% identity; see Figure 5.3B). Unlike *Drosophila* Hbn, Hbn-Am contains an OAR domain that differs in only four positions to the consensus sequence (>69% identity; see Figure 5.3C).

5.2.2.1. Temporal expression patterns of *hbn-Am*

The temporal expression of *hbn-Am* was analysed by probing a developmental northern blot containing mRNA from various coral developmental stages and also by quantitative PCR. For northern analyses, the radioactively labelled probe was a 1475 bp fragment excised from the plasmid using *Spe*I, containing the entire coding region and 478 bp of 3'- UTR. Hybridisation revealed three different size transcripts of approximately 2.2 kb, 1.9 kb and 1.7 kb (Figure 5.4C). The 2.2 kb transcript was observed at all developmental stages examined. The 1.9 kb transcript was present in the prawn chip stage of development but was not detected after the blastopore was closed. The smallest transcript (1.7 kb) was only seen in the egg stage. The various sizes of transcripts may be due to cross-hybridisation from another closely related *paired*-like gene, although high stringency conditions were employed. Alternatively, differences in the size of these transcripts may be due to different processing events following transcription, and/or alternative transcription start sites. Sequence from the *hbn-Am* genomic locus may shed more light on this question.

Temporal expression of *hbn-Am* throughout embryonic and larval development was also examined using real-time RT-PCR on an ABI7700 Sequence Detection System (Applied Biosystems). Primers used to generate a 126 bp amplicon were designed across a region that spans a conserved intron position in other *paired*-like genes, but as no genomic data were available this could not be confirmed for *hbn-Am*. The location of the primers (hbn.F and hbn.R) used in this analysis are indicated in Figure 5.2.

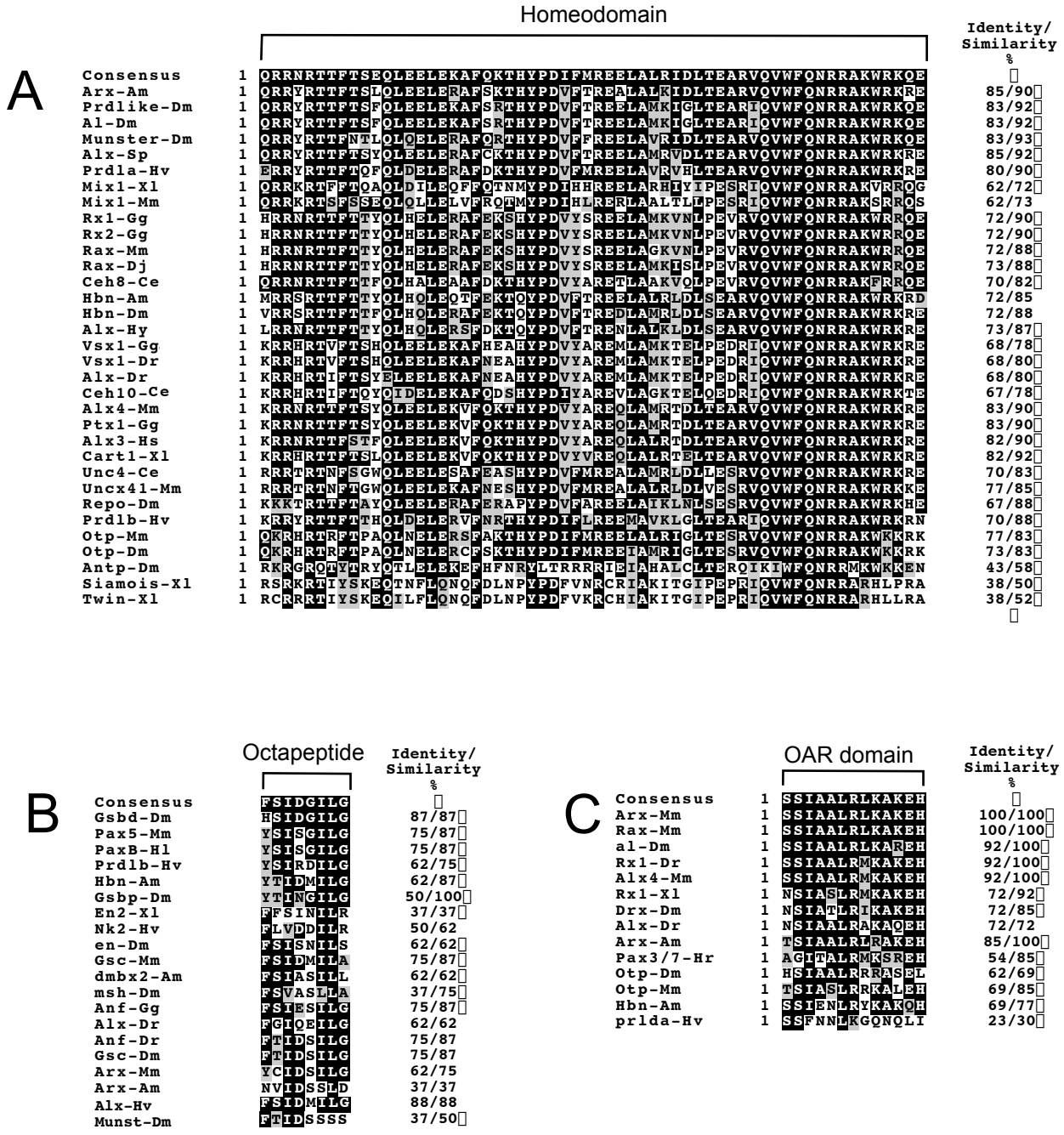


Figure 5.3: A comparison of the homeodomains of *Acropora* Paired-like proteins with other Prd-class family members. (A) Aligned (Clustal W; Thompson et al., 1994) amino acid sequences of the Acropora paired-like homeodomains with that of other Prd-class family members, boxshaded (Boxshade server) to indicate residues shared with the consensus homeodomain sequence. Identical residues are shaded darkly; conserved substitutions are lightly shaded. The column to the right of the alignment indicates the overall identity and similarity of each homeodomain with the consensus sequence. cont/d...

Figure 5.3 (cont/d): A comparison of the homeodomains of *Acropora* Paired-like proteins with other Prd-class family members. (B) Alignment (Clustal W) of the amino acid sequences of the octapeptide motifs of the Prd-class family, boxshaded (Boxshade server) to indicate residues shared with the consensus octapeptide sequence. Identical residues are shaded darkly; conserved substitutions are lightly shaded. (C) Alignment (Clustal W) of the amino acid sequences of the OAR domain motifs of the Prd-class family, boxshaded (Boxshade server) to indicate residues shared with the consensus OAR domain sequence. Identical residues are shaded darkly; conserved substitutions are lightly shaded. The species names (and Genbank Accession No.s) are abbreviated as follows: Am, Acropora millepora; Ce, Caenorhabditis elegans (ceh-10: T34470, unc4: P29506); Dm, Drosophila melanogaster (Prdlike: AAA28840, Al: XP_079165, Munster: NP_477330, Hbn: NP_524903, Repo A54282, Otp: NP_523799, P02833, Gsbd: P09082, Gsbp: P09083, En: O02491, msh: s55392, Gsc: P54366, Drx: CAA11241); Dj, Dugesia japonica (Rax: O97039); Dr, Danio rerio (Vsx1: AAB71611, Alx: AAB66714, Anf: AAB88392, Rx1: O42356); Gg, Gallus gallus (Rx1: Q9PVY0, Rx2: Q9PVX0, Vsx1: Q9IAL2, Ptx1: AAC61772, Anf: P79775); Hl, Hydra littoralis (PaxB: AAB58291); Hr, Halocynthia roretzi (Pax-3/7: BAA12289); Hs, Homo sapiens (Alx: XP_002147); Hv, Hydra vulgaris (prdla: CAA75668, prdla: CAA75669, Nk2: AAB67611); Mm, Mus musculus (Mix1: NP_038757, Rax: AAC53129, Alx4: O35137, Uncx41: CAB09537, Otp: O09113, Pax5: Q02650, Gsc: NP_034481, Arx: O35085); Sp, Strongylocentrotus purpuratus (Alx: Q26657); Xl, Xenopus laevis (Mix1: A32548, Cart1: Q91574, Siamois: A56219, Twin: AAC60331, En2: S19004, Rx1: O42201).

Levels of expression were quantified relative to *integrin-β*, which is expressed at a uniform level throughout development (Hayward et al., unpublished). Developmental stages used in the quantitative PCR analysis were selected based on morphology rather than time into development, as rates of development have been observed to vary with temperature. Figure 5.4A summarises *hbn-Am* expression levels relative to *integrin-β*. The raw data used to construct this histogram are given in Table 5.1. Each time point is the mean of at least two independent assays; analyses were performed in triplicate but outlying data points were excluded from the analysis, as recommended by ABI SDS Compendium 7700 – Version 3.0 (Applied Biosystems).

hbn-Am expression was first detected in the egg (approximately three-fold higher than that of *integrin-β*). At the prawn chip stage, the expression level of *hbn-Am* was increased by more than 300%. Expression levels continued to increase as the cellular bilayer of the prawn chip stage begin to contract and fold inwards, forming the gastral pore. At the donut stage, transcript levels were four times greater than those observed in the egg, and were maintained until the outer cell layer was completely internalised to form the endoderm. After the gastral pore closed and the endoderm and ectoderm were distinct, *hbn-Am* transcript levels dropped below that in the egg. This level of expression was maintained in the remaining developmental stages examined.

