4

tlx-Am, a cnidarian ortholog of the *Drosophila* nuclear receptor gene *tailless*

4.1. INTRODUCTION

Like *empty spiracles*, *tailless* is a head gap gene in *Drosophila* and is now known to directly regulate ems expression (Hartmann et al., 2001). The CNS expression patterns of vertebrate *Tlx* genes resemble that of the *Drosophila* gene; both are involved in the development of the embryonic brain (Monaghan et al., 1995; Rudolph et al., 1997), suggesting conservation of function. During a project aimed at characterising the nuclear receptor complement of *Acropora*, a clear ortholog of the *tll/Tlx* genes was identified (Grasso et al., 2001) and preliminary in situ hybridisation experiments (Grasso, 1999) suggested that this gene might be expressed in an overlapping regions at the aboral end during development. On this basis, and in view of the possibility of conserved interactions with *ems/Emx* genes, the coral *tlx-Am* gene was characterised.

As expected, the *tlx-Am* gene was expressed during early development but, unlike the situation in *Drosophila*, its temporal pattern of expression did not overlap that of the coral *emx-Am* gene, indicating that a regulatory relationship between *tlx-Am* and *emx-Am* genes is unlikely to exist in *Acropora*. The *Acropora tlx-Am* data shed light on several aspects of the evolution of the *tll/Tlx* family, although the in situ hybridisation data obtained for this gene were inconclusive and difficult to distinguish from background staining artefacts.

4.1.1. A superfamily of nuclear receptors

The nuclear receptor superfamily is comprised of transcription factors that regulate a diverse range of biological processes including development, differentiation, homeostasis and metabolism (reviewed in Mangelsdorf and Evans, 1995; Mangelsdorf et al., 1995). Nuclear receptors are characterised by the presence of a highly conserved DNA-binding domain (DBD) that targets the receptor to specific DNA sequences termed hormone response elements. Specific regions within the DBD, termed P-box

and D-box (see Figure 4.2) are important for DNA-binding specificity. The five residue P-box identifies the primary nucleotide in the DNA-binding half site, while the spacing between the half-sites is determined by the five residue D-box (Umesono and Evans, 1989). *Drosophila* Tll is known to bind to a hexameric sequence, AAGTCA, which is present in the gene enhancer region of target genes *Krüppel* and *knirps* (Hoch et al., 1992; Pankratz et al., 1992). The chick Tlx was also shown to bind the same Tll half-site *in vitro*, suggesting conserved genetic programs exist between members of the Tll/Tlx class of nuclear receptors (Yu et al., 1994).

Unlike many transcription factors, they also have a less conserved ligand-binding domain (LBD), which conveys hormone recognition. Lipophilic hormone molecules, such as steroids, thyroid hormones and vitamin D that can diffuse across cell membranes, can differentially regulate gene expression through these ligand-modulated nuclear receptors. Many nuclear receptors are multifunctional, causing transcriptional repression of their target genes, which can be converted to an activation response in the presence of a corresponding ligand. In addition to family members that can bind steroid and non-steroid hormones, other proteins that are clearly nuclear receptors have no known ligand (Mangelsdorf and Evans, 1995). It is unclear whether these 'orphan receptors' function without ligands or if their ligands are still unidentified. In the absence of ligands, some nuclear receptors can undergo modifications such as phosphorylation to become activators (Hammer et al., 1999).

4.1.2. The *tll/Tlx* gene family

The *Drosophila tailless (tll)* gene is a member of the steroid nuclear receptor superfamily (Pignoni et al., 1990) and is classified as an orphan nuclear receptor given that it possesses a hormone binding domain to which no known ligand can bind. As a terminal gene, *tll* affects development at both the anterior and posterior ends of the *Drosophila* embryo and is required for the formation of terminal structures such as the brain, foregut, midgut and hindgut. Transcription of *tll* initiates in the syncytial blastoderm in two caps at the anterior and posterior poles where it is required for terminal pattern formation. During cellularisation, the *tll* expression domain corresponds with the anlage of the entire brain and includes areas from which protocerebral neuroblasts will eventually delaminate during gastrulation. By stage 12, *tll* expression is also present in the primordium of the optic lobe. The posterior

expression domain decreases at gastrulation and terminates by the end of germband extension. Mutations in the *tll* gene result in deletions at the terminal regions of *Drosophila* embryos (Strecker et al., 1986).

A second *Drosophila* gene, *dissatisfaction (dsf)*, encodes a nuclear receptor that is closely related to the Tll transcription factor, and is essential for neural development as well as sexual behaviour in both sexes (Finley et al., 1998). It shares a striking level of similarity with other *tailless* orthologs, but differs in two regions that are critical in DNA binding and dimerisation, although the *in vitro* binding specificity of the Dsf DBD is identical to that of Tll (Pitman et al., 2002). This gene may possibly be an early duplicate of *tll* that has been exposed to significant evolutionary change, or it may be a founding member of another family of nuclear receptors with yet unidentified orthologs in other organisms.

A *tll* ortholog, *Tc-tll,* was also isolated in the short-germband flour beetle, *Tribolium castaneum*. As in *Drosophila*, *Tc-tll* is expressed at the posterior pole but unlike its counterpart in long-germband insects, *Tc-tll* expression is switched off before the activation of genes known to be Tll targets in *Drosophila* (Schröder et al., 2000). In contrast to *Drosophila*, anterior expression is not detected until after cellularisation when *Tc-tll* transcripts appears in the developing head region. These differences may reflect the alternative modes of embryogenesis between *Drosophila* and *Tribolium* and suggest major changes in early *tll* functions and regulatory pathways over a relatively short evolutionary distance.

Vertebrate orthologs of *tll* have been reported in the human (Jackson et al., 1998), mouse (Monaghan et al., 1995), chicken (Yu et al., 1994), frog (Hollemann et al., 1998) and, like *Drosophila,* are expressed in the developing embryonic brain. Common characteristics of embryonic expression patterns are shared by vertebrate homologs of *tll*. All are expressed exclusively in developing fore- and midbrain regions and optic structures. The chick and frog *Tlx* are first detected in the optic vesicles and later in anterior cephalic regions including telencephalic, diencephalic and the eye (Yu et al., 1994). Studies in *Xenopus* and mice have implicated *Tlx* in early vertebrate eye development (Hollemann et al., 1998; Yu et al., 2000). The expression pattern of murine *Tlx* in the ventricular zone of the neuroepithelial layer and undifferentiated

neuroepithelial cells parallels its expression in the protocerebral neuroblast zone of the fly (Daniel et al., 1999). Homozygous mutant mice show a reduced number of cells in the intermediate zone of the cortex at E15.5, consistent with a role of *Tlx* in ventricular zone proliferation and differentiation. In adult *Tlx* mutant mice, this leads to a reduction in the size of the rhinencephalic and limbic structures (including the olfactory region and dentate gyrus) and increased aggressive behaviour (Monaghan et al., 1997). Relative to *Drosophila tll*, the onset of vertebrate *Tlx* expression is delayed during embryogenesis and is spatially restricted to the developing nervous system. Posterior expression zones appear to be an arthropod-specific characteristic, but the later brain-specific expression zone is shared between vertebrates and insects, suggesting that the *tll/Tlx* gene family may be part of an ancient genetic mechanism controlling anterior patterning. The earlier posterior patterning function seen in insects may be a derived characteristic of this phylum, or an earlier function subsequently lost in the vertebrate lineage.

A *tll*-related nuclear receptor was identified in *C. elegans* and is associated with neural development mechanisms including pathway selection by growth cones (Much et al., 2000). Expression of this gene, *fax-1*, was observed in embryonic neurons after neurogenesis as axon extension begins. An association of the *tll/Tlx* gene family with nervous system development in *C. elegans*, *Drosophila* and vertebrates further supports this is as an ancestral function, present in the last common ancestor of these organisms, but these data do not indicate when this function arose. In order to examine the ancestral functions of this gene family, embryonic data from basal metazoans are required.

4.1.3. Statement of goals

The only cnidarian ortholog of *tailless* identified to date is in the coral *Acropora millepora*. *tlx-Am* was first isolated by our collaborators during a search for cnidarian nuclear receptors (Grasso et al., 2001), but was not characterised beyond the sequence level. Characterisation of this gene was proposed to examine its role during coral embryonic development. Studies of the *tll* gene in *Drosophila* have also shown it to be a direct regulator of *empty spiracles* (Hartmann et al., 2001). In order to determine the extent to which this regulatory pathway is conserved, the expression of *tlx-Am* during

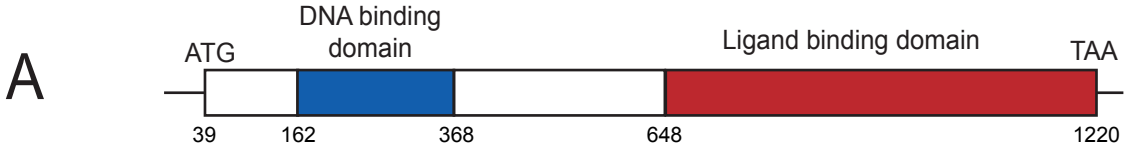
coral embryonic development was investigated and its potential to regulate *emx-Am* during development was analysed.

4.2. RESULTS

4.2.1. The *tlx-Am* cDNA

The *tlx-Am* cDNA clone was isolated and sequenced during a screen for coral nuclear receptors before the start of this project (Grasso et al., 2001). The nucleotide sequence of the *tlx-Am* cDNA is 1353 bp, comprising 38 bp of 5'- untranslated region; 1179 bp of open reading frame and 136 bp of 3'- untranslated region (Figure 4.1B). Conceptual translation of the open reading frame of *tlx-Am* results in a putative protein, Tlx-Am, of 393 amino acids. Translation initiation is predicted to occur at the first methionine starting at base pair 39, although a second methionine codon is located upstream of the cDNA sequence in the genomic sequence starting at –22 bp. Three in-frame termination codons are located in close proximity to this second methionine codon at –67, -56 and –43 bp. A 69 residue DBD is located 42 amino acids from the N-terminal end of this protein. A second conserved region, the 61-residue LBD in the C-terminal half of the protein, is linked to the DBD by a hinge region of 153 amino acids.

