Chapter 3 - Conservation of Dorsal-Ventral Patterning Genes in Cnidaria

3.1 Introduction

3.1.1 Conserved A/P and D/V patterning systems in the Metazoa

In all higher animals so far examined, Hox genes have been shown to play conserved roles in patterning the A/P axis (Slack *et al.*, 1993). Hox genes, a subset of the homeobox gene family, possess several distinguishing factors upon which their classification is based. Different Hox genes have distinct domains of expression along the body axis, and collectively the staggered Hox expression domains effectively subdivide the bilaterian body plan into distinct regions (Lewin, 1997). In addition, Hox genes are arranged in clusters along the chromosome, and their arrangement reflects their spatial and temporal expression; genes located more 5' along the chromosome are expressed earlier and more posteriorly than those at the 3' end of the cluster. An additional characteristic is the association of Hox proteins and co-factors, mediated by the hexapeptide motif N-terminal of the HD which functions to increase DNA-binding specificity (Lewin, 1997). 'Hox-like' genes have been isolated from a number of cnidarians and have been argued as true Hox genes by several authors based on sequence analysis (eg (Martinez *et al.*, 1998; Finnerty, 1999; Gauchat *et al.*, 2000)). Recently, expression data in support of this claim has also been presented (Finnerty, 2004). However in the absence of confirmed chromosomal linkage and correlated patterns of expression, while these genes are clearly 'Hox-like', the issue of whether these are true Hox genes is debatable. In addition to Hox genes, other genes known to have central and conserved roles in patterning the A/P axis of higher animals are also present in cnidarians, including clear homologs of the head patterning genes *otd/otx* and *ems/Emx* (Hislop *et al*, unpublished observations; (Mokady *et al.*, 1998; Muller *et al.*, 1999; Smith *et al.*, 1999).

The specification of the D/V axis is also under the control of a conserved network of genes in bilateral animals, demonstrated by those that pattern the central nervous system

(CNS). Until recently, separate origins were assumed for the arthropod and vertebrate central nervous systems because of major structural differences – for example in vertebrates, the CNS is located dorsally, in arthropods, ventrally. However, despite this apparent axis inversion it is clear that homologous molecular mechanisms pattern the CNS's of the mouse and fly (Holley *et al.*, 1995). In recent years the *Drosophila* ventral nerve cord (VNC) has become a model system for studying the molecular genetic mechanisms that control CNS development, and homologs of many of the major players are found in vertebrates. In *Drosophila* neuroblasts delaminate from the neuroectoderm, subsequently dividing to produce the neurons and glia that comprise the ladder-like CNS. In contrast, the nerve cord of those with a dorsal CNS is derived from the dorsally-located bilaterally symmetric neuroectoderm, where the entire sheet of neuroectoderm folds inwards to ultimately form the brain and spinal cord (see (Arendt and Nubler-Jung, 1999) (see Fig 3.1). Briefly, the development of nerve cord in both insects and vertebrates proceeds as follows. In *Drosophila*, Dorsal protein creates a broad dorsal-ventral gradient (Drier and Steward, 1997), the relative levels of which regulate target genes to establish three embryonic tissues (mesoderm, ectoderm and neuroectoderm) (Stathopoulos and Levine, 2002). The establishment of neurogenic and non-neurogenic sections of ectoderm is achieved by the antagonistic activity of Dorsal target genes in *Drosophila - decapentaplegic (dpp)* which is repressed by high Dorsal levels and *short gastrulation (sog)* which is activated in the presence of low Dorsal levels. Neural progenitor cells (neuroblasts/NBs) arise from the neurogenic portion (see (Arendt and Nubler-Jung, 1999; Cowden and Levine, 2003). In vertebrates, while there is no vertebrate Dorsal homolog, there are homologs of *dpp (BMP-4)* and *sog (chordin)* which play similar roles (Holley *et al.*, 1995). In *Drosophila*, the proneural clusters are arranged into three longitudinal columns (medial, intermediate and lateral) on either side of the midline cells - in vertebrates the orientation of the neural tube means that the medial column comes to lie ventrally, while the intermediate column lies more dorsally (see (Arendt and Nubler-Jung, 1999). Working with the segment polarity genes which govern A/P axis formation, the columnar genes (or D/V specifying genes) subdivides each hemisegment (of ~30 NBs each) into a grid of seven anterior-posterior and three dorso-ventral columns; the induction of the distinct set of downstream genes that specifies the identity of a particular neuroblast in each position is due to the unique combination of expressed genes (see (Skeath and Thor, 2003). The downstream genes effected by expression of each unique combination of segment polarity and columnar

genes are the *AS-C*, *neurogenin* and *atonal* homologs (see (Arendt and Nubler-Jung, 1999).

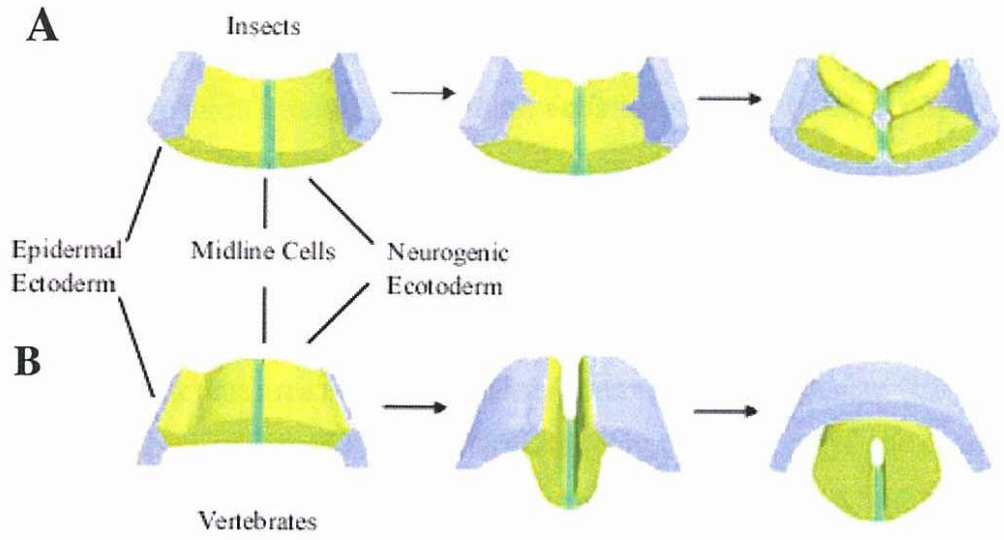


Figure 3.1: Comparison of nerve cord development in insects and vertebrates. Arrows indicate the ontogenic sequence. Epidermal ectoderm (light blue), the midline cells (bright green) and neurogenic ectoderm (khaki green) are shown as indicated (Arendt and Nubler-Jung, 1999).

In *Drosophila,* the genes which subdivide the neuroectoderm into the three longitudinal columns (the columnar genes) are known as *ventral nervous system defective (vnd), intermediate neuroblasts defective (ind)* and *muscle segment homeobox (msh),* and are expressed in the ventral (lateral), intermediate and medial columns respectively. (Skeath *et al.*, 1994; Jimenez *et al.*, 1995; Mellerick and Nirenberg, 1995; D'Alessio and Frasch, 1996; Buescher and Chia, 1997; Isshiki *et al.*, 1997; Mc Donald *et al.*, 1998; Weiss *et al.*, 1998). In vertebrates, where duplication events have resulted in multiple copies of ancestral genes (Holland *et al.*, 1994), the corresponding genes include members of the *nkx2, gsh* and *msx* families. The three Drosophila columnar genes are all activated by the Dorsal protein, but respond differentially – with *ind* requiring a lower level for activation than *vnd* or *msh* (Cowden and Levine, 2003). The expression of *vnd*, *ind*, and *msh* proceeds ventrally to dorsally; *vnd* is expressed first, followed by *ind* and lastly *msh* (von Ohlen and Doe, 2000). A hierarchical cascade of transcription repression also exists between the columnar genes; *vnd* represses *ind* in the ventral

column (Mc Donald *et al.*, 1998; Weiss *et al.*, 1998), and *ind* represses *msh* in the intermediate column (Weiss *et al.*, 1998). Vnd has also been shown to repress *msh* in the procephalic neuroectoderm (Chu *et al.*, 1998; Mc Donald *et al.*, 1998). This cascade of repression is clearly illustrated by experiments in which one of the columnar genes is knocked out - in the absence of *vnd* only 10-20% of the normal number of ventral column neuroblasts form, and those cells express *ind* (Chu *et al.*, 1998; Mc Donald *et al.*, 1998; Weiss *et al.*, 1998). Similarly, in *ind* mutant embryos, only 10% of the normal number of intermediate column neuroblasts form, and their fates change to that of either ventral or dorsal cells (Buescher and Chia, 1997; Isshiki *et al.*, 1997). In *msh* mutant embryos, the neuroblasts of the dorsal column form normally but mis-specification sometimes results in a dorsal-to-intermediate fate change (Buescher and Chia, 1997; Isshiki *et al.*, 1997). In embryos doubly mutant for *vnd* and *ind,* the *msh* expression domain expands ventrally to the midline (von Ohlen and Doe, 2000). However, while the more dorsal gene is affected upon loss of function in any of these genes, there is little effect on transcription of the more ventral gene(s). For example, loss of *msh* function has no effect on *ind* expression and in the *ind* mutants, the *vnd* expression domain expands only slightly to the dorsal side (see (Cornell and Von Ohlen, 2000)). This phenomenon whereby genes expressed in ventral regions repress those expressed in more dorsal regions, but where the repression is not reciprocated, has recently been termed 'ventral dominance'. Ventral dominance also extends to mesodermal and mesectodermal genes, and is in some ways similar to the posterior prevalence model which governs A/P Hox gene expression (Cowden and Levine, 2003). Additional genes most likely regulate these processes, implied by several lines of evidence including the fact that *ind* is not active in the intermediate column after the second wave (SII) of NB formation, (Weiss *et al.*, 1998) and most late-forming medial column NBs develop normally in *vnd* mutant embryos (Chu *et al.*, 1998; Mc Donald *et al.*, 1998). While the identity of these genes is not clear, one candidate regulator is the Sox gene *Dichaete.* Genetic interaction studies show that *Dichaete* enocourages the formation of late-forming NBs in the medial and intermediate columns by working in parallel to *vnd* and *ind* and downstream of *Egfr* (Zhao and Skeath, 2002).

Although expression data are consistent with the notion that the vertebrate D-V columnar genes have similar roles in neural regionalisation as their fly counterparts, unequivocal evidence is often lacking. This is in part due to the presence of gene

duplicates which are often expressed in overlapping domains and thus may be functionally redundant. For example, while the knockout of *msx1* in the mouse neural tube does not result in any remarkable phenotype (Satokata and Maas, 1994), *msx2* and *msx3* are expressed in the same region and thus may compensate for loss of *msx1* function (Wang *et al.*, 1996). In contrast, knockout of mouse *nkx2.2* results in changes in the expression of genes that indicate a ventral-to-intermediate transformation of cell fates (Briscoe *et al.*, 1999), and upon targeted deletion of mouse *nkx2.1*, the resulting phenotypes in the brain are a result of transformation of the palladial primordium (ventral) into a striatal-like anlage (dorsal) (Sussel *et al.*, 1999). Mouse gsh1 and gsh2 (Hsieh-Li *et al.*, 1995; Valerius *et al.*, 1995) and medaka gsh1/2 (Deschet *et al.*, 1998) proteins are all very similar to the Ind protein, and their expression patterns are consistent with roles in specifying neural fates. However, these Gsx genes are expressed after the crucial neural plate stage, suggesting that other genes may have taken over the early *ind* functions in vertebrates. One candidate for this role is *Pax6*, which is expressed in the intermediate neural tube and is repressed by *nkx2.2* (Muhr *et al.*, 2001). Moreover, knockout of *Pax6* function results in dorsal expansion of the *nkx2.2* expression domain (Ericson, 1997). It should be noted that while the *vnd/Nkx*, *ind/Gsh* and *msh/Msx* system appears to be conserved in insects and vertebrates, the upstream patterning pathways differ significantly. For example, in vertebrates, the Sonic Hedgehog (Shh) protein secreted from the midline induces expression of *nkx2.2* and represses *msx*, whereas these interactions do not occur in *Drosophila* (Ericson *et al.*, 1995; Ericson, 1997; Qiu *et al.*, 1998). In addition, while Dpp represses *msh* in *Drosophila*, in vertebrates the Dpp-related proteins BMP2 and BMP4 activate *msx* expression. These (and other) differences have led to the view that the *vnd/ind/msh* system arose early, and is particularly stable in an evolutionary sense - despite major changes in upstream patterning mechanisms, the three regional specification genes have been maintained in both insect and vertebrate evolutionary lineages.

The factors responsible for the initiation of expression and which act to define the dorsal boundaries of the columnar genes are not as well understood. In *Drosophila*, Dpp signalling was thought to be responsible for the repression of *vnd*, and for the establishment of the dorsal border of the *vnd* domain (Mellerick and Nirenberg, 1995), however more recent results directly contradict this and indicate that neither an increase or decrease in Dpp activity affects *vnd* expression (von Ohlen and Doe, 2000). More

likely, both the activation and dorsal limit of *vnd* expression is dictated by a Dorsal protein gradient (von Ohlen and Doe, 2000). As with *vnd,* Dorsal is required for the initiation and maintenance of *ind* expression, however in this case the Egfr signalling pathway is also required (von Ohlen and Doe, 2000). This is evidenced by the fact that ubiquitous ectopic Egfr expression expands the *ind* expression domain dorsally, a phenotype also seen upon elevated Dorsal activity, although in this case this is most likely due to expansion of the *vnd* expression domain (Skeath, 1998; von Ohlen and Doe, 2000). In the absence of Dorsal, the early *vnd* and *ind* expression patterns are abolished (Mellerick and Nirenberg, 1995; von Ohlen and Doe, 2000) and the expression of both *vnd* and *ind* can be induced by ectopic Dorsal expression (von Ohlen and Doe, 2000). Repression by Dpp specifies the dorsal limit of *msh* expression (von Ohlen and Doe, 2000) and recently it has been found that, as with *vnd* and *ind,* an ectopic gradient of Dorsal protein along the A/P axis (generated by expressing a constitutively activated form of the *Toll* receptor (*Toll*^{10B})), can activate *msh* in distinct regions (Cowden and Levine, 2003).

3.1.2 DNA-binding characteristics and mode of repression of vnd/Nk2, ind/Gsh and msh/Msx

Gene-specific transcriptional regulators generally have two activities that determine their function. First, they interact with specific sets of target genes, usually through sequence-specific DNA-binding domains, then they affect the expression of those target genes through activation or repression domains that are often separate and function independently of those involved in DNA binding. Repression of *ind* by Vnd appears to be direct and consensus binding sites for Vnd have been identified in both a putative enhancer (Weiss *et al.*, 1998) and the *ind* promoter (Wang *et al.*, 2002). Repression of *ind* is likely to be mediated by an Eh1 domain present in Vnd, via recruitment of the Groucho (Gro) corepressor. In support of this, Gro has been shown to bind to Vnd *in vitro* (Smith and Jaynes, 1996; Cowden and Levine, 2003). However there are conflicting data on the importance of the Eh1 domain; one report states that deletion of the Eh1 motif of Vnd does not influence the repression of *ind* or *msh* (Koizumi *et al.*, 2003), while another reports that mutations in the Eh1 domain of Vnd abolishes the repression of *ind* and *msh* (Cowden and Levine, 2003). As with Vnd, both Ind and Msh

contain putative Eh1 domains N-terminal of their homeodomains, suggesting that they may also function as Groucho-dependent repressors. The vertebrate Vnd homolog, Nkx2.2 has also been shown to be able to recruit members of the Groucho family of corepressors *in vivo* in the ventral tube (Muhr *et al.*, 2001). Since Nkx2.2 and other vertebrate Gsh and Msx proteins also contain Eh1 motifs, it is possible that Gro-mediated repression may be evolutionarily conserved in CNS patterning.

The repression of *msh* by Ind has not yet been shown to be direct, and the lack of known early Msh target genes complicates investigating its role as a transcriptional repressor. However, both *in vivo* and *in vitro* studies show that the Msx proteins of vertebrates act as potent transcriptional repressors, but this activity is independent of the presence of Msx DNA binding sites; instead, repressive activity appears to be mediated via protein-protein interactions, for example with members of the basal transcription machinery or with other homeodomain proteins (Catron *et al.*, 1995; Semenza *et al.*, 1995; Catron *et al.*, 1996; Zhang *et al.*, 1996; Newberry *et al.*, 1997; Zhang *et al.*, 1997). The nature of these interactions are unclear, however one study implicates the N-terminal arm of the homeodomain. The HD may therefore have functions in protein-protein interaction and/or transcriptional repression, as well as in DNA binding (Zhang *et al.*, 1996). In the case of MSX-1, a number of regions both N- and C-terminal of the homeodomain have also been implicated in transcription repression, including several that are rich in alanine, glycine and proline residues (Catron *et al.*, 1995).

In addition to the Eh1 domain, some NK2-class proteins including those related to Vnd, also contain a second domain located C-terminal of the homeobox termed the NK2-specific domain (NK2-SD). This domain is unique to the NK2-class and contains a proline-rich region, a hydrophobic core with valine or isoleucine in every second position, ((V/I)n(V/I)PVnV) and is flanked by basic amino acids (see (Harvey, 1996). While it does not appear to be necessary for sequence-specific DNA binding *in vitro* (Guazzi *et al.*, 1990; Damante *et al.*, 1994; Watada *et al.*, 2000), the primary structure of the NK2-SD suggests that it might function as an accessory DNA-binding domain or as a protein-protein interface (see (Harvey, 1996). Functional studies indicate this domain masks a transactivation potential of NK2 proteins, supporting a model in which the ability of NK-2-class proteins to activate specific genes during development is dependent upon interactions mediated via the NK2-SD (Watada *et al.*, 2000). The

NK2-SD is separated from the HD by a linker of 9 – 32AA that shows some conservation amongst vertebrate members (Harvey, 1996).

The three-dimensional crystal structure of the Vnd/NK2 HD both alone and bound to its target DNA has been established by 2-D and 3-D nuclear magnetic resonance (NMR) spectroscopy (Tsao *et al.*, 1994; Tsao *et al.*, 1995; Gruschus *et al.*, 1997; Gruschus *et al.*, 1999). A unique feature of the Vnd/NK2 homeodomain is the presence of a tyrosine residue at position 54 (Y54) which seems to be of primary importance in terms of preference of this class of HDs for the consensus sequence 5'-T(T/C)AAGTG(G/C) (Damante *et al.*, 1994; Chen and Schwartz, 1995; Harvey, 1996; Wang *et al.*, 2002) rather than the canonical homeodomain recognition sequence (5'-TAATGG). Mutations at Y54 always decrease affinity of Vnd/Nkx2.2 to the consensus DNA binding site *in vitro* (Weiler *et al.*, 1998) and abolish repression of *ind* and *msh in vivo* (Koizumi *et al.*, 2003). The core recognition sequence (5'-AAGT) plays a major role in determining affinity of binding and is recognised by all members of the NK2 family; it has been suggested that specificity is determined by the sequences flanking the core (Wang *et al.*, 2002). The DNA-binding characteristics of the msh/Msx homeodomain are less well defined, although the crystal structure of the Msx-1 homeodomain/DNA complex suggests that the HD recognises the canonical 5'-TAAT sequence via interactions typical of other classes of homeodomain, and that DNA specificity may be conferred by the bases flanking the core recognition sequence (Hovde *et al.*, 2001). No data is yet available for specifics of the Ind/Gsh homeodomain/DNA complex.