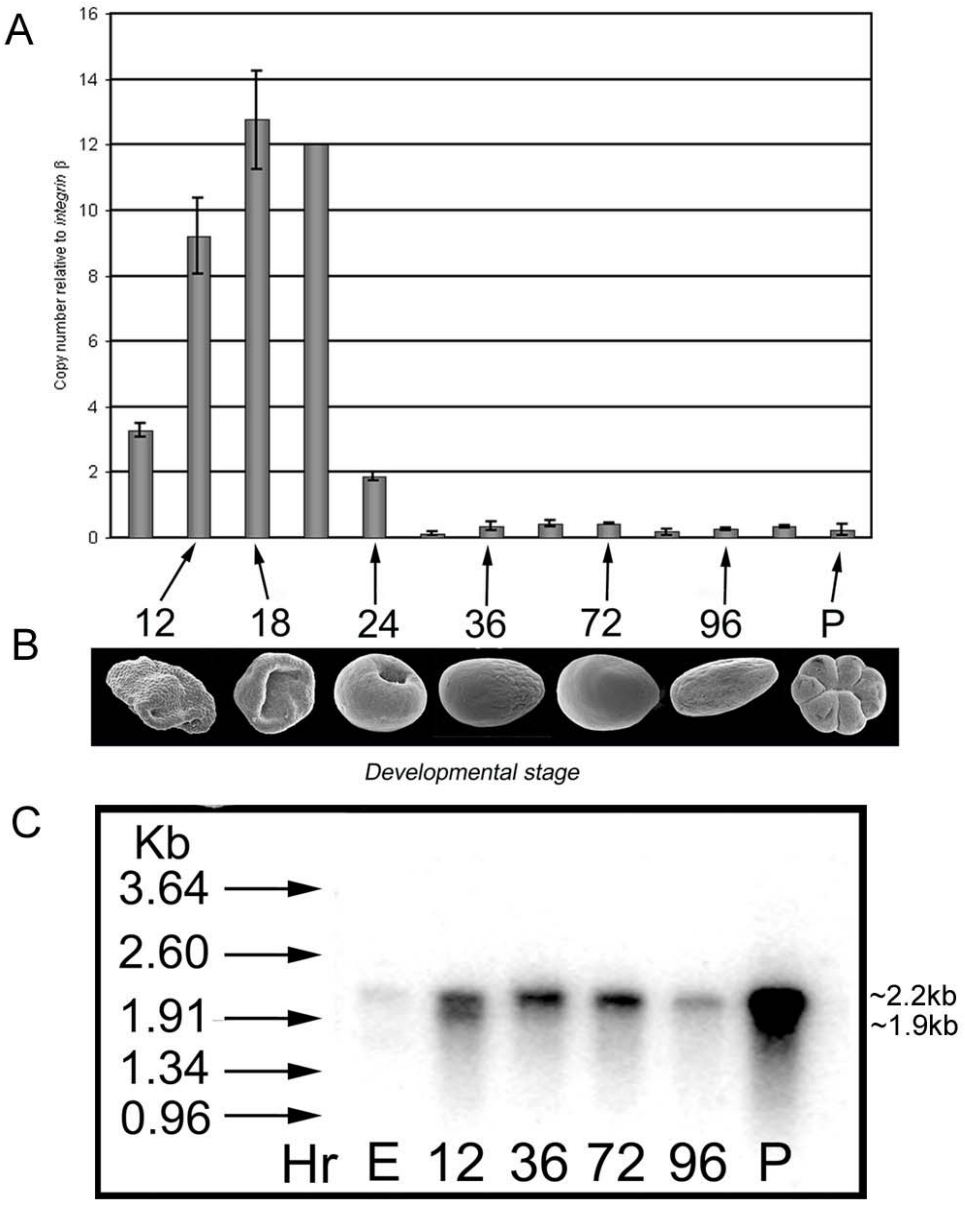


Figure 5.4: Temporal patterns of *hbn-Am* expression during development.
(A) Graphical representation of *hbn-Am* expression levels during development, expressed relative to the levels of the ubiquitously expressed *integrin*-β. Numbers on the y-axis indicate the copy number of hbn-Am relative to the copy number of *integrin*-β. Thirteen stages were examined. (B) Electron micrographs depicting the relative morphologies of coral embryos and larvae from the same stages of development used in the real-time PCR analysis and/or mRNA extraction for northern blotting (not to scale). (C) Hybridisation pattern observed after probing a northern blot of mRNA extracted from *Acropora millepora* fertilised eggs (E) and embryos of various developmental stages (numbers indicate hours post fertilisation) with a radioactively labeled fragment of the *hbn-Am* cDNA. Equal amounts of RNA, as measured spectrophotometrically, were loaded in each lane. However, differing amounts of contaminating ribosomal RNA meant that the relative amounts of messenger RNA loaded per lane were E (egg)=1, 12 h=2, 36h=2, 72h=2, 96 h=1, P (post-settlement)=4.

Table 5.1. Temporal expression patterns of *hbn-Am* during development

Numerical data from RT-PCR of *hbn-Am* expression. The morphological appearance of each of developmental stages is described in the column one. The mRNA copy number for both *hbn-Am* and *integrin-β* for each stage are listed with an average figure and standard deviation. Levels of *hbn-Am* expression relative to *integrin-β* were determined in the final column and used to construct the graph in Fig 5.4A

Developmental Stage	*hbn-Am* mRNA copy number	*integrin-b* mRNA copy number	*hbn-Am* Normalised to *integrin-b*
Egg	5.5	1.7	
	6.0	1.8	
	-	-	
Average	5.75±0.353	1.75±0.071	3.285±0.207
Prawn Chip	5.5	0.59	
	6.7	0.79	
	-	0.61	
Average	6.1±1.15	0.663±0.110	9.20±1.158
Donut	20	1.5	
	17	1.4	
	-	-	
Average	18.5±2.12	1.45±0.071	12.759±1.493
Post Donut	3.9	0.33	
	3.9	0.32	
	-	-	
Average	3.9±0.0	0.325±0.007	12.000±0.005
Sphere with deep pore	12	7.0	
	11	5.9	
	-	5.5	
Average	11.5±0.707	6.13±0.777	1.876±0.145
Sphere with pore closing	0.25	2.8	
	0.46	2.7	
	-	-	
Average	0.355±0.148	2.75±0.071	0.129±0.054
Pre-Pear 1	1.7	4.4	
	1.1	5.2	
	2.3	-	
Average	1.7±0.6	4.8±0.566	0.354±0.129
Pre-Pear 2	5.4	15	
	6.3	16	
	8.4	16	
Average	6.7±1.539	15.6±0.577	0.429±0.099
Pears	15	36	
	16	37	
	-	37	
Average	15.5±0.707	36.6±0.577	0.423±0.019
Late Pear	2.6	12	
	1.3	10	
	-	-	
Average	1.95±0.919	11±1.414	0.177±0.086
Planulae	3.3	14	
	3.8	15	
	-	11	
Average	3.55±0.353	13.3±2.08	0.267±0.033
Pre-settlement	6.2	15	
	5.6	18	
	-	18	
Average	5.9±0.424	17±1.732	0.347±0.028
Post-settlement	0.12	1.1	
	0.38	0.98	
	-	-	
Average	0.25±0.183	1.04±0.084	0.240±0.177

5.2.2.2. Spatial patterns of expression

The spatial distribution pattern of *hbn-Am* expression was examined at early developmental stages, corresponding to quantitative PCR data, using *in situ* hybridisation (Figure 5.5). A DIG-labelled riboprobe, generated from linearised plasmid containing the complete *hbn-Am* cDNA sequence, was used in the hybridisation reaction. Note that the in situ results are not completely consistent with temporal data from northern analyses, in that the transcripts seem to be present for much more of development in the northern blotted material.

A detailed description of *Acropora* embryology is presented in Figure 1.6. A schematic representation of the morphological changes that occur during gastrulation is presented in Figure 1.7. Expression of *hbn-Am* is initially detected at the late prawn chip stage (15 h into development) as the edges of the bilayer begin to condense and fold inwards (Figure 5.5A). This mottled pattern of expression is restricted to the ectoderm on the convex surface of the embryo. Expression becomes more intense and solid, maintaining a similar distribution pattern, and is excluded from an irregular circular zone surrounding the deepening gastral pore. The exclusion zone around the gastral pore decreases in size as the embryo takes on a more spherical shape and is most apparent in transverse sections of embryos (Figure 5.5 B'C'D'). At the late spherical stage, just before the gastral pore closes, *hbn-Am* expression is limited to the presumptive ectoderm, and is excluded from the presumptive endoderm. Once the gastral pore has closed and the endoderm and ectoderm are quite distinct, *hbn-Am* expression can no longer detected by in situ hybridisation.

Whilst in situ data are consistent with quantitative PCR data, it is difficult to correlate these with the results of northern analyses. Two different sized transcripts were identified in the northern blot analysis using the complete *hbn-Am* cDNA as a probe; the small transcript is detected only in the 12 h stage, while the larger transcript was ubiquitously expressed throughout development. Data from quantitative PCR and in situ hybridisation experiments agree only with the temporal expression of the 1.9 kb transcript seen in the northern. As mRNA samples from each of these stages were depleted, northern blot experiments could not be repeated, however the specific nature of the primers used in the quantitative PCR suggests the larger transcript seen in the northern analysis may be due to cross-hybridisation with another closely related

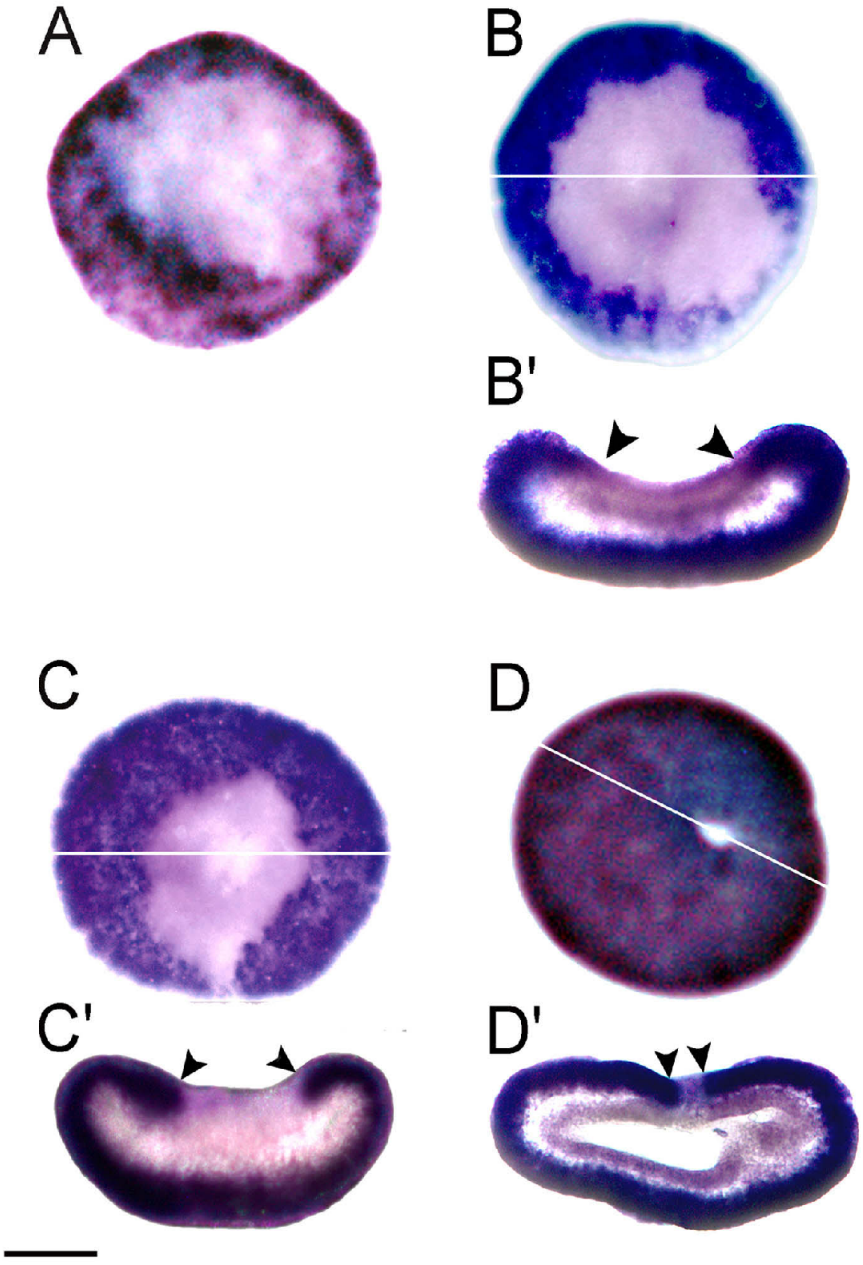


Figure 5.5: Spatial patterns of *hbn-Am* expression during development.
In situ hybridisation pattern observed when embryos of various developmental stages, between prawn-chip stage and late donut, were probed with a DIG-labeled *hbn-Am* probe. (A-D) Whole-mount embryos at various stages of development after fertilisation: (A) ~20 h; (B) ~30 h; (C) ~36 h; (D) ~48 h after fertilisation. The white line indicates the plane of section of embryos in A, B, C. Corresponding sectioned embryos are shown in A, B, C. All embryos are orientated such that the concave surface is facing upwards. In all sectioned embryos, arrowheads indicate the limit of *hbn-Am* expression in the presumptive ectoderm. (Scale bar = 100∝m).

homeobox gene. The same region of the *hbn-Am* cDNA used to probe the northern blot was also used for in situ hybridisation but similar temporal expression patterns not observed. Repeated northern blot experiments using only the 3'-UTR of the *hbn-Am* cDNA may provide more information, but require the availability of fresh material for mRNA extraction.

5.2.3. The *arx-Am* cDNA

The *arx-Am* cDNA clone was isolated from a λZAP-II *A. millepora* pre-settlement stage library. Under my supervision, Ms. Dalma Seboek screened the *A. millepora* pre-settlement stage cDNA library using a 350bp PCR probe amplified from the λPC1.1 genomic clone. The probe included 138bp of the homeobox region and 212 bp 5'- of the homeobox. Of 50,000 plaques screened, two positive clones were isolated that corresponded to the amplicon from the genomic clone. Both cDNA clones were identical and are herein referred to as *arx-Am*.

The nucleotide sequence and conceptual translation of the *arx-Am* cDNA are presented in Figure 5.6. The *arx-Am* cDNA is 1721 bp, comprising 790 bp of open reading frame and 931 bp of 3'- UTR. No in-frame methionine codon is present in the cDNA sequence 5'- of the homeobox, however an in-frame ATG codon is present in the genomic sequence, 32 bp 5'- of the end of the cDNA sequence. It is preceded by an in-frame stop codon (TAA) a further 18 bp upstream of the cDNA, suggesting that this is the initiating methionine residue. Conceptual translation of this open reading frame results in a putative protein, Arx-Am, of 274 amino acids.