The Tlx-Am DBD is clearly related to the Tll/Tlx nuclear receptor family and is highly conserved with identities ranging from 65 to 80% compared with the DBD of other Tll/Tlx family members. The coral sequence most closely resembles vertebrate Tlx orthologs and the fly Dsf (>81% identity; see Figure 4.2A). Relative identity with invertebrate orthologs was somewhat lower. The Tlx-Am DBD contains a single amino acid insertion (Asp) between amino acids 40 and 41 of the tll/Tlx consensus. This insertion position is based on the alignment of the vertebrate Tlx DBD sequences, but cannot be confirmed unequivocally. A three amino acid residue (Gly, Ser, Pro) insertion is found at the same position in *C. elegans* Fax-1. *Drosophila* Dsf, which shares 78% identity with the coral sequence in the DBD, also contains a two-residue (Gly, Asp) insertion at this same position within the D-box region of the second zinc finger motif; this region is involved in dimerisation in some nuclear receptors (Figure 4.2C) (Glass, 1994). It is possible that these proteins with the additional residues are no longer able to form dimers and have acquired new functions.



B

```
                                                                  *         *                  M
-89           ccaataagctgcgagctttatctaaggtttccctgatatttgctgttagccggcctctcggaacaagatgcactgttcttctgaaacat   -1

                                                                                  M  W  K  F  R  V  V  T  K  L  S  S  K  M  N  S  D  Q  Q  L  D   21
  1  TGCTTTCCCTTAGTGCAGTACCAGAAATCTCAAGACCAATGTGGAAGTTCCGCGTAGTAACAAAGCTTTCATCTAAAATGAACTCTGATCAGCAACTAGA
       C  T

                 D  V  V  V  D  N  N  D  K  S  K  P  N  G  R  L  L  D  I  P  C  K  V  C  G  D  R  S  S  G  K  H  Y   54
101  CGACGTGGTTGTTGATAACAACGACAAATCGAAACCTAATGGCCGACTTCTGGATATTCCTTGCAAAGTCTGCGGGGATCGATCTTCTGGGAAACATTAC
          A                                                                                                  T

                 G  I  Y  A  C  D  G  C  S  G  F  F  K  R  S  I  R  R  N  R  S  Y  T  C  R  A  T  N  G  K  G  N  C   87
201  GGAATTTATGCTTGTGACGGCTGCAGTGGCTTTTTCAAGAGAAGTATTCGCAGAAATCGTTCTTACACTTGCAGAGCTACCAATGGCAAGGGCAACTGTC

                 P  V  D  K  I  H  R  N  Q  C  R  S  C  R  L  K  K  C  F  E  V  T  M  N  K  D  A  V  Q  H  E  R  G  P  121
301  CGGTTGATAAAATCCACAGAAACCAGTGCCGATCCTGTCGGTTGAAGAAATGCTTTGAAGTCACCATGAACAAAGATGCTGTTCAACATGAACGCGGGCC
                                                                  T

                    R  S  S  T  L  R  K  Q  K  M  L  K  E  A  Q  E  R  M  E  I  A  T  T  A  H  A  Q  Q  S  P  S  G  F  154
401  TCGAAGCTCCACTCTACGGAAACAAAAGATGCTGAAGGAAGCTCAGGAGAGGATGGAGATTGCTACAACTGCGCATGCCCAGCAAAGTCCTAGTGGATTC

                    I  N  T  L  L  A  A  E  T  C  M  E  G  C  Q  L  P  E  N  D  I  D  L  G  L  D  F  K  P  I  R  F  Q   187
501  ATCAACACGTTGCTTGCAGCCGAGACCTGCATGGAGGGCTGCCAGCTTCCAGAAAATGATATTGACCTGGGTTTGGATTTTAAACCAATCAGATTCCAGT
                                   T

                 S  D  L  S  I  S  M  Y  Y  S  H  P  E  A  V  C  E  A  A  A  K  L  L  F  M  S  V  K  W  A  R  N  I  P  221
601  CAGACTTGTCCATATCGATGTATTATTCCCACCCAGAAGCTGTATGCGAGGCCGCTGCCAAGTTATTGTTCATGTCTGTTAAATGGGCCCGGAATATCCC

                    S  F  M  S  L  P  F  R  D  Q  V  I  L  L  E  E  G  W  R  E  L  F  V  L  G  A  S  Q  W  S  M  P  L   254
701  GTCATTCATGAGTCTACCTTTCAGGGATCAGGTTATCCTCCTGGAAGAGGGTTGGCGCGAGCTGTTTGTTCTTGGAGCGAGTCAGTGGTCCATGCCTTTG

                    E  I  A  P  I  L  A  A  A  G  M  H  V  D  T  T  P  P  E  T  I  V  D  V  M  A  T  V  R  L  L  Q  E   287
801  GAAATCGCGCCAATTTTGGCAGCCGCGGGTATGCACGTTGACACTACTCCACCCGAGACGATCGTTGATGTCATGGCAACAGTACGCTTGCTACAAGAAA

                 T  V  N  K  F  K  A  A  N  V  D  S  T  E  Y  A  C  L  K  A  I  V  L  F  K  S  T  A  C  G  L  K  D  P  321
901  CGGTTAACAAGTTCAAAGCAGCCAATGTTGACTCGACTGAATACGCATGTCTGAAAGCGATAGTTCTGTTCAAGTCAACCGCGTGCGGGCTCAAAGACCC
                                       Tlx.F                                  Tlx.R

                    E  Q  V  E  S  T  Q  D  Q  A  Q  L  M  L  G  E  Y  V  R  A  Q  C  P  T  Q  L  A  R  F  G  R  L  L   354
1001 CGAGCAAGTGGAATCTACTCAGGACCAGGCCCAACTCATGCTCGGAGAGTATGTTCGTGCCCAGTGCCCAACACAACTAGCACGGTTTGGACGCCTTTTA

                    L  L  L  P  A  L  R  R  I  S  T  K  E  I  E  E  L  F  F  K  K  T  I  G  A  V  P  I  E  R  L  L  S   387
1101 CTTTTACTACCCGCTTTGCGGCGAATAAGCACCAAGGAAATTGAGGAGCTTTTCTTTAAGAAGACAATAGGAGCTGTTCCCATCGAGAGACTTCTGAGCG

                 D  M  M  K  N  D  *                                                                          393
1201 ACATGATGAAAAATGATTGATAAGTGTAGCAATGAAAATAAGTGATTTTGTTTGTTATAGACGAACGGCCCTGATTCCTCTTTGAGTCTGTATATTCAGA

1301 AGTTCATTAAAGTATTTTGTTTTATTACTAGTTGAAAAAAAAAAAAAAAAAAA
```

Figure 4.1: The *tlx-Am* cDNA. *cont'd.../*

C

```
   Exon 1                  Intron 1                  Exon 2
CGAAACCTAATG gtatttctcgtt..(143bp)..tacgctgcgaag GCCGACTTCTGG

   Exon 2                  Intron 2                  Exon 3
TGAACAAAGATG gtaagagacaaa.(6772bp)..gaccattttcag CTGTTCAACATG

   Exon 3                  Intron 3                  Exon 4
TTCAGGGATCAG gtaagtgcagct..(598bp)..tgtttctaacag GTTATCCTCCTG

   Exon 4                  Intron 4                  Exon 5
TGTTCAAGTCAA gtgagtattgaa..(592bp)..ttcattctacag CCGCGTGCGGGC
```

Figure 4.1: The *tlx-Am* cDNA. (A) Schematic representation of the *tlx-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region, the DNA-binding domain (DBD) is indicated by the blue box and the ligand-binding domain (LBD) is indicated by the red box. The numbers beneath the box represent the nucleotide sequence. (B) The nucleotide sequence of the *tlx-Am* cDNA and predicted amino acid sequence of the protein. The DBD residues are highlighted in blue and the LBD residues are shown in red. An asterisk indicates the 3'- in-frame stop codon. Each of the four intron positions is indicated by a red triangle. Six base pair polymorphisms between the cDNA and genomic sequence are indicated by alternative nucleotides beneath the cDNA sequence at each polymorphism position. Numbers on the left side represent the nucleotide sequence; numbers on the right side represent the amino acid sequence. Primers (tlx.F and tlx.R) used in the real-time quantitative PCR are shown by arrows beneath the nucleotide sequence. (C) The intron-exon boundary of the four introns in *tlx-Am*. The GT-AG consensus splice sites are indicated in red. The total size of each intron is given in brackets.

Levels of identity within the LBD of Tll/Tlx proteins are much lower than within the DBD. However, the coral LBD most closely resembles those in the vertebrate Tlx orthologs (>46% identity; see Figure 4.3), relative identity with invertebrate sequences being much lower. Vertebrate Tlx proteins are highly conserved; human and chick LBD sequences are identical while the mouse and frog sequences differ by one and nine residues respectively. The *D. melanogaster* and *D. virilis* sequences both contain an centrally located 22-residue insertion within the LBD, and differ from each other at only fourteen residues. As it is also from an insect, it is surprising that the *Tribolium* Tlx sequence possesses a truncated LBD, which more closely resembles vertebrate sequences than other invertebrate sequences. The putative Tlx protein from *C. elegans* does not posses a LBD. Relative identity between Tlx sequences outside these conserved domains reflects phylogenetic conservation. For example, both *D. melanogaster* and *D. virilis* contain long peptide insertions between the DBD and LBD that are not present in *Tribolium* (Figure 4.3).