3.1.3 Cnidarian genes related to vnd, ind and msh

Although cnidarians may lack a true Hox system, genes clearly related to each of those responsible for D/V patterning in higher animals are present. The best characterised of these belong to the *ind/Gsh* and thus far, *ind/Gsh*-like genes have been identified from three out of four of the classes of Cnidaria, the anthozoans, hydrozoans and sycphozoans (see Table 3.1). The classification of this gene family as orthologous to *ind/Gsh* has sometimes been contentious (Schummer *et al.*, 1992; Schierwater and Kuhn, 1998), however many studies conclude that the cnidarian *cnox2* genes are orthologous to the *Drosophila* and vertebrate *ind/Gsh*-family, strongly supported by

thorough phylogenetic analyses (Finnerty, 1999; Gauchat *et al.*, 2000; Hayward *et al.*, 2001; Finnerty *et al.*, 2003). Spatial expression patterns of these genes have been investigated in six phylogenetically diverse cnidarians, and with the exception of the colonial zooids, these genes are clearly differentially expressed along the oral-aboral axis. However, considering the conserved expression of *ind/Gsh* orthologs, it is surprising that in some studies expression of *ind/Gsh* orthologs is biased to the aboral region, in others in the oral region (Schummer *et al.*, 1992; Shenk *et al.*, 1993a; Shenk *et al.*, 1993b; Cartwright *et al.*, 1999; Gauchat *et al.*, 2000; Hayward *et al.*, 2001; Yanze *et al.*, 2001). In the anthozoans *Acropora millepora* and *Nematostella vectensis,* expression of the *ind/Gsh* orthologs (*cnox-2Am* and *anthox2*), is preferentially expressed at the oral end of the planula larvae (Hayward *et al.*, 2001; Finnerty *et al.*, 2003), and in *Acropora* the *cnox2*-expressing cells have been putatively identified as neurons on morphological grounds (Hayward *et al.*, 2001). The identity of *ind/Gsh*-expressing cells in other cnidarians have not yet been ascertained. One important aspect of the *Acropora cnox-2Am* expression pattern is that it resembles the expression pattern of its orthologs in *Drosophila* and vertebrates. In *Acropora*, *cnox-2Am* expression is restricted along the O/A axis of planulae; it is most abundant along the sides of the larva and is partially excluded from the oral region, and completely excluded from the aboral region (Hayward *et al.*, 2001). This suggested that the O/A axis of cnidarians might be analogous to the D/V axis of bilaterians, and more intriguingly that the *ind/vnd/msh* repressive cascade might exist in cnidarians and contribute to patterning the nervous system.

Probable orthologs of *vnd* and *msh* have also been identified in cnidarians; a *msh*-like gene is known from *Hydra viridissima* and *Hydra vulgaris* (Gauchat *et al.*, 2000) (Schummer *et al.*, 1992) and a *vnd*-like gene from *Nematostella vectensis* (NK-2) is in Genbank (accession #AAP88430) but expression data are not yet available for these genes. In *Acropora millepora*, two tightly-linked *msx* loci have been identified, but these are likely to represent pseudogenes (see section 3.2.5).

Table 3.1. Cnidarian orthologs of *ind/Gsh* triploblastic genes

Class	Organism	Gene Name	Reference	GenBank Accession #
Anthozoa	*A. millepora*	*cnox-2Am*	(Hayward *et al.*, 2001)	AAK28380
	N. vectensis	*anthox2*	(Finnerty and Martindale, 1997; Finnerty *et al.*, 2003)	AAG37793
Hydrozoa	*E. dichotoma*	*cnox-2-Ed*	(Schierwater *et al.*, 1991)	AAB20573
	P. carnea	*gsx*	(Yanze *et al.*, 2001)	AAG09805
	H. vulgaris	*cnox2*	(Gauchat *et al.*, 2000)	CAB87555
	H. symbiolongicarpus	*cnox2*	(Cartwright *et al.*, 1999)	AAB87073
	H. viridissima	*cnox2*	(Schummer *et al.*, 1992)	CAA45909
Scyphozoa	*A. aurita*	*scox2*	(Schierwater and Kuhn, 1998)	BAA88624
	C. xamachana	*scox2*	(Schierwater and Kuhn, 1998)	AAD32576

n.b: As with the inconsistent nomenclature of other cnidarian gene families, not all genes designated as *cnox2* (for example the *cnox2* gene of *Podocoryne* (Masuda-Nakagawa *et al.*, 2000)) or *anthox2* (for example the *anthox2* gene of *Metridium* (Finnerty and Martindale, 1997)) are orthologous to the *ind/Gsh* families.

3.1.4 Statement of goals

Data from *Drosophila* and vertebrates implies that the *vnd/ind/msh* system arose early in animal evolution (Skeath, 1999). The presence of clear homologs of these genes in cnidarians, and the striking similarity of expression patterns of the *Acropora* and *Drosophila ind* genes (Hayward *et al.*, 2001) suggests that this system may also be present in cnidarians, possibly functioning in patterning along the O/A axis.

To clarify the possibility of such a system predating the cnidarian-"higher" animal divergence, probable orthologs of *vnd* and *msh* were cloned from *Acropora*, and their expression patterns studied both pre- and post-settlement. In addition, the post-

settlement expression pattern of the Acropora *ind* homolog, *cnox-2Am* was studied, and aspects of *cnox-2Am* function studied via transgenic expression in *Drosophila.*

3.2 Results

3.2.1 Isolation of vnd1-Am, vnd2-Am and vnd3-Am

In the course of attempting to identify a *vnd/NK2*-like gene from *Acropora*, three such genes were found. Initially, the isolation of a *vnd/NK2*-like gene from *Acropora* was attempted by PCR using degenerate primers based on conserved segments of NK2-type homeodomain sequences from other organisms. At the time, the *vnd/NK2* ortholog from *Nematostella vectensis (NvNK2)* was not in the database, and therefore could not be used as the basis for primer design. Instead, an alignment was made of the homedomains and NK2-SDs encoded by *vnd/NK2* genes from *Drosophila (*#CAA60619), mouse (#2207236A, #NP_033411), *Caenorhabditis* (#P41936; note this protein does not contain a NK2-SD), zebrafish (#Q90481) and *Xenopus* (#AAB28271); for completeness, the Hydra NK2 (#AAB67611) sequence was also included in the alignment, although the corresponding gene probably belongs to the *tinman/Nkx2.5* gene family (Grens, 1996). Two forward (NKHDfor_A 5'-gcicaracitaygarytrga; NKHDfor_B 5'-gcicaracitaygarytyga) and one reverse (NKHD_rev 5'-taickrtgrttytgraacca) degenerate primer were designed from the homeobox while two reverse primers (NK2SDrev_A 5'-caiggytticcrtciykrac; NK2SDrev_B 5'-caiggytticcrtciykyac) were designed from the NK2-SD. In addition, a pair of degenerate general homeobox primers were employed (genHDfor 5'-gaattcgarytcgaraargarttycat; genHDrev 5'-ggatccttrttytgraaccagatyttdat) (Fig 3.2). PCR experiments were carried out using all possible combination of these primers on each of four stage-specific cDNA libraries (prawn chip, pre-settlement, post-settlement and adult): in addition, first-strand cDNA transcribed from RNA from various stages, and both genomic DNA and genomic library were used as templates. A subset of the NK2-specific primers were also used in combination with vector primers when libraries served as templates. Despite employing a large number of different annealing temperatures and cycling conditions, and wide ranges of Mg^{2+}, template and primer concentrations no specific fragments corresponding to NK2 gene(s) could be amplified. Fortuitously, mixed stage cDNA library for *Nematostella vectensis* then became

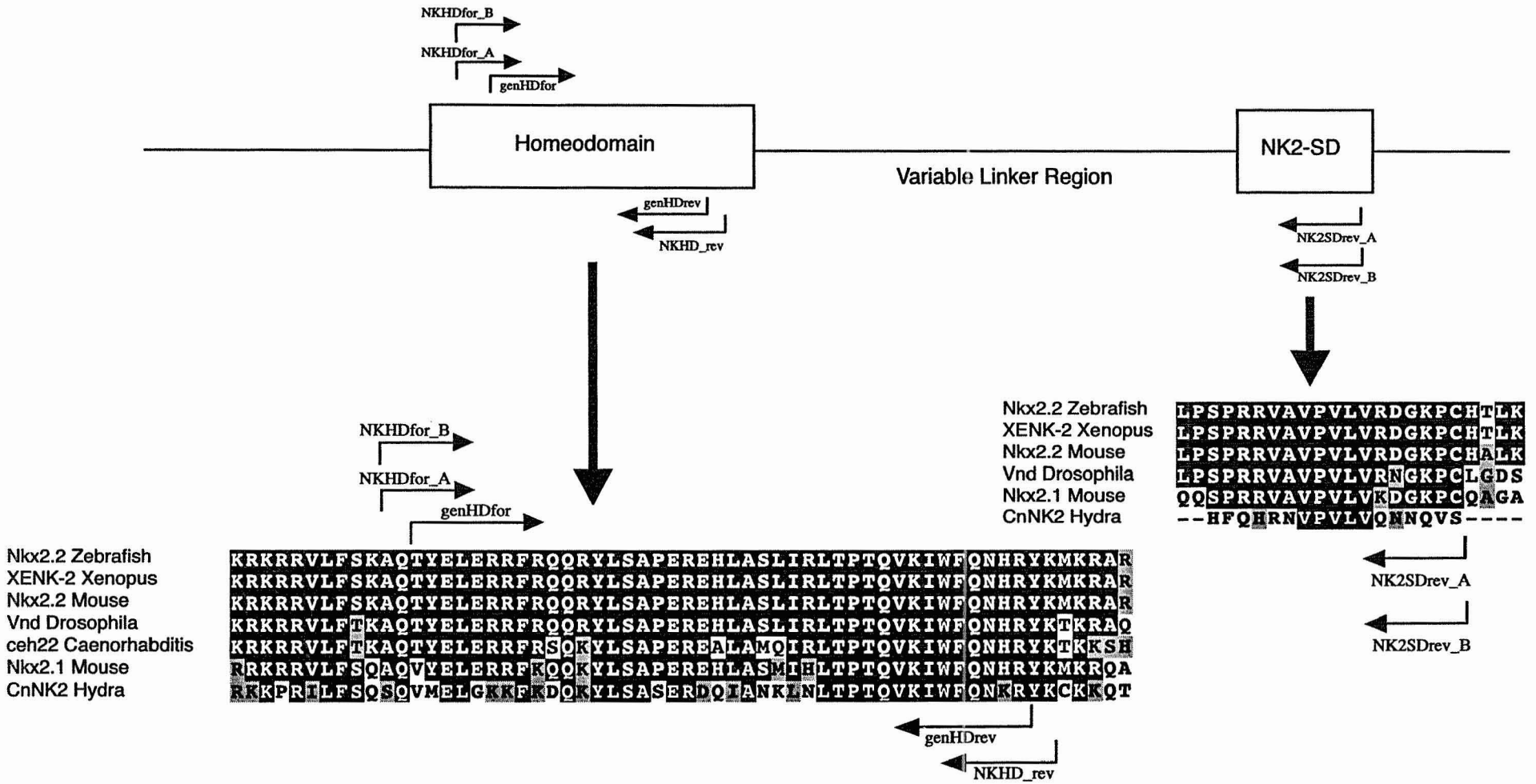
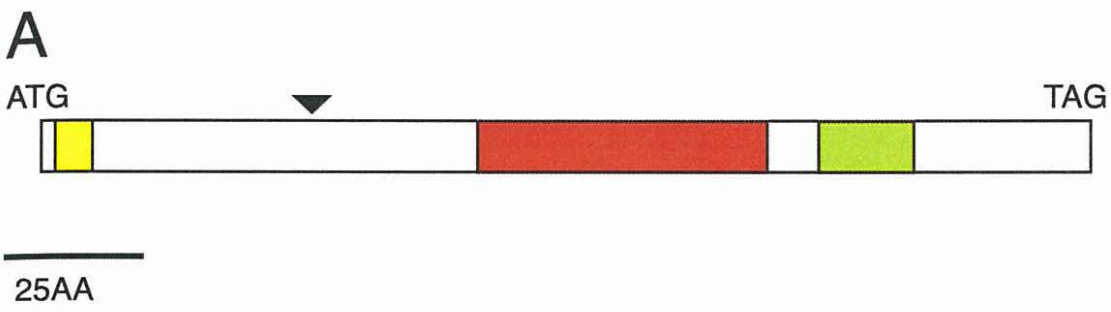


Figure 3.2: Degenerate primers for amplification of a NK2-related gene from *Acropora*. Boxshaded proteins show the alignments used in the design of degenerate primers from either homeodomain or NK2-SD. The schematic diagram shows the relative positions of the protein domains in the hypothetical *Acropora* gene. The primer sequences are as follows; NKHDfor_A 5' gcicaracitaygarytrga, NKHDfor_B 5' gcicaracitaygarytyga, NKHD_rev 5' taickrtgrttytgraacca, NK2SDrev_A 5' caiggytticcrtciykrac, NK2SDrev_B 5' caiggytticcrtciykyac, genHDfor 5' gaattcgarytcgaraargarttycat, genHDrev 5' ggatccttrttytgraaccagatyttdat. Note that figures are not to scale.

available (kindly provided by John Finnerty, Boston University) and the redundant PCR approach was successful when using this as a template. Nested PCR using forward homeobox primer (NKHD_forA) and reverse homeobox primer (NKHD_rev) on bands originally amplified using a forward homeobox primer (NKHD_forA) and reverse NK2SD primer (NK2SD_revA), yielded a homeodomain fragment of the NK2-type. In addition, a second fragment obtained from the first round of PCR was found to correspond to a larger fragment of the gene that encoded the homeodomain, NK2SD and the 11AA linker between these. Several attempts were made to isolate the full length *Nematostella* cDNA clone by screening, however no positive plaques were observed in any of the primary screens. From the *Nematostella* sequences, more specific primers were designed with which to search for vnd/NK2-type genes in *Acropora* based on the assumption that the *Acropora* and *Nematostella* sequences would be more similar to each other than to other *vnd/NK2* genes from higher animals. These primers (NK_for 5'-agacgatttcgcagcaac; Nkfor_C 5'-aargcicaracitaygaryt; NK_rev 5'-attttcacttgtgtaggagtg; Nv_revNK2SD#1 5'-tcagggtttgccgtcgtt; Nv_revNK2SD#2 5'-accagaacgggcacggg; Degen_rev 5'-caiggytticcrtciykrac) were then used on the *Acropora* templates as outlined above. Ultimately experiments using the NKfor_C and NK_rev primers yielded single specific fragments from both the *Acropora* prawn chip and pre-settlement cDNA libraries. Sequencing analysis of these (cloned) PCR products revealed that the fragments were identical, and contained a 120bp fragment from a *vnd/NK2*-like homeobox (from AA position 20 to position 60 of the homeodomain). This fragment was used to screen both cDNA (prawn chip and pre-settlement) and genomic libraries. Repeat screens of the cDNA libraries (~$5x10^5$ pfu's each) using the entire fragment as probe initially proved unsuccessful however a genomic clone was successfully isolated, enabling determination of significantly more of the coding sequence (see section 3.2.3) of the locus hereafter known as *vnd1-Am.* Fragments from the genomic clone were then employed as probes for further screening of cDNA libraries and genomic sequence data used to design primers to enable determination of the 5' and 3' ends of the cDNA by RACE (Rapid Amplification of cDNA Ends). A Virtual northern analysis using a *vnd1-Am* fragment (see Fig 3.3) as probe revealed that the gene was expressed most heavily post-settlement; interestingly, two bands were detected on the blot at positions corresponding to sizes of approximately 1kb and 1.5 kb (Fig 3.4). On the basis of the virtual northern result,



B

```
ATAAGTACTGAAACGTGGGTCGACAGGAGCACGGTAATGAAGGCAATTTGAGATCCAT
GAACGCACGCGATCGAGCGCTTGAGATGACTTCGTTCTCGGTCAAGGATATTCTTAAT
                           M  T  S  F  S  V  K  D  I  L  N

CTTCCAGAAAACAGTATAGGTTTACCCAAAAGTGAAACTACACTTGAATCAACGCAA
 L  P  E  N  S  I  G  L  P  K  S  E  T  T  L  E  S  T  Q

TGCTCCGAAGGAAATTCTCGAAGCCCATCATGTTGTCCTCTCGTATTTAATCAAGTC
 C  S  E  G  N  S  R  S  P  S  C  C  P  L  V  F  N  Q  V

CAAGAGTGCTCCTACACAAATGGACTTCTGAGCCCTACAGAAAAACCTACGACTTCA
 Q  E  C  S  Y  T  N  G  L  L  S  P  T  E  K  P  T  T  S

GCCACAAACGAAGAAATAACAGACCAAGAAGAATCTCAGGGCGATCTGAATAATCCG
 A  T  N  E  E  I  T  D  Q  E  E  S  Q  G  D  L  N  N  P

CCAAAGAAACGCAAACGCCGTGTGCTGTTTAGCAAAGCGCAGACGTACGAGTTTGAG
 P  K  K  R  K  R  R  V  L  F  S  K  A  Q  T  Y  E  F  E

AGACGGTTTCGTCAACAGCGCTATTTATCGGCGCCGGAGCGGGAACAACTCGCACGG
 R  R  F  R  Q  Q  R  Y  L  S  A  P  E  R  E  Q  L  A  R

ATTATAAACTTGAGCCCGACACAAGTCAAAATATGGTTTCAGAATCATCGCTACAAG
 I  I  N  L  S  P  T  Q  V  K  I  W  F  Q  N  H  R  Y  K

TACAAGAAACAGAGTTTTGAAAAAGGAGTGTTAAGTGATGGCGCACCTTTGTTCGCA
 Y  K  K  Q  S  F  E  K  G  V  L  S  D  G  A  P  L  F  A

CCAAGAACTGTACCGGTGCCTGTTTTAGTGCGCGATGGACAGCCATGTGCCCCAGTT
 P  R  T  V  P  V  P  V  L  V  R  D  G  Q  P  C  A  P  V

AATAATTTTGGATATTTACGGCACGATCCTTCAGACTTTTATCATTTTCCCCCTGTA
 N  N  F  G  Y  L  R  H  D  P  S  D  F  Y  H  F  P  P  V

ACATCATCGTACAGCGCCCCTTCAAATCCATATTGGAACTGGTAGCGTAATTAGAGA
 T  S  S  Y  S  A  P  S  N  P  Y  W  N  W  .

AAATATCGTGTGTGTGATTTAGCGAACAAAGCAAGAACAATGGCAGTGCTGTTTCTAT
TGTCGCTTCCGTGAATCTAAAGTCAATTTTAAACCTATACGCCAAAAAAAAAAAAAAA
AAAA
```

Figure 3.3: The *vnd1-Am* cDNA. In both (A) and (B) yellow shading represents the Eh1 motif, red shading the homeodomain and green shading the NK2-SD. The intron position is indicated by a black triangle (A) Schematic representation of the *vnd1-Am* ORF. The start (ATG) and stop (TAG) codons are shown at the top. (B) The nucleotide sequence of the *vnd1-Am* cDNA and the predicted amino acid sequence of the protein. The thick black line indicates the position of probe used for virtual northern blotting and cDNA screening.