5.2.4. The Arx-Am protein

The homeodomain in the Arx-Am protein is located in the N-terminal half of the protein, from amino acid residues 55 to 114. The Arx-Am homeodomain is clearly related to the Paired-class, as it encodes all six invariant residues diagnostic of this gene family. A glutamine residue at position 50 of the homeodomain also supports its classification in the Q_{50} Prd-like gene subclass. Overall, the Arx-Am homeodomain most closely resembles those encoded by *Drosophila aristaless*, *paired-like* and *munster*, and *Alx-Sp* in the sea urchin (> 83% identity; see Figure 5.3A).

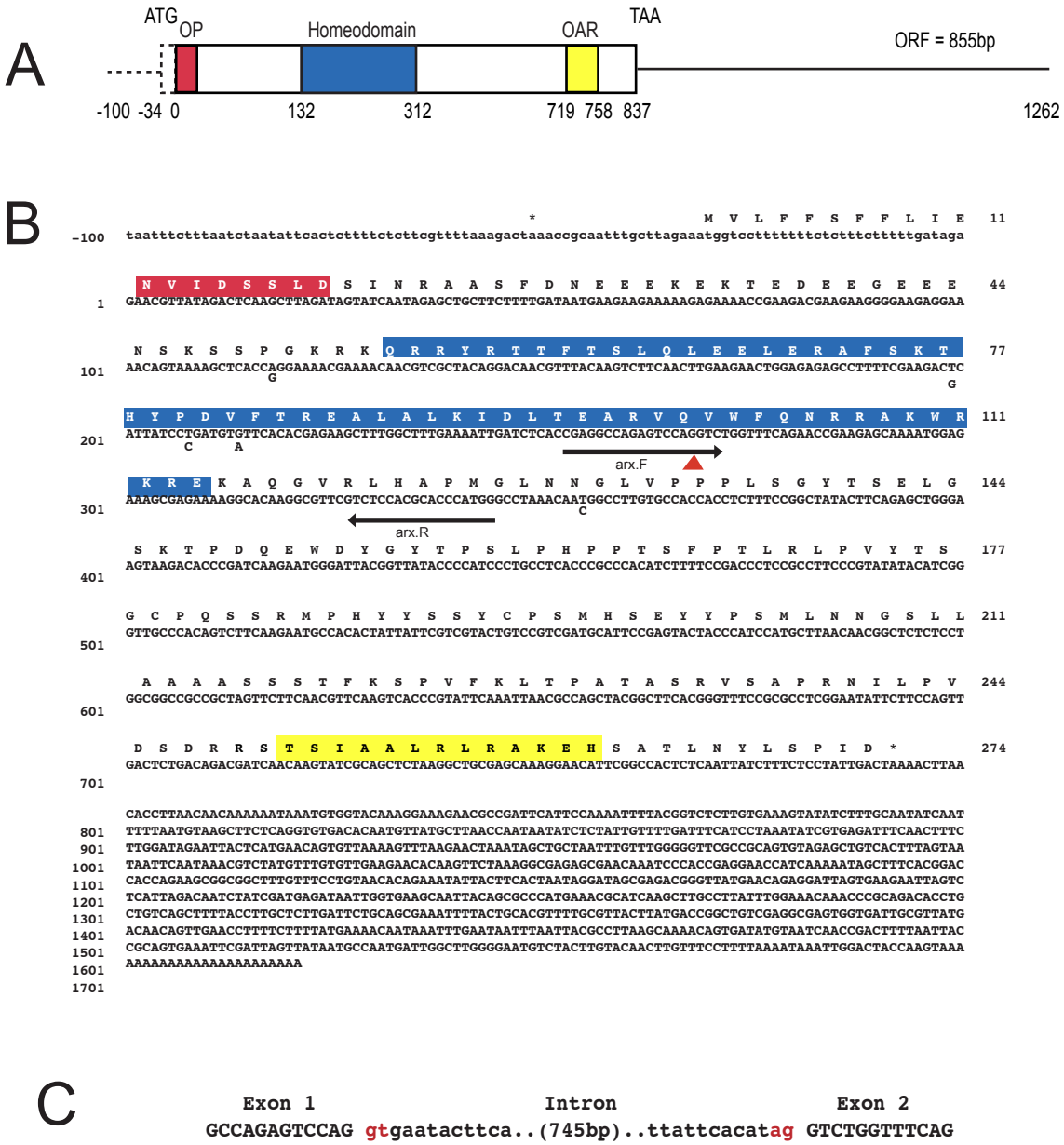


Figure 5.6: The *arx-Am* cDNA. (A) Schematic representation of the *arx-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region and the homeodomain region is indicated by the blue box. The octapeptide and OAR domains are indicated by red and yellow boxes respectively. Dashed box and line indicate the genomic sequence included to complete the ORF. The start (ATG) and stop (TAA) codons are shown at the top. The numbers beneath the box represent the nucleotide sequence. (B) The nucleotide sequence of the *arx-Am* cDNA and predicted amino acid sequence of the protein. The genomic sequence included to complete the ORF is in lowercase. Residues of the octapeptide, homeodomain and OAR domain are highlighted in red, blue and yellow respectively. An asterisk indicates the 3'- in-frame stop codon. The intron position is shown by a red triangle. Nucleotide polymorphisms between the cDNA and genomic sequence are indicated beneath each position in the cDNA sequence. Numbers on the left side represent the nucleotide sequence; numbers on the right side represent the amino acid sequence. Primers (arx.F and arx.R) used in the real-time quantitative PCR are shown by arrows beneath the nucleotide sequence. (C) The intron-exon boundary of the single intron in *arx-Am*. The GT-AG consensus splice sites are indicated in red. The total size of the intron is shown in brackets.

A putative octapeptide sequence, located 36 residues N-terminal of the homeodomain, shares >37% identity with the octapeptide consensus (FSIDGILG; see Figure 5.3B). An OAR domain is also present in Arx-Am, 10 residues from the carboxyl terminus, and closely matches to the consensus (TSIAALRLRAKEH; see Figure 5.3C).

5.2.5. The *arx-Am* genomic locus

The genomic structure of a gene is an important characteristic to examine in evolutionary comparisons. For this reason, the structure of the *arx-Am* genomic locus was determined.

A λGEM-12 genomic library, generated using high molecular weight DNA isolated from *A. millepora* sperm, was screened with an 850 bp PCR amplicon generated using degenerate primers (Prd-like HD.F2: 5' MRIACICAYTAYCCNGAYGTNTT 3'; Prd-like HD.R1: 5'CCAYTTIGCICKICKRTTYTGRAACCA – 3'), designed based on paired-like protein sequences from representative organisms. 50,000 plaques were screened using this probe and three positive clones were isolated. A single clone, λPC1.1 was selected for further analysis. Using the 'primer-walking' strategy, 10.4 kb of this genomic clone was sequenced including the *arx-Am* genomic locus, 3.8 kb of upstream sequence and 4.2 kb of 3' sequence. A BlastX (Altschul et al., 1990) database search of the sequence outside the *arx-Am* locus failed to find any ORF with identity to a previously characterised protein. The longest ORF identified was 573 bp. Sequencing of the genomic locus revealed that the *arx-Am* gene contains a single intron of 745 bp located between codons 46 and 47 (VQ-VWF) of the homeobox (Figure 5.7) and possesses donor and acceptor splice sites that conform to the consensus GT-AG sequence (Figure 5.6C). This is a conserved intron position in many *paired*-type homeobox genes (reviewed in Gauchat et al., 2000).

5.2.6. Temporal expression patterns of *arx-Am*

Expression of *arx-Am* was examined by probing a northern blot containing mRNA from coral embryos at various stages of development and also by quantitative PCR. For northern analyses, the radioactively labelled probe was a 1701 bp fragment excised from the plasmid using *Xho*I and *Xba*I, containing the complete *arx-Am* coding sequence including the 3'- UTR. Hybridisation revealed several *arx-Am* transcripts of

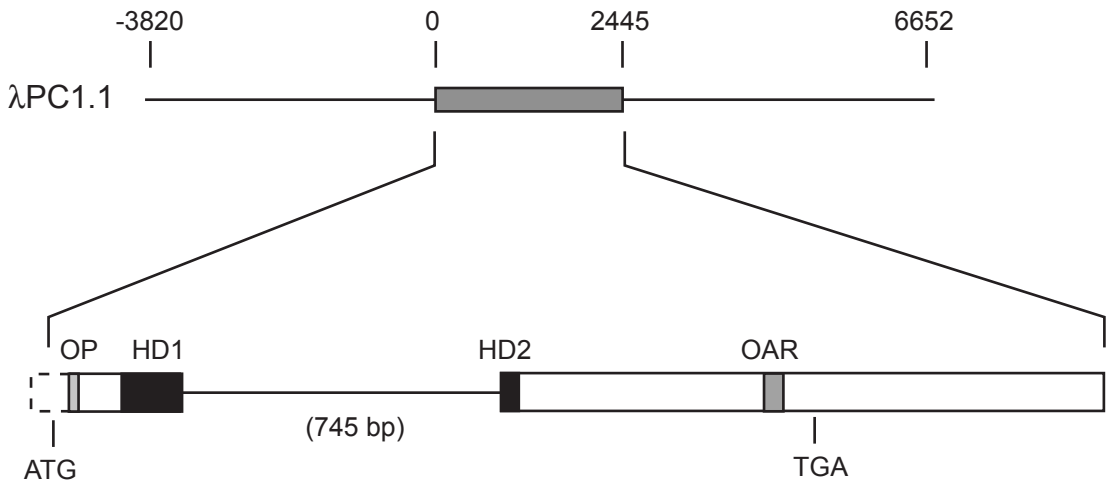


Figure 5.7: Genomic organisation of the *arx-Am* locus.
Schematic representation of the λPC1.1 genomic clone. The shaded boxes indicate the region of the genomic clone containing exons. Numbers refer to the nucleotide sequence of the genomic clone with zero defined as the 5'-end of exon 1. The expanded view shows the relative sizes of the intron and exons with start and stop codons marked and the homeodomain, ocatapeptide and OAR domain represented by a labeled boxes coloured blue, red and yellow respectively. (B) The intron-exon boundary of the single intron in *arx-Am*. The GT-AG consensus splice sites are indicated in red. The size of this intron is shown in brackets.

different sizes (Figure 5.8C), expressed at two different stages during early coral development when significant morphological change is occurring.

The first wave of *arx-Am* expression was detected in the cellular bilayer stage and transcripts at this stage were estimated to be approximately 2.3 kb. A smaller transcript of approximately 1.9 kb was present in the donut stage of development (24 h in Figure 5.8), after the bilayer has separated and the two tissue layers are first differentiated. The pear stage of development at 72 h was not included on this northern analysis, but a second wave of *arx-Am* expression was apparent in the 96 hr sample of pre-settlement planulae. This transcript was also approximately 1.9 kb. Expression of *arx-Am* was also detected in the post-settlement polyp sample but a clear band is not apparent and estimates of transcript size are difficult.

These transcripts of different sizes may represent alternative splice variants or be transcribed from different start sites. The 1.9 kb transcript most likely corresponds to the cDNA isolated, but the 2.3 kb transcript may initiate at an alternative first exon that was not apparent from the cDNA or genomic sequences.

The temporal expression of *arx-Am* throughout embryonic and larval development was also examined using real-time RT-PCR on an ABI 7700 Sequence Detection System (ABI Applied Biosystems). Primers used to generate a 92 bp amplicon were designed to span an intron so that any genomic contamination would be immediately apparent. The location of each primer (arx.F and arx.R) used in this analysis is shown in Figure 5.6B. Developmental stages used in the RT-PCR and analysis of the data were as previously discussed (section 5.2.3). Figure 5.8A summarises *arx-Am* transcript levels relative to *integrin-β*. The raw data used to construct this histogram are given in Table 5.2.

arx-Am expression occurred in two waves, but detected levels were never significantly higher than that of *integrin-β* (< 0.5%). The first wave of *arx-Am* expression occurred very early in development, spanning the formation of the endoderm from the egg to late donut stages. The second expression wave initiated at the pear stage as the larvae began active locomotion. *arx-Am* expression was reduced 10-fold in planulae prior to settlement and was negligible after settlement.

