Identification and expression analysis of two novel members of the Mesp family in zebrafish

STEPHEN J. CUTTY*1, RITA FIOR2,3, PEDRO M. HENRIQUES2,3, LEONOR SAÚDE2,3 and FIONA C. WARDLE*1

1Randall Division of Cell and Molecular Biophysics, New Hunt's House, King's College London, Guy's Campus, London, UK, 2Instituto de Medicina Molecular e Instituto de Histologia e Biologia do Desenvolvimento, Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal and 3Instituto Gulbenkian de Ciência, Oeiras, Portugal

ABSTRACT Mesp proteins play crucial roles in the formation of heart, vasculature and somites during vertebrate embryogenesis. We have used phylogenetic and genomic analysis, combined with qRT-PCR and in situ hybridization, to characterize two novel additional mesp genes in zebrafish, mesp-ab and mesp-bb, and describe their expression pattern in wild type and segmentation mutants. Both mesp-ab and mesp-bb are expressed in early mesoderm with mesp-ab expression initiating later, at the end of gastrulation. During somitogenesis, both mesp genes are expressed dynamically in the anterior presomitic mesoderm. mesp-ab is expressed in presumptive somites S-I and S-II, while mesp-bb is detected in S-I, S-II and S0, with expression restricted to the rostral compartment of presumptive somites. We show that the segmentation clock program regulates expression of these newly identified zebrafish mesp genes in a similar manner to their ohnologs, mesp-aa and mesp-ba. We also present evidence that zebrafish, minnow and salmon retained these additional mesp genes after the teleost whole genome duplication, while medaka, stickleback, fugu and tetraodon did not. Finally we show that although expression and regulation of zebrafish mesp genes appears highly comparable, there is no conservation in non-coding regions with other teleosts. In this study we have completed the description of the Mesp family in zebrafish, which will enable correct genome annotation and facilitate further functional studies on the role of these proteins in zebrafish.

KEY WORDS: mesp, zebrafish, presomitic mesoderm, conserved regulation

Introduction

mesp genes, which encode basic helix-loop-helix (bHLH) transcription factors, play important roles in formation of mesodermal tissues during development. In humans, mutations in MESP2 are associated with skeletal defects (Whittlock et al., 2004). Mesp2 knockout mice also show severe axial skeletal defects that result from problems in the formation of the somites, which are the building blocks of the axial skeletal elements of the vertebrate body. In particular, Mesp2 was shown to be important for somite border formation and their rostral-caudal identity (Morimoto et al., 2005, Saga et al., 1997, Takahashi et al., 2000). This crucial role in somitogenesis is conserved in all non-mammalian species examined, such as chick, Xenopus and zebrafish (Saga and Takahashi, 2008). Knockout of Mesp1 in mouse, on the other hand, leads to defects in heart formation, probably through perturbed migration of progenitor cells during gastrulation (Saga et al., 1999). Interestingly, it has also been shown that Mesp1 is able to influence cardiovascular cell fate: overexpression in mouse embryonic stem cells leads to differentiation of cardiovascular cell types (Bonde et al., 2008, David et al., 2008, Lindsley et al., 2008). However, as shown by double knockout and rescue studies in mouse, Mesp1 and Mesp2 together play partially redundant roles not only in cardiac progenitor migration but also in somite formation (Kitajima et al., 2000, Saga, 1998). Interestingly, the single Ciona mesp gene has also

Abbreviations used in this paper: bHLH, basic helix-loop-helix; HCNE, highly conserved non-coding element; PSM, presomitic mesoderm; RACE, rapid amplification of cDNA ends; WGD, whole genome duplication; WISH, whole mount in situ hybridization.

*Address correspondence to: Fiona Wardle. Randall Division of Cell and Molecular Biophysics, New Hunt’s House, King’s College London, Guy’s Campus, London, SE1 1UL, UK. Tel: +44-20-7848-6469. Fax: +44-20-7848-6435. e-mail: fiona.wardle@kcl.ac.uk

Web: http://www.kcl.ac.uk/biohealth/research/divisions/randall/sections/signalling/wardle/index.aspx

Supplementary Material (4 Figures and 3 Tables) for this paper is available at: http://dx.doi.org/10.1387/ijdb.113447sc

Accepted: 31 January 2012. Final, author-corrected PDF published online: 18 April 2012.