4.2.2. Evolutionary relationships amongst the tll/Tlx nuclear receptor family

To better understand the evolutionary relationship between *tll/Tlx* family members, maximum likelihood phylogenetic analyses were carried out using MolPhy Version 2.3 (Adachi and Hasegawa, 1996) based on regions of the DBD and LBD that could be aligned unambiguously. These sequence comparisons included the DBD, with and without the insertion region in the D-box, and alpha helices 3-5 and 7-11 (Williams and Sigler, 1998) of the LBD (Figure 4.4 A,C). The optimal tree topographies for each analysis are shown in Figure 4.4 (B,D,E).

Phylogenetic analyses of the Tll/Tlx DBD sequences indicate that the coral Tlx-Am is more closely related to the vertebrate sequences than to that of *Drosophila, Tribolium* or *C. elegans*, which appear to be more derived. The coral Tlx-Am sequence consistently clusters with the *Drosophila* Dsf sequence, whether or not the D-box inserted region is included in the analyses (Figure 4.4 D,E). The inclusion of this region does, however, influence which sequence is basal. Phylogenetic analyses of the LBD clearly show that the coral sequence is basal and more closely related to the vertebrate sequences than to the *Drosophila* sequences. The highly derived nature of the *Drosophila* sequences is indicated by the length of those branches.

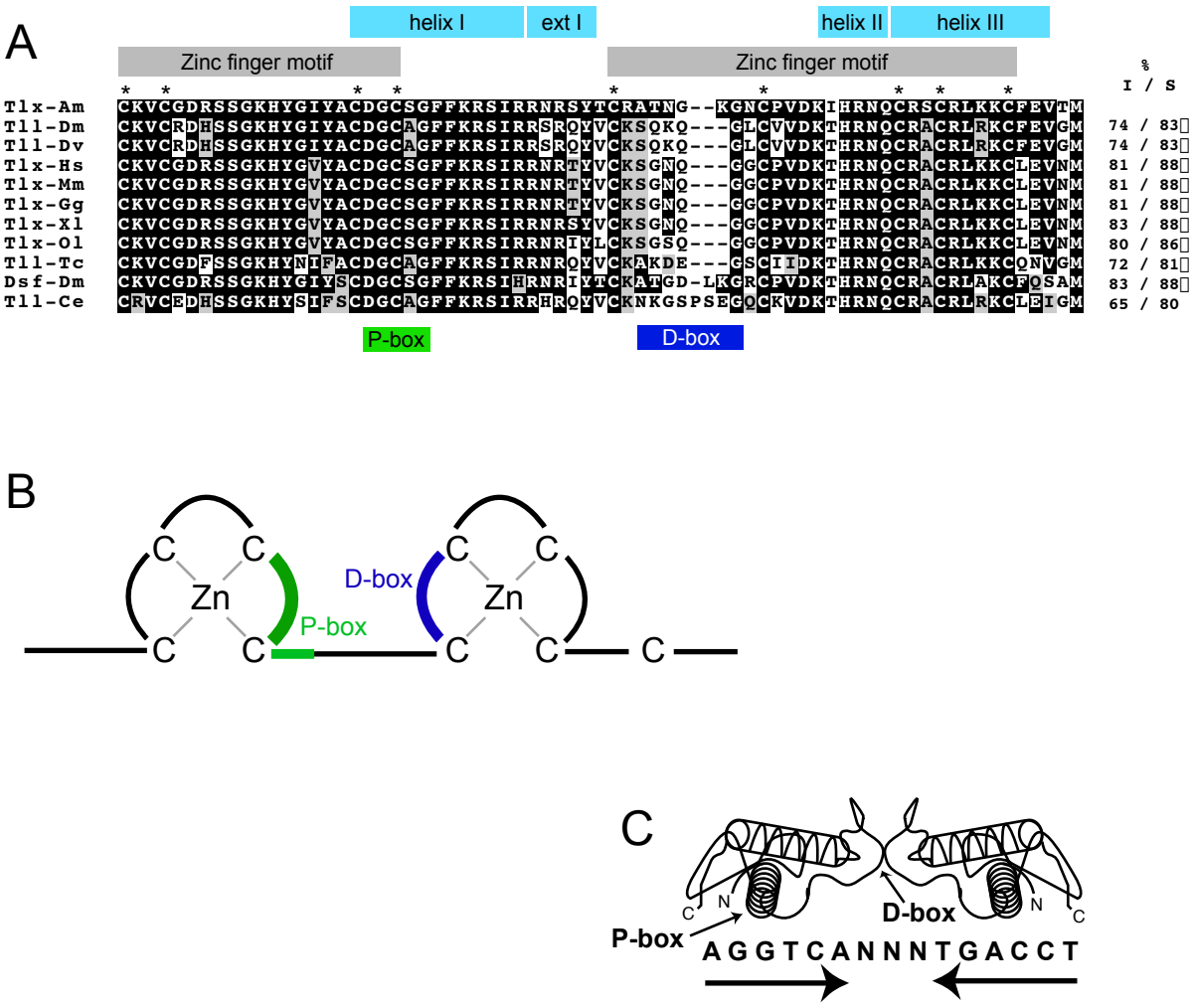


Figure 4.2: A comparison of Tll/Tlx DNA-binding domains. (A) Aligned (Clustal W; Thompson et al., 1994) amino acid sequences of the Tlx-Am DNA-binding domain (DBD) with that of other Tll/Tlx family members, boxshaded (Boxshade server) to indicate residues shared with Tlx-Am DBD. Identical residues are shaded darkly; conserved substitutions are lightly shaded. Secondary structures within this domain are indicated above the alignment with three alpha helices and one extension region represented by light blue boxes. Each of the two zinc finger regions is indicated by a grey box. Nine cysteine residues involved in forming two zinger finger domains are indicated by asterisks. Two additional regions, the P-box and D-box, are indicated below the alignment by green and blue boxes respectively. The column to the right of the alignment indicates the overall identity and similarity of each DBD with the Tlx-Am motif. The species names (and Genbank Accession No.) are abbreviated as follows: Am, *Acropora millepora*; Dm, *Drosophila melanogaster* (AAB71371); Dv, *Drosophila virilis* (AAB71370); Hs, *Homo sapiens* (Q9Y466); Mm, *Mus musculus* (Q64104); Gg, *Gallus gallus* (Q91379); Xl, *Xenopus laevis* (P70052); Ol, *Oryzias latipes* (Q9YGL3); Tc, *Tribolium castaneum* (AAF74116); Ce, *Caenorhabditis elegans* (CAA97428). (B) Schematic diagram of the zinc finger binding motif visualized in expanded format. Colours correspond to (A). (C) Schematic diagram illustrating the point of contact between two DNA-binding domains via the D-box region in the glucocorticoid receptor (derived from Glass, 1994). Assembled as a dimer on a palindromic binding site.

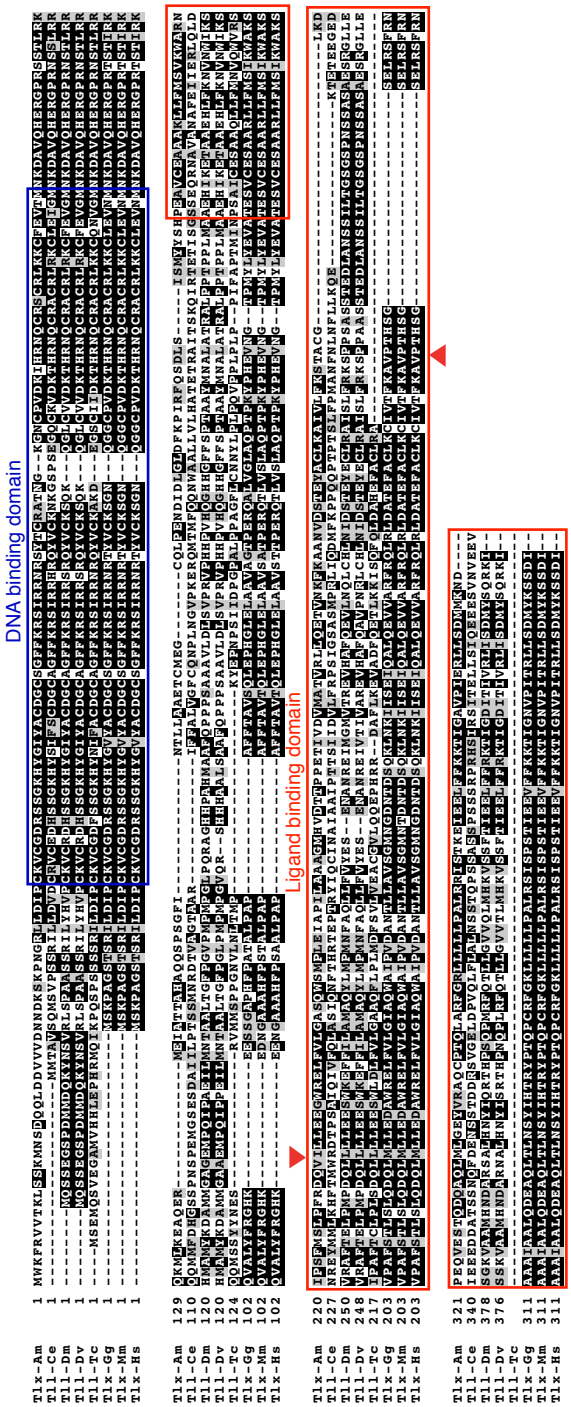
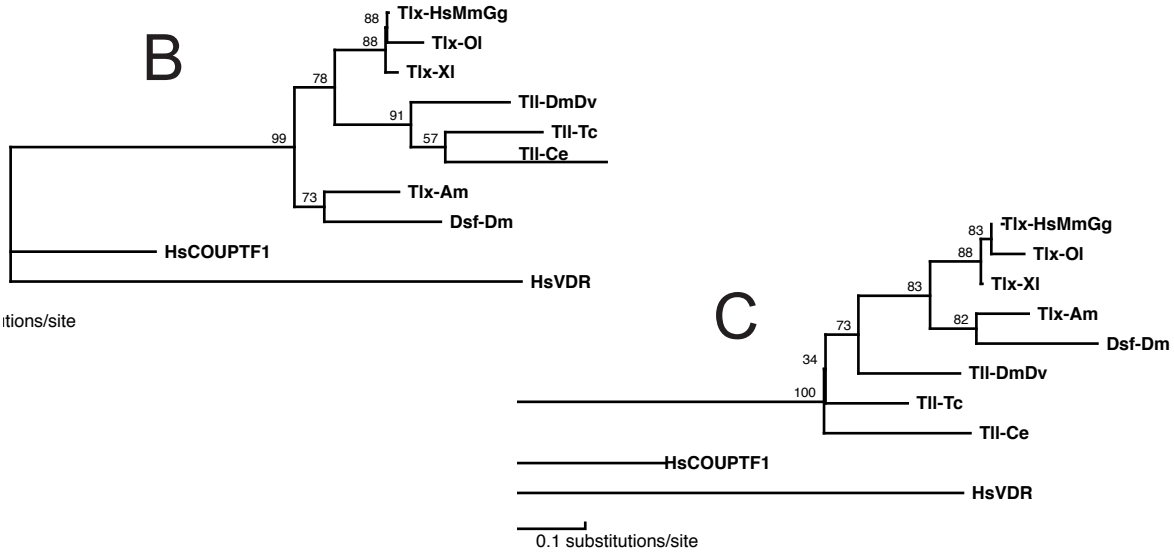