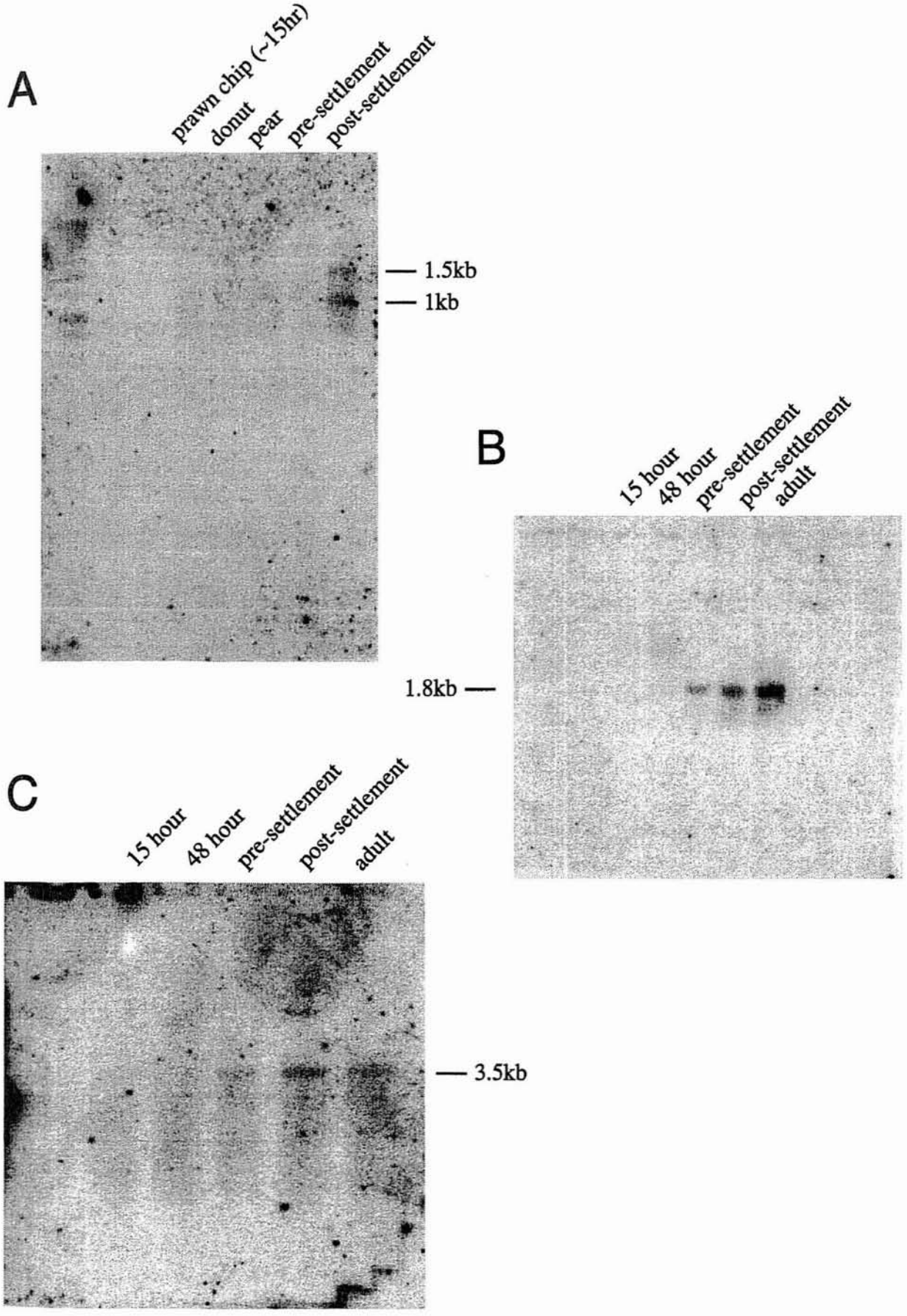
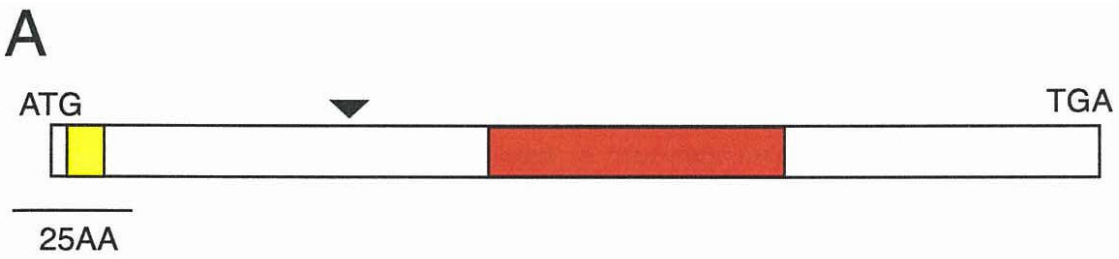


Figure 3.4: Virtual Northern data for *vnd1-Am*, *vnd2-Am* and *vnd3-Am*. (A) Virtual northern analysis of *vnd1-Am*. Two transcripts are visible in pre-settlement and post-setllement stage larvae. (B) Virtual northern analysis of *vnd2-Am*. A single hybridising band of ~1.8kb in pre-settlement, post-settlement and adult stages is visible. (C) Virtual northern analysis of *vnd3-Am* shows a single hybridising band at ~3.5kb in pre-settlement, post-settlement and adult stages.

$5x10^5$ plaques of the post-settlement cDNA library were screened with the 390bp genomic fragment (see Fig. 3.3). Simultaneously, RACE was performed on 5' and 3' RACE-ready post-settlement cDNA kindly provided by Ms Lauretta Grasso (RSBS, ANU). The RACE PCR reactions were performed according to recommendations and reagents provided in the BD SMART™ RACE cDNA Amplification Kit, using specific 5' RACE (vnd_SP1 5' tggggcacatggctgtccatc; vnd_SP3 5' acggcgtttgcgtttctttggc) and 3' RACE (vnd_SP4 5' ccgccaaagaaacgcaaacgcc) primers. From the cDNA screen, two putative positives were isolated and sequenced; one of these was shown to be a *vnd/NK2*-like gene, but was different to *vnd1-Am* and is hereafter referred to as *vnd3-Am,* while the other was a *Hsp70*-like gene not related to the NK2 class. As described in section 3.2.3, a third *vnd/NK2*-like gene, known as *vnd2-Am,* was also isolated in the course of this project by genomic walking from the *vnd1-Am* locus.

Sequence analysis of the RACE products enabled assembly of the complete cDNA sequence of *vnd1-Am* (shown as Fig. 3.3). The *vnd1-Am* cDNA sequence is 839bp, consisting of a predicted ORF of 645bp with 83bp 5'UTR and 114bp 3' UTR. Conceptual translation of the *vnd1-Am* ORF results in a putative protein Vnd1-Am of 215 amino acids. Comparison to other cnidarian *vnd/NK2*-like genes (*Nematostella* NK2 and the other *Acropora vnd* genes) implies that translation is initiated at nucleotide 84 of the contig, despite the fact that the ORF extends further upstream. An Eh1 motif is located near the N-terminus of the predicted protein. The homeodomain is located at positions 75 to 134, and an NK2SD at positions 145 to 164. The ORF is terminated by a stop codon 37AA C-terminal of the NK2SD (see Fig 3.3). The size of the cDNA contig implies that the 1kb signal on the virtual northern blot (Fig 3.4) is likely to correspond to *vnd1-Am*; the identity of the second (~1.5kb) band is unclear, but may be due to cross hybridisation.

In the course of primer walking from the *vnd1-Am* locus, another *vnd/NK2*-type homeobox was identified; hereafter this locus is known as *vnd2-Am.* Virtual northern analysis using a 332bp fragment of the putative *vnd2-Am* ORF as probe (see section 3.2.3 and Fig 3.5), identified a 1.8kb band in pre-settlement, post-settlement and adult stages (Fig 3.4). A *vnd2-Am* cDNA clone was isolated by screening approximately 5 $x10^5$ plaques from the adult cDNA library using the 332bp genomic fragment as probe.



B

```
  GGCGTGTGACCTGTTTCAATCATGATATCCGAGCTCCCGACACAATGGGAGCCCAT
TTGAGACCGTTTACACAAAACGCGAGCAACGCCGGGGATCAAAATGACTTCCTTTTCT
                                                M  T  S  F  S

ATAAAAAACATTCTTAATCTTCCAGAAGACACAATTCGCTCTTTGAATTCCAAAGGC
 I  K  N  I  L  N  L  P  E  D  T  I  R  S  L  N  S  K  G

GAAGATCAGATGTCGTTACAAGAAAGAAGAAGCCTTCTAATTTGTTCCCCGAGCTCA
 E  D  Q  M  S  L  Q  E  R  R  S  L  L  I  C  S  P  S  S

GACGAGCAAGAAGAGGACTCAAGCACGCAAGAAATAGCGAAATCAAGCGGTCTACAA
 D  E  Q  E  E  D  S  S  T  Q  E  I  A  K  S  S  G  L  Q

GTGCTGTCCTCTACAACATCGAGCGCTCAACTTGAAACGAGCAAGAAAGAGCATTCG
 V  L  S  S  T  T  S  S  A  Q  L  E  T  S  K  K  E  H  S

GAGTCTAACAAGAAACGAAAGAGGCGGGTTTTGTTCACGAAAGCGCAGACGTTTGTC
 E  S  N  K  K  R  K  R  R  V  L  F  T  K  A  Q  T  F  V

TTAGAAAAAAGATTTCAACAGCAGCGCTACCTTTCAGCACCCGAGCGGGAAGAGTTG
 L  E  K  R  F  Q  Q  Q  R  Y  L  S  A  P  E  R  E  E  L

GCAAGGATTGTCCATCTCACACCGGCCCAAGTTAAAATTTGGTTTCAAAATCACAGA
 A  R  I  V  H  L  T  P  A  Q  V  K  I  W  F  Q  N  K  R

TATAAATACCGCAAGCAAATTAGCGAACGAGGTCCTTGGGACAAGCCATATTCACCT
 Y  K  Y  R  K  Q  I  S  E  R  G  P  W  D  K  P  Y  S  P

CTATCGCAACCAGCGTTACTTAGCGGACCGGGACAATTTTGCAGCTGCGTAAGCCCA
 L  S  Q  P  A  L  L  S  G  P  G  Q  F  C  S  C  V  S  P

AGTTGCAGCTACCTTGGAAGCGCGTCATATCAAAGCGACTACCTTCACTATGCGACG
 S  C  S  Y  L  G  S  A  S  Y  Q  S  D  Y  L  H  Y  A  T

CCGGTGAATTCATCTTATCTTTCACCGACCTTTTGGTGAATCCGAACTCACGCGCTT
 P  V  N  S  S  Y  L  S  P  T  F  W  .

TGAACTACTCAGTACACTGAACAAATTGTACATAGCTTCTTGTATTGTTTAAAAGGAG
TCATAGTTTATTTATCAAACCGTTTGTCTGTGTTTTAAAAGTAATTGGAGTAGTAAAT
GTCCATTCCTTTTGCGTTTGACTAAAAATAGACTAAGACGATCGTACAAAATTTTGGT
GTAGTGGCGCTATGAAGTAAAAATACACTTGTAAATAGGGTCTGAGAAAACCAAACCA
CCAAAACTCGCAATAACTTTCTAAAAAAGTTTGGCTGTTGTAAACCTAACTCTTGCGG
TTTCCTTGTACAAAGCTATTGAATTCGCGTAGAAAATCTGCAATTTTGCTGGAACTGG
AAAAATAATTTCTTGTACACGGCCATGAAATCAAAGATGAATGATTTGAAGTGCAGAT
TTACCTTCTTCGTTTAAATTTTCCTTGAAGGCGGATATAGATGGAAATTTACTGCGTG
TAGTCAGTCGGTAGACACGTCTTGTGTAATCATCTGTCTTGGTGGTGTTAAAAAGAAC
AGAAAAACCGTCAGTAATTTTCATTATGGGCAGCAATAAACAGAATTCATTTGAAAAA
AAAAAAAAAAAAAAAAA
```

Figure 3.5: The *vnd2-Am* cDNA. In both (A) and (B) yellow shading represents the Eh1 motif and red shading the homeodomain. The intron position is indicated by a black triangle (A) Schematic representation of the *vnd2-Am* ORF. The start (ATG) and stop (TGA) codons are shown at the top. (B) The nucleotide sequence of the *vnd2-Am* cDNA and the predicted amino acid sequence of the protein. The thick black line indicates the position of probe used for virtual northern blotting and cDNA screening.

The single *vnd2-Am* cDNA clone isolated is probably artifactual – the nucleotide sequence consists of two cDNAs cloned in a 'head-to-head' arrangement with the poly-A tails at either end of the insert (see Fig 3.6). This organisation has been reported for other cDNA clones in the library (D. Hayward, personal observations) although the exact cause is not clear. The part of the clone corresponding to the *vnd2-Am* transcript is at the T3 end of the vector, and extends to approximately the centre of the clone. The artifactual clone appears not to encode the entire Vnd2-Am protein; the N-terminal region of the protein is missing, including the start codon. For this reason, 5' RACE was performed using cDNA prepared from mixed developmental stages as template with specific primers designed to *vnd2-Am* (vnd2_SP1 5' ggtgaatatggcttgtcccaagg; vnd2_SP2 5' tcccgctcgggtgctgaaagg) and the reagents supplied in BD SMART™ RACE cDNA Amplification Kit. This approach yielded a fragment of ~400bp which could be aligned with the known cDNA sequence, resulting in a contig (see Fig 3.5) corresponding to near full length *vnd2-Am* transcript. The contig consists of a 99bp 5' UTR followed by an ORF of 624bp ORF and a 590bp 3' UTR. However, the size of the *vnd2-Am* contig assembled from the 5'-RACE and cDNA sequences (1313 bp) differs significantly from the size of the transcript implied from virtual northern analysis (~1.8kb), suggesting the 5'RACE product does not accurately reflect the actual size of the *vnd2-Am* 5'UTR. Conceptual translation of the *vnd2-Am* ORF results in a predicted protein of 208 amino acids. As in the Vnd1-Am protein, an Eh1 motif is located at the N-terminus of the Vnd2-Am protein. However, unlike Vnd1-Am and most other proteins of the Vnd/NK2 type, the Vnd2-Am protein does not contain an NK2SD.

As described above, the *vnd3-Am* clone was serendipitously isolated when a cDNA library was screened using a *vnd1-Am* fragment as probe. The 1.9 kb *vnd3-Am* cDNA clone consists of an 12bp 5' UTR followed by the presumed coding sequence of 648bp (encoding a predicted protein of 216 amino acid residues) and 3' UTR of 1200bp (Fig 3.7). Note that although there are no in frame stop codons upstream of the putative translation start site, comparison with the *Nematostella* NvNK2 and *Acropora* Vnd1-Am and Vnd2-Am proteins implied that the start codon corresponds to nt 19-21 in the contig. The Vnd3-Am protein contains both an N-terminal Eh1 motif and a C-terminal NK2SD as well as the NK2-type homeodomain (Fig 3.8). Virtual northern analysis of *vnd3-Am* (Fig 3.4) using the entire cDNA clone as probe shows expression in pre-

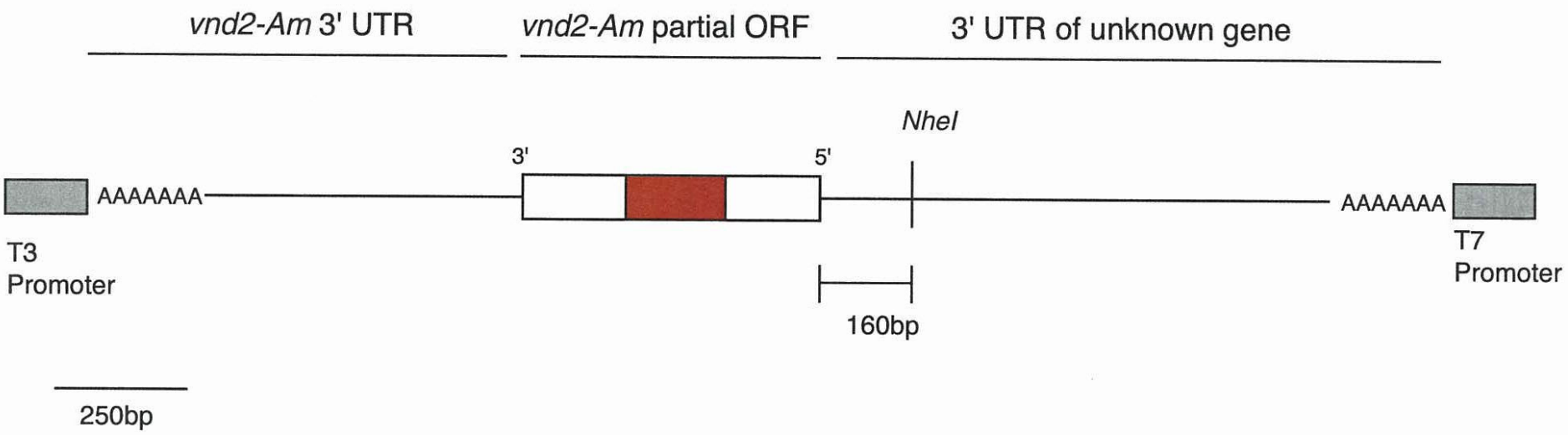


Figure 3.6: Organisation of the *vnd2-Am* cDNA clone isolated from the post-settlement cDNA library. The red shaded box indicates the *vnd2-Am* homeobox while the grey shaded boxes indicated the T3 and T7 promoter sequences showing the orientation of the clone. The two poly-A tails of the clone is shown, the one closest to the T3 promoter belongs to *vnd2-Am* and the other to an unknown gene. The *Nhe I* site used to linearise the plasmid in preparation for DIG-labelling is shown; 160bp of the 5' region of the unknown gene was included by necessity when making the DIG-labelled RNA probe. (Scale bar = 250bp).

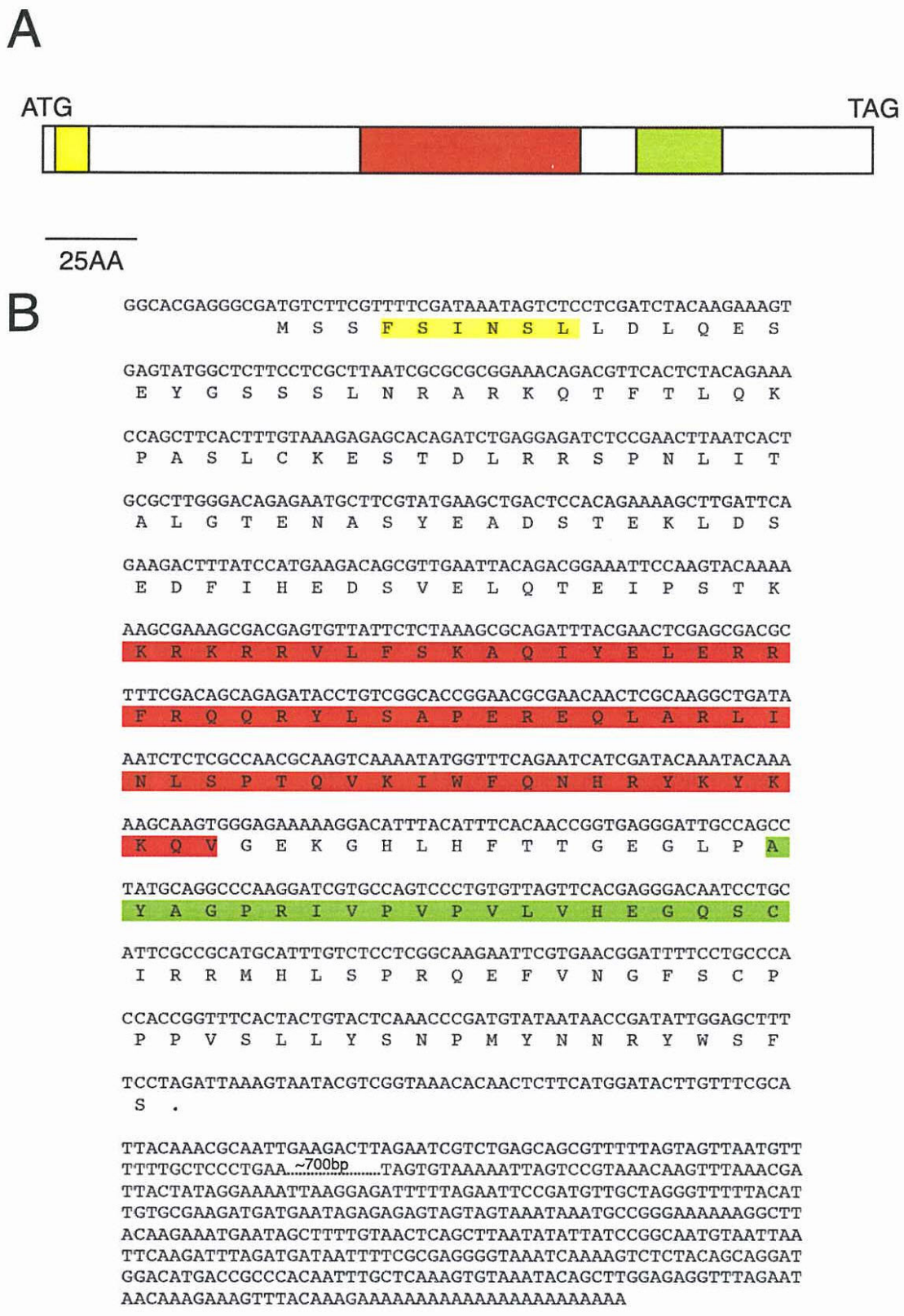


Figure 3.7: The *vnd3-Am* cDNA. In both (A) and (B) yellow shading represents the Eh1 motif, red shading the homeodomain and green shading the NK2-SD. The intron position is indicated by a black triangle (A) Schematic representation of the *vnd3-Am* ORF. The start (ATG) and stop (TGA) codons are shown at the top. (B) The nucleotide sequence of the *vnd3-Am* cDNA and the predicted amino acid sequence of the protein.

settlement, post-settlement and adult stages as indicated by a hybridising band at ~3.5kb; the size of this band implies that the *vnd3*-Am UTRs may be substantially longer than those represented in the cDNA clone.