ISSN: Online 1696-3547, Print 0214-6282 © 2012 UBC Press
Printed in Spain
been shown to have a role in heart formation (Davidson et al., 2005, Satou et al., 2004), suggesting a conserved ancestral role for mesp genes.

In zebrafish, two mesp genes, mesp-a and mesp-b arranged head-to-tail on chromosome 7 were characterized previously (Durbin et al., 2000, Sawada et al., 2000). Zebrafish mesp-a, for instance, is expressed in the germ ring during gastrulation and was shown to be required for initiation of haematopoiesis (Hart et al., 2007, Sawada et al., 2000). Later in development, mesp-a is expressed in a dynamic manner in the anterior presomitic mesoderm (PSM), from where the somites will form, and plays a role in rostral-caudal somite polarity and boundary formation (Lee et al., 2009, Sawada et al., 2000). mesp-b is also expressed in the anterior PSM and knockdown studies in zebrafish suggest that it has a conserved role with Mesp2 during somite segmentation, being involved in the rostro-caudal polarity of the somites and their epithelialisation (Kawamura et al., 2005, Lee et al., 2009, Sawada et al., 2000). In addition, when zebrafish mesp-b is knocked into the mouse Mesp2 locus it is able to mostly rescue the Mesp2 deficiency, suggesting mesp-bis functionally homologous to Mesp2 (Nomura-Kitabayashi et al., 2002). Interestingly, the regulatory regions that govern expression in the anterior PSM also appear to be conserved between fish mesp-b and mammalian Mesp2, since a medaka mesp-b cis-regulatory module is able to functionally replace the mouse Mesp2 PSM enhancer (Yasuhiko et al., 2008).

Here we report the existence of two additional mesp genes, mesp-ab and mesp-bb on zebrafish chromosome 25. These additional genes likely have arisen as a result of the teleost-specific whole genome duplication (WGD) event that occurred ~350 million years ago (Crow et al., 2006, Hoegg et al., 2004). From our genomic analysis and in accordance with the zebrafish rules we propose a new nomenclature of mesp genes in zebrafish. We also describe the spatio-temporal expression of these genes during embryogenesis and in segmentation mutants. We show that expression of all four zebrafish mesp genes (mesp-aa, mesp-ba, mesp-ab and mesp-bb) is very similar, and that T-box and RBPJ-k binding sites can be found in the genomic sequence surrounding these genes, suggesting that core regulatory elements have been conserved.

Results

Isolation of mesp-ab and mesp-bb

We scanned databases for cDNA, EST, protein and genomic sequences related to mesp-aa and mesp-ba and found two genes which map to zebrafish chr.25. We then isolated full-length cDNA for these genes using PCR and RACE (Supplementary Table S1). Based on sequence, phylogenetic and syntenic analysis presented below and in consultation with the zebrafish nomenclature committee we have named these genes mesp-ab and mesp-bb. The two mesp genes previously identified on chr.7 are now named mesp-aa and mesp-ba.

Full-length mesp-ab encodes a protein of 240 amino acids (compared to 223 aa encoded by mesp-aa), whilst full-length mesp-bb encodes a protein of 244 amino acids (compared to 236 aa encoded by mesp-ba). It has been noted previously that Mesp-related proteins are highly conserved in the bHLH domain, but share little sequence similarity in the rest of the protein, and this is also the case for zebrafish Mesp-ab and Mesp-bb (Supplementary Table S2). For instance Mesp-ab is only 41.2% identical to Mesp-aa overall but across the bHLH they share 87% identity, whilst Mesp-bb is 42.2% identical to Mesp-ba, with 88.9% identity in the bHLH domain. When compared to human MESP1 and MESP2, Mesp-ab and Mesp-bb share between 73.5% - 83.3% identity across the bHLH domains but very little similarity outside this domain (Supplementary Table S2).