Figure 4.3: A comparison of complete Tll/Tlx proteins from representative organisms.
Aligned (Clustal W -Thompson, 1994) amino acid sequences of the entire Tlx-Am protein sequence with that of other Tll/Tlx family members, boxshaded (Boxshade Server) to indicate residues shared with the Tlx-Am protein. Identical residues are shaded black, conserved substitutions are shaded grey. Numbering in the first column refers to the protein sequence of each organism. The conserved DNA-binding domain is boxed in blue, the semi-conserved ligand-binding domain is boxed in red. The column to the right at the end of the alignment indicates the overall identity and similarity of each LBD with that of Tlx-Am. The species names (and Genbank Accession No.) are abbreviated as follows: Am, *Acropora millepora*; Dm, *Drosophila melanogaster* (AAB71371); Dv, *Drosophila virilis* (AAB71370); Hs, *Homo sapiens* (Q9Y466); Mm, *Mus muscularis* (Q64104); Gg, *Gallus gallus* (Q91379); Tc, *Tribolium castaneum* (AAF74116); Ce, *Caenorhabditis elegans* (CAA97428).


A
P box
D box
Tlx-Am CKVCGDRSSGKHYGIYACDGCSGFFKRSIRRNRSYTCRATNG--KGNCPVDKIHRNQCRSCRLKKCFEVTM
Tll-Ce CRVCEDHSSGKHYSIFSCDGCAGFFKRSIRRHRQYVCKNKGSPSEGQCKVDKTHRNQCRACRLRKCLEIGM
Dsf-Dm CKVCGDRSSGKHYGIYSCDGCSGFFKRSIHRNRIYTCKATGD-LKGRCPVDKTHRNQCRACRLAKCFQSAM
Tll-Dm CKVCRDHSSGKHYGIYACDGCAGFFKRSIRRSRQYVCKSQKQ---GLCVVDKTHRNQCRACRLRKCFEVGM
Tll-Dv CKVCRDHSSGKHYGIYACDGCAGFFKRSIRRSRQYVCKSQKQ---GLCVVDKTHRNQCRACRLRKCFEVGM
Tll-Tc CKVCGDFSSGKHYNIFACDGCAGFFKRSIRRNRQYVCKAKDE---GSCIIDKTHRNQCRACRLKKCQNVGM
Tlx-Ol CKVCGDRSSGKHYGVYACDGCSGFFKRSIRRNRIYLCKSGSQ---GGCPVDKTHRNQCRACRLKKCLEVNM
Tlx-Xl CKVCGDRSSGKHYGVYACDGCSGFFKRSIRRNRSYVCKSGNQ---GGCPVDKTHRNQCRACRLKKCLEVNM
Tlx-Gg CKVCGDRSSGKHYGVYACDGCSGFFKRSIRRNRTYVCKSGNQ---GGCPVDKTHRNQCRACRLKKCLEVNM
Tlx-Mm CKVCGDRSSGKHYGVYACDGCSGFFKRSIRRNRTYVCKSGNQ---GGCPVDKTHRNQCRACRLKKCLEVNM
Tlx-Hs CKVCGDRSSGKHYGVYACDGCSGFFKRSIRRNRTYVCKSGNQ---GGCPVDKTHRNQCRACRLKKCLEVNM


B
88
Tlx-HsMmGg
88
Tlx-Ol
Tlx-Xl
78
Tll-DmDv
91
99
57
Tll-Tc
Tll-Ce
73
Tlx-Am
Dsf-Dm
HsCOUPTF1
HsVDR
itions/site
C
83
Tlx-HsMmGg
88
Tlx-Ol
Tlx-Xl
83
82
Tlx-Am
Dsf-Dm
73
34
Tll-DmDv
100
Tll-Tc
Tll-Ce
HsCOUPTF1
HsVDR
0.1 substitutions/site


D
Helix 3
Helix 4
Helix 5
Tlx-Am PEAVCEAAAKLLFMSVKWARNIPSFMSLPFRDQVILLEEGWRELFVLGASQ
Tll-Dm AEHIKETAAEHLFKNVNWIKSVRAFTELPMPDQLLLLEESWKEFFILAMAQ
Tll-Dv AEHIKETAAEHLFKNVNWIKSVRAFTELPMPDQLLLLEESWKEFFILAMAQ
Tlx-Ol TESVCESAARLLFMSIKWAKSVPAFSTLPLSDQLILLEDAWRELFVLGIAQ
Tlx-Xl TESVCESAARLLFMSIKWAKSVPAFSTLSLQDQLMLLEDAWRELFVLGIAQ
Tlx-Mm TESVCESAARLLFMSIKWAKSVPAFSTLSLQDQLMLLEDAWRELFVLGIAQ
Tlx-HsGg TESVCESAARLLFMSIKWAKSVPAFSTLSLQDQLMLLEDAWRELFVLGIAQ
Helix 7
Helix 8
Helix 9
Helix 10
Helix 11
Tlx-Am MATVRLLQETVNKFKAANVDSTEYACLKAIVLFKS-TA CGLKDPEQVESTQDQAQLMLGEYVRAQCPTQLARFGRLLLLLPALRRISTKEIEELFFKKTIGAVPIERLLSDMMKND--
Tll-Dm TREVHAFQEVLNQLCHLNIDSTEYECLRAISLFRK-SP RGLLESGKVAAMHNDARSALHNYIQRTHPSQPMRFQTLLGVVQLMHKVSSFTIEELFFRKTIGDITIVRLISDMYSQRKI
Tll-Dv AREVHAFQAVPNRLCHLNIDSTEYECLRAISLFRK-SP RGLLESSKVAAMHNDARNALHNYISRTHPNQPLRFQTLLGVVTLMHKVSSFTIEELFFRKTIGDITIVRLISDMYSQRKI
Tlx-Ol MAEIQALQEVVTRFRQMRLDATEFACLKCIVTFKA-VP RAFRNASAIAALQDEAQLTLNSYIHTRYPTQPCRFGKLLLLLPALRSVSPSTIEEVFFKKNIGNVPITRLLSDMYKSSDI
Tlx-Xl ISEIQALQDVVSRFRQLRLDATEFACLKCIVTFKAGVS RNFRNAAAISALQDEAQLTLNSYIHTRYPTQPCRFGKLLLLLPALRSINPSTIEEVFFKKTIGNVPITRVLSDMYKSSDI
Tlx-Mm ISEIQALQEVVARFRQLRLDATEFACLKCIVTFKA-VP RSFRNAAAIAALQDEAQLTLNSYIHTRYPTQPCRFGKLLLLLPALRSISPSTIEEVFFKKTIGNVPITRLLSDMYKSSDI
Tlx-HsGg ISEIQALQEVVARFRQLRLDATEFACLKCIVTFKA-VP RSFRNAAAIAALQDEAQLTLNSYIHTRYPTQPCRFGKLLLLLPALRSISPSTIEEVFFKKTIGNVPITRLLSDMYKSSDI

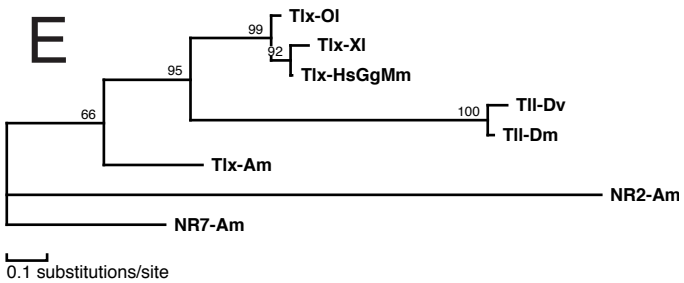

E
99
Tlx-Ol
92
Tlx-Xl
Tlx-HsGgMm
95
66
100
Tll-Dv
Tll-Dm
Tlx-Am
NR2-Am
NR7-Am
0.1 substitutions/site