3.2.2 The Vnd1-Am, Vnd2-Am and Vnd3-Am proteins – assignment to the Vnd/NK2 family using phylogenetic analysis

The homeodomains of Vnd1-Am, Vnd2-Am and Vnd3-Am unambiguously belong to the NK2 class, and phylogenetic analysis revealed they are most closely related to the Vnd/NK2 type. The HDs in these three *Acropora* proteins are all very similar; Vnd1-Am and Vnd3-Am show 93/97% identity/similarity, and the corresponding figures for Vnd1-Am versus Vnd2-Am and Vnd2-Am versus Vnd3-Am are 78/88% and 77/92% respectively.

In addition to the homeodomain, these *Acropora* proteins contain motifs characteristic of the Vnd/NK2 class. As in *Drosophila* Vnd and its vertebrate homologs Nkx2.1/2.4 and Nkx2.2/2.9, each of the *Acropora* Vnd proteins contains an Eh1 motif at the N-terminus, and in each case the phenylalanine residue known to be important for interactions with Groucho (Cowden and Levine, 2003) is present (Fig 3.8). An NK2-SD is present in Vnd1-Am and Vnd3-Am, but interestingly is absent from the Vnd2-Am protein (Figs 3.3, 3.5 and 3.7).

To better understand the evolutionary relations between the Vnd/NK2 type proteins, the homeodomain sequences were subjected to maximum likelihood phylogenetic analysis using MolPhy Version 2.3 (Adachi and Hasegawa, 1996). In these analyses, there is a clear clade consisting of the three *Acropora* Vnd proteins and *Nematostella* NK2, but the cnidarian clade is clearly grouped with arthropod and vertebrate Vnd/Nkx2.2 proteins (see Fig. 3.8) and these are well-resolved from the Tinman/Nkx2.5 clade (which includes the *Hydra* NK2 protein, CnNK2). The *Acropora* Vnd2-Am and *Nematostella* NK2 proteins are more similar to each other than either are to Vnd1-Am or Vnd3-Am, suggesting that these may be orthologous. If this hypothesis is correct, the absence of an NK2-SD in Vnd2-Am, but not in *Nematostella* NK2 suggests secondary loss in *Acropora*.

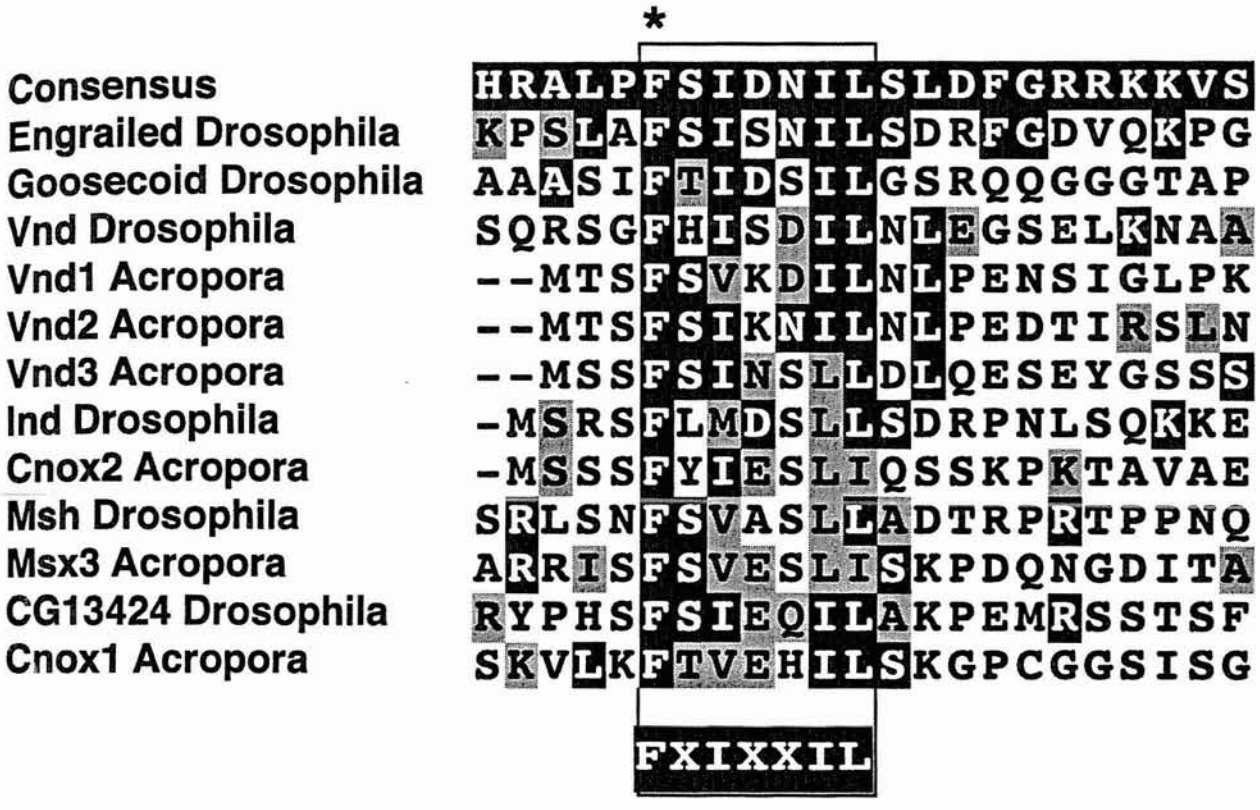
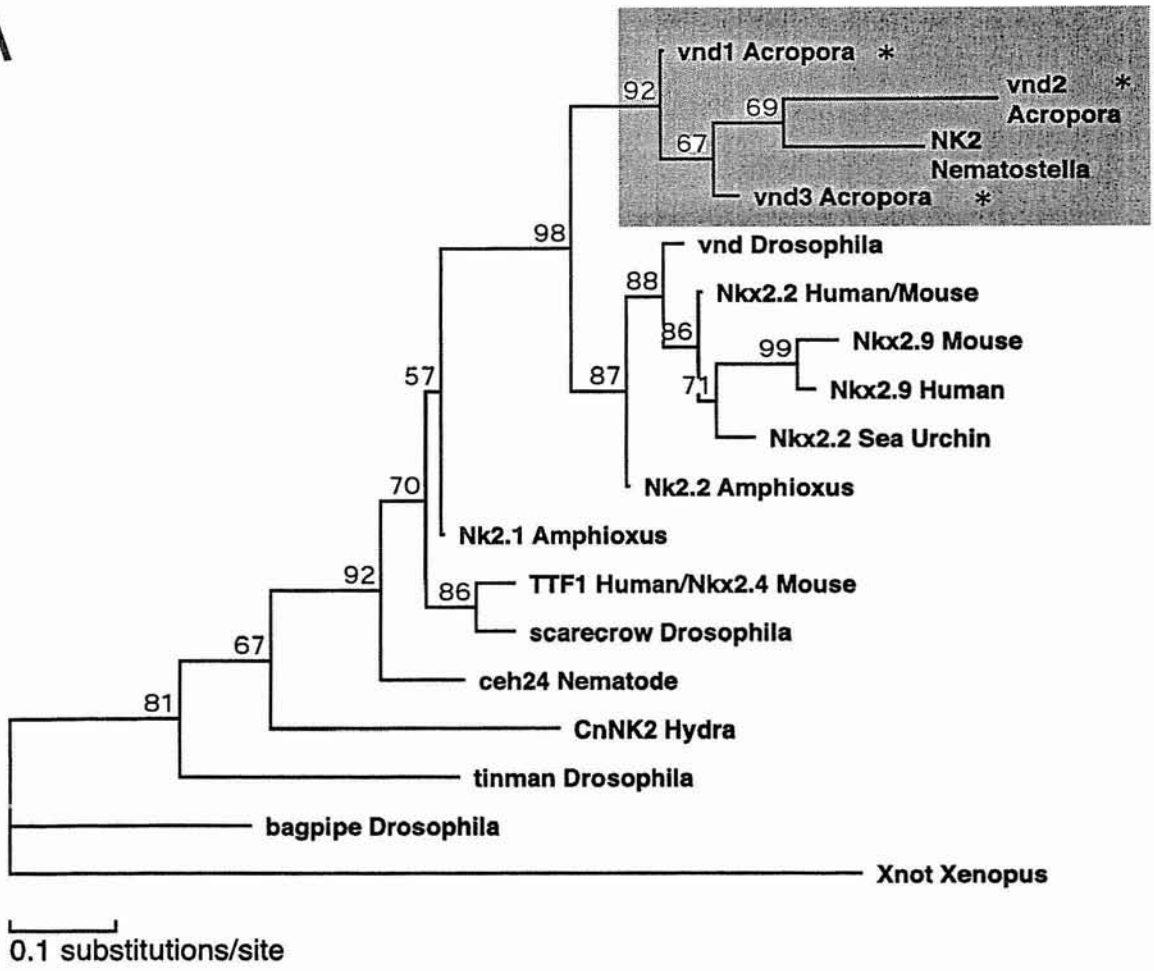


Figure 3.8: Comparison of Eh1 motifs from *Acropora* proteins with related *Drosophila* proteins. A Boxshade alignment of the Eh1 motif sequences from the *Acropora* Vnd, Msx3-Am and Cnox1-Am proteins as compared to the consensus Eh1 motif. The asterisk represents the phenylalanine residue important for interactions with Groucho. Below the alignment, the core consensus of the Eh1 motif is shown with X's indicating variable amino acid positions. Database accession numbers of the sequences are as follows: *Drosophila melanogaster,* Engrailed (P02836), Goosecoid (P54366), Vnd (CAA60619), Ind (AAK77133), Msh (Q03372), CG13424 (NP_611491). (Adapted from Cowden and Levine, 2004).

Figure 3.9: Comparison of the *Acropora* Vnd proteins to other Vnd/Nk2-type proteins. (A) Vnd1-Am, Vnd2-Am, Vnd3-Am and related Vnd/NK2 homeodomains were analysed by Maximum-Likelihood phylogenetic analysis in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff model of protein evolution and local rearrangement of the NJ trees. Numbers against branches indicate the percentage of 1000 bootstrap replicates supporting topology. The shaded box indicates the clade consisting of cnidarian Vnd-related sequences, while stars indicate the *Acropora* proteins. The Xnot homeodomain from *Xenopus* served as an outgroup. (B) A Boxshade alignment of the homeodomain sequences used in the phylogenetic analyses. The star indicates the invariant tyrosine residue at position 54 of the homeodomain (Y54) characteristic of all NK2 proteins. (C) A boxshade alignment of the NK2-SD's of various Vnd/NK2-type proteins. The stars indicate the residues that make up the hydrophobic core of the NK2-SD. In both (B) and (C) identical residues are shaded black and conserved substitutions are shaded grey. The column to the right of the alignments indicates the overall identity and similarity of each protein with the Vnd1-Am protein. The species name and GenBank Accession number of each protein used are as follows: *Nematostella vectensis* NvNK2 (AAP88430); *Drosophila melanogaster* Vnd (CAA60619), Scarecrow (AAF26436), tinman (AAQ75417), bagpipe (AAQ74415); *Homo sapiens* Nkx2.2 (O95096), Nkx2.1/TTF1 (NM_003317), Nkx2.9 (AF000297); *Mus musculus* Nkx2.2 (2207236A), Nkx2.9 (O70584), Nkx2.4 (AF202039); *Strongylocentrotus purpuratus* (sea urchin) Nkx2.2 (AAS58444), Nkx2.1 (AAM94862); *Branchiostoma floridae* (amphioxus) Nkx2.2 (AAD01958), Nkx2.1 (AAC35350); *Caenorhabditis elegans* (nematode) ceh24 (NM074018); *Hydra vulgaris* CnNK2 (AAB67611); *Xenopus laevis* Xnot (CAA79629), NK2 (AAB28271); *Gallus gallus* Nk2.2 (AAD04630); *Danio rerio* (zebrafish) Nk2.2 (Q90481). Note that human TTF1/Nkx2.1, mouse Nkx2.1, human Nkx2.4 and mouse Nkx2.4 all have identical sequences.

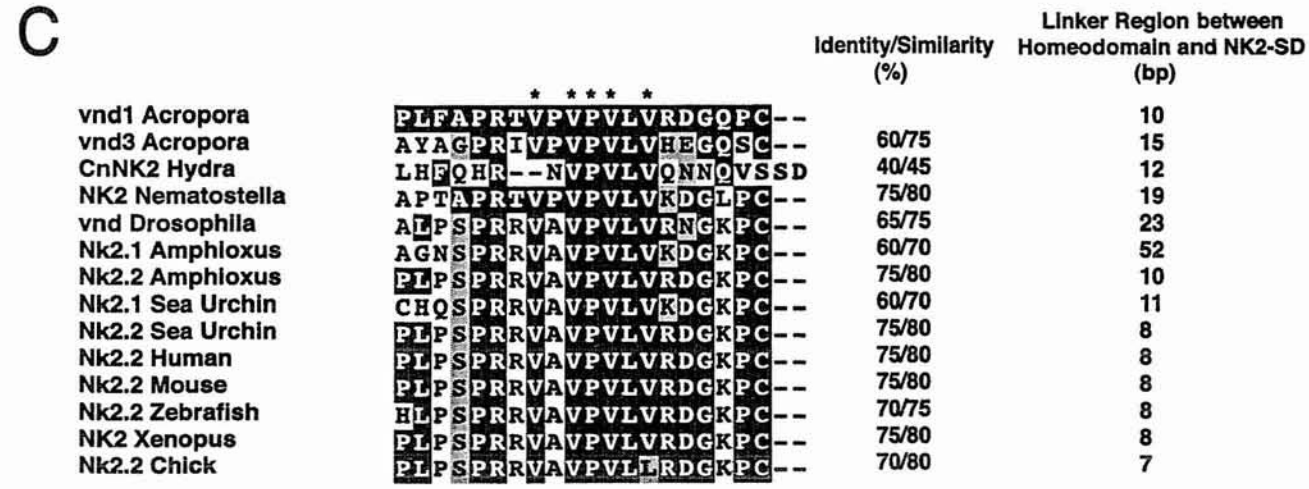
A


vnd1 Acropora *
vnd2 Acropora *
NK2 Nematostella
vnd3 Acropora *
vnd Drosophila
Nkx2.2 Human/Mouse
Nkx2.9 Mouse
Nkx2.9 Human
Nkx2.2 Sea Urchin
Nk2.2 Amphioxus
Nk2.1 Amphioxus
TTF1 Human/Nkx2.4 Mouse
scarecrow Drosophila
ceh24 Nematode
CnNK2 Hydra
tinman Drosophila
bagpipe Drosophila
Xnot Xenopus
92
69
67
98
88
86
99
71
57
87
70
92
86
67
81
0.1 substitutions/site

B


Identity/Similarity (%)
vnd1 Acropora
vnd2 Acropora
vnd3 Acropora
NK2 Nematostella
vnd Drosophila
Nkx2.2 Human
Nkx2.2 Mouse
Nkx2.9 Mouse
Nkx2.9 Human
Nkx2.2 Sea Urchin
Nkx2.2 Amphioxus
Nkx2.1 Amphioxus
Nkx2.1 Human
Nkx2.4 Mouse
scarecrow Drosophila
ceh24 Nematode
CnNk2 Hydra
tinman Drosophila
bagpipe Drosophila
78/90
93/95
80/93
83/90
85/90
85/90
83/92
80/90
88/90
88/93
82/92
78/90
78/90
80/92
78/88
62/85
63/75
60/80

C


Identity/Similarity (%)
Linker Region between Homeodomain and NK2-SD (bp)
vnd1 Acropora
vnd3 Acropora
CnNK2 Hydra
NK2 Nematostella
vnd Drosophila
Nk2.1 Amphioxus
Nk2.2 Amphioxus
Nk2.1 Sea Urchin
Nk2.2 Sea Urchin
Nk2.2 Human
Nk2.2 Mouse
Nk2.2 Zebrafish
NK2 Xenopus
Nk2.2 Chick
60/75
40/45
75/80
65/75
60/70
75/80
60/70
75/80
75/80
75/80
70/75
75/80
70/80
10
15
12
19
23
52
10
11
8
8
8
8
8
7

3.2.3 The vnd1-Am and vnd2-Am genomic loci

Screening approximately 50 000 pfu's of the *Acropora* genomic library with a 120bp *vnd1-Am* fragment (see section 3.2.1) yielded a large number (60) of primary positives, one of which was chosen at random and purified for sequencing. Sequencing of the clone was initiated by the use of both internal primers and vector primers, and was continued using a primer walking strategy. During the course of sequencing the clone, a second NK2-type homeobox gene (*vnd2-Am)* was identified approximately 9kb upstream of *vnd1-Am.* The two genes are orientated in a 'head-to-head' arrangement, with the start methionines (see above) approximately 7kb apart (see Fig 3.10). Each gene contains a single intron; *vnd1-Am* has a 406bp intron located 101bp upstream of the homeobox, while *vnd2-Am* has a 652bp intron located 80bp upstream of the homeobox. Both introns conform to the GT-AG rule. Screening (~50 000 pfu's) was also conducted for the *vnd3-Am* locus, using the cDNA as a probe. A number (19) of putative positives were isolated although time constraints precluded their characterisation.