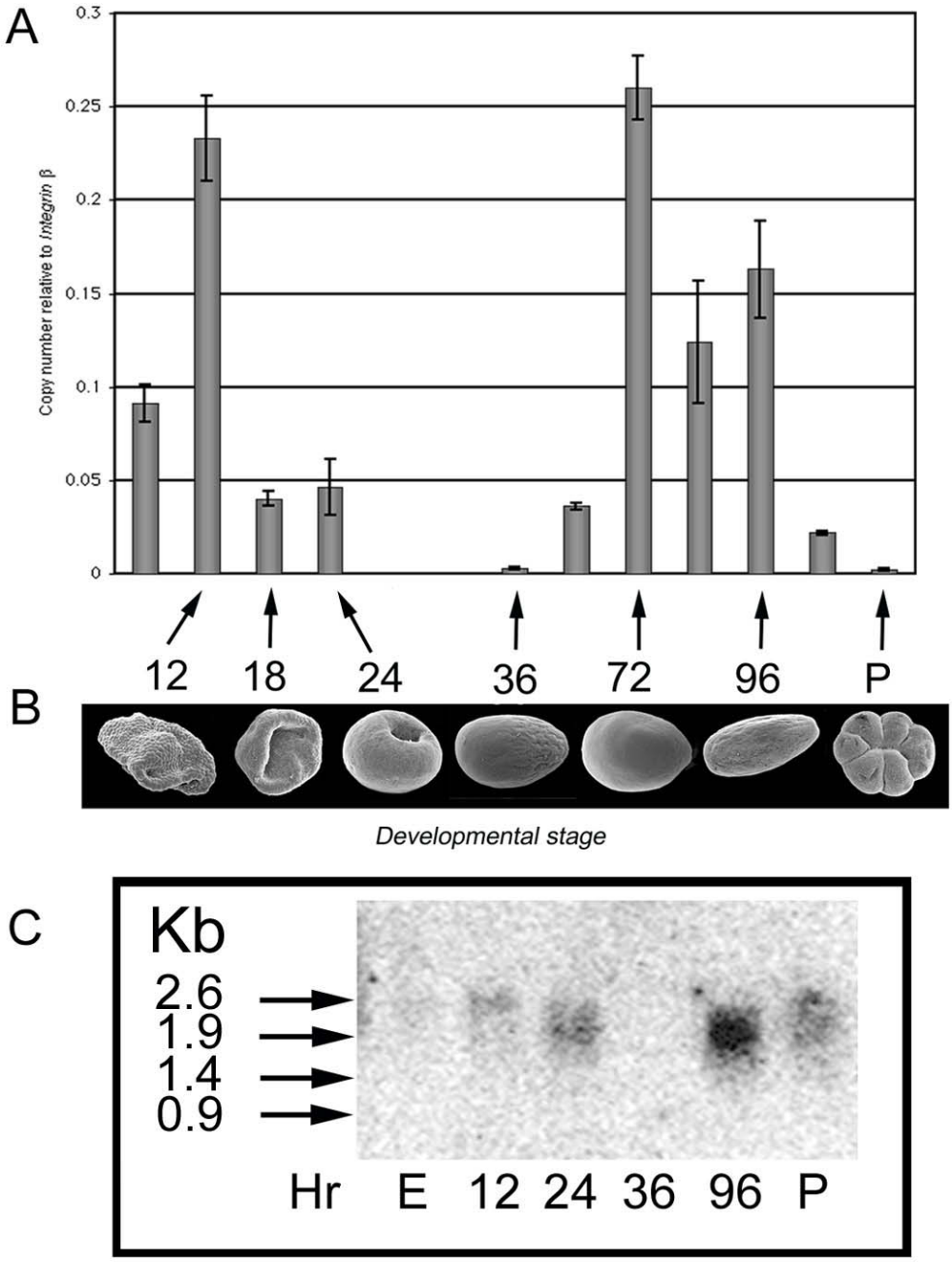


Figure 5.8: Temporal patterns of *arx-Am* expression during development. (A) Histogram representing *arx-Am* expression levels during development, expressed relative to the levels of the ubiquitously expressed *integrin*-β. Numbers on the y-axis indicate the copy number of arx-Am relative to the copy number of *integrin*-β. Thirteen stages were examined that corresponded to the stage indicated by arrows beneath. Not all stages are represented; only those stages also examined in the northern blot are indicated. (B) Electron micrographs depicting the relative morphologies of coral embryos and larvae from the same stages of development used in the real-time PCR analysis and mRNA extraction for northern blotting (not to scale). (C) Hybridisation pattern observed after probing a northern blot of mRNA extracted from *Acropora millepora* fertilised eggs (E) and embryos of various ages (numbers indicate hours post fertilisation), and post-settlement polyps (P), with a radioactively-labeled fragment of the *arx-Am* cDNA.

Table 5.2. Temporal expression patterns of *arx-Am* during development

Numerical data from RT-PCR of *arx-Am* expression. The morphological appearance of each of developmental stages is described in the column one. The mRNA copy number for both *arx-Am* and *integrin-β* for each stage are listed with an average figure and standard deviation. Levels of *arx-Am* expression relative to *integrin-β* were determined in the final column and used to construct the graph in Fig 5.8A

Developmental Stage	***arx-Am* mRNA copy number**	***integrin-b* mRNA copy number**	***arx-Am* Normalised to *integrin-b***
Egg	0.18	1.7	
	0.15	1.8	
	0.15	-	
Average	0.16±0.017	1.75±0.071	0.091±0.010
Prawn Chip	0.15	0.59	
	0.13	0.61	
	-	-	
Average	0.14±0.01	0.60±0.10	0.233±0.023
Donut	0.057	1.5	
	0.064	1.4	
	0.052	-	
Average	0.0577±0.006	1.45±0.071	0.040±0.004
Post Donut	0.011	0.33	
	0.019	0.32	
	-	-	
Average	0.015±0.005	0.325±0.007	0.046±0.015
Sphere with deep pore	0.005	7.0	
	0.006	5.9	
	-	5.5	
Average	0.0055±0.0002	6.13±0.777	0.0001±5.24E-06
Sphere with pore closing	0	2.8	
	0	2.7	
	0	-	
Average		2.75±0.071	0
Pre-Pear 1	0.013	4.4	
	0.020	5.2	
	-	-	
Average	0.0165±0.005	4.8±0.566	0.003±0.0009
Pre-Pear 2	0.58	15	
	0.52	16	
	0.57	16	
Average	0.557±0.032	15.6±0.577	0.036±0.002
Pears	9.8	36	
	10.0	37	
	8.8	37	
Average	9.53±0.643	36.6±0.577	0.260±0.017
Late Pear	1.4	12	
	1.7	10	
	1.0	-	
Average	1.367±0.351	11±1.414	0.124±0.033
Planulae	2.5	14	
	2.0	15	
	2.0	11	
Average	2.17±0.289	13.3±2.08	0.163±0.026
Pre-settlement	0.37	15	
	0.39	18	
	0.36	18	
Average	0.373±0.015	17±1.732	0.022±0.001
Post-settlement	0.0017	1.1	
	0.0026	0.98	
	-	-	
Average	0.0021±0.0006	1.04±0.084	0.002±0.0005

5.2.7. Spatial patterns of expression

The spatial patterns of *arx-Am* expression were examined by in situ hybridisation of embryos and planulae at various developmental stages. A DIG-labelled riboprobe, generated using linearised plasmid containing the complete *arx-Am* cDNA, was used in the hybridisation reaction, in the hydrolysed and unhydrolysed state. Several attempts to identify a pattern of distribution of *arx-Am* transcripts were unsuccessful despite alterations to experimental protocols, including increased concentrations of riboprobe and secondary anti-DIG antibody, without any observable difference. This is a problem, which is mostly likely due to the low level of *arx-Am*, as revealed by quantitative PCR data. Similar in situ hybridisation results have also been experienced with other coral genes (including *tlx-Am*) that are expressed during embryonic development, and a solution is yet to be found. Without spatial expression data, it is difficult to interpret what role *arx-Am* might be playing during coral embryonic development.

5.2.8. Paired-like genes: the K_{50} subclass

The *dbmx1-Am* and *dbmx2-Am* cDNA clones were isolated from a λZAP-II *A. millepora* pre-settlement stage cDNA library after a single genomic clone, λPC2.3, was isolated, sequenced and shown to contain two different homeobox regions encoding Prd-class homeodomains. A 700 bp amplicon showing identity to other Prd-class proteins was generated using degenerate primers (Prd-like HD.F2 and Prd-like HD.R1; as previously described in section 5.2.6). Using this 700 bp amplicon as a probe, 50,000 plaques of the *A. millepora* genomic library were screened and a single clone, λPC2.3, was isolated. Within this clone, two paired-like genes of the K50-type (Galliot et al., 1999) are tightly linked (within approximately ~10 kb) and are organised in tandem in the *Acropora* genome (see Figure 5.11). The two genes are expressed, although in both cases only incomplete cDNA clones have so far been isolated. The isolation and characterisation of the cDNA clones corresponding to these genes is discussed below.

5.2.8.1. The *dmbx1-Am* cDNA

PCR primers were designed to amplify 200 bp of the λPC2.3 genomic sequence, corresponding to the homeobox region and adjacent 20 bp of 5'- sequence of each gene.

Each fragment was then used as a probe to screen the *A. millepora* pre-settlement stage cDNA library. From a screen of 600,000 plaques, two partial clones (A3 and A5) corresponding to homeobox gene A were isolated and found to be identical. Comparisons with the BlastX database (Altschul et al., 1990) found this sequence most closely resembled mouse *Dmbx1* and is herein referred to as *dmbx1-Am*.

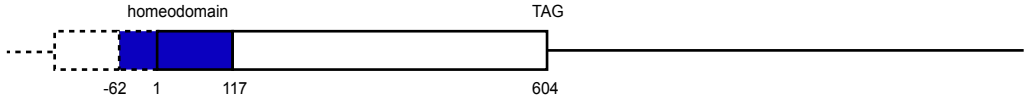
The nucleotide sequence and conceptual protein translation of *dmbx1-Am* cDNA are presented in Figure 5.9. The *dmbx1-Am* cDNA is a partial clone and lacks the 5'- end of this gene, including one-third of the homeobox. The nucleotide sequence of the *dmbx1-Am* cDNA is 1346 bp, comprising 603 bp of open reading frame and 743 bp of 3'-UTR. Conceptual translation of the ORF of the *dmbx1-Am* cDNA starts at amino acid 22 of the homeodomain. The first 100 bp of genomic sequence immediately upstream of the cDNA sequence were included in the deduced translation to complete the Dmbx1-Am homeodomain sequence for Prd-class comparisons and phylogenetic analyses. Additional library screens of 500,000 plaques in both the pre-settlement and prawn chip stage cDNA libraries failed to isolate a full-length *dmbx1-Am* cDNA.

5.2.8.2. The *dmbx2-Am* cDNA

A partial cDNA clone, B3, corresponding to homeobox gene B was isolated from a screen of 600,000 plaques of the *A. millepora* pre-settlement cDNA library using a probe containing the homeobox region and adjacent 20 bp of 5'- sequence of the second *paired*-like homeobox. BlastX database (Altschul et al., 1990) comparison of this sequence revealed that, like *dmbx1-Am*, it most closely resembled the mouse *Dmbx1* gene and is herein referred to as *dmbx2-Am*.