Phylogenetic analysis of mesp proteins

We next searched databases for cDNA, EST, protein and genomic sequence for evidence of mesp-related genes in four other teleost species (medaka, fugu, stickleback and tetraodon). In addition to the two published medaka sequences (Terasaki et al., 2006) we found sequences in fugu, tetraodon and stickleback, which encode mesp-a and mesp-b genes based on the sequence and phylogeny analysis presented here (Fig. 1A; Supplementary Table S1). A neighbour joining tree of these mesp-related genes is shown on Fig. 1A and indicates that mesp-ab is mostly closely related to mesp-aa and mesp-bb to mesp-ba. Interestingly we were only able to find evidence for one mesp-a and one mesp-b in fugu, medaka, tetraodon and stickleback (see below and discussion). This could be explained by two scenarios: either the mesp-a/mesp-b locus duplication only occurred in zebrafish or it was specifically lost in the other four fish species after the teleost WGD. However, if we could find other teleost fish containing a mesp locus duplication, then this would argue for the second scenario-retention in zebrafish and loss in the other fish. To test this we further searched trace archives and found sequences related to mesps in minnows (Pimephales promelas and Poeciliopsis turneri) and atlantic salmon (Salmo salar). The genomes and EST projects of these fish are still preliminary and we were not able to identify all likely mesp genes in these fish. However our analysis indicates that the mesp-bb duplication is not unique to zebrafish, but is also present in fathead minnow (Pimephales promelas), suggesting that fugu, medaka, tetraodon and stickleback may have lost this extra set of mesp genes (Fig. 1B). Salmon, which has undergone a 4th round of genome duplication after the teleost WGD, also show a mesp locus duplication (Fig. 1B).

Syntenic analysis of mesp proteins

Zebrafish mesp-aa and mesp-ba are found in tandem on chr.7 flanked by an sv2b homologue and an anpep homologue (ENS-DARG00000089708) and in a landscape containing akap13 and igf1rb. We find mesp-ab and mesp-bb in a similar landscape on chr.25 in Zv9 (Fig. 1C). We also find fugu, medaka, tetraodon and stickleback mesps surrounded by genes homologous to svb2, anpep, akap13 and igf1r in addition to ap3s2. Interestingly this synthony is partially maintained in tetrapods, with the Mesp locus being surrounded by Anpep and Ap3s2 to one side, but Wdr93 and Gmnr2 on the other side.

Medaka chr.3 and chr.6 share a common ancestor with zebrafish chr.7 and chr.25 (Kasahara et al., 2007, Woods et al., 2005) and since medaka mesp-a and mesp-b are mapped to chr.3 we also looked at chr.6 for evidence of additional mesp-related genes that would support the hypothesis that mesp-ab and mesp-bb arose during the teleost WGD. Although we were able to find similar surrounding genes (sv2b and anpep) on the medaka chr.6 assembly we were unable to find any mesp-related gene sequence in this region (Fig. 1C). However the genomic sequence in this area is
Fig. 1. Phylogenetic analysis and genomic organization of mesp-related genes. (A) Optimal Neighbour-Joining tree showing inferred evolutionary relationships between zebrafish (Dr), fugu (Tr), medaka (Ol), tetraodon (Tn) and stickleback (Ga) mesp genes. The tree is rooted with Amphioxus mesp (Bf mesp). Bootstrap percentages over 50% are shown next to the branches. The scale bar infers the number of amino acid substitutions per site. mesp-a genes are shown in dark green, mesp-b genes in light green. (B) Neighbour-joining tree including minnow (Pm, Pp) and salmon (Ss) sequences. (C) Arrangement of genes surrounding mesp-related genes in zebrafish, medaka, fugu, tetraodon and stickleback. mesp genes are found in tandem and surrounded by akap13, sv2b, anpep, ap3s2 and igf1r in all fish (except zebrafish chr. 7 which does not have ap3s2 in this region). Homologous genes are shown in the same colour across species, and annotated with gene name or gene identifier. In the case of zgc:73349 homologues of this gene are not annotated in medaka and stickleback but inspection of the genomic suggests the sequence is present. Additional inserted genes on zebrafish chr25 are shown in grey. Dotted line indicates incomplete genomic sequence. Genomic coordinates are given for each region shown based on genome sequence versions Zebrafish Zv9, MEDAKA1, FUGU4, TETRAODON8, Stickleback BROADS1.
incomplete and thus we can not rule out that additional mesp genes exist in medaka. A search for additional sv2b, anpep, or other flanking genes in fugu, tetraodon and stickleback did not reveal any evidence of a duplicated mesp-containing region in these species. Although more comprehensive genomic sequence for other teleosts will be needed to confirm it, the results presented above support a view that mesp-ab and mesp-bb are the result of a duplication of an ancestral chromosomal region, which now lies on chr.7 and chr.25 in zebrafish.