3.2.4 Spatial expression patterns of vnd1-Am, vnd2-Am and vnd3-Am

The spatial expression patterns of each of the three *Acropora vnd* genes were analysed by *in situ* hybridisation. DIG-labelled riboprobes were generated from the appropriately linearised plasmids (see section 2.5.4). For *in situ* hybridisation of *vnd1-Am*, a cocktail of two antisense probes was used. This consisted of riboprobes transcribed from linearised plasmids containing the 5'-RACE or 3'-RACE products using either SP6 or T7 RNA polymerase as appropriate. A *vnd2-Am*-specific riboprobe was generated using T3 RNA polymerase after first digesting the plasmid containing the partial cDNA clone with *NheI* so that transcription would be halted before continuing into the second cDNA located on the same clone (see Fig 3.6). The *vnd3-Am* riboprobe was generated from the cDNA clone transcribed from the T7 promoter. Only pre-settlement and post-settlement embryos were used for *in situ* hybridisation experiments, as only these stages gave clear hybridisation signals on virtual northern blots (see Fig 3.4). In addition to routine use of DIG-labelled riboprobes in combination with either NBT/BCIP or BM

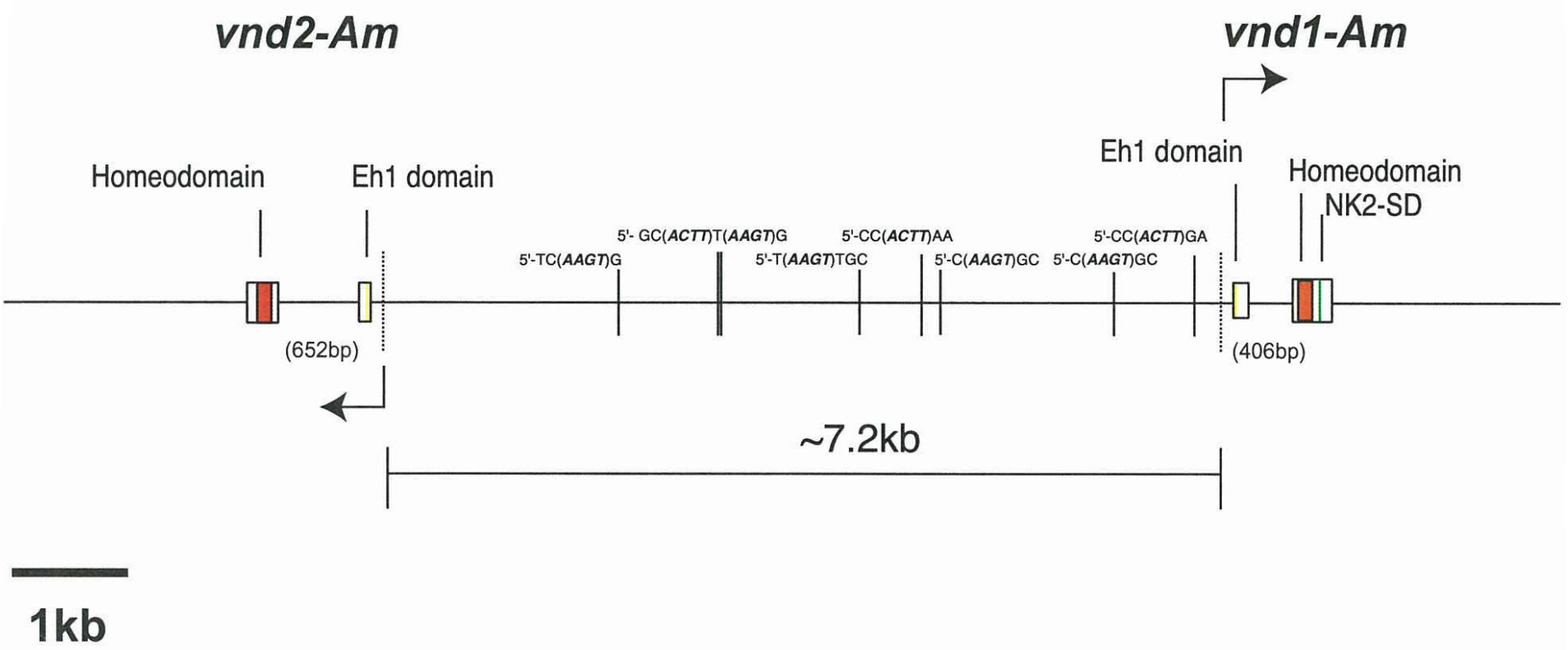


Figure 3.10: Genomic Organisation of the *vnd1-Am* and *vnd2-Am* locus. Schematic representation of the lambda 3.2 genomic clone. Red shaded boxes indicate homeodomains, the yellow line the Eh1 motifs and the green line the NK2-SD. The putative transcription start sites are marked with dashed lines. Coding regions are shown as boxes. *Vnd1-Am* and *vnd2-Am* are interrupted by one intron each. The total size of each intron is shown in brakcets. Direction of transcription is indicated by the arrows. The length between the putative transcription start sites (see section 3.3.2) is shown (~7.2kb). Lines between the putative transcription start sites indicate consensus NK2 binding sites on either strand (with the relevant sequence shown) which concur fully with the consensus binding site (5' T(T/C)AAGTG(G/C) or those which differ in only one of the nucleotides which flank the 5' AAGT core (see section 3.3.2).

purple substrates, other approaches (Fl-labelled riboprobes and the Sigma® Red substrate) were also employed in the case of *vnd3-Am,* which showed the most consistent pattern of expression. Note that for cases where probe and substrate were compatible (see section 3.3.4), results obtained were consistent between all three genes.

The expression patterns seen for each of the three *Acropora vnd* genes were essentially indistinguishable, and are therefore described together. In pre-settlement embryos, expression is first seen in the invaginated ectoderm that represents the inward opening oral pore (Fig 3.11A(i)–(ii)). As development proceeds, the embryo flattens to create the characteristic disc-shaped post-settlement form. At this stage of development, expression continues in the ectoderm surrounding the oral pore. In some embryos with the multi-lobed 'flower-like' appearance caused by formation of the mesenteries, in addition to the usual pattern around the oral pore (Fig 3.11C(i)-(ii)), expression can also be seen in a band extending partially on each side of the embryo (Fig 3.11B(i)–(iii)). Sectioned embryos show staining in the ectodermal tissue more clearly (Fig 3.11D). Virtual northern analysis indicated that the *vnd* genes are also expressed in adult material (this is true at least for *vnd2-Am* and *vnd3-Am*; for *vnd1-Am* the blot used did not include adult cDNA), however the existing *in situ* hybridisation protocols are not applicable to adult colonies. Specific cell types could not be distinguished in these *in situ* experiments, although this does not rule out the possibility of expression in the nervous system. It may be that expression is not limited to nerve cells, but also occurs in other cell types that define the oral pore.

3.2.5 Isolation of msx3-Am

Prior to the work described here, a tightly-linked pair of *msh*-like homeobox genes, (*msx1-Am* and *msx2-Am*) had been identified in Acropora (Hislop *et al.*, 2005). During the course of this project, a substantial period of time was spent unsuccessfully screening for cDNAs corresponding to these loci. Several other lines of evidence as well as these failed screening attempts lead us to believe that these loci represent pseudogenes. Quantitative RT-PCR using mRNA from all major developmental stages did not yield signals significantly above background. Moreover, the predicted amino

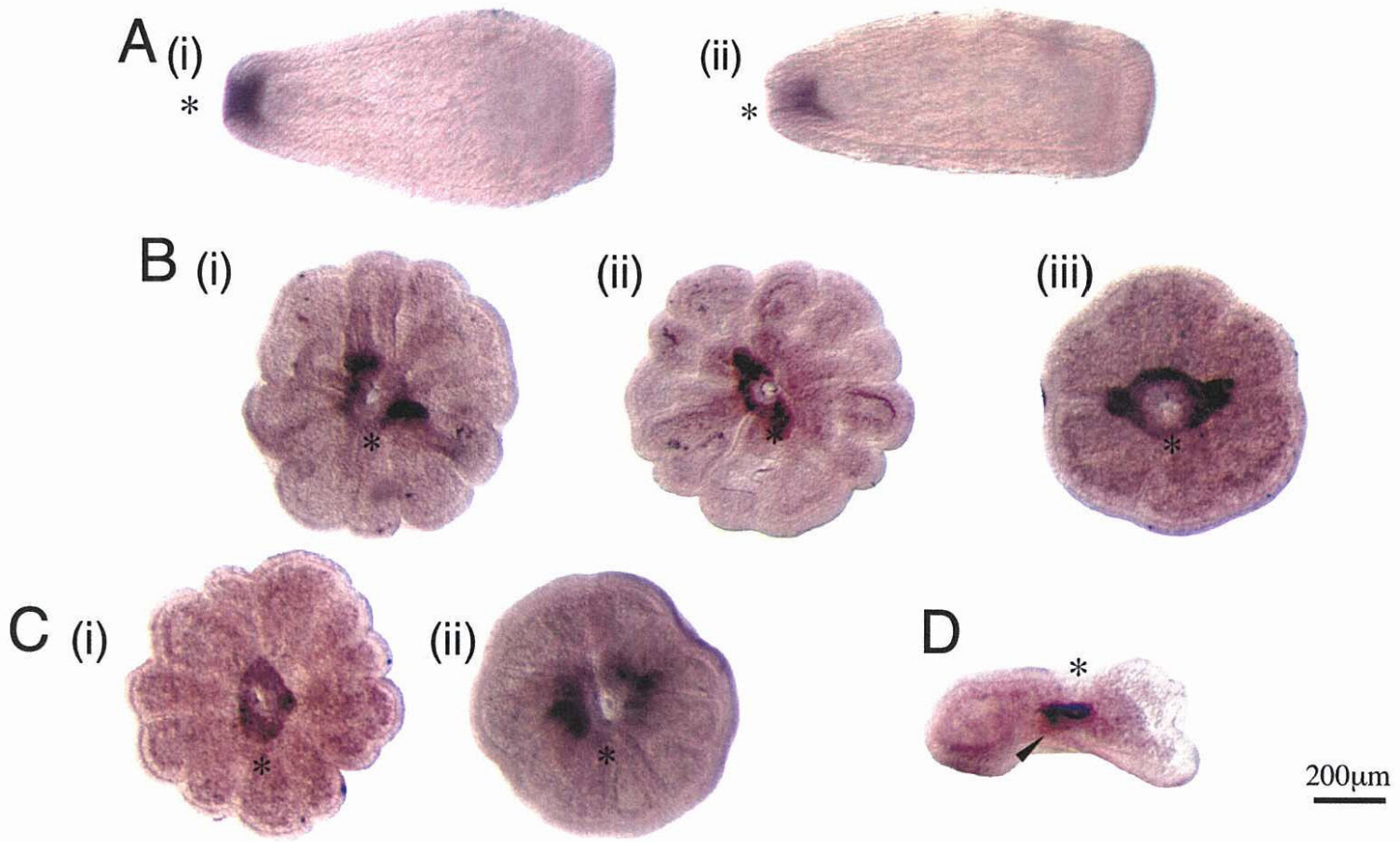


Figure 3.11: Spatial expression patterns of the *vnd* genes in *Acropora* pre-settlement and post-settlement. *In situ* hybridisation pattern observed when embryos are hybridised with either *vnd1-Am, vnd2-Am* or *vnd3-Am* DIG-labelled riboprobes. An asterisk marks the position of the oral pore in each example. Expression initiates in pre-settlement planula larvae (A(i) - (ii)) and is restricted to the oral pore. A(i) shows expression in a whole mount embryo, while A(ii) shows expression in a cross-section of a larvae of similar age which more clearly shows expression in the ectodermal tissue surrounding the oral pore. Expression in post-settlement polyps varies from an eye-shaped region of expression (B(i) - (iii)) to a tight band of expression (C(i)) or a wider band (C(ii)). In all studied, expression is in the ectoderm surrounding the oral pore, best shown in a cross-section of a post-settlement polyp (D). In this example, some bleeding of substrate from cells results in a smear of signal as shown by the arrowhead.

acid sequences of the *Acropora* Msx HDs are highly derived and feature some dramatic amino acid substitutions that are likely to be structurally important (Hislop *et al.*, 2005).

In order to investigate whether additional Msx loci are present in *Acropora*, PCR was conducted using degenerate primers designed from an alignment of Msx-type homeodomains. As in the case of those designed for *vnd* genes (see above), degeneracy was minimised and inosines used in positions of full redundancy. Forward primers were designed to a conserved 8AA region from 4AA N-terminal of the homeodomain to AA position 4 (msx_def1 5'-aarcayaargciaaymgiaarcc; msx_def2 5'-aarcayaaraciaaymgiaarcc). In addition, two further forward degenerate primers were designed from a comparison of codon usage in *Acropora* in order to lower degeneracy further (msx_def3 5'-aarcataargcwaaymgraarcch; msx_def4 5'-aarcataaracwaaymgraarcch). A single reverse primer was utilised in the amplification reactions (msx_der2 5'-ckrttytgraaccadatytt) (Fig 3.12). After numerous attempts on various templates and differing conditions, a single 180bp homeobox fragment was obtained when using first-strand cDNA from post-settlement larvae as template. This fragment was found to correspond to a novel *msx*-related gene hereafter known as *msx3-Am*. Several attempts to isolate an *msx3-Am* clone by screening pre- and post-settlement cDNA libraries ($5x10^5$ pfu's of each) using the 180 bp fragment as probe failed. However a genomic clone was successfully identified using the same screening strategy, enabling the generation of a larger (380 bp) putative ORF fragment (by PCR) for use in cDNA library screening (see Fig 3.13). Screening $5x10^5$ plaques of the post-settlement cDNA library allowed the identification of 5 putative positives from the primary screen. Upon sequencing, two of these clones gave clear matches to the *msx3-Am* fragment. However, although the longer of the clones was 1.3kb, it did not appear to contain the complete coding sequence, as no in frame start codon was present upstream of the homeobox. For this reason, RACE was performed on 5' RACE ready cDNA generated from mixed stage embryos using *msx3-Am* specific primers (msx_SP1 5'-catgagcataaggctggaaagcg; msx_SP4 5'-cttcaggcattcgtccaaggctg). A 750bp 5' RACE fragment was cloned and sequenced, and the contiguous alignment of this fragment with the longest cDNA clone obtained resulted in an *msx3-Am* cDNA contig, consisting of 145bp 5' UTR, 774bp ORF and 650bp 3' UTR (Fig 3.13). Translation of the putative ORF gives a protein Msx3-Am of 258AA, containing an Eh1 motif and a Msx-

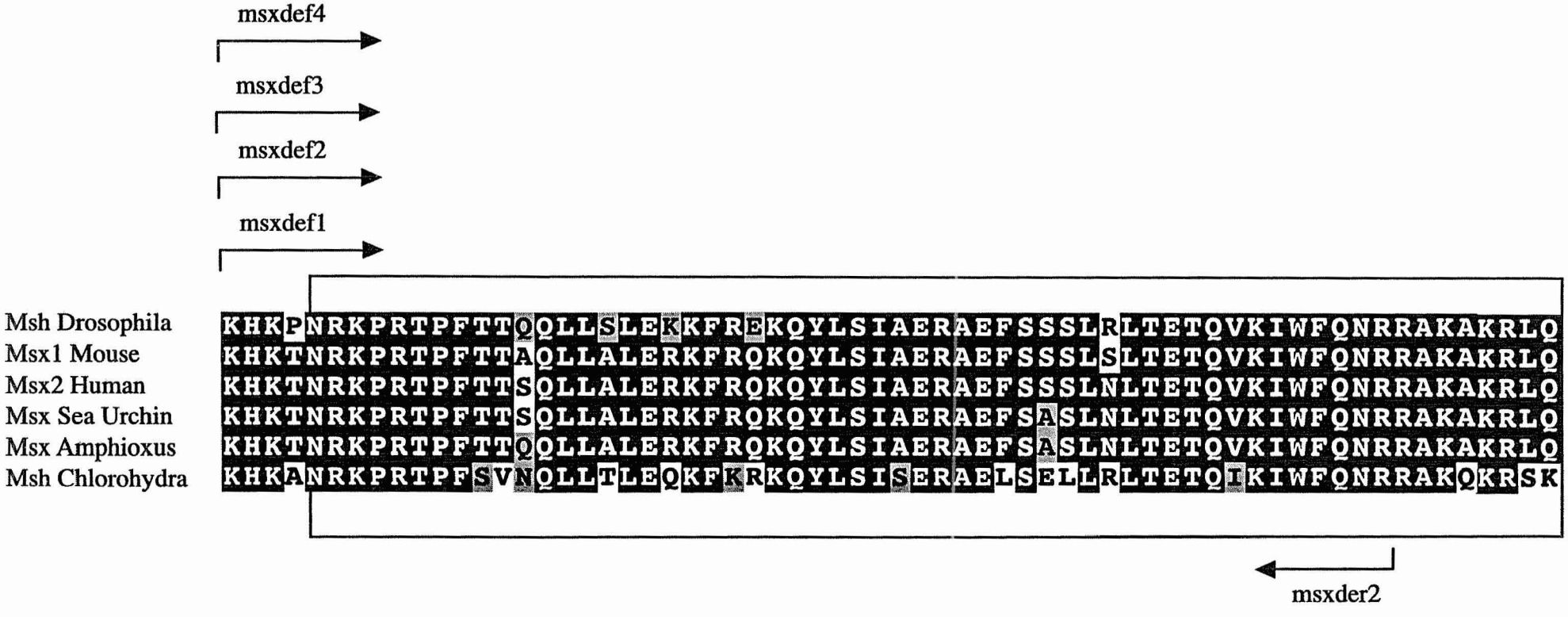


Figure 3.12: Degenerate primer positions for amplification of a *msx*-related gene from *Acropora*. The boxshaded protein depicts an alignment of msx-type homeodomains and a portion N-terminal of this from various organisms. The boxed section indicates the homeodomain. Primer positions are shown. The primer sequences are as follows; msx_def1 5'aarcayaargciaaymgiaarcc, msx_def2 5' aarcayaaraciaaymgiaarcc, msx_def3 5' aarcataargcwaaymgraarcch, msx_def4 5' aarcataaracwaaymgraarcch, msx_der2 5' ckrttytgraaccadatytt. Accession numbers are as follows; *Drosophila melanogaster* msh (#Q03372); *Mus musculus* (mouse) Msx1 (#P13297); *Homo sapiens* Msx2 (P35548); *Strongylocentrotus purpuratus* (Sea Urchin) Msx (AAB97688); *Branchiostoma floridae* (Amphioxus) Msx (#CAA10201); *Chlorohydra viridissima* Msh (S20896).