The nucleotide sequence and conceptual protein translation of *dmbx2-Am* are presented in Figure 5.10. Although this clone lacks the extreme 5'- end of the gene, sequence encoding the homeodomain and an octapeptide motif is present. The nucleotide sequence of *dmbx2-Am* cDNA is 1530 bp, comprising 733 bp of ORF and 797 bp of 3'-UTR. Conceptual translation of the ORF of the *dmbx2-Am* cDNA indicates it is only a partial clone. While a methionine residue is encoded at base pairs 137 to 139, a putative octapeptide motif encoded by base pairs 32 to 56 suggests the translation initiation codon is located further upstream. A full-length cDNA containing a complete open

A



B

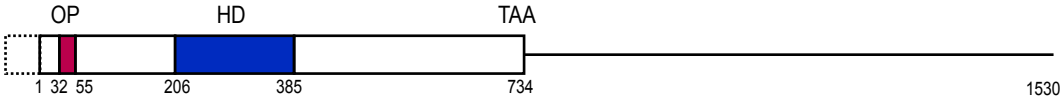
```
-200 cactggagtagtgatacatccactgttgtgaagttttggaacacaagagacctattcagtaaagtttacacgaatatatcttctgcttccgtttcagata

     I  I  L  E  V  H  R  N  P  T  R  A  K  Q  R  R  S  R  T  A  F  T  L  Q  Q  L  R  L  L  E  S  T  F  S
-100 taatcttagaagtccatcgtaacccaacgagggcaaaacaacgtagatctcgaacagctttcacccttcaacaactgaggctattggagagcactttctc

     K  T  H  Y  P  D  V  A  M  R  E  Q  L  A  S  W  T  N  L  P  E  S  R  I  Q  V  W  F  K  N  R  R  A
   1 AAGACGCATTACCCGGATGTTGCGATGAGGGAACAACTTGCAAGTTGGACAAATCTTCCTGAATCAAGAATTCAAGTCTGGTTCAAAAACAGAAGAGCCA

     K  H  R  K  Q  D  R  G  A  Q  Y  A  Q  S  D  E  N  G  E  P  F  Y  E  S  D  N  N  L  A  F  S  A  G  S
 101 AGCATAGGAAGCAAGACAGAGGAGCACAATACGCTCAGTCTGATGAGAATGGCGAACCTTTTTATGAAAGTGACAACAATTTGGCCTTCAGTGCCGGATC

       R  Q  C  P  F  P  L  P  S  T  L  L  F  P  T  E  N  N  F  A  L  G  T  R  M  F  T  S  Q  L  R  P  E
 201 GCGACAGTGCCCATTTCCACTTCCATCGACATTATTATTTCCAACCGAAAACAACTTTGCATTGGGCACAAGAATGTTTACATCGCAACTAAGACCAGAG

       E  S  M  G  S  S  S  T  S  N  Y  Q  V  F  S  D  K  G  E  S  F  P  L  N  Q  H  P  I  G  T  F  N  S
 301 GAGAGCATGGGAAGTTCATCAACATCAAATTATCAGGTGTTCTCTGATAAGGGCGAATCCTTCCCACTTAATCAGCATCCCATCGGTACGTTCAATTCTC

     P  R  Y  C  Y  P  T  R  V  D  N  S  P  A  V  P  T  L  E  V  S  C  H  N  Y  P  V  G  G  D  D  V  E  C
 401 CACGTTATTGTTATCCCACAAGAGTGGACAACTCTCCGGCGGTCCCAACATTAGAAGTCTCCTGCCACAATTACCCTGTCGGTGGTGATGACGTTGAGTG

       L  A  R  K  N  E  T  Y  H  K  V  W  K  P  T  I  S  S  V  D  Q  W  R  L  K  V  S  G  H  N  N  S  C
 501 TCTTGCTCGTAAAAATGAAACTTATCATAAAGTATGGAAACCAACAATTTCTAGTGTCGACCAATGGCGACTGAAAGTGAGTGGTCACAATAACAGCTGT
                                                                                                   A

     Q  *
 601 CAGTAGGACGATGACACTAGATGCAGGCCACACTAAGCTTAATTTTAATTGAGGAGAAATATGTTTACCAACAGAAGAGGCCATTAACATAAATATCATA
                 T

 701 GCACGCAGTCCTCACAATCAAACAGAAAAAAAAAAACAAGCAAACAAATAAAAGAGTACACATGCAGTACAGATGAGACAAGAAATTTCCCCACCAGAGCC
                                                                      A                 G

 801 CCTTAGAACAGATGTAATGTGGAAAACAAGCAAGACACTTGTTACAAAGAAAGATCCTTCTTGTCAATTATGAAACTCTTTTGTTGAGAAGTACCTTCAG
                                                                                                   A

 901 TCAACCGAGTGTGCGCTGCCTTTAAAGATGATAAGAGATTAATAATACGTAATTAACTGTCATGGAAAGTTAATATTTGGCCTTACTAAAAACTTAGGTT

1001 TGCACGTCTCTGGATTTCTGAAACTTATCGCCTTTAAACTAGCCTGCCGTACCCTAGTCTACCCTAGTTCCATTATCAATGACTGTGGTTGAAGTCTTTC
                          G

1101 TTTCAACATTCGTACCCGAGTGCGCAGGCACACTGAAGTAAATTTGGTAATTACGTAACATGAAAGTTCCCTGAAAAATTTTTAACCCTCGGCTAAAATG
                                                                          A           C     A

1201 TCTTAATTTACGTCTCCTTGGCAAGTGTAAGAGACTTAAGTGTGGCCCAAATGTTATGAAAGATGTTCAATAACCGCCAGACCCCATTAAATATATTAAA
                                    A                                     G           TA

1301 TATTTCAGCGACCCTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

C

```
Exon 1                          Intron                          Exon 2
TCAAGAATTCAA gtatgtgaaacaa...(461bp)...gtttacttacag GTCTGGTTCAAA
```

Figure 5.9: The *dmbx1-Am* cDNA. (A) A schematic representation of the *dmbx1-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region and the homeodomain region is indicated by the blue box. The stop (TAG) codon is shown at the top; the location of the start codon could not be determined. The numbers beneath the box represent the nucleotide sequence of the cDNA. (B) The nucleotide sequence of the *dmbx1-Am* cDNA and predicted amino acid sequence of the protein. The genomic sequence included to complete the homeodomain-encoding region is in lowercase. Residues of the homeodomain are highlighted in blue. An asterisk indicates the 3' in-frame stop codon. The intron position is shown by a red triangle. Nucleotide polymorphism between the cDNA and genomic sequence are indicated beneath the appropriate position in the cDNA sequence. Numbers on the left side represent the nucleotide sequence. (C) The intron-exon boundary of the single intron in *dmbx1-Am*. The GT-AG consensus splice sites are indicated in red. The total size of the intron is shown in brackets.

A



B

```
    G T R E T K D H S S F S I A S I L L D D S F V P N R G H F I G A T
  1 CGGCACGAGGGAAACGAAAGACCATTCATCATTTAGCATTGCGAGTATCTTACTAGACGATAGTTTCGTTCCTAATCGGGGTCACTTCATCGGTGCCACG

    A H Q A E I Y P S G A T M S L A E R L A D V I L E N Q D D K K S G    21
101 GCTCACCAGGCAGAAATTTATCCCTCAGGAGCCACGATGTCGTTAGCGGAGCGCTTAGCAGATGTTATTCTGGAGAACCAAGATGATAAGAAAAGCGGAG

    G K P R R M R T C F S P R Q L Q I L E Q A F S K T H Y P D V L L R E    55
201 GAAAGCCACGAAGAATGCGCACTTGTTTCTCCCCTCGACAACTTCAAATACTGGAGCAGGCCTTTTCTAAGACGCATTATCCGGATGTGCTTTTAAGGGA
                                  A

    Q L S N F A N I P E S R I Q V W F K N R R A K Y R K H E K S G A E    88
301 ACAACTATCCAATTTTGCCAACATTCCAGAATCAAGAATACAGGTGTGGTTCAAAAATCGACGAGCAAAGTATCGAAAGCATGAAAAGTCGGGAGCAGAA

    R T L R E P E S L K P R T E S S T L S P L P I F Q S S R T A N Q P   121
401 AGGACATTACGCGAACCAGAAAGTTTAAAACCAAGGACTGAAAGTTCAACCCTATCTCCTCTTCCAATTTTCCAAAGTTCAAGAACAGCCAACCAACCTG

    V L R S V P R K P N L E H L Y N D T E R L P N V S T F L P P P F R E   155
501 TTTTAAGAAGCGTCCCCAGAAAACCAAACTTAGAACATCTTTACAACGACACTGAACGTCTTCCAAACGTTTCAACCTTTTTACCACCACCATTTCGTGA
                              A           T

    I L S H N M V A A S P F H S F P C T P E S Y R Q Q Q F A G C S L R   188
601 AATTCTTTCACATAATATGGTTGCCGCATCTCCTTTCCACTCATTCCCTTGCACGCCTGAATCTTATCGACAGCAGCAATTTGCTGGCTGTTCTCTGAGG
                          C

    N Y T D L P T F G Y I *                                           199
701 AATTACACTGACCTACCCACATTCGGATACATCTAAGACCACTGAGGAAATGAGGTATTAAAAGCGAGACTACGTTATTTTTTTTCAAAATGCAAGGCAA

801 AGGCTTCTCTCCTCGTGGTTTTAACAACCTTTCAACATAATTTCTTTCATAATATTAAGCTTCATTACGGCTCAGATTAGGGAAATAATGAATATTCCAT
                                                              G

901 TTATTTGCTACTTTATATGGTACGTAAATGAAATACATTTCAAAACATCGCTTCAGTGTCTTTTTTCATAGTTTTCTACAAGCAAAATAAAACATAGTCC
                                                                 G

1001 CAAATGTGACAATTTCAAAAAAGTTTACAGAAAGCGAAGTAAGGGTTTATATTTCAAGTAATGTGCAGTGTGCCACTAGCCACGTGGGAGACATAATTGT
                                        T

1101 CAAATAACAGCTACTAGGCGCCGAGCAATTCTCGGACGAGTGTTGCAGCGAGCAGAGAGGAAGCATTTCCATAGAATGTAGCTATTTTTATTGCCATCAA

1201 TAATACGAATAAAGTGTTATTATCTTTATTATCATTTTTATAATAATTATTATTATTATCACTTTCTGTGAGATCTCATCTATTTTTGTGCACTTGACAG

1301 TTACCAAAATTAAGAGCTTTTGAAATAATATGTTTTCGTAGTAAGGTTGGTTGGTTAATTATCAAATTGTATTTAAGTAATGAATAATAATAGTAATGAA

1401 CTATGATTAGATAATTGCACAGTATCTTTATTCACCGCTGGCAATGAATATTTGCTATCATTAATTGGTGTAAATAACTAGTAAATAAAGACGAACGGTT

1501 CCAGCTTTTGGCAAAAAAAAAAAAAAAAA
```

C

```
Exon 1                    Intron 1                    Exon 2
AGCGCTTAGCAG gtttgtgctttc...(>3.5 kb)..ccttacgtgcag ATGTTATTCTGG