Expression of mesp-related genes in wild-type embryos

We next characterized the temporal and spatial expression of the newly identified mesp-ab and mesp-bb genes in early zebrafish embryos. qPCR analysis indicates that mesp-ab and mesp-bb have very similar temporal expression profiles to mesp-aa and mesp-ba, respectively (Fig. 2A). The expression of both mesp-aa and mesp-ab initiates at around 40% epiboly stage, with mesp-ab being more highly expressed than mesp-aa. Expression of both genes peaks at shield stage and continues at low levels until 26-somite stage. In contrast the expression of mesp-ba and mesp-bb initiates later in development at bud stage, and continues at similar levels until late somite stages. Expression of all four genes is gone by 24 hours-post-fertilization (hpf; prim 5) when somitogenesis completes.

To analyse spatial expression we generated whole mount in situ hybridization (WISH) probes, which recognize the 3’ end of each RNA (in order to avoid the bHLH domain which may cross-react between paralogues) and compared these with probes which recognize the full-length RNA. We found no difference in expression patterns between full-length and bHLH-free probes (Supplementary Fig. S1 and data not shown) and since the full-length probe gave a stronger signal we used these for the following analyses (Figs. 2-4).

Fig. 2. Embryonic expression of mesp-ab and mesp-bb. (A) qPCR time course of mesp gene expression from 1-cell to prim-5 (24 hours-post-fertilization). Expression is shown in units relative to total 18S RNA expression at the stage shown. (B,C) Expression of mesp-ab is not detected at 1-cell or dome stage. Animal to the top. (D,E) mesp-ab expression initiates in the ventral and lateral margin at 40% epiboly and persists in a similar pattern at shield stage. Dorsal view; animal to the top. (F) Animal view of shield stage mesp-ab expression, ventral to the top. (G,H) As gastrulation proceeds and mesoderm ingresses, expression persists in the lateral mesoderm (arrows), and is down-regulated on the ventral side (arrowhead). (G) Shows dorsal view; animal to the top, (H) shows animal view, ventral to the top. (I) At the end of gastrulation stripes are seen in the paraxial mesoderm. Dorsal view, anterior to the top. (J) Anterior view, ventral to the top, of bud stage mesp-ab expression in paraxial mesoderm. (K-O) mesp-ab expression persists in paraxial mesoderm during somitogenesis. (K) Dorsal view, anterior to the top, at 3-somite stage showing one stripe of expression. (L) Lateral view, anterior to the top left, at 3-somite stage showing one stripe of expression. (L’-L”*) Enlarged dorsal view of mesp-ab expression in paraxial mesoderm at 3-somite stage showing one or two stripes of expression (asterisks). (M) Lateral view, anterior to the top left, at 12-somite stage. (M’-M”*) Enlarged dorsal view of mesp-ab expression in paraxial mesoderm at 12-somite stage showing one or two stripes of expression (asterisks). (N) Lateral view, anterior to the top, of mesp-ab expression in the tail at 26-somite stage. (O) Dorsal view, anterior to the top, of mesp-ab expression in the tail at 26-somite stage. (P,Q) Expression of mesp-bb is not detected at 1-cell stage or throughout early development; 70% epiboly shown as an example. Animal to the top. (R) mesp-bb expression is first detected at bud stage when stripes are seen in the paraxial mesoderm. Dorsal view, animal to the top. (S) Animal view, ventral to the top, of bud stage mesp-bb expression in the paraxial mesoderm (arrows). (T-W) mesp-bb expression persists in paraxial mesoderm during somitogenesis. (T) Dorsal view, anterior to the top, at 3-somite stage showing two stripes of expression. (T’-T”*) Enlarged dorsal view of mesp-bb expression in paraxial mesoderm at 3-somite stage showing two or three stripes of expression (asterisks). (U) Lateral view, anterior to the top left, at 12-somite stage showing two stripes of expression. (U’-U”*) Enlarged dorsal view of mesp-bb expression in paraxial mesoderm at 8-12-somite stage showing one or two stripes of expression (asterisks). (V) Lateral view, anterior to the top, of mesp-bb expression in the tail at 26-somite stage. (W) Dorsal view, anterior to the top, of mesp-bb expression in the tail at 26-somite stage.
Spatially mesp-ab expression also closely resembles mesp-aa expression (Durbin et al., 2000, Sawada et al., 2000). In accordance with our qPCR results, mesp-ab is not detected maternally or at early blastula stages by WISH (Fig. 2 B,C) but is first detected at 40% epiboly in the blastoderm margin where it is excluded from the dorsal region (Fig. 2D). At shield stage expression continues to be seen in the ventral-lateral margin (Fig. 2 E,F). At 70% epiboly expression is seen in the mesodermal layer of cells (Fig. 2 G,H; arrows), with expression beginning to recede from the ventral region (Fig. 2H; arrowhead). From tailbud stage through somite formation stages transverse stripes of expression are seen in the anterior region of the PSM (Fig. 2 I-O). In the PSM, mesp-ab expression is dynamic and either one or two sets of stripes can be observed in embryos of the same stage until near the end of somitogenesis (26-somite stage), when only one stripe is observed (Fig. 2 K-O).