Figure 4.4: How is Tlx-Am related to other Tll/Tlx proteins? Conserved regions of Tlx-Am and related Tll/Tlx domains that could be unambiguously aligned were subjected to phylogenetic analyses using MolPhy version 2.3 (Adachi and Hasegawa, 1996). (A) Alignment of Tlx-Am DNA-binding domain and related regions. The location of the P-box and D-box motifs are indicated above the alignment. The D-box region containing the inserted amino acids is boxed. (B) Optimal tree topology resulting from the phylogenetic analysis of the entire DBD alignment presented in A. (C) Optimal tree topology resulting from the phylogenetic analysis of the alignment of the DBD without the ambiguous D-box region (boxed) is presented in A. (D) Alignment of the ligand-binding domain of Tlx-Am and related regions. Only the regions that could be aligned unambiguously were included, including helicies 3 - 5, 7, 8 and helicies 9 – 11. (E) Optimal tree topology resulting from the phylogenetic analysis of the LBD alignment presented in D. All trees shown are a result of Maximum Likelihood distance analyses; numbers against the branches indicate the percentage of 1000 bootstrap replicates supporting the tree topology. The species names (and Genbank Accession No.) are abbreviated as follows: Am, *Acropora millepora*; Dm, *Drosophila melanogaster* (AAB71371); Dv, *Drosophila virilis* (AAB71370); Hs, *Homo sapiens* (Q9Y466); Mm, *Mus musculus* (Q64104); Gg, *Gallus gallus* (Q91379); Xl, *Xenopus laevis* (P70052); Ol, *Oryzias latipes* (Q9YGL3); Tc, *Tribolium castaneum* (AAF74116); Ce, *Caenorhabditis elegans* (CAA97428). The outgroups used in these analyses are members of different NR subfamilies and classes to that of the Tlx subfamily (2E) as defined by the NR Nomenclature Committee (1999). Outgroup HsCOUPTF1 (X16155 – subfamily 2F) and HsVDR (J03258 – subfamily 1) were used in analyses of the DBD; outgroups AmNR2 (AAL29194 – subfamily 4) and AmNR7 (AAL29200 – subfamily 2F), both with complete LBD, were used in analyses of the LBD.

4.2.3. Expression patterns of *tlx-Am*

Temporal and spatial distribution of the *tlx-Am* mRNA was investigated during embryonic, larval and post-settlement stages of development. It is important to establish the patterns of *tlx-Am* expression during development to help understand its function.

4.2.3.1. Temporal distribution of the *tlx-Am* transcript

The temporal expression of *tlx-Am* was analysed by Dr. David Hayward (ANU) using northern blot analysis. Messenger RNA from coral embryos at various stages of development was included on a northern blot, which was probed with a radioactively labelled fragment excised from the pBluescript plasmid using *Xba*I and *Xho*I and which contained the complete *tlx-Am* cDNA. Hybridisation revealed a single band of around 1.5 kb (Figure 4.5C), approximately 150 bp longer than the *tlx-Am* cDNA. The *tlx-Am* message was first detected 24 h into development and was still detectable at 48 h. No stages between 48 h and 96 h of development were examined, but expression levels were very low in the stage immediately before settlement (96 h). *tlx-Am* expression could not be detected in the adult tissue (although note that the quantity of mRNA used in this lane was approximately half that of the remaining samples).

The temporal distribution of *tlx-Am* expression was also analysed during embryonic and larval development using quantitative PCR on an SDS 7700 (Applied Biosystems) real-time PCR machine. Expression levels were quantified relative to the levels of the ubiquitously expressed *integrin-β* (Hayward, et al., unpublished). Numerical data from this analysis is given in Table 4.1 and is graphically represented in Figure 4.5A. In general, the relative expression level of *tlx-Am* was 10-fold lower than the transcript levels of another *Acropora* gene, *emx-Am* (see Figure 3.5A). Low levels of *tlx-Am* first were detected in the egg stage but could not be detected again

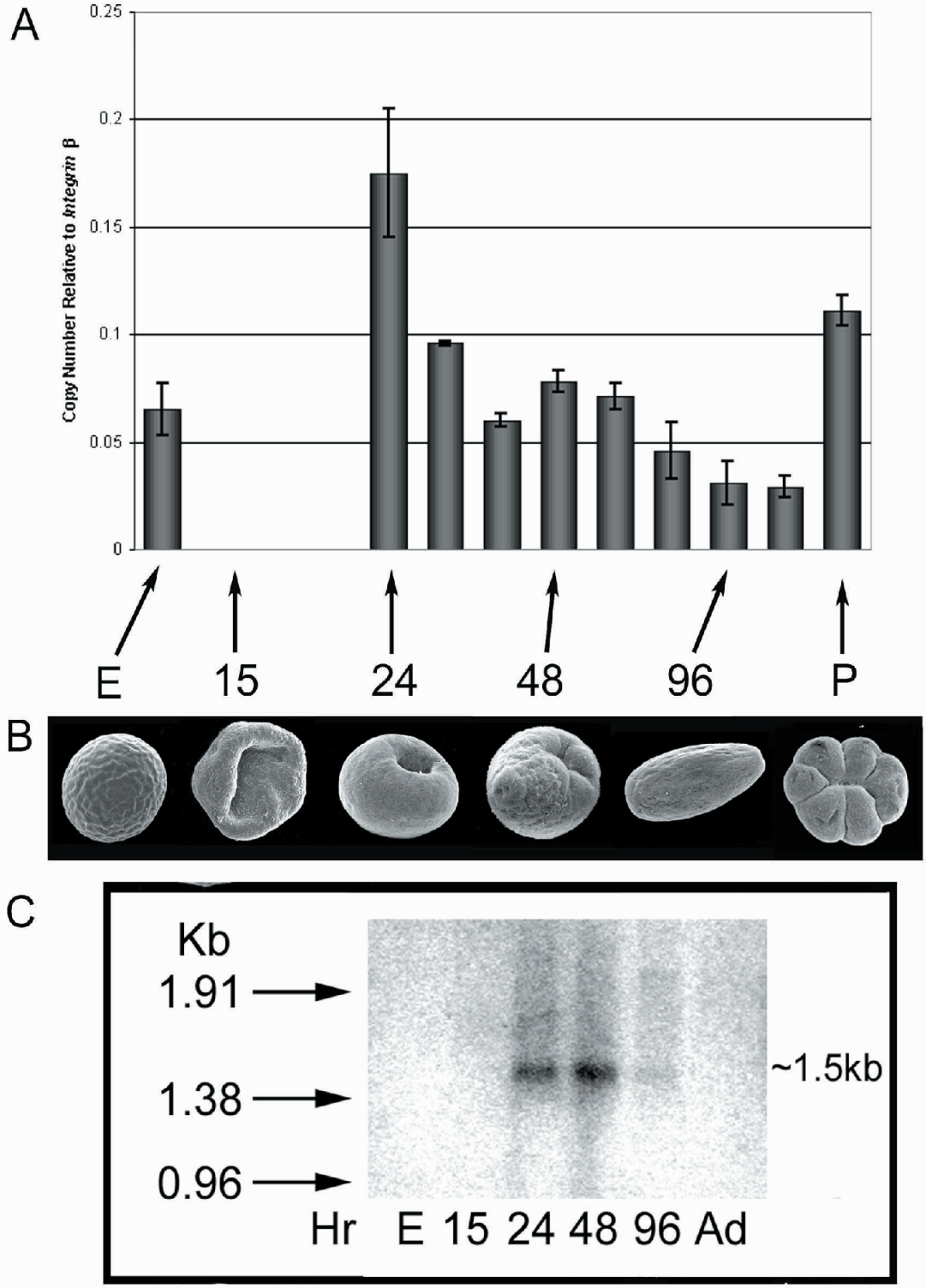


Figure 4.5: Temporal patterns of *tlx-Am* expression during development.
(A) Graphical representation of *tlx-Am* expression levels during development using real-time PCR data, expressed relative to the levels of the ubiquitously expressed *integrin*-β. Numbers on the y-axis indicate the copy number of *tlx-Am* relative to the copy number of *integrin*-β. Thirteen stages were examined. The morphology of stages also studies by northern blotting are shown in (B) (not to scale). (B) Electron micrographs depicting the relative morphologies of coral embryos and larvae from a selection of developmental stages examined in the real-time PCR analysis and mRNA extraction for northern blotting (not to scale). (C) Hybridisation pattern observed after probing a northern blot of mRNA extracted from *Acropora millepora* fertilised eggs (E) and embryos of various ages (numbers indicate hours post fertilisation) with a radioactively-labelled fragment of the *tlx-Am* cDNA. Equal amounts of RNA, as measured spectrophotometrically, were loaded in each lane. However, differing amounts of contaminating ribosomal RNA meant that the relative amounts of messenger RNA loaded per lane were E (egg)=1, 15 h=2, 24 h=2, 48 h=2, 96 h=2, Ad (Adult)=1.

Table 4.1. Temporal patterns of *tlx-Am* expression during development.