Exon 3                    Intron 2                    Exon 4
TCAAGAATACAG gtgacaaacagt...(591 bp)...cttgtcgtatag GTGTGGTTCAAA
```

Figure 5.10: The *dmbx2-Am* cDNA. (A) Schematic representation of the *dmbx2-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region, the octapeptide is shown by a red box and the homeodomain region is indicated by the blue box. The stop (TAA) codon is shown at the top; the location of the start codon could not be determined. The numbers beneath the box represent the nucleotide sequence of the cDNA. (B) The nucleotide sequence of the *dmbx2-Am* cDNA and predicted amino acid sequence of the protein. Residues of the homeodomain are highlighted in blue. An asterisk indicates the 3' in-frame stop codon. The two intron positions are shown by red triangles. Nucleotide polymorphisms between the cDNA and genomic sequence are indicated beneath the appropriate position in the cDNA sequence. Numbers on the left side represent the nucleotide sequence; numbers on the right side represent the amino acid sequence. (C) The intron-exon boundary of both introns in *dmbx2-Am*. The GT-AG consensus splice sites are indicated in red. The total size of each intron is shown in brackets.

reading frame could not be isolated in subsequent screening attempts using 500,000 plaques of both the pre-settlement and prawn chip stage cDNA libraries.

5.2.9. The genomic structure of *dmbx1-Am* and *dmbx2-Am*

In order to establish the extent to which the genomic organisation of *dmbx1-Am* and *dmbx2-Am* has been conserved, the structure of each locus was determined. A schematic representation of the *dmbx1-Am* and *dmbx2-Am* genomic loci is presented in Figure 5.11. Comparisons of each cDNA sequence with the corresponding genomic sequence revealed the presence of an intron between codons 46 and 47 of the homeodomain (Q-VWFK) at both loci; this intron was 461 bp in *dmbx1-Am* and 591 bp in *dmbx2-Am*. This position, which is common in many *paired*-type genes, is also shared with another *Acropora paired*-like gene, *arx-Am*, but has not been confirmed for *hbn-Am*. A second intron was also identified at the *dmbx2-Am* locus in phase one of the aspartate codon, 15 residues N-terminal of the homeodomain. This intron was at least 3.5 kb in length but its exact size is not known as this region was not fully sequenced. The presence of a second intron at the *dmbx1-Am* locus could not be confirmed, as the cDNA is incomplete. All introns possess donor and acceptor splice sites that conform to the GT-AG consensus (Figure 5.9C and 5.10C).

5.2.10. The Dmbx1-Am and Dmbx2-Am proteins

The Dmbx-1 and Dmbx2-Am proteins are highly divergent, sharing only 33% overall identity (<41% similarity) (Figure 5.12A). Each protein contains a Paired-class homeodomain in its N-terminal half and each possesses the six invariant residues characteristic of this class (see Introduction). Both proteins are further classified into the K_{50} Prd-like gene sub-class based on the residue at position 50 of the homeodomain, a lysine.

The homeodomains of Dmbx1-Am and Dmbx2-Am are 65% identical (>78% similarity; see Figure 5.12B). They most closely resemble that of Dmbx1 in the mouse and Manacle in *Hydra* (>65% identity). The Dmbx2-Am protein also contains an octapeptide motif (62% identity to consensus; see Figure 5.3B), which is separated from the homeodomain by 50 residues, in the N-terminal half of the protein. A C-terminal OAR domain is not present in either protein.

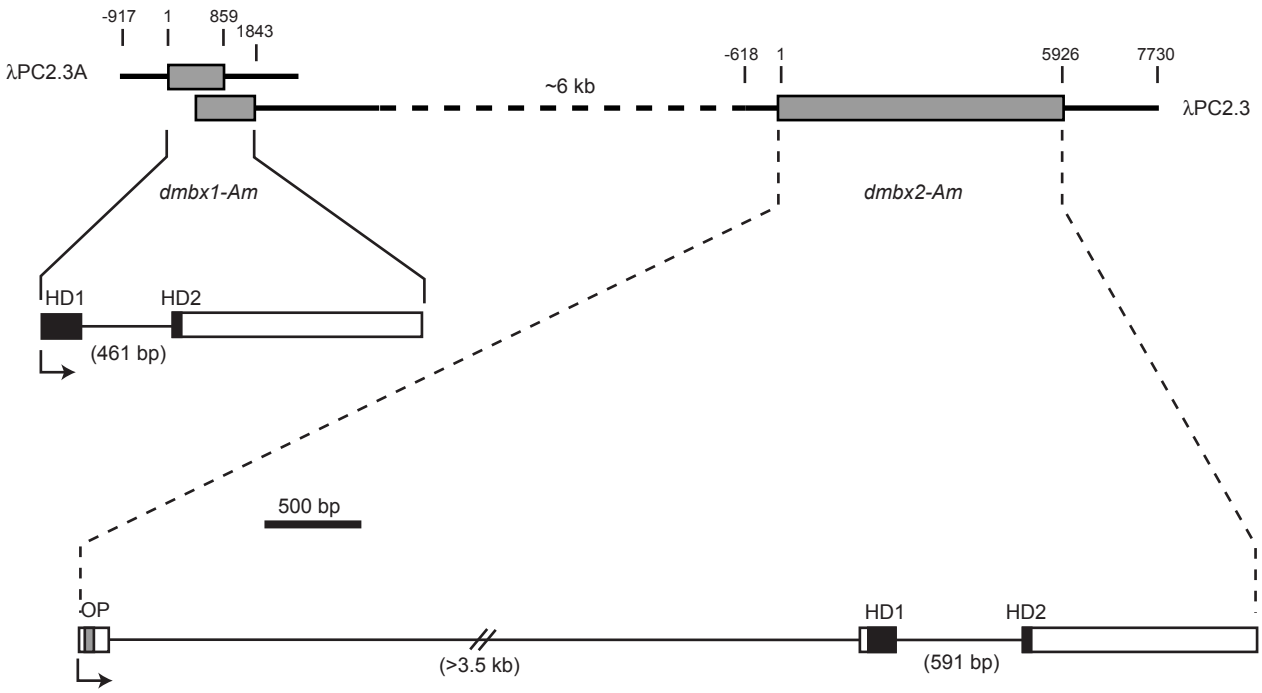


Figure 5.11: Genomic organisation of the *dmbx1-Am* and *dmbx2-Am* genomic loci.
Schematic representation of the overlapping genomic clones, λPC2.3 and λPC2.3A. The lightly shaded boxes indicate the exon-contain regions of the genomic clones. Numbers above refer to the nucleotide sequence of each genomic clone with zero defined as the 5'- end of the cDNA sequence. The expanded view shows the relative sizes of the introns and exons with the homeodomains (HD1 and HD2) and octapeptide (OP) regions indicated by labeled boxes. Intron sizes are given in brackets.

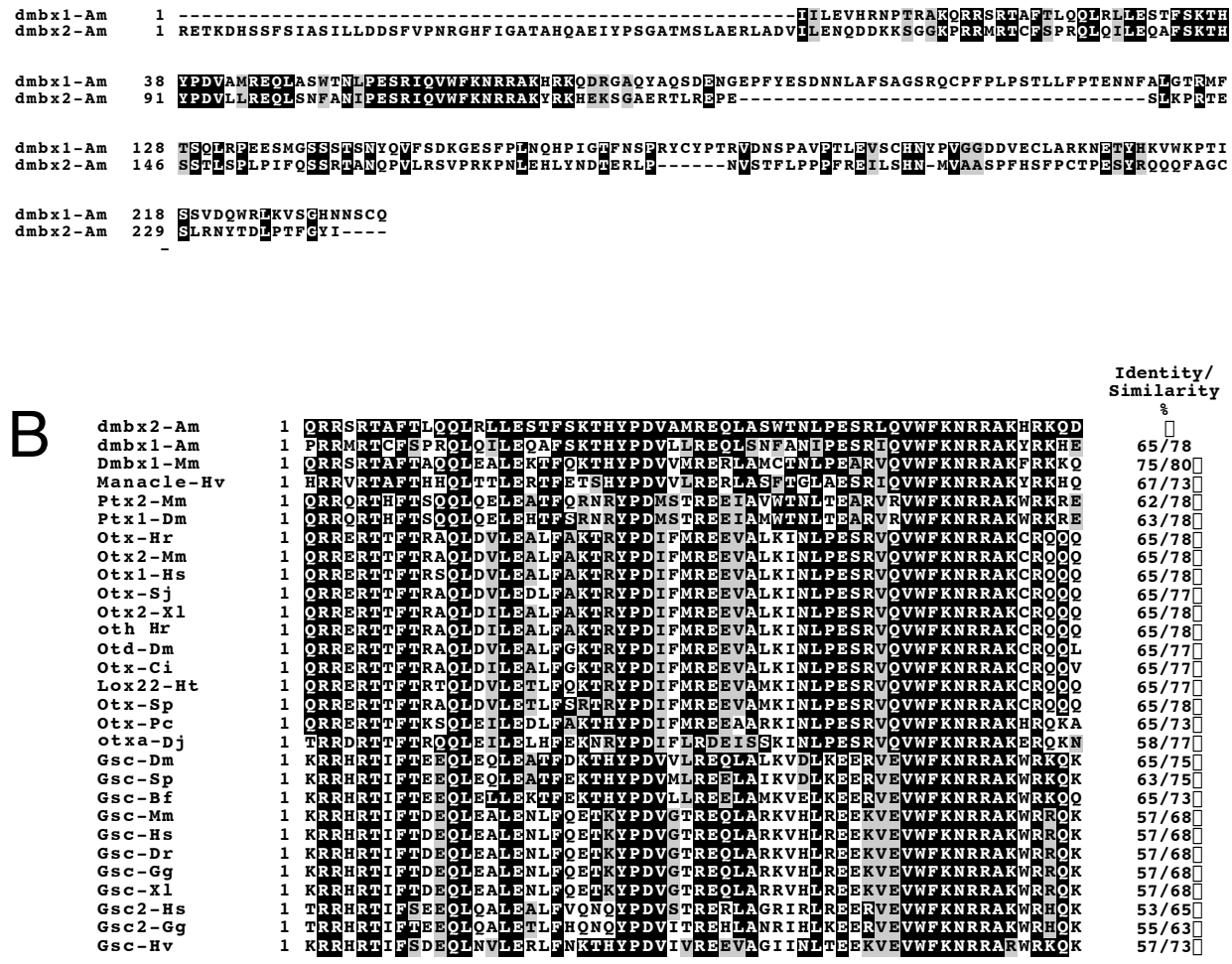


Figure 5.12: A comparison of the Dmbx1-Am and Dmbx2-Am homeodomains with other K50 Prd-type homeodomains. (A) Aligned (Clustal W; Thompson et al.., 1994) amino acid sequences of the Dmbx1-Am and Dmbx2-Am proteins, boxshaded (Boxshade server) to indicate residues shared by both proteins. (B) Aligned (Clustal W; Thompson et al., 1994) amino acid sequences of the Dmbx1-Am and Dmbx2-Am homeodomains with that of other K50 Prd-type homeodomains, boxshaded (Boxshade server) to indicate residues shared with the Dmbx2-Am homeodomain. Identical residues are shaded darkly; conserved substitutions are lightly shaded. The column to the right of the alignment indicates the overall identity and similarity of each homeodomain with the Dmbx2-Am motif. The species names (and Genbank Accession No.s) are abbreviated as follows: Am, *Acropora millepora*; Bf, *Branchiostoma floridae* (Gsc: AAF97935); Dm, *Drosophila melanogaster* (Ptx1: NP_477285, Otd: A35912, Gsc: P54366); Dr, *Danio rerio* (Gsc: AAA50028); Gg, *Gallus gallus* (Gsc: CAA49897, Gsc2: CAA70980), Hr, *Halocynthia roretzi* (; oth: BAA24679); Hs, *Homo sapiens* (Otx1: S39406, Gsc: P56915, Gsc2: AAC39544); Hv, *Hydra vulgaris* (Manacle: AAD30998, Gsc: AAF14575); Ht, *Helobdella triserialis* (leech- Lox22: AAB61443); Mm, *Mus musculus* (Dmbx1: AF421858, Ptx2: NP_035228, Otx2: P80206, Gsc: NP_034481); Pc, *Podocoryne carnea* (Otx: AAF04002); Asc, *Herdmania curvata* (Otx: AAD30504); Sj, *Stichopus japonicus* (sea cucumber – Otx: BAB16104), Sp, *Strongylocentrotus purpuratus* (Otx: Q26417); Ci, *Ciona intestinalis* (Otx: AAG59802), Xl, *Xenopus laevis* (Otx2: Q91813, Gsc: B42768).

5.3. DISCUSSION

5.3.1. Conserved protein motifs

The octapeptide (Schneitz et al., 1993), or eh1/GEH domain (Hemmati-Brivanlou et al., 1991), found in a number of Paired-class and non-Paired-class proteins, has been shown to function as a transcriptional repressor in triploblastic organisms (Mailhos et al., 1998; Smith and Jaynes, 1996). Most members of the Rx (Furukawa et al., 1997; Mathers et al., 1997); Ceh-10 (Liu et al., 1994; Svendsen and McGhee, 1995) and Anf ((Hermesz et al., 1996; Kazanskaya et al., 1997; Zaraisky et al., 1992) families encode octapeptides, but this feature is less common in the Al, Gsc and Pax-3/7 families, and has not been detected in the Ptx, Otx, Pax-4/6, Unc-4, Mix, Prx, Otp, Cart-1 or Arx homeoproteins isolated to date (reviewed in Galliot et al., 1999). The OAR domain is found in representatives of all three subclasses of Prd-type superclass proteins. The OAR domain is a conserved feature of Cart-1, Prx and Og12 proteins, but they never possess an octapeptide. Similarly Arx-related proteins in bilaterians possessing both an octapeptide and OAR domain are uncommon, but two exceptions are Arx from the mouse and zebrafish (Miura et al., 1997).

hbn-Am and *arx-Am* are the first reported cnidarian *paired*-like genes to encode both octapeptide and OAR motifs, in addition to the homeodomain. A putative OAR domain was reported in *Hydra* prdl-a (23% identity to consensus; see Figure 5.3C). N-terminal octapeptides are present in a number of cnidarian Paired-type proteins including *Hydra* Alx, Pax-B and prdl-b (88%, 75% and 62% identity to consensus respectively; see Figure 5.3B). In 1999, Galliot et al. suggested that the homeodomain, octapeptide and OAR domains, present in differing combinations in various Paired-type proteins, were acquired at different stages in evolution. However the domain structures of the proteins encoded by *hbn-Am* and *arx-Am* suggest that the ancestral Prd-class protein possessed all three structural features and that more derived representatives of several classes frequently display secondary loss of some domains. It is therefore surprising that none of the *Hydra* homologs of these proteins share a similar domain structure to that of the *Acropora* proteins. More than likely this is due to secondary loss of these domains in the *Hydra* proteins.

5.3.2. Evolutionary relationships amongst Paired-like homeodomains

On the basis of BlastX (Altschul et al., 1990) database comparisons, the homeodomains of Hbn-Am, Arx-Am, Dmbx1-Am and Dmbx2-Am proteins are clearly members of the Prd-like gene class and can be further classified into the Q_{50} *Prd*-like gene sub-class (Hbn-Am and Arx-Am) and the K_{50} *Prd*-like gene subclass (Dmbx1-Am and Dmbx2-Am) respectively.

In each *Acropora paired*-like gene for which the genomic structure was determined, a conserved intron was found between codons 46 and 47 of the homeodomain. *arx-Am*, *dmbx1-Am* and *dmbx2-Am* all share this intron, although its size varies from 461 bp to 745 bp. This is a common intron insertion position in many Paired-type homeodomains (reviewed by Gauchat et al., 2000), suggesting it is evolutionarily conserved and of functional importance to the splicing mechanisms and subsequent processing prior to protein translation.