Similarly mesp-bb expression closely resembles mesp-ba expression (Sawada et al., 2000). In accordance with our qPCR results, we find mesp-bb is not detectable by WISH maternally or during blastula and gastrula stages (Fig. 2 A,P-Q). Expression is detected at bud stage when transverse stripes are seen in the PSM (Fig. 2 R-S; arrows). These stripes are expressed dynamically and at bud stage either one, two or three stripes are observed, while either two or three stripes persist throughout the majority of somitogenesis (Fig. 2 T-U”). However, by 26-somite stage, as somite formation is finishing, only one stripe of mesp-bb expression is seen (Fig. 2 V-Y).

Previous studies have shown that mesp-aa is expressed at the anterior PSM in a territory that corresponds to the prospective somites S-I (restricted to the rostral compartment) and S-II (Durbin et al., 2000, Sawada et al., 2000). To define the precise location of mesp-ab expression within the prospective somites we compared its expression with mesp-aa and with the caudal somite compartment marker myod1, using double WISH. This analysis revealed that mesp-aa and mesp-abexpression overlaps during somitogenesis (Supplementary Fig. S2). Moreover, when compared with myod1 we see that the most anterior mesp-ab stripe is positioned immediately adjacent and posterior to the last myod1 stripe present in prospective somite S0 (Fig. 3A-C). Thus mesp-ab expression is located in the rostral compartment of prospective somites in the anterior PSM.

Comparison of mesp-bb expression with mesp-ba and myod1 (Fig. 3D-F; Supplementary Fig. S2) also confirms that mesp-bb

Figure 3. Expression of mesp-ab and mesp-bb in the anterior presomitic mesoderm. (A) Double in situ showing expression of myod1 (red) and mesp-bb (blue) at 8-somite stage. (B,C) Fluorescent double in situ showing expression of myod1 (red) and mesp-ab (green) at 8-somite stage. The most anterior stripe of mesp-ab in S-I is immediately adjacent to the caudal myod1 expression in S0. Dynamic mesp-ab expression also overlaps with myod1 expression in the adaxial cells (white arrows). (D) Double in situ showing expression of myod1 (purple) and mesp-bb (blue) at 8-somite stage. (E,F) Fluorescent double in situ showing expression of myod1 (red) and mesp-bb (green) at 8-somite stage. Expression of mesp-bb in S-I abuts myod1 expression in S0. Dynamic mesp-bb expression also overlaps with myod1 expression in the adaxial cells (white arrows).