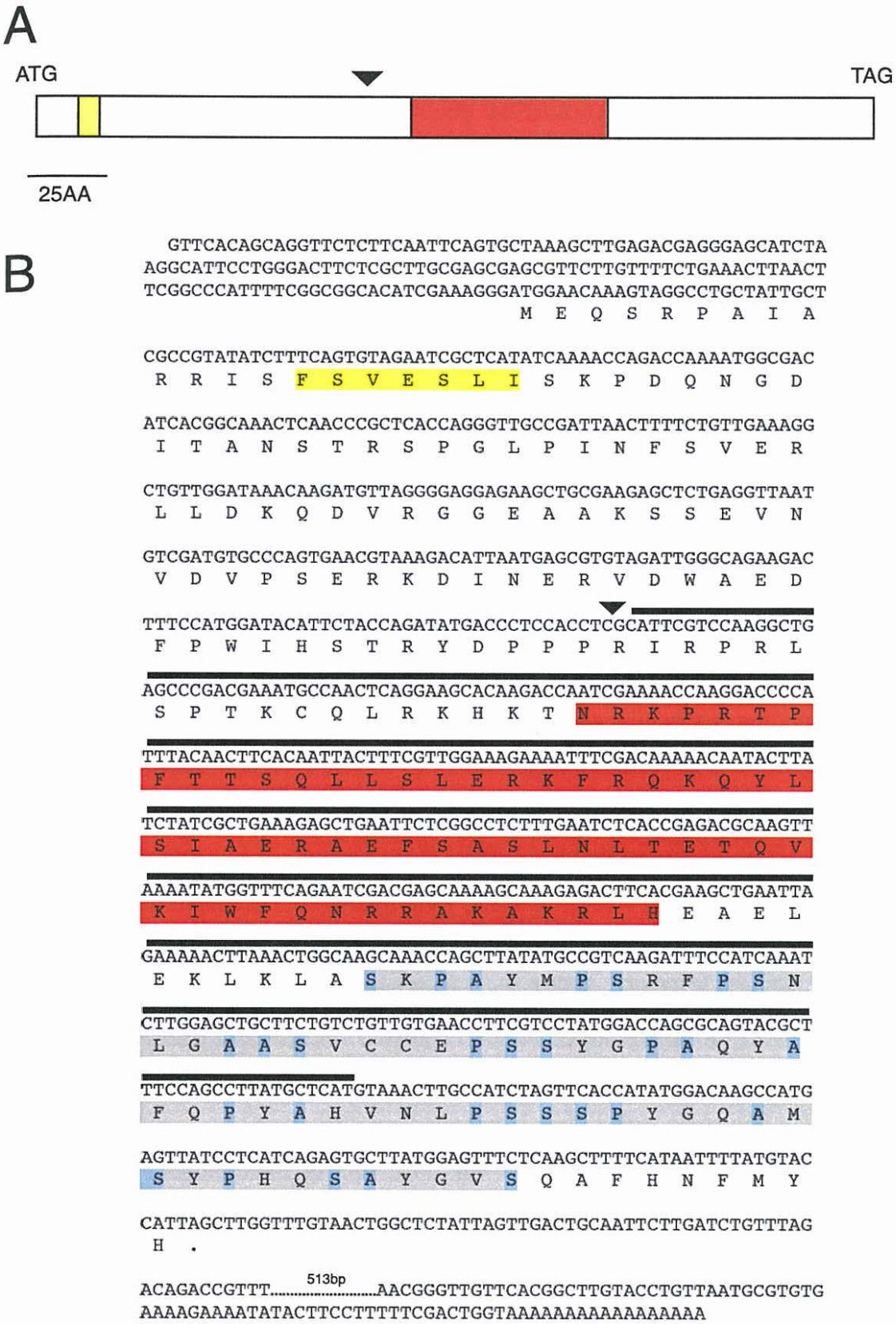


Figure 3.13: The *msx3-Am* cDNA. In both (A) and (B) yellow shading represents the Eh1 motif and red shading the homeodomain. The intron position is indicated by a black triangle (A) Schematic representation of the *msx3-Am* cDNA. The start (ATG) and stop (TAG) codons are shown at the top. (B) The nucleotide sequence of the *msx3-Am* cDNA and the predicted amino acid sequence of the protein. The thick black line indicates the position of probe used for virtual northern blotting and cDNA screening. The grey shaded area indicates the proline/serine/alanine-rich region with the proline, serine and alanine residues indicated in blue.

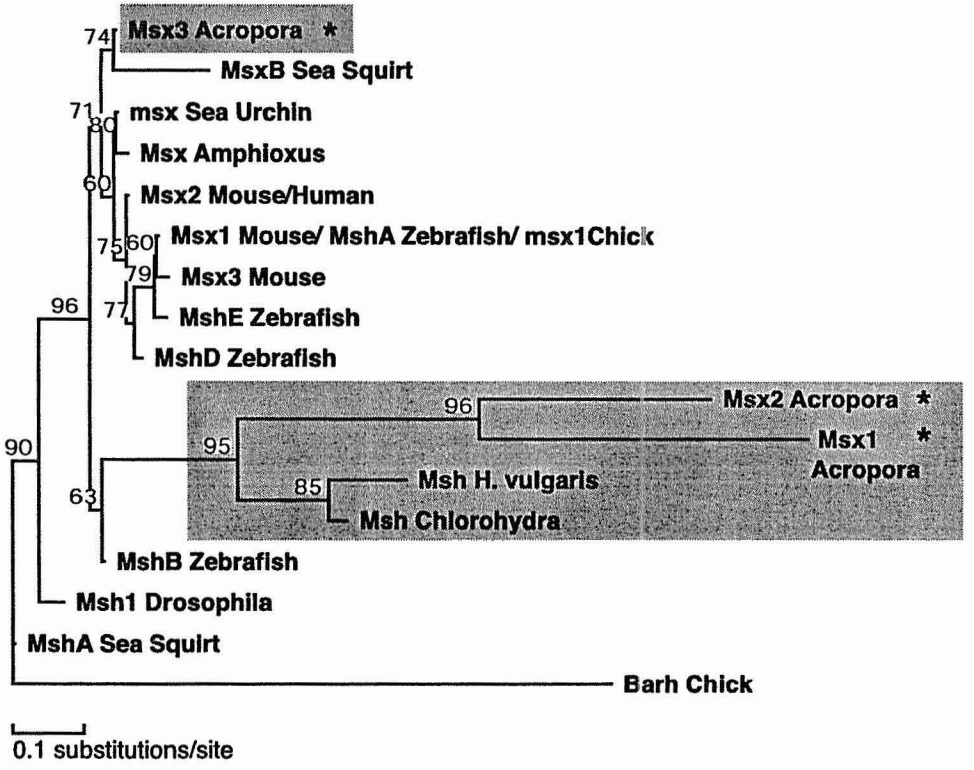
type homeodomain. The homeodomain lies between AA residues 113 and 172 of the Msx3-Am protein and is located 81 residues N-terminal of the stop codon. In addition to the Msx-type homeodomain, Msx3-Am contains an Eh1 motif as do other members of the msh/Msx family (Fig 3.13 and Fig 3.8). The motif displays the conserved phenylalanine residue essential for interactions with the Groucho family of corepressor proteins (Fig 3.8) (Cowden and Levine, 2003).

3.2.6 The Msx3-Am protein – assignment to the Msh/Msx family using phylogenetic analysis

The homeodomain of the Msx3-Am protein is a clear member of the msh/Msx family, and is most similar to that present in the sea urchin *Strongylocentrotus purpuratus* SpMsx (96.6%/98.3% identity/similarity). Surprisingly, in general the Msx3-Am HD is more similar to those in vertebrate msh/Msx proteins (eg with mouse and human Msx1 (92/97% identity/similarity)) than to the invertebrate members of this group including its nominal orthologs in the hydrozoan cnidarians *Chlorohydra viridissima* (73/83% identity/similarity) and *Hydra vulgaris* (72/80% identity/similarity) (see Fig 3.14B). In order to better understand the evolutionary position of Msx3-Am, maximum likelihood phylogenetic analyses were performed using representative vertebrate and invertebrate msh/Msx-like sequences (Adachi and Hasegawa, 1996). The HDs encoded by the two putative *Acropora* pseudogenes (Hislop *et al.*, 2005), Msx1-Am and Msx2-Am were also included in these analyses. From these analyses (Fig 3.14A) it can be seen that the vertebrate Msx sequences form a well supported and distinct clade and the *Acropora* Msx3-Am homeodomain groups with the msxB HD of the sea squirt (*Ciona intestinalis*). The Msx1-Am and Msx2-Am sequences are clearly well diverged and group with the two Hydra msh proteins; the length of the branches containing Msx1-Am and Msx2-Am suggests divergence and reinforces their assignment as pseudogenes.

Figure 3.14: Comparison of the *Acropora* Msx proteins to other msh/Msx-type proteins. (A) Msx1-Am, Msx2-Am, Msx3-Am and related msh/Msx homeodomains were analysed by Maximum-Likelihood phylogenetic analysis in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff model of protein evolution and local rearrangement of the NJ trees. Numbers against branches indicate the percentage of 1000 bootstrap replicates supporting topology. The shaded box indicates the clade consisting of cnidarian Msx-related sequences, while stars indicate the *Acropora* proteins. The Barh homeodomain from chick served as an outgroup (accession # NP_989524). (B) A Boxshade alignment of the homeodomain sequences used in the phylogenetic analyses. Identical residues are shaded black and conserved substitutions are shaded grey. The column to the right of the alignment indicates the overall identity and similarity of each protein with the Msx3-Am homeodomain. The species name and GenBank Accession number of each protein used are as follows: *Ciona intestinalis* (sea squirt) MsxB (CAB42631), MsxA (CAD56691); *Strongylocentrotus purpuratus* (sea urchin) msx (AAB97688); *Branchiostoma floridae* (Amphioxus) Msx (CAA10201); *Mus musculus* Msx2 (Q03358), Msx1 (P13297), Msx3 (P70354); *Homo sapien* Msx2 (P35548); *Danio rerio* (zebrafish) MshA (NP_571349), MshE (NP_571348), MshD (NP_571351), MshB (Q03356); *Gallus gallus* msx1 (P28361); *Hydra vulgaris* Msh (CAB88390); *Chlorohydra viridissima* Msh (S20896); *Xenopus laevis* Xhox7.1 (P35993); *Drosophila melanogaster* Msh1 (Q03372).

A



B

		Identity/Similarity %
msx3 Acropora	NRKPRTPFTTSQLLSLERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKRLH	
msx2 Acropora	DRSPRIPFSSSQLASFEAKFLQTQYLSGSEVHDLSSWLSVSEHRVKIWFQNRRAREKKTK	53/72
msx1 Acropora	SGSPRVPFTAEQLAKLENSYKESHYVTSSQATQLSSMLKLPESRVKIWFQNRRAREKKKR	42/63
msh H. vulgaris	NRKPRTPFSVNQLLALEQKFKRKQYLSISERSELSVMLRLTETQRKIWFQNRRAKQKRRT	72/80
msh Chlorohydra	NRKPRTPFSVNQLLTLEQKFKRKQYLSISERAELSELLRLTETQIKIWFQNRRAKQKRSK	73/83
msx Sea Urchin	NRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKRLQ	97/98
Xhox7.1 Xenopus	NRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSSSLNLTETQVKIWFQNRRAKAKRLQ	95/98
Msh1 Drosophila	NRKPRTPFTTQQLLSLEKKFREKQYLSIAERAEFSSSLRLTETQVKIWFQNRRAKAKRLQ	90/95
Msx2 Mouse/Human	NRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSSSLNLTETQVKIWFQNRRAKAKRLQ	95/98
Msx1 Mouse	NRKPRTPFTTAQLLALERKFRQKQYLSIAERAEFSSSLSLTETQVKIWFQNRRAKAKRLQ	92/97
MshE Zebrafish	NRKPRTPFSTAQLLALERKFRQKQYLSIAERAEFSSSLSLTETQVKIWFQNRRAKAKRLQ	90/97
MshD Zebrafish	NRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSSSLTLTETQVKIWFQNRRAKAKRLQ	93/97
Msx3 Mouse	NRKPRTPFTTAQLLALERKFHQKQYLSIAERAEFSSSLSLTETQVKIWFQNRRAKAKRLQ	90/97
Msx Amphioxus	NRKPRTPFTTQQLLALERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKRLQ	95/97
MshB Zebrafish	NRKPRTPFSTSQLLSLERKFRQKQYLSIAERAEFSNSLNLTETQVKIWFQNRRAKAKRLQ	95/97
MshA Sea Squirt	NRKPRTPFTTQQLNSLEKKFREKQYLSIAERAEFSNSLSLTETQVKIWFQNRRAKSKRLQ	87/93

3.2.7 Isolation of a msx3-Am genomic clone

Screening of the genomic library (~1x10^4 pfu's) with a 180bp *msx3-Am* PCR product (see section 3.2.5) led to the identification of 23 putative positives, one of which was isolated, purified and subjected to DNA sequencing. The DNA sequence of the entire ~14kb λmsx2.2 clone was achieved via primer walking, initiated with both vector and internal *msx3-Am* primers. The complete *msx3-Am* ORF was present in the clone, and was found to contain a single (233bp) intron between nt 295 and 296 in the cDNA contig (see Fig 3.13). This position corresponds to between AA residues 97 and 98 in the Msx3-Am protein (i.e. 18AA residues N-terminal of the homeodomain). The intron conforms to the GT-AG rule.

3.2.8 Spatial expression pattern of msx3-Am

The expression pattern of *msx3-Am* in *Acropora* embryos was analysed by *in situ* hybridisation. A DIG-labelled riboprobe was generated from a linearised plasmid template encoding the longest *msx3-Am* clone, and used in hybridisation on pre- and post-settlement larvae. Expression was seen in the ectoderm, but does not appear to be limited to a specific cell type (Fig 3.15). As in the case of the *Acropora vnd*-like genes, this does not discount the possibility of expression in the nervous system, but rather indicates that the transcript is not restricted to neurons. In pre-settlement stages, *msx3-Am* is expressed in the ectoderm covering approximately two-thirds of the embryo from the oral end, leaving approximately one-third of the aboral end of the embryo completely free of expression (Fig 3.15A-C). Expression in some embryos appears to be restricted from the oral pore/gastric cavity (Fig 3.15A, E-E(i), G). In post-settlement larvae *msx3-Am* is expressed in the ectoderm on the oral surface but the transcript appears to be absent from the aboral surface (Fig 3.15E-E(i)). Expression appears to be concentrated in the growing tip of the polyp, which continues to grow upwards to ultimately form the adult colony (Fig 3.15F, G).

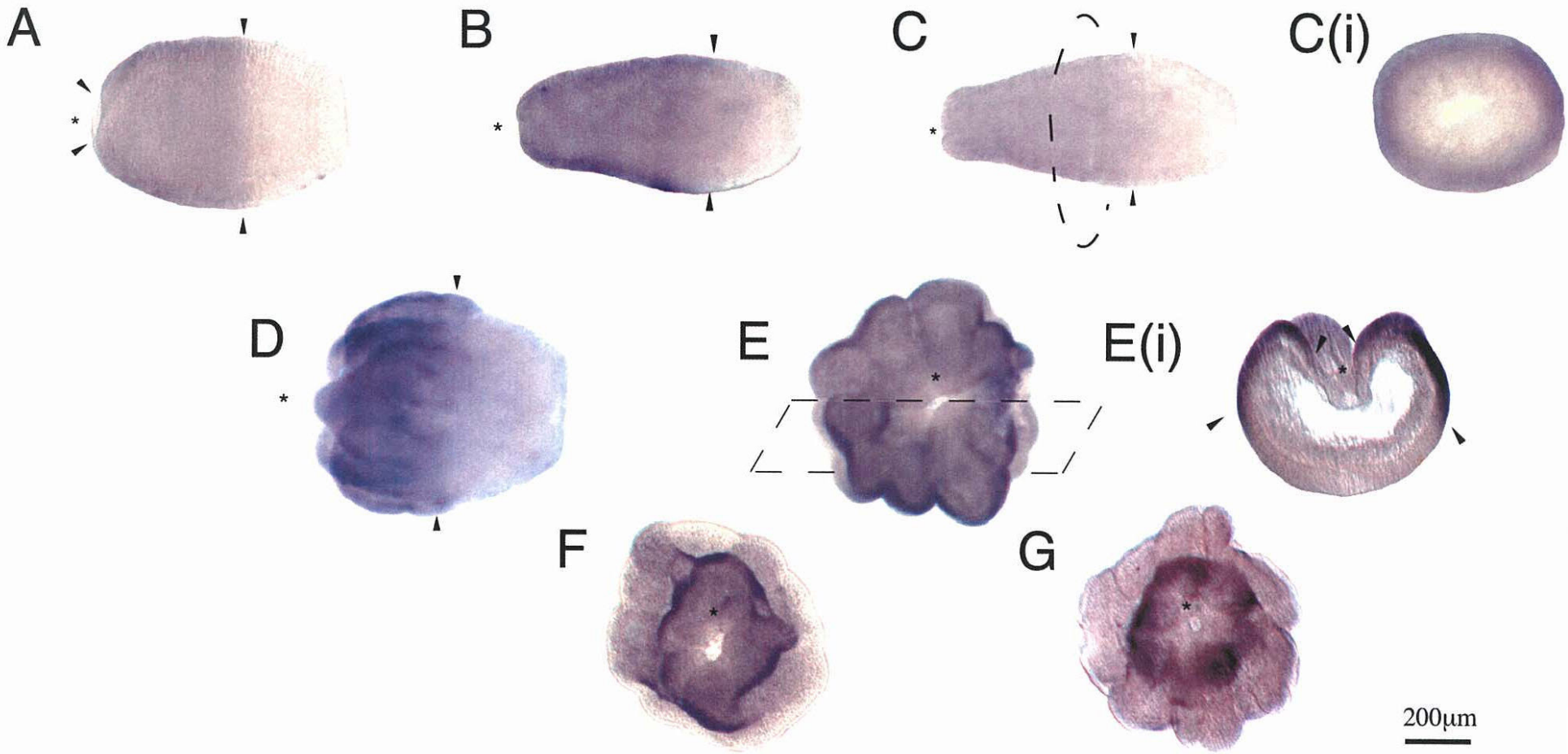


Figure 3.15: Spatial expression pattern of *msx3-Am* during development. *In situ* hybridisation pattern observed in pre- and post-settlement embryos when hybridised with a DIG-labelled *msx3-Am* probe. Asterisks show the oral pore where appropriate. Cross-sections are shown by the dotted lines in (C) and (E). Corresponding sectioned embryos are shown in (C)(i) and (E)(i) which show the ectodermal expression more clearly. A-C show expression pre-settlement whereas D shows expression in the 'mushroom' stage, and E-G show expression post-settlement. The arrowheads show the limits of *msx3-Am* expression which in some cases appears to be restricted from the oral pore (see (A), E(i) and (G)). Older post-settlement polyps with a growing tip showing concentrated *msx3-Am* expression are shown as (F) and (G).

3.2.9 Spatial expression pattern of cnox-2Am in post-settlement embryos

As discussed above, the pre-settlement expression pattern of the *Acropora* ind/Gsx ortholog *cnox-2Am* has previously been described by our group (Hayward *et al.*, 2001). However, nothing was known about the post-settlement expression of this gene. In order to enable comparisons to be made with data for the *Acropora vnd* and *msx* genes, *in situ* hybridisation experiments were performed on post-settlement stages using a DIG-labelled *cnox-2Am* using riboprobe kindly supplied by Dr David Hayward (RSBS, ANU).

Cnox-2Am expressing cells are seen in the ectoderm surrounding the oral pore in the region of the growing tip (see Fig 3.16A-D). These cells were of the same morphology as those previously described in pear stage embryos and planula larvae; one type is bipolar with its nucleus located approximately halfway across the ectoderm, connected to both the surface ectoderm and basement membrane by thin cytoplasmic projections (Fig 3.16E), while a second type of cell is seen to contain the nucleus and bulk of the cytoplasm near the basement membrane, clearly connected to the surface of the cytoplasm (Fig 3.16F(i)-(iii)) – both cell types are consistent with previous descriptions of putative neurons (Hislop, unpublished observations; (Chia and Koss, 1979; Fautin and Mariscal, 1991; Grimmelikhuijzen and Westfall, 1995).