Numerical data from RT-PCR of *emx-Am* expression. The morphological appearance of each of developmental stages is described in the column one. The mRNA copy number for both *emx-Am* and *integrin-* for each stage are listed with an average figure and standard deviation. Levels of *emx-Am* expression relative to *integrin-* were determined in the final column and used to construct the graph in Fig 4.5A

Stage	*tlx-Am* copy number	*Integrin-* copy number	*tlx-Am* normalised to *Integrin*
Egg	0.13	1.7	
	0.10	1.8	
	-	-	
Average	0.115±0.021	1.75±0.071	0.065±0.012
Prawn Chip	0	0.59	
	0	0.79	
	0	0.61	
Average		0.663±0.110	0
Donut	0	1.5	
	0	1.4	
	0	-	
Average		1.45±0.071	0
Post Donut	0	0.33	
	0	0.32	
	0	-	
Average		0.325±0.007	0
Sphere with deep pore	0.88	5.5	
	0.92	5.9	
	1.2	-	
Average	1.0±0.174	5.7±0.2	0.175±0.030
Sphere with pore closing	0.27	2.8	
	0.26	2.7	
	-	-	
Average	0.265±0.005	2.75±0.071	0.096±0.001
Pre-Pear 1	0.30	4.4	
	0.28	5.2	
	-	-	
Average	0.29±0.01	4.8±0.566	0.060±0.003
Pre-Pear 2	1.16	15	
	1.28	16	
	-	16	
Average	1.22±0.084	15.6±0.577	0.078±0.005
Pears	2.46	36	
	2.77	37	
	-	37	
Average	2.615±0.219	36.6±0.577	0.071±0.006
Late Pear	0.53	12	
	0.36	10	
	0.64	-	
Average	0.51±0.141	11±1.414	0.046±0.013
Planulae	0.53	14	
	0.43	15	
	0.29	11	
Average	0.417±0.121	13.3±2.08	0.031±0.010
Pre-settlement	0.55	15	
	0.43	18	
	-	18	
Average	0.49±0.084	17±1.732	0.029±0.005
Post-settlement	0.12	1.1	
	0.11	0.98	
	-	-	
Average	0.115±0.007	1.04±0.084	0.111±0.007

until 24 h into development when the embryo resembled a flatten sphere with a depression on one side. Following this stage, expression levels fell approximately 50% but were maintained at this level until after settlement. This data (with the exception of the egg stage) agrees with the expression data from the northern analysis. Also note that post-settlement mRNA was not included on the northern blot.

4.2.3.2. Spatial distribution of the *tlx-Am* message

Attempts to analyse the spatial distribution of *tlx-Am* transcripts in embryos and planula larvae at various development stages were unsuccessful. DIG-labelled riboprobes, generated from linearised plasmid encoding the complete *tlx-Am* cDNA, were used in either the hydrolysed or unhydrolysed state. Repeated experiments using a number of different histochemical kits and variable reaction conditions were unable to detect the presence of *tlx-Am* transcripts. This is an issue experienced with other coral genes, known to be expressed during embryonic development, and a solution is yet to be found. In brief, background staining and staining artefacts make it difficult to interpret the staining pattern for some coral genes.

4.2.4. The *tlx-Am* genomic locus

In order to establish the extent to which the genomic structure of *tlx-Am* has been conserved in evolution, the structure of the *tlx-Am* locus was determined.

An *A. millepora* λGEM12 genomic library was screened under my supervision by Ms. Christine Dudgeon, using the complete 1367 bp *tlx-Am* cDNA fragment as a radioactively labelled probe. Ms Dudgeon screened 150,000 plaques and isolated seven clones of which one, λTLL-1, was sequenced under my supervision by Ms. Dalma Soebok as part of an honours project. This clone, however, did not contain the complete cDNA sequence, due to a large intron preceding the third exon. As part of this project, I screened 80,000 plaques of the genomic library with a 229 bp *EcoR*I/*Pst*I restriction fragment containing the 5'- end of the *tlx-Am* cDNA. One positive clone, λTLL-6, contained the first and second exons of *tlx-Am* and upstream promoter sequence (Figure 4.6). The *tlx-Am* genomic locus contains five exons and four introns, all of which possess donor and acceptor splice junctions that conform to the GT-AG rule (Figure 4.1C). Five polymorphisms were detected between the cDNA and genomic

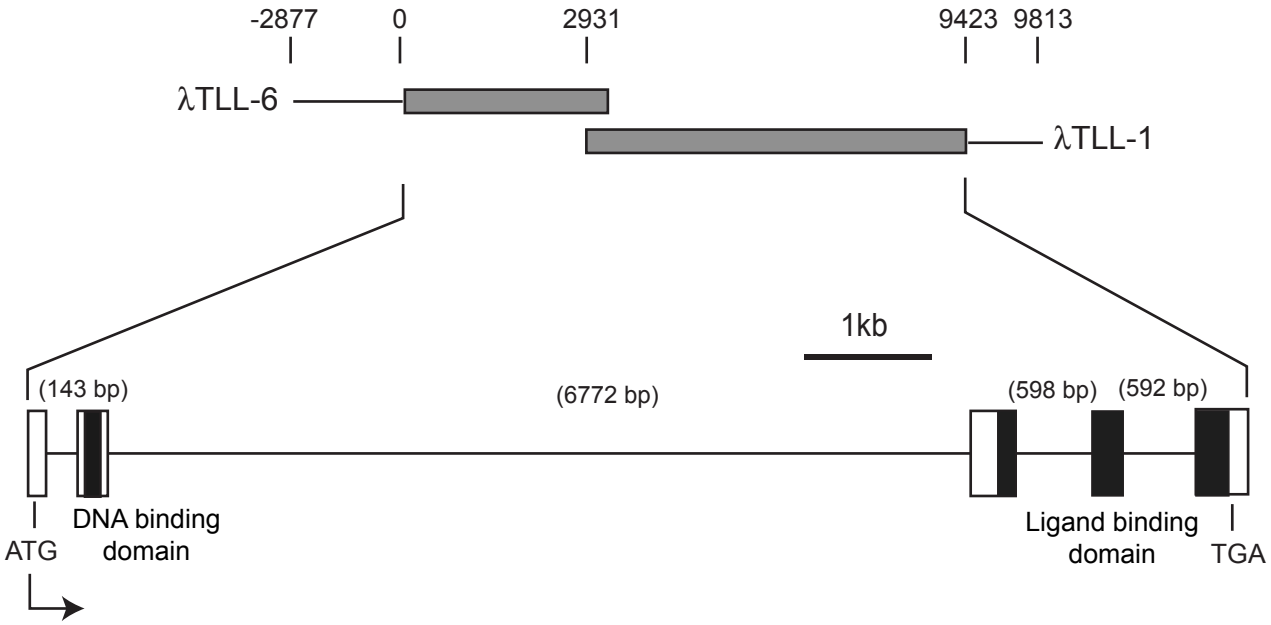


Figure 4.6: Genomic organisation of the *tlx-Am* locus.
(A) Schematic representation of two genomic clones, λTLL-1 and λTLL-6. Grey shaded boxes indicate the region of the genomic clone containing exons. Numbers refer to the nucleotide sequence of the overlapping genomic clones with zero defining the 5'-end of the first exon. The expanded view shows the relative sizes and positions of the exons. The start and stop codons are indicated and the DBD and LBD are represented by blue and red boxes respectively. (B) The intron-exon boundaries of the four introns in *tlx-Am*. The GT-AG consensus splice sites are indicated in red. The total size of each intron is shown in brackets.

sequences (Figure 4.1B), and all but one encodes silent changes in the protein translation. A single polymorphism, located at nucleotide 109, results in either a valine (GTT) or an asparagine (GAT) residue at position 24 in the N-terminal region of the protein (i.e. outside the DNA binding domain).

4.2.5. Characterisation of the *tlx-Am* promoter region

Potential regulatory regions located upstream of the first exon were identified on the basis of sequence identity and comparisons with elements previously identified in other organisms. Nucleotide sequence of λTLL-6 yielded 2.8 kb of sequence upstream of the cDNA sequence. Comparison of this sequence with the BlastX database (Altschul et al., 1990) revealed no open reading frames with significant identity to known proteins, while a BlastN database (Altschul et al., 1990) search showed no identity with previously identified promoter sequences. Possible regulatory regions of *tlx-Am* were identified using transcription factor databases to screen the 2.8 kb of promoter sequence, proximal to the first exon (Figure 4.7). For clarity, the first nucleotide of the cDNA is labelled +1 and the first adjacent nucleotide in the promoter sequence is labelled –1. Putative binding sites of transcription factor and known regulators of Tll/Tlx proteins were identified using MatInspector 2.2 (Quandt et al., 1995) and the TFSEARCH binding site database (Heinemeyer et al., 1998).

Putative basal transcriptional regulator elements including TATA and CAAT boxes were identified using MatInspector 2.2 (Quandt et al., 1995). Northern blot results estimate the size of the *tlx-Am* transcript to be approximately 1.5 kb. Therefore, as the *tlx-Am* cDNA clone is 1.35 kb, the estimated start site of *tlx-Am* transcription must lie within 200 bp of the first exon. Given that functional TATA and CAAT boxes must lie within 50 bp and 100 bp of the initiation start site respectively (Lewin, 1997), those TATA and CAAT motifs lying within –1 bp to –400 bp were identified. The TATA and CAAT boxes located at –144 and – 178 bp respectively fit the spacing requirements for a start site located between –130 and –80 bp.