To further clarify evolutionary relationships, maximum likelihood phylogenetic analyses based on the homeodomain sequences were carried out in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff substitution matrix. The dataset used in a previous study (Galliot et al., 1999) was updated to include several new sequences obtained for cnidarians, including the Arx-Am, Hbn-Am, Dmbx1-Am and Dmbx2-Am sequences described above, as well the *Acropora millepora* Pax proteins Pax-Bam and Pax-Dam (Miller et al., 2000). Representatives of other families cloned from other cnidarians were also included in the analyses – Manacle (Bridge et al., 2000), HyAlx (Smith et al., 2000a) and Cngsc (Broun et al., 1999). Although analyses using the large dataset were informative, the highly divergent nature of specific Q_{50} genes and gene families led to low levels of support for specific aspects of tree topology. The approach taken was one of selective removal of highly divergent types, so that analyses included the approximately twelve types with robust relationships together with subsets of three (Ptx, Anf and Unc-4) divergent families. Figures 5.13 show a representative tree generated using this approach.

The maximum-likelihood analyses have several major implications for relationships between Prd-type homeobox genes. Firstly, the three K_{50} homeodomain classes recognised previously (the Gsc, Otx and Ptx families; (Galliot et al., 1999) consistently

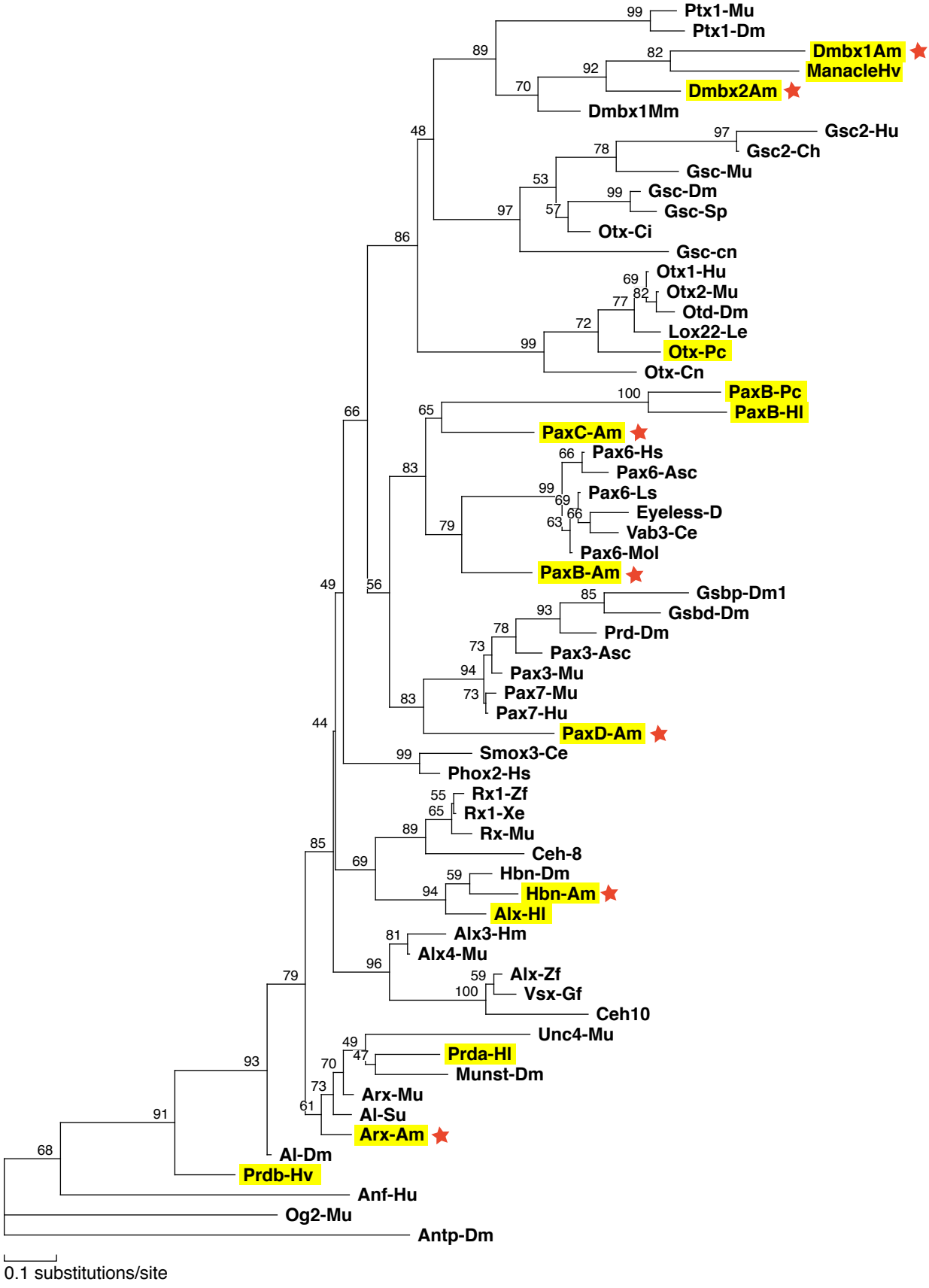


Figure 5.13: Relationships within the Prd-type superclass. *cont'd.../*

Figure 5.13: Relationships within the Prd-type superclass.

Prd-type homeodomains from *Acropora* and other organisms were subjected to phylogenetic analysis using MolPhy (Adachi and Hasegawa, 1996). The tree shown is a result of Maximum Likelihood distance analyses; numbers against the branches indicate the percentage of 1000 bootstrap replicates supporting the topology. Cnidarian sequences are highlighted in yellow, and *Acropora* sequences are marked by a red asterisk. The species names (and Genbank Accession No.s) are abbreviated as follows: Am, *Acropora millepora* (Pax-B: AAF64460, Pax-C: AAC15711); Bf, *Branchiostoma floridae* (Gsc: AAF97935); Ce, *Caenorhabditis elegans* (Vab3: S60252, Otd: NP_509860, ceh8: Q94398, ceh10: T34470, unc4: P29506); Ca, *Carassius auratus* (Vsx1: Q90277); Dm, *Drosophila melanogaster* (Al: XP_079165, Ptx1: NP_477285, Otd: A35912, Eyeless: I45557, Toy: NP_524638, Gsc: P54366, Munster: NP_477330, Otp: NP_523799, Hbn: NP_524903, Repo: A54282, GSbd: P09082, Prd: AAA28840, Antp: P02833); Dj, *Dugesia japonica* (otx: BAA90698, Rax: O97039) Dr, *Danio rerio* (Gsc: AAA50028, Otp: AAD42021, Rx1: O42356, Vsx1: AAB71611, Alx: AAB66714); Gg, *Gallus gallus* (Gsc: CAA49897, Gsc2: CAA70980, Rx1: Q9PVY0), Hr, *Halocynthia roretzi* (oth: BAA24679); Hs, *Homo sapiens* (Otx1: S39406, Otx2: P32243, Pax6: A56674, Gsc: P56915, Gsc2: AAC39544, Alx3: XP_002147, Prop1: O75360); Hv, *Hydra vulgaris* (Manacle: AAD30998, Gsc: AAF14575, prdla: CAA75668, prdla: CAA75669, Alx: AAG03082); Ht, *Helobdella triserialis* (leech- Lox22: AAB61443); Ls, Cynops pyrrhogaster (Pax6: BAA24022); Lj, *Lampetra japonica* (Pax6: BAB62531); Mm, *Mus musculus* (Arx: O35085, Alx4: O35137, Dmbx1: AF421858, Ptx2: NP_035228, Og2: AAC52828, Otx2: P80206, Otp: O09113, Gsc: NP_034481, Mix1: NP_038757, Pax7: P47239, Prx2: CAA37055, Prop1: P97458, Rx: AAC53129, Unc41: CAB09537); Pc, *Podocoryne carnea* (Otx: AAF04002); Asc, *Phallusia mammilata* (Pax-6: CAA71094); Sj, *Stichopus japonicus* (sea cucumber – Otx: BAB16104), Sp, *Strongylocentrotus purpuratus* (Alx: Q26657, Otx: Q26417, Gsc: AAG31170); Ci, *Ciona intestinalis* (Otx: AAG59802), Xl, Xenopus laevis (Cart1: Q91574, Otx2: Q91813, Gsc: B42768, Pitx1: Q9W751, Siamois: A56219, Twin: AAC60331, Mix12: A32548, AAC60020, Milk: AAC60376).

formed a distinct monophyletic group with high bootstrap support, implying a common origin. Within the K_{50} clade, the *Cionia* 'Otx' sequence consistently clustered in the Gsc clade rather than with the true Otx sequences. As both clades are supported by high bootstrap values, the *Ciona* 'Otx' should probably be reclassified as a *Gsc* gene. The *Hydra manacle* gene is most likely orthologous with *Acropora dmbx1; dmbx2* is more closely related to the vertebrate *dmbx1* gene than are *manacle* and *dmbx1*. This group of sequences appear to define a new family of K_{50} Paired-type sequences, whose closest relatives are probably the *Ptx* genes (see Figure 5.13).

A second important implication of the maximum-likelihood analyses is that the two S_{50} Prd-type homeodomain families, the Pax-4/6 and Pax-3/7 homeodomain types, consistently cluster together with the K_{50} families as a distinct monophyletic group emerging from the Q_{50} types (Figure 5.13). Depending upon the particular dataset analysed, the two S_{50} classes formed either a monophyletic group distinct from the K_{50} classes, or branched off sequentially within the K_{50}/S_{50} clade; either topology was relatively weakly supported.

Within the Q_{50} Prd-type homeodomain families, *Hydra* Alx consistently clustered with the Hbn sequences of *Drosophila* and *Acropora*, with high bootstrap support. Although vertebrate orthologs of these genes are yet to be isolated, these sequences may define a new subfamily of Q_{50} Prd-type homeodomains, whose closest relatives appear to be the *Rx* genes (see Figure 5.13).

Contrary to previous analyses (Galliot et al., 1999), the overall implication of the maximum-likelihood analyses presented here is that the residue at homeodomain position 50 is a good indicator of evolutionary relatedness in the Paired-type superclass. The likely pattern during evolution was that ancestral Q_{50} types gave rise to an ancestral K_{50} gene from which the three to four K_{50} types diverged. The Q_{50} to K_{50} transition probably occurred via an S_{50} intermediate. The Q_{50} paired-like sequences are highly diverse and for this reason, patterns of relatedness between these remain unclear.

Together with mouse Dmbx1 and *Hydra* Manacle, Dmbx1-Am and Dmbx2-Am appear to form a novel paired-like homeoprotein subfamily. *Dmbx1*, which is also conserved in zebrafish and human, is expressed in the diencephalon and mesencephalon of the

mouse embryonic brain and is a marker for the developing anterior vertebrate nervous system. Comparisons with the hydrozoan member of this novel sub-family are complex; *manacle* in *Hydra* is expressed in the differentiating basal disk ectoderm of the adult *Hydra,* which is not easily compared with the developing vertebrate nervous system. Investigations of *manacle* expression during embryogenesis would be more appropriate for comparison with the Bilateria, but this is difficult to accomplish in *Hydra* as its mode of sexual reproduction is unpredictable and inaccessible. Investigations of the expression patterns of *dmbx1-Am* and *dmbx2-Am* in the coral during embryogenesis and post-settlement stages will help elucidate the role of this *paired*-like subfamily of proteins during development in the Cnidaria.

5.3.3. *arx-Am* expression during coral development

Arx-related genes in bilaterians are often involved in brain development during embryogenesis in vertebrates and flies alike. In *Acropora*, the spatial distribution of *arx-Am* could not be determined, but temporally this gene is expressed in two waves that occur during periods of significant morphological development. The first wave of *arx-Am* expression is detected as the cellular bilayer is beginning to form two separate tissue layers. The second wave occurs later in development, around the time when the nervous system first appears, and decreases dramatically immediately prior to settlement, when the nervous system is observed to disintegrate. It is possible *arx-Am* may have a conserved role in nervous system development, but more informative spatial data are required.

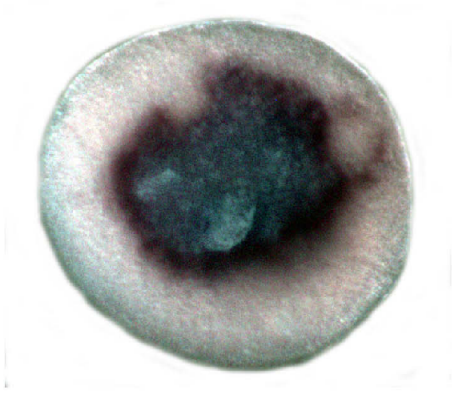
5.3.4. *hbn-Am* expression during coral development

The *hbn* gene in *Drosophila* is expressed in the developing brain and central nervous system; expression initiates at the blastoderm stage in the anterior dorsal head primordia and is maintained in cells fated to form the brain during later embryonic stages (Walldorf et al., 2000). Likewise, the *Acropora hbn-Am* gene is expressed relatively early in development, but is less restricted in its pattern of distribution. In the syncytial blastoderm, the broad band of *Drosophila hbn* expression resembles that of other head gap genes in the dorsal head region. Later, this expression domain is retracted and reorganised, eventually corresponding to each neuromere of the CNS. A secondary expression domain is also found posteriorly at the midgut/hindgut boundary during germband retraction and in the VNC during late stages of embryogenesis (Walldorf et

al., 2000). This resembles the expression pattern of *aristaless* in the fly intestinal tract (Schneitz et al., 1993). In contrast, while *Acropora hbn-Am* is expressed ubiquitously in the presumptive ectoderm, it is not specifically restricted within this region. The role of *hbn* in the posterior region of the fly may resemble a similar expression zone of *aristaless* in the intestinal tract (Schneitz et al., 1993), but does not parallel *hbn-Am* expression in *Acropora*.