Figure 4. Expression of mesp-ab and mesp-bb in segmentation mutants. (A,H) Wild-type. (B,I) fss/tbx24-/-, (C,J) bea/deltaC-/-, (D,K) aei/deltaD-/-, (E,L) wit/mib-/- embryos at the 8-somite stage hybridized with mesp-ab (A-E) and mesp-bb (H-L) in situ probes. (G,N) wit/mib-/- and (F,M) wild-type or wit/mib-/- sibling embryos at 17-somite stage hybridized with mesp-ab (F,G) and mesp-bb (M,N). S0, S-I and S-II mark the position of presumptive somites. All embryos were flat-mounted and are shown in dorsal views, anterior to the top.
expression, similarly to mesp-ba is found at the rostral compartment of prospective somites S0, S-I and S-II (Sawada et al., 2000; Fig. 3D-F).

The similarity in expression patterns between all mesp zebrafish genes suggest that, although their coding sequences have changed since duplication, the regulatory elements are more constrained and have remained alike.

**The new mesp genes are also downstream of the segmentation program**

It was shown previously that mesp genes are downstream of the segmentation program (Durbin et al., 2000, Sasaki et al., 2011, Sawada et al., 2000, Yasuhiko et al., 2006). In zebrafish two classes of mutants that show a striking segmentation phenotype were identified in the large scale genetic screens: fused somites (fss), in which the T-box family transcription factor tbx24 is mutated, and several Notch pathway genes (Holley, 2007). In fss/tbx24 mutants no somitic boundaries are formed, while in Notch signalling mutants such as bea/deltaC-/- and wit/mib-/- the segmentation clock is desynchronized and irregular somitic boundaries are made (reviewed in Lewis et al., 2009).

Our results and others show the expression of mesp-aa is absent from fss/tbx24-/- mutants and very weakly expressed in bea/deltaC-/-, aei/deltaD-/- and wit/mib-/- mutants (Supplementary Fig.S3 B-E; Durbin et al., 2000, Sawada et al., 2000). In addition mesp-ba expression is also absent in fss/tbx24-/- mutants, while expressed in a diffuse pattern in the bea/deltaC-/-, aei/deltaD-/- and wit/mib-/- mutants (Supplementary Fig.S3 G-J; Sawada et al., 2000). These expression data suggest that while mesp-aa activation is shared by Tbx24 and Notch signalling, mesp-ba expression, although clearly synchronized by the Notch pathway to make perfect stripes, is primarily dependent on Tbx24 activity.

The newly identified members mesp-ab and mesp-bb show a similar expression regulation by the segmentation program to their ohnologues. Similar to mesp-aa, the expression of mesp-ab is absent in fss/tbx24-/- mutants (Fig. 4B) and very weak or absent in the strongest Notch signalling mutant the wit/mib-/- (Fig. 4 E-G). In the mutants for the Notch ligands bea/deltaC-/- and aei/deltaD-/-, the expression of mesp-ab is weak and diffuse (Fig. 4 C,D), suggesting a redundant requirement of the two ligands. The expression of mesp-bb is also absent in fss/tbx24-/- mutants (Fig. 4I) and weak and diffuse in the Notch signalling mutants bea/deltaC-/-, aei/deltaD-/- and wit/mib-/- (Fig. 4 J-N). In contrast to mesp-ab and similar to its ohnologue mesp-ba, the expression in the strongest Notch signalling mutant wit/mib-/- is still detected (Fig. 4N), suggesting that the mesp-b group is mainly regulated by Tbx24.

**T-box and RBPJ-κ binding sites are present in the mesp locus regulatory regions**

Previous studies have shown T-box and RBPJ-κ binding sites are important in regulating somite expression of mesp genes in medaka and mouse (Terasaki et al., 2006, Yasuhiko et al., 2006, Yasuhiko et al., 2008). Since the expression patterns of the mesps are highly similar we asked whether these binding sites are present in teleost mesp loci using the matrix-scan program within RSAT (Thomas-Chollier et al., 2008, Turatsinze et al., 2008; Fig. 5). This revealed many T-box and RBPJ-κ binding sites in all teleost species, and identified the previously discovered T-box and RBPJ-κ binding sites upstream of medaka mesp-b which regulate gene expression in the somites (Terasaki et al., 2006, Yasuhiko et al., 2008; Fig. 5). Clusters of T-box and RBPJ-κ binding sites are also seen upstream of the teleost mesp-a genes, except in stickleback