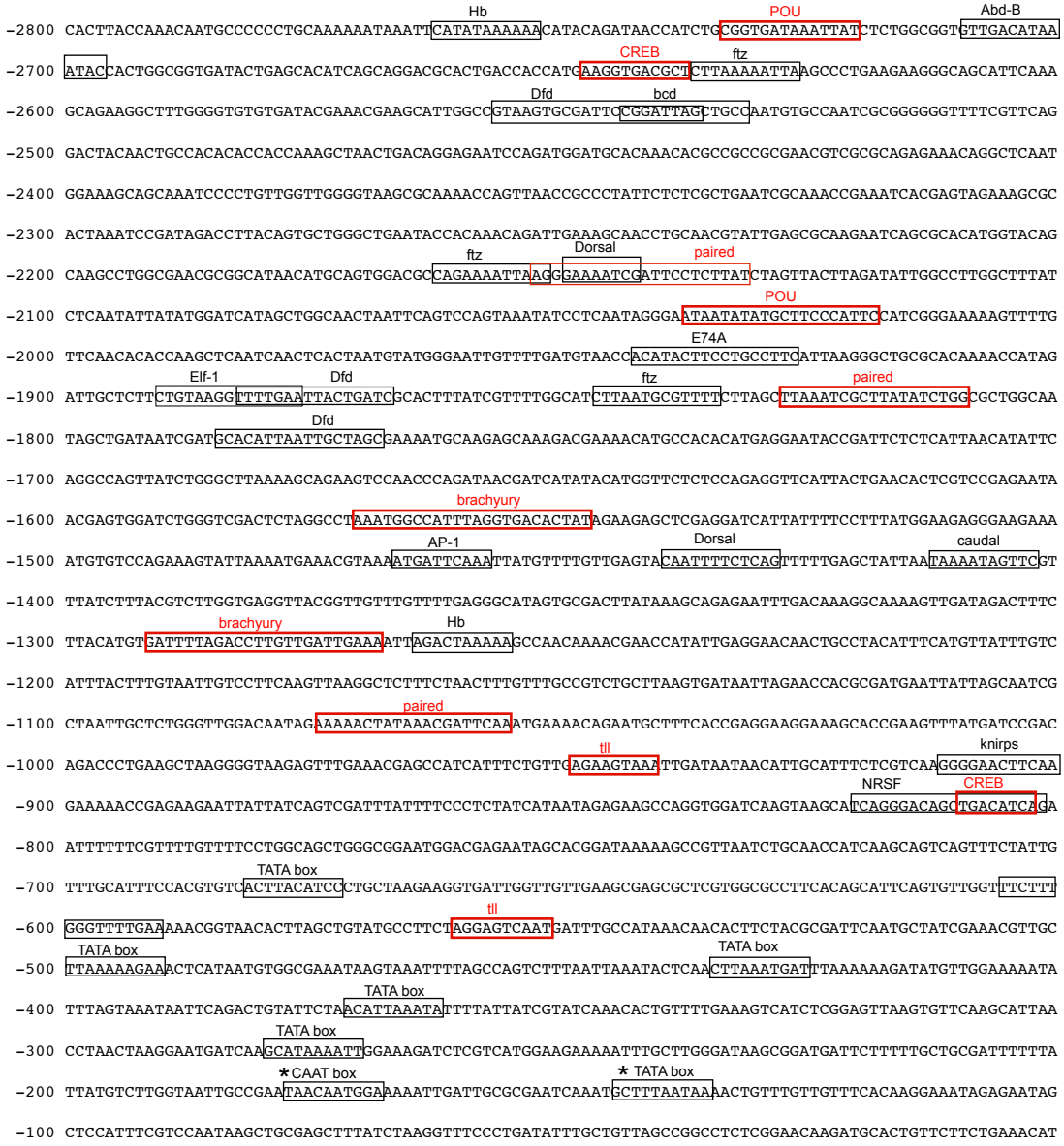


Figure 4.7: The basal promoter of *tlx-Am*.
Nucleotide sequence of the proximal 2.8 kb of genomic clone λTLL-6 upstream of exon 1. Putative binding sites of potential regulating proteins of tlx-Am and basal transcription elements are boxed. Possible regulators include the Tailless nuclear receptor (Tll), homeoproteins (Dorsal, Ftz, Dfd, Hb, Abd-B, Brachyury), other transcription factors (Elf-1), paired-like homeoproteins (paired), POU domain proteins (POU), and proteins responding to external stimuli such as cAMP (CREB, NRSF) and hormones (E74A, AP-1). TATA and CAAT boxes identified within the proximal 500 bp are boxed and labelled. All sites were identified using MatInspector 2.2 (Quandt et al., 1995) and the TRANSFAC 4.0 (Heinemeyer et al., 1998) binding site databases. Numbers to the left of the nucleotide sequence refer to the genomic clone with zero defined as the 5'-end of exon 1.

4.3. DISCUSSION

4.3.1. Structural differences in the DNA-binding domain

The Tlx-Am protein shares a high level of identity with other Tlx proteins, vertebrate homologs in particular, in the DBD and LBD. However, Tlx-Am differs from Tlx proteins in the critical D box region, which may alter DNA binding or dimerisation properties. This region is known to affect preferential spacing between DNA binding half-sites of individual receptors when forming dimers (Luisi et al., 1991; Perlmann et al., 1993; Umesono and Evans, 1989; Rastinejad et al., 1995). Tlx-Am has eight amino acids in this region, whereas *Drosophila* and vertebrate Tlx proteins have seven. This may lead to differences in target site recognition however the isolated DBD of the Dsf protein (which has 9 amino acids in this region) also binds to the Tll AAGTCA half-site (Pitman et al., 2002). Given that this is the case, in the absence of empirical data it is unclear whether the Tlx-Am DBD will bind consensus Tlx targets.

4.3.2. Expression patterns of *tlx-Am*

In the absence of spatial expression data for *tlx-Am*, it is not possible to predict a role for this gene. The temporal expression data provided by northern blot and quantitative PCR analyses indicate *tlx-Am* expression is initiated during gastrulation, as the presumptive ectoderm and endoderm are formed and differentiated, and continues through the planula stage and until after settlement. *tlx-Am* expression was not detected in the adult stage.

Activation of *tlx-Am* during gastrulation suggests its function differs from the neural role of its orthologs in higher animals. Indeed, *tlx-Am* expression is highest during the early stages when neural cell types are still undifferentiated. It is possible that it has a later secondary function in the nervous system but it is premature to predict such a role without spatial expression data.

4.3.3. Regulation of *tlx-Am*

The complex regulation of *tll* in *Drosophila* has been extensively studied and this gene is found to be differentially regulated in the anterior and posterior expression domains. Regulators of vertebrate *Tlx* are less well characterised, but it is possible that regulatory mechanisms of *tll/Tlx* transcription are not highly conserved between vertebrates and

flies. As only a small number of orthologs of developmentally important bilaterian genes have been characterised in *Acropora*, it is difficult to assess the extent of conserved regulatory relationships between coral, flies and vertebrates.

Expression of *Drosophila tll* in different expression zones is mediated by a complex set of regulatory elements. Transcription is both activated and repressed by the activity of Bicoid; repressive activity by Dorsal has also been identified (Liaw and Lengyel, 1992; Pignoni et al., 1992). Two Dorsal binding sites were identified in the coral *tlx-Am* promoter, although an orthologue of this gene is yet to be cloned from any cnidarian.

Binding sites for other transcription factors known to be active during early *Drosophila* development were also identified in the *tlx-Am* promoter, including Deformed, Fushi tarazu, Paired, and Hunchback. Two Tll binding sites were also identified, suggesting the possibility of an autoregulatory feedback loop for *tlx-Am* expression, although there is no precedence for this in *tll/Tlx* genes. A single Caudal binding site is also present in this promoter; in *Drosophila* this gene is expressed during early embryogenesis in a gradient along the A/P axis and is likely to be involved in establishing/maintaining the hindgut primordium (Wu and Lengyel, 1998). While the ancestral origins of anterior patterning are still unclear, it seems likely that regulators of head patterning pathways in the Bilateria are also present in radially symmetrical metazoans.

4.3.3.1. Possible interactions of *tlx-Am* and *emx-Am*

Like its complimentary head gap gene *ems*, *tll* is expressed during early embryonic patterning and brain development in the fly head. Based on *tll* functional mutations, a direct transcriptional regulatory relationship has been demonstrated between these two gap genes in the fly; *tll* directly inhibits *ems* expression in the developing embryonic head via TLL binding sites in a distal *ems*-enhancer element (Hartmann et al., 2001). While similarities are apparent in the expression pattern and function of these two genes between their *Drosophila* and vertebrate homologs, little is known of the *cis*-acting regulatory elements of vertebrate *Emx* genes. As orthologs of both these genes are present in the coral, the question was asked whether this regulatory relationship observed in the fly is also present in the coral.

Investigations to establish whether a regulatory interaction exists between *tlx-Am* and *emx-Am* first focused on the expression patterns of each gene. *tlx-Am* expression is initiated very early in embryonic development, while *emx-Am* was not expressed until much later in development, after the planula had acquired locomotion. Given the non-synchronous timing of expression of these genes, it appeared unlikely that these two genes are regulatory interactors. In the absence of spatial expression data for *tlx-Am*, it was difficult to determine whether similar cell types expressed both genes, such as in the neural cell populations expressing *emx-Am*.

Recently, Tlx was shown to regulate mouse *Pax-2* via a binding site (AAGTCA; (Mangelsdorf and Evans, 1995), which is conserved in both the human and mouse promoters (Yu et al., 2000). This consensus sequence was also present in TLL binding sites identified in the fly *ems* promoter (Hartmann et al., 2001). Investigations failed to locate TLL binding sites (AAGTCA), as previously identified, in the 3 kb *emx-Am* promoter. These data, together with evidence to support a mutually exclusive temporal expression pattern of *tlx-Am* and *emx-Am*, suggest the regulatory relationship observed in *Drosophila* is not present in the coral, and may not have ancestral origins. Comparative studies in vertebrates may determine if similar *cis*-regulatory elements exist in *Emx* genes.