In adult *Hydra*, *HyAlx* was expressed in ectodermal epithelial cells at the base of each tentacle; it is specifically associated with tentacle formation, rather than head formation during both budding and head regeneration (Smith et al., 2000a). *Acropora hbn-Am* expression during embryogenesis does not resemble that of *HyAlx*. While it is possible that this is a derived function specific to *Hydra*, expression in *Acropora* adult stages is yet to be examined.

Based on in situ hybridisation data, the distribution of *hbn-Am* mRNA during coral embryonic development is consistent with a role in ectoderm specification. Its exclusive expression in the presumptive ectoderm suggests Hbn-Am may act as a repressor to specify ectodermal tissue and prevent its differentiation to endodermal-type tissue. Alternatively, it may activate ectoderm-specifying genes. Interestingly, the *Acropora snail* gene has recently been shown to be expressed at the same time and in a mutually exclusive pattern to *hbn-Am* during coral embryonic development (see Figure 5.14; Ball, unpublished). In a variety of organisms, *snail* genes are expressed in cells undergoing an epithelial-mesenchyme transition (Alberga et al., 1991; Carver et al., 2001; Essex et al., 1993; Hammerschmidt and Nusslein-Volhard, 1993; Langeland et al., 1998; Nieto et al., 1992; Paznekas et al., 1999; Sefton et al., 1998; Smith et al., 2000b; Spring et al., 2002; Wada and Saiga, 1999), and the expression patterns of the *Acropora* and *Drosophila snail* genes are strikingly similar. *snail* is a mesodermal determinant in the fly, acting as a repressor of neurectodermal genes such as *single-minded* (Kasai et al., 1992), *vnd/NK-2* (Mellerick and Nirenberg, 1995), *rhomboid* (Ip et al., 1992a; Ip et al., 1992b) and components of the Notch signalling pathway (Cowden and Levine, 2002); Snail also activates genes in the mesoderm primordia including *zfh-1* (Lai et al., 1991) and *vestigial* (Fuse et al., 1996). There are no precedents for a direct interaction between a *paired*-like gene (such as *hbn-Am*) and *snail*, but their mutually exclusive expression patterns imply that this might be the case in *Acropora*.



snail-Am

Figure 5.14. The spatial expression pattern of *snail-Am* in an *Acropora* embryo. One example (at approximately 36 hours post fertilisation) of the in situ hybridization pattern observed when embryos of early developmental stages were probed with a DIG-labeled *snail-Am* probe (Ball, unpublished).

5.3.5. Duplicates of *dmbx*-related homeobox genes

During vertebrate evolution, there are likely to have been two round of genome-wide duplication, which have lead to (for example) four Hox clusters in mammals (reviewed in Ruddle et al., 1999) that correspond to single clusters in arthropods and nematodes (reviewed in Finnerty and Martindale, 1998). Broad scale duplications events have been suggested to be of greater evolutionary significance than tandem duplication events as they enable the duplication of entire gene pathways (Ohno, 1970). Without the constraints imposed by selective pressure, these duplicated pathways can more easily acquire mutations and novel functions, which may have facilitated the morphological diversification of the vertebrates.

Although not universally accepted (see for example, Hughes et al., 1998, 1999; Skrabanek et al., 1998; Martin et al., 2001), the idea that genome-wide duplications have occurred in vertebrates is widely held. Comparative data from cnidarians is limited but there is evidence of individual gene duplications in the Cnidaria, including the *Hydra nanos* genes (Mochizuki et al., 2000) and the *Smad 1/5* genes in *Acropora* (Samuel et al., 2001). Given the time since the divergence of the Cnidaria from the line leading to the triploblasts, it is not surprising that some genes have been independently duplicated in each lineage.

During the course of this project, three linked pairs of duplicated homeobox genes have been identified in *Acropora*; the first of which was the pair of *dmbx*-related genes discussed in this chapter. Subsequently, a pair of *msx*-related genes was identified (Hislop et al., in prep), and a pair of *not* class homeobox genes were also identified on one genomic clone (Hayward et al, unpublished).

These cases imply that a significant proportion of the *Acropora* genes may have been duplicated; of the 13 distinct classes of homeobox genes thus far identified in *Acropora* (excluding *Pax* genes), at least three exist as duplicated gene pairs. For *Acropora cnox2-Am* and *Hex* genes for which the structure is known (Hayward et al., 2001; Samuel et al., unpublished), the possibility of linked duplicates cannot be ruled out, while no genomic data are available for other genes. Therefore, the true extent of duplication may more encompassing that the cases documented here. Based on this preliminary data it appears likely that tandem duplication events occurred frequently in

Acropora; no comparative genomic data are available for any other cnidarians, but it is possible these duplication events are a common feature of cnidarian genomes and may have contributed to the large genome size estimated for some members of this phylum. A short paper is in preparation discussing this data, but is not included in this thesis.

5.4. CONCLUSIONS

During a search for *Acropora* orthologs of *Hydra prdl-a*, a total of four *paired*-like homeobox genes were identified in the coral; two of the Q_{50} subclass (*hbn-Am, arx-Am*) and two of the K_{50} subclass (*dmbx1-Am, dmbx2-Am*). Both *arx-Am* and *hbn-Am* encoded an octapeptide motif and OAR domain, in addition to the homeodomain, suggesting that these features were present in the ancestral Prd-type gene, and that secondary loss of one or both motifs has occurred in many Prd-type genes. Surprisingly, no *Hydra paired*-like genes possess both features, but this may reflect the derived nature of the Hydroza.

The Hbn-Am homeodomain most closely resembles those of *Drosophila* Hbn and *Hydra* HyAlx. While *hbn* is involved in brain specification in the fly and HyAlx functions during *Hydra* tentacle formation, neither role is easily correlated to that of *hbn-Am* in *Acropora*. In situ hybridisation results suggest a role for *hbn-Am* in early ectoderm specification; the *Acropora snail* gene is expressed in a mutually exclusive and complementary domain during coral developing, suggesting the possibility of an interaction between *hbn-Am* and *snail-Am*.

arx-Am expression occurs in two waves during coral embryonic development, but the spatial distribution of the transcripts could not be determined by in situ hybridisation. The two time windows of expression may represent two diverged roles for this gene during coral development, but expression data are needed to clarify this. In situ data from adult *Acropora* stages may also provide more information.

Phylogenetic analyses indicated the homeodomain of Dmbx1-Am is probably orthologous to *Hydra* Manacle, while Dmbx2-Am is more closely related to the vertebrate Dmbx1 than either Manacle or Dmbx1-Am; together this group of *Dmbx-*

related genes appear to define a new family of K_{50} Prd-type sequences, most closely related to the Ptx family (Figure 5.13). An ortholog of these genes is yet to be identified in *Drosophila* or *C. elegans*.

While vertebrate orthologs are yet to be identified, Hbn-Am consistently clusters with the *Drosophila* Hbn and *Hydra* HyAlx sequences. None of the *Acropora* genes, however, appears to be strictly orthologous with *Hydra prdl-a*.

5.5. FUTURE DIRECTIONS

As a first stage in examining the possibility of a genetic interaction between *Acropora hbn-Am* and *snail*, it might be worthwhile searching the *hbn-Am* 5'- genomic region for snail binding sites, and likewise the snail 5'- genomic region for Hbn-Am binding sites. The presence of hbn-Am or snail binding sites in either region will provide the first steps in understanding the regulation of these genes.

In situ hybridisation studies did not reveal the spatial distribution of *Acropora arx-Am* transcripts. To determine what role *Arx*-related genes play in simple metazoans, cnidarian data from embryonic and adult stages are still required. Isolation of an *Arx* ortholog in another anthozoan, such as *Nematostella*, may provide more promising data; shared patterns of expression between *Nematostella* and *Acropora* may suggest conserved functional roles.

It is highly likely that *dmbx1-Am* and *dmbx2-Am* are novel members of a new K_{50} Prd-type family. The role of the *Hydra manacle* gene in differentiating basal disk ectoderm may reflect the derived nature of hydrozoans, but in vertebrates *Dmbx* genes are involved in anterior brain formation. In the absence of a ortholog of this gene from the fly, expression data for the *Acropora* genes would help determine if conserved roles are apparent in this gene family.