3.2.10 Conservation of function between cnox-2Am and Drosophila ind

Mutant rescue experiments – in which a mammalian gene is expressed in a fly line that is defective in the corresponding gene – have often been used to investigate conservation of function (for examples see (Leuzinger *et al.*, 1998; Rincon-Limas *et al.*, 1999)). There are also a few examples of this approach being used to investigate relationships between cnidarian genes and their nominal *Drosophila* counterparts – eg. the hydra *nanos* (Mochizuki *et al.*, 2000) and *achaete-scute* (Grens *et al.*, 1995) genes. As described above, comparison of expression patterns implies that *cnox-2Am* may be orthologous with the *Drosophila ind* and its vertebrate counterparts, the *Gsx* genes (Hayward *et al.*, 2001). To investigate the possibility of conservation of function, the

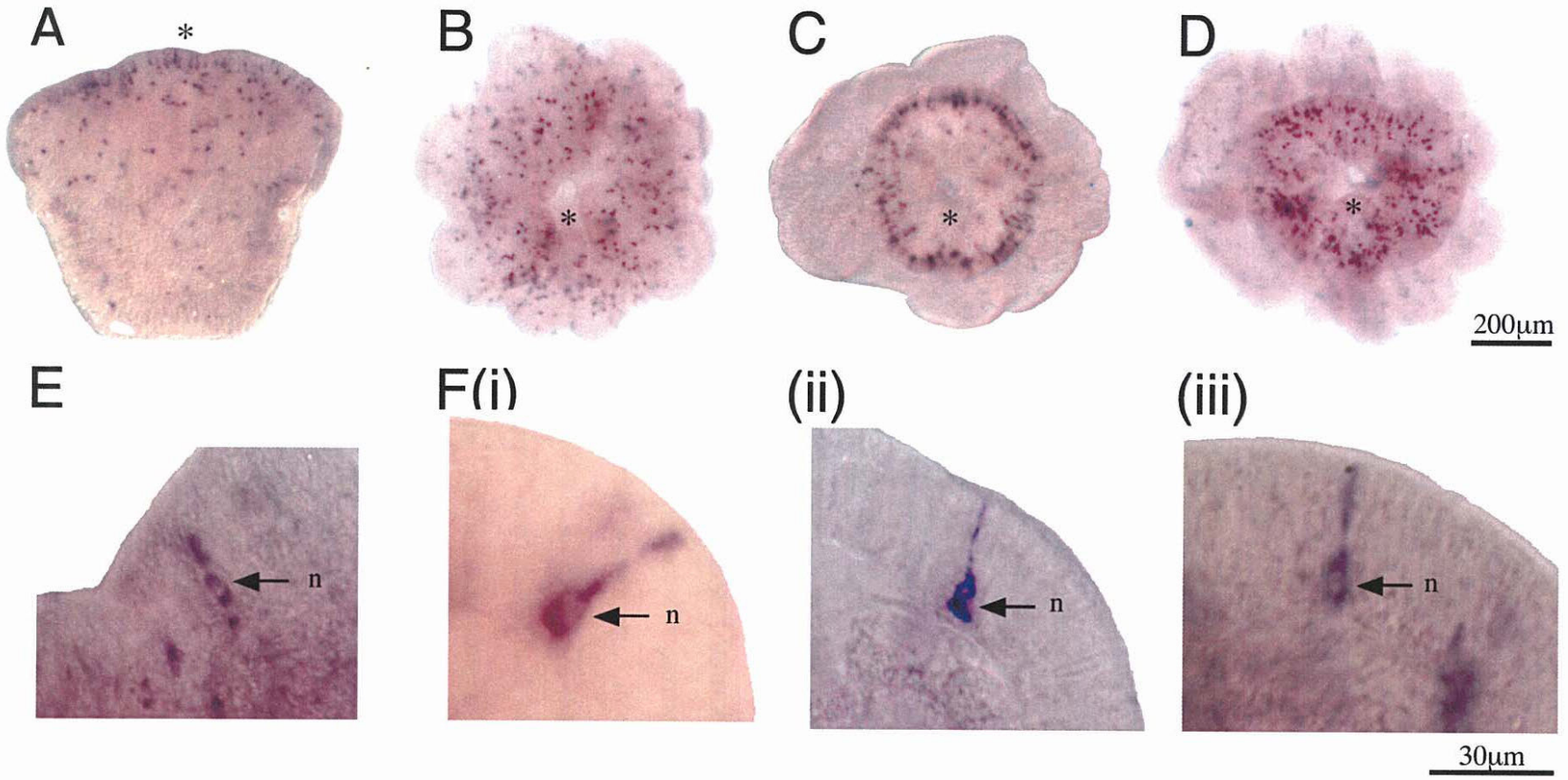


Figure 3.16: Spatial expression patterns of *cnox-2Am* in post-settlement larvae. *In situ* hybridisation pattern observed when post-settlement embryos were hybridised with *cnox-2Am* DIG-labelled riboprobe. (A-D) shows whole-mount embryos of various post-settlement morphologies. The oral pore is marked by an asterisk in each. (A) Indicates the mushroom-shaped embryo occasionally seen in post-settlement larvae collections, while (B-D) shows the progression of the post-settlement form from a flat disc (B) to those with a growing tip (C-D) where *cnox-2Am* expression is clearly concentrated. (E-F) shows high power magnification of *cnox-2Am* expressing cells, the morphology of which is consistent with previous descriptions of putative neurons (see section 3.2.9). The 'n' inidicates the nucleus of the cells.

consequences of expressing *cnox-2Am* in *Drosophila* were explored by making use of the GAL4-UAS system (Brand and Perrimon, 1993). The initial approach used was to drive *cnox-2Am* expression using GAL4 lines known to result in near ubiquitous expression early in *Drosophila* development. However, preliminary attempts were also made toward the generation of an ind-GAL4 line, which should ultimately enable expression of *cnox-2Am* in a pattern closely mimicking that of *ind* itself; these preliminary attempts are presented as Appendix B.

A *cnox-2Am* pUAST construct had been generated in our laboratory prior to the start of this project. However, the corresponding control (an *ind*-pUAST construct) was not available. To enable the generation of UAS-*ind* fly lines, full-length *ind* cDNA (kindly supplied by Prof G Technau, University of Mainz) was cloned into pUAST. *XhoI* and *KpnI* restriction sites were added to the primer sequences in order to facilitate directional cloning into pP(UAST); the primers used were ind_for_XhoI 5'-ccgctcgagatgtcgcgttcatttttgatg and ind_rev_KpnI 5'-ggggtaccctacgcctcaaccttcaatt. The insert was initially cloned into pGEM-T and the sequence confirmed prior to restriction digestion and cloning into pP(UAST). The construct was purified using a QIAGEN midiprep purification kit before 6µg of plasmid was injected into early *white*$^-$ (w^{1118}) *Drosophila* embryos as described in section 2.2.3. 250 *Drosophila* embryos were injected and 14 survived to adulthood. These were crossed back to w^{1118} adult flies and the progeny of three crosses had red/orange eyes indicating integration of the P-element insert into the genome. Those flies that had successfully integrated the construct were crossed against appropriate balancer stocks to create stable lines, and the chromosomal location of the transgenes determined by back crossing an *ind*-GAL4 balanced stock to wild type virgin females and observing the phenotype (see section 2.2.3). Fly crosses were performed by Ms Lucija Tomlenovic.

For the initial rescue experiments, a UAS-*cnox-2Am* line was crossed with the *armadillo*-GAL4 line, resulting in near ubiquitous early expression, and the *even-skipped (eve)* phenotype was used to assess the extent of rescue. Normally, *eve* is expressed in the two progeny of intermediate NB 4-2, and these are completely absent in the *ind* mutant lines (Weiss *et al.*, 1998). In preliminary rescue experiments, these

two *eve*-positive progeny cells were clearly present in a significant number (approximately 10% of several hundred flies examined) of *arm*-GAL4 x UAS-*cnox-2Am* flies (see Fig 3.17). Thus *cnox-2Am* appears to be capable of rescuing aspects of the *ind*⁻ phenotype, albeit partially.

Whilst encouraging, these preliminary results should be treated with caution for several reasons. The GAL4 driver (*arm*-GAL4) causes ectopic expression of *ind*, often resulting in secondary effects on CNS organisation and thus complicating interpretation of the *eve* phenotype. In addition, the results of *arm*-GAL4 x UAS-*ind* crosses are not yet available for comparison, so the statistical significance of the UAS-*cnox-2Am* results is unclear. These results must therefore be confirmed statistically; moreover, the intention is to make use of other driver lines (crosses with *sca*-GAL4 are underway) and also to repeat the *arm*-GAL4 crosses. The ultimate goal of this line of investigation is to carry out crosses using an *ind*-GAL4 line; preliminary work towards this end was carried out during this project, and is described in Appendix B. Further work directed at this aim is being carried out in collaboration with Prof G Technau's group at the University of Mainz.

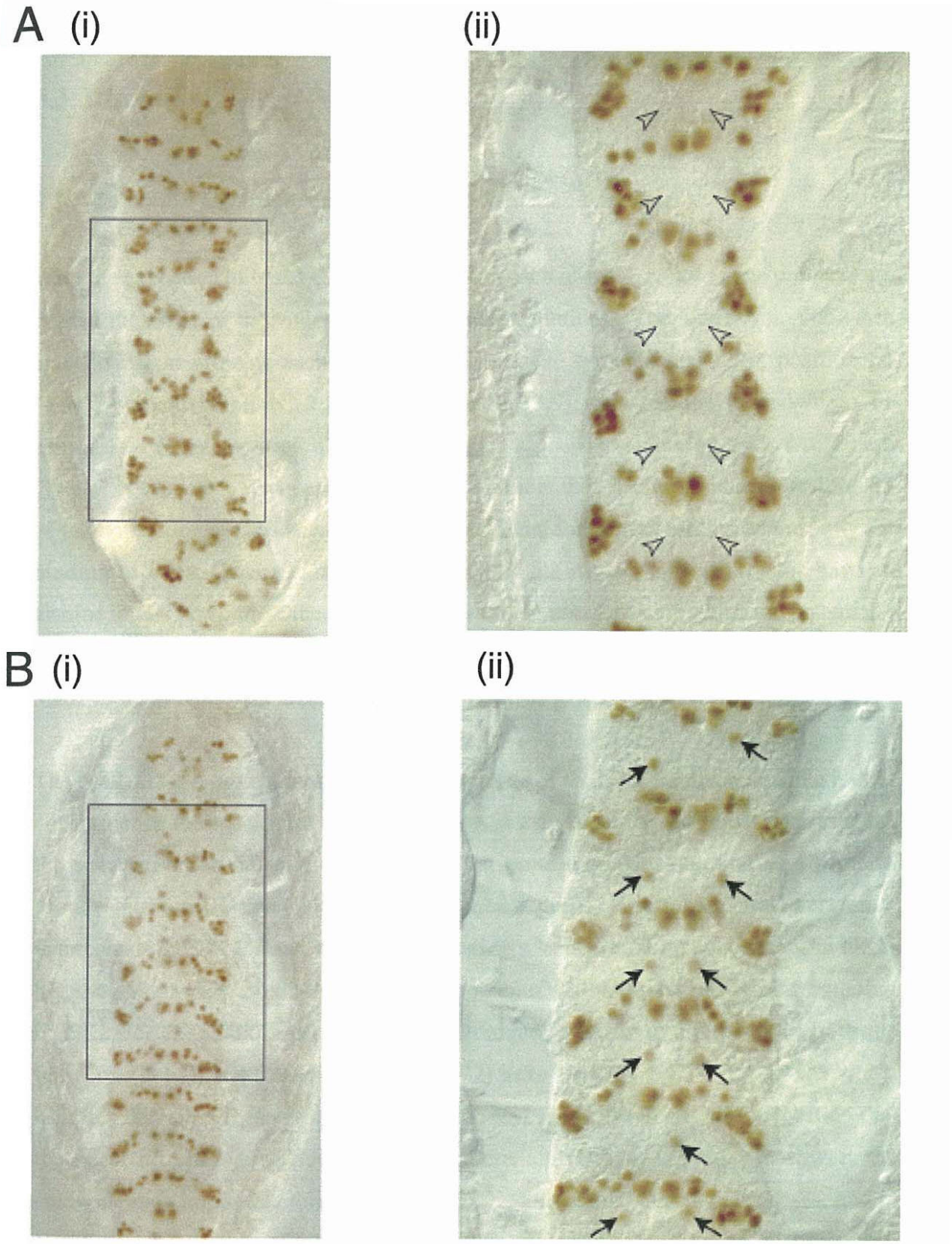


Figure 3.17: Even-skipped antibody staining in arm-GAL4 x UAS-cnox-2Am in an ind mutant background. In each case the boxed section in (i) is enlarged to show (ii). (A)(i) and (ii) shows the ind mutant phenotype where the position of missing RP2 neurons is indicated by arrow heads. (B) (i) and (ii) Expression of UAS-cnox-2Am construct under the control of arm-GAL4 driver. Black arrows show rescue of eve expression in RP2 neurons. This phenotype was seen in approximately 10% of individuals.

3.3 Discussion

3.3.1 Structural features of the cnidarian Vnd and Msx homologs

With the exception of Msx1-Am and Msx2-Am, each of the *Acropora* gene products are typical members of the vnd/Nk2.2 and msh/Msx families. The Vnd1-Am, Vnd2-Am and Vnd3-Am proteins each contain the diagnostic tyrosine residue at position 54 (Y54), present in all NK2 members (see Fig 3.9); this residue is known to be responsible for recognition of the unique DNA sequence containing a 5'-AAGT core (Weiler *et al.*, 1998; Wang *et al.*, 2002). In addition, the Alanine residue at position 35 of the homeodomain has been shown to be essential for correct folding and affinity for binding to DNA (Xiang *et al.*, 1998) – each of the *Acropora* Vnd proteins have an alanine in this position. Other residues known to contact DNA, or which are potentially able to (K3, R5, V6, L7, F8, Y25, K46, I47, Q50, N51) are also all conserved in the Vnd1-/Vnd2-/Vnd3-Am proteins (Gruschus *et al.*, 1997; Weiler *et al.*, 1998).

The NK2-SD is a protein motif unique to members of the NK2 family, and while it does not influence the specificity of DNA binding, it may function with a cofactor to mask the transactivation potential of the particular protein or may function as a protein-protein interface (Harvey, 1996; Watada *et al.*, 2000). It is found in all vertebrate members of the NK-2 class, although interestingly Vnd2-Am does not possess an NK2-SD, but Vnd1-Am, Vnd3-Am and the probable Vnd2-Am *Nematostella* ortholog NvNK2 do. The Vnd2-Am protein is not unique in lacking an NK2-SD – both the flatworm (Dth-2) and Caenorhabditis (ceh-22) have Vnd/NK-2 class of proteins which lack this domain. These proteins also lack an Eh1 motif, leading to the suggestion that the ancestral NK2 gene possessed neither domain (Harvey, 1996). However, the cnidarian data imply that this interpretation is incorrect, and that in these cases, and in Vnd2-Am, the absence of NK2-SD's reflects secondary losses. The fact that this gene is still clearly expressed indicates that if a new function arose from the loss of the NK2-SD, it has been advantageous to the organism. Alternately, it is possible that the NK2-SD has no function in the other Vnd proteins in *Acropora* because the cofactors necessary to interact with it are not present. While murine Nkx2.2 has been shown to

have a powerful C-terminal transactivation domain, it has not been ascertained exactly which residues are important in determining the transactivation potential of the protein, and the entire region C-terminal of the NK2-SD may be necessary for full activation (Watada *et al.*, 2000). The C-termini of vertebrate NK2 proteins range in size from 54AA (Human/Mouse Nkx2.9) to 97AA (Human/Mouse Nkx2.1) residues, while that of *Drosophila* Vnd is 72AA residues (and contains Arginine and Histidine-rich regions). The C-termini of the *Acropora* proteins are significantly shorter (Vnd1-Am 32AA; Vnd3-Am 35AA; Vnd2-Am 32AA), suggesting that these proteins may lack transactivation potential. The size of linker region from the NK2-SD to the homeodomain is within the currently known range of 9-32 amino acids for Vnd1-Am (11AA) and Vnd3-Am (16AA), although there is no conservation in these linker regions as seen in those from some vertebrates (Harvey, 1996).

Unlike the dramatic amino acid substitutions exhibited by the Msx1-Am and Msx2-Am homeodomains, the Msx3-Am homeodomain contains each of the residues known to be important for DNA binding specificity (R2, R58), core stability (L16) and for the formation of salt bridges which also stabilise the core (K23/E30; R31/E42) (Hovde *et al.*, 2001). However, there is evidence to suggest that the Msx-type homeodomain may serve only as a protein-protein interface to mediate repression, by interacting with other members of the core transcription machinery such as TATA-binding protein (TBP) (Catron *et al.*, 1995; Zhang *et al.*, 1996), thus Msh/Msx function may not require DNA binding. The N-terminal arm of the MSX-1 HD (Zhang *et al.*, 1996) and regions N- and C- terminal of the HD have been suggested as being responsible for this interaction (Catron *et al.*, 1995). Notably, Msx3-Am is identical to MSX-1 at 13 of the 14 residues at the N-terminal end of the homeodomain. Although unlike MSX-1, there are no glycine-rich regions in Msx3-Am, residues 187 to 248 of the Msx3-Am protein (C-terminal of the homeodomain) are relatively rich in alanine (16%), serine (19%) and proline (14%) residues (see Fig 3.13). The homologous *Hydra vulgaris* and *Chlorohydra* sequences are not full length so it is not clear if this is a conserved feature of other cnidarian Msx proteins, although it does not appear to be present in any other Msx proteins isolated so far. No function has been reported for alanine, serine and proline-rich regions in the literature thus far.

Eh1 motifs are present in Cnox-2Am, Msx3-Am and each of the *Acropora* Vnd proteins; this motif is also present in most of their homologs in higher animals (Fig 3.8), implying that it has some basic function in each case. The phenylalanine residue present in each case is known to be important for interaction with the Groucho family of co-repressors (Smith and Jaynes, 1996). At least one Groucho class co-repressor is represented in the *Acropora* EST collections, so it seems likely that the Eh1/Groucho interaction predates the cnidarian / higher metazoan split, and that the cnidarian Eh1 domains function to mediate transcriptional repression.

3.3.2 Regulation of the vnd1/2/3-Am, msx3-Am and cnox-2Am genes

In both vertebrates and insects, the spatial pattern of expression of *vnd, ind* and *msh* homologs is achieved by a repression cascade. The interaction between members of this cascade has not yet been proven to be direct, although Vnd binding sites present in an *ind* putative enhancer region 3' of the *ind* transcription unit can function as transcriptional mediators *in vivo* if placed adjacent to an heterologous enhancer (Cowden and Levine, 2003). It has also been shown that the Vnd protein is required for maintenance of *vnd* gene expression, although again it is not clear if this autoregulation is direct; however, many consensus binding sites for Vnd are present in the 5' flanking region of *vnd*, and this region has also shown to be sufficient to generate the normal pattern of *vnd* gene expression (Saunders *et al.*, 1998). While 5'-genomic sequence is not yet available for *vnd3-Am*, the corresponding sequences are available for both *vnd1-Am* and *vnd2-Am*. The start methionines of each of these genes are located only 9kb apart, indicating that they might share common regulatory elements. 5' RACE results indicate that transcription starts at base –83 (relative to the start codon) for *vnd1-Am* and –99bp (relative to the start codon) for *vnd2-Am* (although these putative transcription start sites should ideally be confirmed by other methods). Given that functional TATA and CAAT boxes must lie within 50bp and 100bp of the initiation start site respectively, (Lewin, 1997), TATA and CAAT motifs lying between –183 and –1bp for *vnd1-Am* (relative to the start codon) and between –199bp and –1bp for *vnd2-Am* were identified. Two CAAT boxes at –125 and –39bp and two TATA motifs at –120 and –83bp of *vnd1-Am* support the notion transcription starts in this area. In the region of the putative *vnd2-Am* transcription start site however, while three CAAT boxes were found (-182, -

82bp and –56bp) no TATA motifs were identified, nor was a CATAA box which has been reported in literature as a non-canonical TATA box (Pailhoux *et al.*, 1992).

A search for the consensus recognition sequence of NK2 proteins (5' AAGT) in the intervening region between the two *vnd*-like genes (from –83bp of *vnd1-Am* to –99bp of *vnd2-Am,* relative to the start codons of each) revealed a total of 85 putative NK2 binding sites (on both strands), two of which match perfectly with the Vnd consensus sequence (5' T(T/C)AAGTG(G/C)), and one of which is located just 59bp 5' of the putative *vnd1-Am* transcription start site. Six others differ from the consensus at just one of the nucleotides which flank the central 5'-AAGT sequence (i.e. one of those indicated in bold 5'-**T(T/C)**AAGT**G(G/C)**) suggesting the possibility of autoregulation for both *vnd1-Am* and *vnd2-Am* (see Fig 3.10).

The availability of genomic data for *cnox-2Am* enabled a search for NK2 binding sites in the 2.8kb proximal 5'-region. A total of 36 sites were found containing the core 5'-AAGT sequence, however none of these were perfect matches to the consensus Vnd binding sequence. It is still possible that Vnd1-/Vnd2-/Vnd3-Am proteins bind to the *cnox-2Am* promoter, as the binding specificity of the cnidarian Vnd proteins is unknown, and may have diverged from its fly and mouse counterparts.