4.3.4. Genomic structure and its evolutionary implications

The genomic structure of a gene is a significant characteristic to consider when interpreting the evolutionary history of a gene family. The intron/exon structures of *tlx/Tll* genes from the human and two insects have been determined and were compared with that of the *Acropora tlx-Am* locus to examine the level of structural conservation (Figure 4.8). The *tlx-Am* genomic locus contains four introns located throughout the coding region, each of which is shared with at least one other organism. Interestingly, *Acropora* has more introns at this locus than *Drosophila* (one) or *Tribolium* (three) but less than in humans (eight). The first of the *Acropora* introns is located in the region upstream of the DNA binding domain, and is conserved in human, *Drosophila* and *Tribolium* loci. In all cases, the intron occurs in phase one (between the first and second base) of the Gly/Ser residue (see Figure 4.8), suggesting that this intron was present in the *tll/Tlx* gene of the last common ancestor.

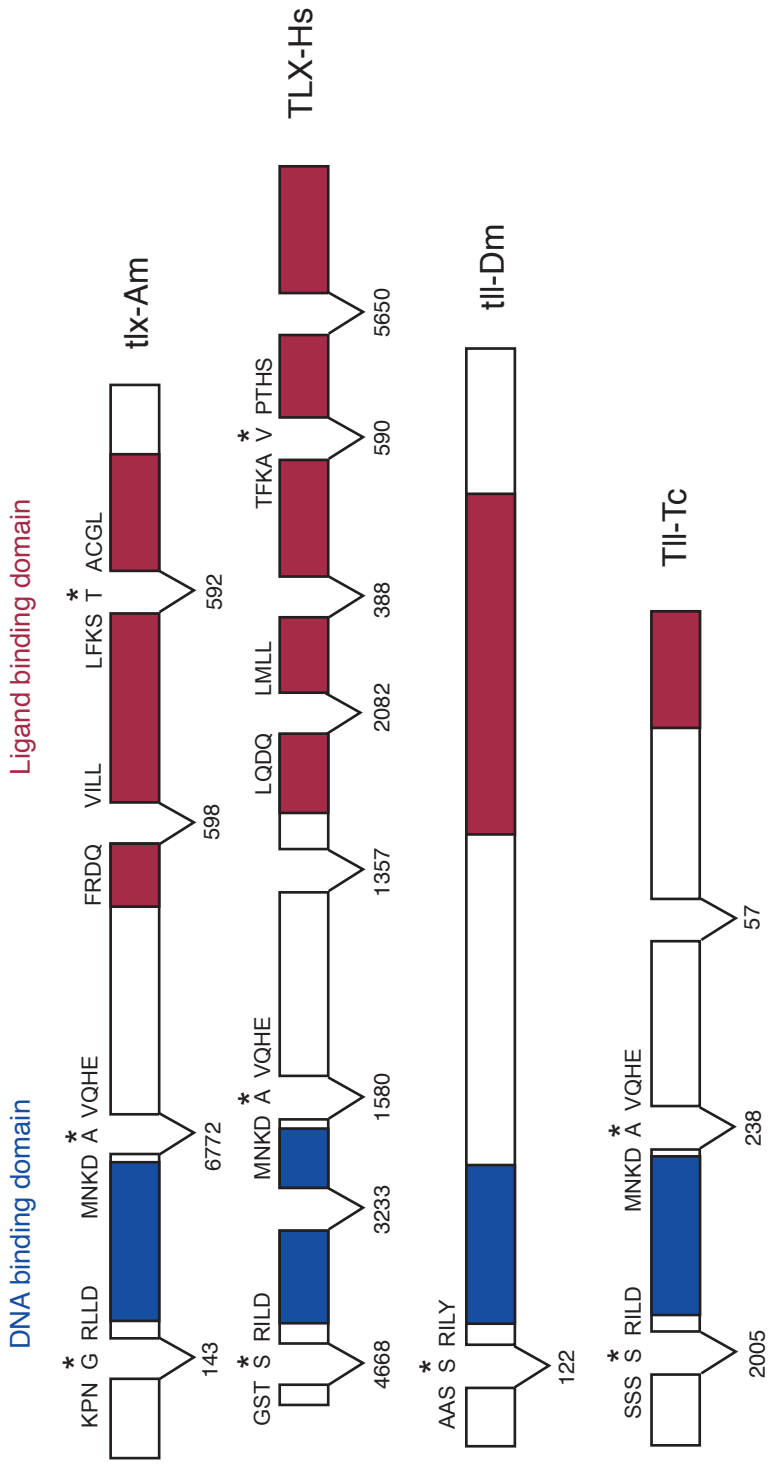


Figure 4.8: Comparative genomic structure of *tll/Tlx* family representatives.
Schematic representation of the intron/exon structure of representative *tll/Tlx* genes from four species. Open boxes indicate the coding regions of each gene, the DNA-binding domain is indicated by blue boxes and the ligand-binding domain is indicated by red boxes. Intron positions are shown by an open triangle beneath each gene and the size of each intron is shown by the number below each triangle (in bp). Letters above each intron location show the encoded residues flanking the splice sites. Residues that are interrupted by an intron are indicated by an asterisk. Species names (and Genbank accession numbers) are abbreviated as follows: Am; *Acropora millepora*; Hs, *Homo sapiens* (AL078596); Dm, *Drosophila melanogaster* (P18102); Tc, *Tribolium castenum* (AAF74116).

There is, however, substantial variation in the size of this intron, which is very small in *Acropora* (143 bp) and *Drosophila* (122 bp), but significantly longer in humans (~4.6 kb) and *Tribolium* (~2 kb). The second intron at the *Acropora tlx-Am* locus is situated in close proximity to the 3' end of the DBD. It occurs between the first and second base of the Ala codon, and is present in both the *Tribolium* and human loci, where protein sequence identity in the surrounding region is 100%. Intron length at this position is also highly variable ranging from 6.7 kb in *Acropora* to 1.6 kb in humans and only 238 bp in *Tribolium*.

The remaining two introns in *tlx-Am* are located within the LBD, where sequence identity is more variable than the DBD. However, the level of conservation is still high enough to indicate these two intron positions are also conserved in the human *Tlx* locus. The first of these LBD introns occurs between codons in conserved regions in both species (between Q and V in *Acropora* and between Q and L in the human) but vary in size (598 bp in *Acropora* and ~2.1 kb in human). Figure 4.8 indicates the position of the second intron in this region; it occurs in phase one of the Thr residue in the coral and of the Val residue in humans. The size of this intron is almost identical at 590 bp and 592 bp in humans and coral respectively. Sequence divergence outside the DBD in *Tribolium* is too high to allow the position of its third intron to be compared with its *tll/Tlx* counterparts. The remaining four introns in the human *Tlx* locus are not shared with any other *tll/Tlx* locus examined.

Given the conserved position of all introns at the *tlx-Am* genomic locus, the most plausible evolutionary explanation is that all introns present in *tlx-Am* were also present in the last common ancestor and that the additional introns observed in the human locus have accumulated during evolution. Alternatively, it may be that all human *Tlx* introns have ancestral origins and that secondary loss is responsible for the reduced intron density in *Acropora, Drosophila* and *Tribolium*.

Sizes of homologous introns in *tll/Tlx* genes vary greatly, as is the case among many homologous gene families, and evidence from *Drosophila* suggests intron size distribution is a process of natural selection (Carvalho and Clark, 1999). In genomic regions of low recombination, intron size is more likely to be very small (<60 bp) or very large (>800 bp) as natural selection is less effective in these regions, but evidence

also suggests transcription is less efficient when large introns interrupt coding regions (Carvalho and Clark, 1999). Transcription of these regions, as well as additional processing such as splicing, is costly and is reflected by a decrease in size and density of introns in highly expressed genes in both humans and *C. elegans* (Castillo-Davis et al., 2002).

4.4. CONCLUSIONS

The *tlx-Am* gene is clearly an orthologous member of the *tll/Tlx* nuclear receptor gene family, but unlike the regulatory relationship observed in *Drosophila* (Hartmann et al., 2001), Tlx-Am is not a direct regulator of *emx-Am* in *Acropora*. Temporal expression of *tlx-Am* from both northern analysis and quantitative PCR data excludes this possibility. Spatial expression data was inconclusive and unable to provide insights into the functional role of *tlx-Am* during coral development.

Comparison of the structure of the *tlx-Am* locus with those of other *tll/Tlx* family members reveals that all four introns present in *tlx-Am* are also conserved in human *Tlx*, while the single intron present in *Drosophila tll* is also conserved in *Acropora*. This suggests *Acropora* has maintained the ancestral intron compliment while secondary loss/gain has accounted for the introns present in other metazoans *tll/Tlx* genes.

Alignment of the DBD of Tll/Tlx proteins also revealed that *Acropora* Tlx-Am possesses an amino acid insertion in a region involved in protein dimerisation in related protein families. This insertion position is also shared in the *Drosophila* Dsf protein and in a related *C. elegans* protein. The functional significance of these extra residues is, however, not known. Sequence conservation is not limited to the DBD; relatively high identity is also apparent in the ligand-binding domain, suggesting this domain is still of functional significance despite the classification of the *tll/Tlx* family as 'orphan' nuclear receptors.

4.5. FUTURE DIRECTIONS

Future studies of the cnidarian tlx genes are open to several lines of investigation. In situ hybridisation studies were unable to determine the spatial distribution pattern of *Acropora tlx-Am*. To determine what role *Tlx* genes play in simple metazoans, cnidarian embryonic data is still required. Isolation of an ortholog in another anthozoan, such as the sea anemone, may provide more promising in situ hybridisation data to reveal possible functions of this gene family in cnidarians.

A wealth of data is available on the DNA binding specificity of the DBD of Tlx proteins. It would be of interest to assess the affinity of the *Acropora* Tlx-Am DBD for DNA elements identified in other animals, and determine if this is affected by the amino acid insertion in the D-box region of this domain. By the same token, a transgenic approach using the coral *tlx-Am* gene to rescue the *Drosophila* null phenotype may help assess the level of functional conservation of this family.