In *Drosophila*, genetic studies have shown that a Sox gene *Dichaete/fishhook* functions in parallel to *vnd* and *ind* in nerve cord development. For this reason, when a fragment of a gene encoding part of a Sox-domain from *Acropora* was discovered as part of the preliminary EST project (Technau *et al.*, submitted), the complete cDNA was isolated in order to investigate whether this gene might interact with coral *vnd* and *ind*. From a screen of 500 000 plaques of a pre-settlement cDNA library with a 150bp fragment of EST lacking the HMG domain, 2 positives were isolated which were found to encode a HMG domain containing protein, which was more similar to the HMG20 proteins from vertebrates than a Dichaete-like protein from *Drosophila*. In addition, as with the HMG20-like proteins, the *Acropora* gene was expressed at constant levels throughout development (as seen on a virtual northern blot) and as such would be unlikely to play a specific role in nervous system development (results not shown).

The transcription start site of *msx3-Am* was tentatively identified to be 146bp upstream of the start codon. A search for TATA and CAAT sequences in the –246 to –46bp region (relative to the translation start site) was undertaken; three CAAT motifs were found at –126bp, -179bp and –209bp and a single TATA box was found at –176bp, all of which fall in the spacing specifications where they might be functional. The murine *Msx-1* promoter, and the *Ciona intestinalis Ci-msxb* promoter are the only two that have been so far analysed in detail; in *Msx-1* a proximal element (at –2198bp) and a distal element (-4006bp) were both found to be essential for correct expression (MacKenzie *et al.*, 1997) and a 3.8kb putative promoter region was found to contain potential *cis*-acting elements that regulate *ci-msxb* tissue-specific expression, (Russo *et al.*, 2004). These two regions do not show any significant homology with one another, nor did they share any common binding sites for any known transcription factors (Russo *et al.*, 2004), suggesting that in these two distantly related species, and most probably in *Acropora*, different transcription factors regulate the expression patterns of *msx*-like genes depending upon context.

3.3.3 Genomic organisation of vnd/Nk-2 and msx/Msh loci

The apparent duplication of both *msx* and *vnd*-like genes in *Acropora* is not a novel phenomenon. In the course of characterising the gene complement of *Acropora*, several cases of tightly linked paralogs have been identified (Hislop *et al*, 2005 – Appendix D) (Grasso *et al.*, 2001; Samuel *et al.*, 2001; Ball *et al.*, 2004), and furthermore it seems that a significant number of regulatory genes in other cnidarian species have been independently duplicated including the *nanos* and *paired-like* genes in *Hydra* (Gauchat *et al.*, 1998; Mochizuki *et al.*, 2000), and the *snail* and *mox* genes in *Nematostella* (Martindale *et al.*, 2004). In addition, at least some of the duplicated genes in *Acropora* are also present in *Hydra* (Go *et al.*, unpublished) (Mochizuki *et al.*, 2000) and therefore precede the Anthozoa/Hydrozoa divergence. It is possible that these duplication events may have contributed significantly to the large genome size estimated for some cnidarians (Zacharias *et al*, in press) (David and Campbell, 1972). Independent duplication of some genes is perhaps to be expected, as the Cnidaria may have been distinct from the bilaterian stem since deep in Pre-Cambrian time (Conway Morris,

2000), however the extent of gene duplication and organisation of the duplicated genes within the Cnidaria is unknown.

Duplication events have been postulated to be a major factor in the evolution of body plans, and the duplication of the Hox genes for example, the reason for an increase in complexity of the higher animals (Lundin, 1999). Initially, a duplicated gene pair may have fully overlapping, redundant functions (Force *et al.*, 1999) and this appears true for *vnd1-/2-/3-Am,* as each have fully overlapping expression patterns. In the absence of selection pressure, duplicated genes are predicted to become pseudogenes in the first few million years after duplication (Lynch and Conery, 2000). However, some duplicated genes may acquire novel functions and thus escape this conversion (Prince and Pickett, 2002) (Yokoyama and Yokoyama, 1989) (Locascio *et al.*, 2002). We have no idea how fast these processes occur in cnidarians, or indeed if at all, and thus no speculation can be given as to how recent the duplication event and subsequent pseudogenisation of the *msx1-/msx2-Am* genes may have been.

There are precedents for the linked organization of the *vnd1-Am* and *vnd2-Am* genes, suggesting the possibility of synteny, as is the case for mice and humans where the four *vnd*-type genes are organised as two linked paralogous pairs; *Nkx2.2/Nkx2.4* and *Nkx2.1(TTF1)/Nkx2.9* in both species (Wang *et al.*, 2000). This apparent conservation of linkage might be functionally important, although the two pairs of paralogous genes have complementary and overlapping expression patterns in the CNS (Price *et al.*, 1992; Pabst *et al.*, 1998). The fact that the genes are more similar to their duplicated partner on another chromosome than to their linked partner indicates that an ancestral NK2-type gene was duplicated to form a tandem gene pair, and the pair then duplicated to another chromosomal locus at a later time (Wang *et al.*, 2000). However, phylogenetic analysis shows that each of the *Acropora* sequences fall into a clade that includes *Drosophila* Vnd and mouse and human Nkx2.2 and Nkx2.9, but within this the cnidarian sequences form a distinct monophyletic group; thus it seems that *vnd1-Am* and *vnd2-Am* have been independently duplicated within the Cnidaria, and their organisation does not reflect synteny between cnidarians and mammals. Other than the data presented in this study on *Acropora*, primitive organisms appear to have only one *msh*-like gene whereas in mammals there are at least three family members. This suggests a duplication of the *msh/Msx* family members around the origin of the vertebrate lineage, supported by the

fact that *Amphioxus* possesses only one *msh* gene (Holland, 1991; Holland *et al.*, 1994). Unless one considers the unlikely possibility that significant gene loss has occurred in all other lineages leading to the vertebrates, the most plausible explanation for the presence of the three *msx*-like genes in *Acropora*, is that this is a cnidarian-specific duplication.

The chromosomal organisation of the *vnd3-Am* and *msx3-Am* loci relative to *vnd1/2-Am* and *msx1/2-Am* is not known, thus the mode of duplication of these genes is not clear. *Vnd1-Am* is more similar to *vnd3-Am* than it is to its linked partner *vnd2-Am,* while *msx1-Am* and *msx2-Am* are clearly more similar to each other than to *msx3-Am.* Although the former situation in some respects parallels that of the vertebrate Vnd loci (see above), it is unclear at this stage whether other *vnd* genes lie proximal to *vnd3-Am.* While highly speculative, it is possible that as with the human and murine *vnd/Nkx2* genes, *vnd3-Am* and *msx3-Am* may each be one of a pair of linked genes and that the paralogous gene has not yet been identified

Conserved intron-exon structure is often used as an indication of orthology between genes. *Vnd1-Am* and *vnd2-Am* each contain single introns, located 101bp (33AA) and 80bp (26AA) upstream of the homeobox respectively. The genomic structure of the *vnd3-Am* locus has not yet been ascertained. *Drosophila vnd* contains two introns, located 137AA and 357 AA N-terminal of the homeodomain. Vertebrate *Nkx2.1* genes typically also contain two introns at approximately equivalent positions (see Fig 3.18A) (Hamdan *et al.*, 1998). This suggests that either intron loss has occurred in the Cnidaria or intron gain has occurred in the vertebrate lineage. As in both murine *Msx-1* and *Drosophila msh, msx3-Am* contains a single intron N-terminal of the homeodomain whereas the *Ciona intestinalis Ci-msxb* gene contains four introns, one in an equivalent position to that in *Acropora*, two located further upstream and one interrupting the homeobox, and appears therefore to have undergone significant gain of introns (see Fig 3.18B).

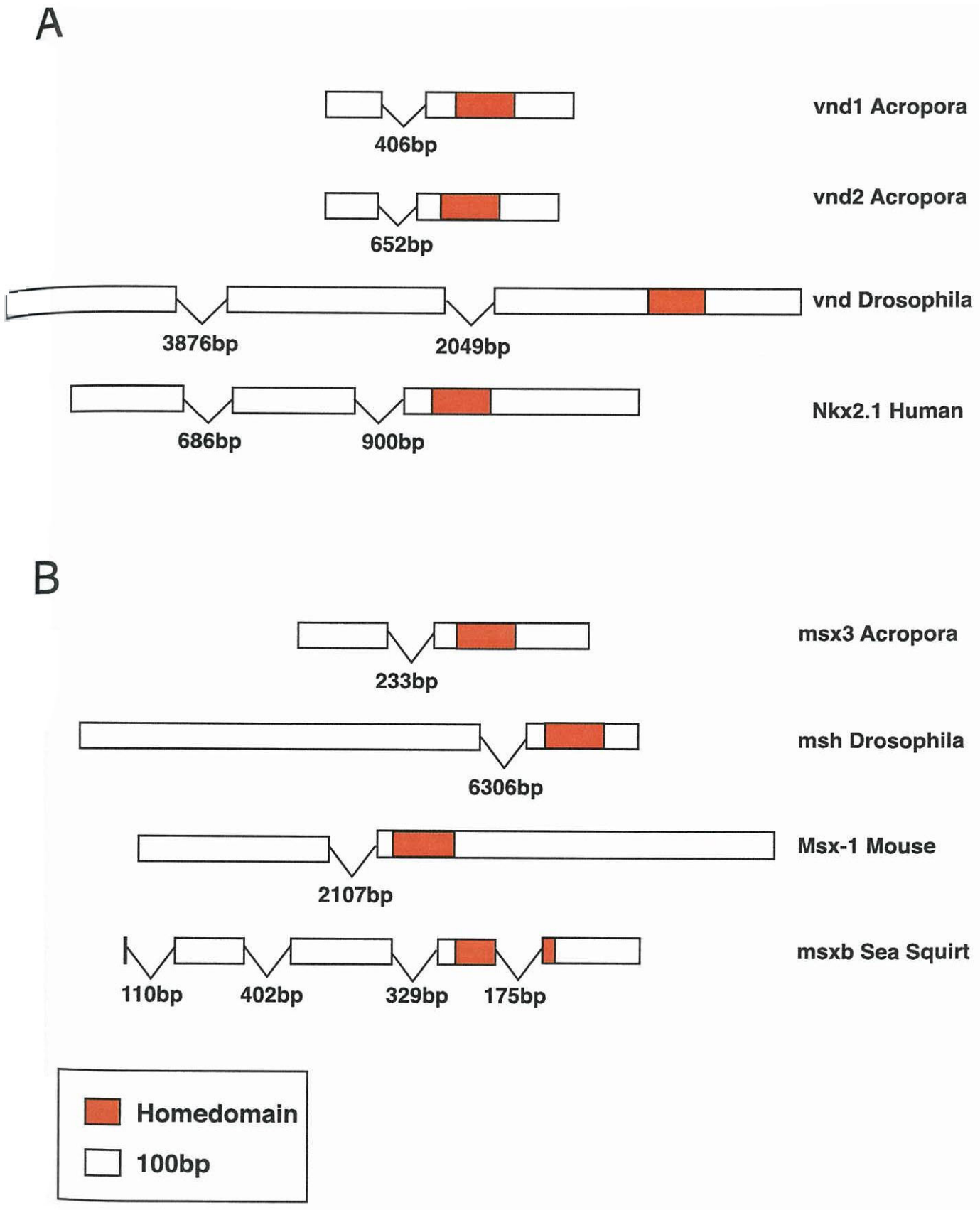


Figure 3.18: Comparative genomic structure of *vnd* and *msh* genes. Schematic representation of the intron/exon structure of representative (A) *vnd* and (B) *msh* genes. In both (A) and (B) open boxes represent the coding region of each gene and red shading indicates the homeobox of each gene. Introns are indicated by open triangles with the size of each intron shown below the triangle. Drawings are to scale (box indicating 100bp is given in the key).

3.3.4 Expression patterns and possible roles of Acropora vnd1-/vnd2-/vnd3-Am, msx3-Am and cnox-2Am

The expression patterns of the *Acropora vnd1-/vnd2-/vnd3-Am*, *msx3-Am* and *cnox-2Am* genes were investigated in pre- and post-settlement material by *in situ* hybridisation during the course of this study, with a view to understanding the roles these genes might play in coral development and for comparison with their counterparts in higher animals. Traditionally the O/A axis of 'radially-symmetric' animals such as the cnidarians was assumed to correspond to the A/P axis of bilaterians (Freeman, 1981a; Freeman, 1981b), however recent evidence, including that presented in this study supports the notion that bilateral symmetry arose prior to the cnidarian/bilaterian split, and that the radial symmetry is a derived state (Ball *et al.*, 2004; Finnerty, 2004). Previous studies on *cnox-2Am* expression in pre-settlement stage larvae, show localisation of *cnox-2Am* transcript in a subset of putative neurons in a spatially restricted pattern along the sides of the larvae, with message rare in the oral end and absent from the aboral end, resembling that of *ind/Gsx* genes (Hayward *et al.*, 2001). Prior to this study, no spatial expression data were available for any cnidarian *msx* or *vnd* gene.

The expression patterns of the three *Acropora vnd* genes were indistinguishable, each gene being expressed in the ectoderm surrounding the oral pore in both pre- and post-settlement larvae; *in situ* experiments were not carried out on earlier material, based on virtual northern data (see Fig 3.4). In general, *vnd3-Am* exhibited the strongest and most reliable expression; it is not known whether this is a quality of the transcript itself, or due to an exogenous factor such as a more efficiently labelled riboprobe. While clearly restricted to the ectoderm, expression in specific cell types could not be ascertained for any of the *vnd* genes. The *msx3-Am* gene is also expressed in a spatially restricted pattern in the ectoderm of both pre-settlement and post-settlement larvae. Expression appears to be restricted from around the oral pore, but this is often not clear due to 'bleeding' of the substrate from cells.

On the basis of the conserved functions of both *vnd* and *msh* genes in nervous system patterning throughout the higher Metazoa, our working hypothesis was that homologs

of these genes might also be involved in development of the cnidarian nervous system. In apparent contradiction to this, the expression of *vnd1-/vnd2-/vnd3-Am* and *msx3-Am* does not appear to be restricted to specific cell types. However, as these genes are likely to be expressed in neurons within their general domains of expression, the possibility of genetic interactions of the type seen in *Drosophila* cannot be discounted.

In order to relate the expression patterns of *vnd1-/vnd2-/vnd3-Am* and *msx3-Am* with that of *cnox-2Am*, *in situ* hybridisation with a DIG-labelled *cnox-2Am* riboprobe was conducted on post-settlement material. As in pre-settlement larvae (Hayward *et al.*, 2001), *cnox-2Am* was expressed in putative neurons (Fig 3.16). Although not ideal, morphology alone has often been used for the identification of neurons in other cnidarians (Chia and Koss, 1979; Fautin and Mariscal, 1991; Grimmelikhuijzen and Westfall, 1995). Note that cells visualised using an antibody against RFamide, the most abundant cnidarian neuropeptide, are morphologically indistinguishable to the cells detected by *in situ* hybridisation using *cnox-2Am* probes, supporting their assignment as neurons (Hislop *et al*, unpublished) (Hayward *et al.*, 2001).

The expression patterns of these three *Acropora* genes are similar in two important respects. First, although *cnox-2Am* expression begins somewhat earlier (Hayward *et al.*, 2001), all three genes are expressed simultaneously. Second, expression of each of the genes is restricted along the O/A axis; in each case expression is biased towards the oral end, and absent from the aboral end. While the expression domains of *msx3-Am* and *cnox-2Am* overlap, there are indications that representatives of the three groups of genes may be expressed in mutually exclusive regions during development. In both pre- and post-settlement material, the three *vnd* genes are expressed in the ectoderm around the oral pore, while *cnox-2Am* is excluded from this zone, and *msx3-Am* may also be excluded. These latter genes are expressed in similar regions in the ectoderm during both pre- and post-settlement stages, but *cnox-2Am* is seen only in a subset of neurons while *msx3-Am* is more generally expressed – it is thus possible that *msx3-Am* is restricted from the *cnox-2Am*-expressing cells, but is expressed in other cells of the ectoderm, including other neurons. One interpretation of these data is that some of the genetic interactions known from higher animals occur in Acropora; *vnd1-/vnd2-/vnd3-Am* may repress both *cnox-2Am* and *msx3-Am,* and *cnox-2Am* may in turn repress *msx3-Am.* In order to determine if the transcript distribution was convincingly consistent with

this hypothesis, attempts were made to carry out double *in situ* hybridisation experiments using *vnd3-Am* and *msx3-Am* riboprobes. These experiments require two differentially labelled riboprobes (DIG or FL), and generally employ either Vector BCIP/NBT or Roche BM purple for one assay, and Sigma Red for the other. Unfortunately, these attempts were unsuccessful; appropriate combinations of riboprobe and substrate could not be developed, as some probe/substrate combinations were incompatible for reasons unknown. *In situ* hybridisation with *msx3-Am* could only be achieved using a DIG-labelled riboprobe and BM purple as a substrate, while for *vnd3-Am* DIG-labelled riboprobes could be stained with either Sigma Red or Vector BCIP/NBT, but the FL-labelled riboprobe was only successful with Vector BCIP/NBT substrate.

3.4 Conclusions

The results presented detail the isolation and expression pattern of members of the *vnd, msx* and *ind* families in the anthozoan cnidarian *Acropora millepora*. Expression of members of each of these gene families is axially restricted, and the *in situ* data are consistent with repression of both *Msx/msh* (*msx3-Am*) and *Gsh/ind* (*cnox-2Am*) genes by the Vnd family members.

Whereas these genes are expressed differentially along the D/V axis in higher animals, in *Acropora* their expression is restricted along the O/A axis. In addition to these members of the D/V cascade, core components of the A/P pattering system are also expressed in axially restricted patterns in *Acropora* (Hislop *et al.*, submitted) as are homologs of *forkhead* (Martindale *et al.*, 2004), *emx* (Mokady *et al.*, 1998), *aristaless* (Gauchat *et al.*, 1998; Bridge *et al.*, 2000), *goosecoid* (Broun *et al.*, 1999), *brachyury* (Technau and Bode, 1999), *wnt* (Hobmayer *et al.*, 2000) and *nanos* (Mochizuki *et al.*, 2000) in other cnidarians. These data imply that the molecular machinery to enable the specification of two axes predates the cnidarian/higher animal split, and that true bilaterality was achieved by shifting the expression of some of these genes through 90^0. However, there are intriguing hints that the story is not that simple (Hayward *et al.*, 2002); Grasso *et al.*, unpublished). Understanding axis specification (and many other aspects of development and patterning) in cnidarians awaits the development of reliable

methods for interfering with ('knocking down') gene activity, which should be achievable within the next few years.

3.5 Future Directions

The data presented in this chapter represents only initial characterisation of three gene families in *Acropora*; aspects of their expression patterns reman to be clarified, and this work could form the basis of functional analyses. Double *in situ* methods are required to test the idea that *vnd* and *msx* genes are expressed in mutually exclusive regions, and to investigate whether these genes are expressed in the nervous system. The rescue experiments in *Drosophila* being carried out with *cnox-2Am,* could be extended to the *vnd* genes and *msx3-Am* to inestigate conservation of function, and morpholino methods should be applied to investigate loss of function effects in *Acropora*. The possibility of repressive interactions between these genes could be tested by the application of chromatin immunoprecipitation techiniques (ChIP) and DNA binding assays such as electrophoretic mobility shift assays (EMSAs) or the yeast one-hybrid system. The genomic organisation of *vnd3-Am* in relation to *vnd1-/2-Am* and *msx3-Am* in relation to *msx1-/2-Am* should also be determined by chromosome walking.