---

**Fig. 5. Analysis of regulatory regions across teleost mesp loci.** Regions encompassing 5 kb upstream of mesp-b to 2 kb downstream of mesp-a in the mesp loci of zebrafish chromosome 7 and 25, medaka, fugu, tetraodon and stickleback are shown with T-box (blue) and RBJP-κ binding sites (red). Previously identified T-box (T1, T2, T3) and RBPJ-κ (W1) binding sites that regulate medaka mespb expression in somites are indicated. Regions boxed in purple are conserved between the two zebrafish loci while regions boxed in orange are conserved between medaka, fugu, tetraodon and stickleback. See Supplementary Fig. S4 for conservation details.
which appears to lack T-box binding sites in this region.

We also compared these genomic regions using VISTA which detects regions of high conservation (HCNEs; Frazer et al., 2004; Supplementary Fig. S4). This analysis revealed limited conservation in non-coding regions between the two zebrafish loci, other than in the intron of mesp-ab/bb, in the intergenic region between the two mesp genes, and downstream of mesp-aa/ab. Surprisingly there was no homology in non-coding regions between zebrafish and the mesp loci in other teleosts, suggesting that if regulatory sites are conserved between zebrafish and other teleosts they do not fall in a region of high conservation (Supplementary Fig. S4A). When medaka was compared to the other teleosts conservation in non-coding regions was seen between fugu, tetraodon and stickleback, but not zebrafish, in the region upstream of mesp-b and between the two mesp genes (Supplementary Fig. S4B). Interestingly the regulatory T-box (T1-3) and RBPJ-k (W1) binding sites identified upstream of medaka mesp-b fall in or close to these regions of conservation, suggesting this regulation may be conserved in percomorphs (Supplementary Fig. S3B; Fig. 5). Short regions of high conservation were also found upstream of the mesp-a genes, although these regions were not conserved between zebrafish and other teleosts (Fig. 5; Supplementary Fig. S4).

These results indicate that genomic sequence within the mesp loci has also diverged, although T-box binding sites upstream of mesp-b genes have remained in a conserved region in medaka, fugu, tetraodon and stickleback, but not zebrafish.

Discussion

We have isolated two mesp-related genes in zebrafish, mesp-ab and mesp-bb. The phylogenetic and syntenic analysis presented here suggests that these are ohnologues of mesp-aa and mesp-ba, generated by the teleost whole genome duplication. This duplication appears to have been retained in zebrafish and at least one other cyprinid (fathead minnow) but lost in percomorphs (medaka, fugu, tetraodon and stickleback). However more complete EST and genomic sequences in other fish will be needed to examine this issue fully.

After genome duplication there are several hypotheses for the fates of the new pair of genes: one of the new pair can either be lost; can take on a new function (neo-functionalization); may share the function with the ancestral gene (sub-functionalization); or a combination of these fates (neo-sub-functionalization). Functional change may be brought about by changes in protein coding sequence, or through changes in regulatory elements that cause divergence in spatio-temporal expression of duplicated genes (Jimenez-Delgado et al., 2009).

Our expression analysis suggests that mesp-ab and mesp-bb have close to identical expression patterns with mesp-aa and mesp-ba, respectively. We do not believe these similar expression patterns are due to probe cross-reactivity between the bHLH domains of the mesp-a genes (or mesp-b genes) because, for instance, the bHLH sequence between mesp-aa and mesp-ba is more similar than between the two mesp-a genes (80.2% vs 75.3% identity respectively; Supplementary Table S2) yet mesp-aa and mesp-ba give different temporal and spatial expression patterns (Sawada et al., 2000; data not shown). This suggests that differences seen in expression are specific to that gene.

The similarity in expression patterns between the ohnologues suggests that the core regulatory elements of these genes have yet to evolve in zebrafish, which is highly unusual for ohnologues (Kassahn et al., 2009) and suggests these regulatory sequences are in someway constrained. Another explanation would be that these genes are not ohnologues, but the result of more recent duplication which would mean regulatory sequence has not yet had time to evolve. However, we did not find a great deal of conservation in non-coding regions between the two zebrafish loci, which would argue against this possibility. However since the mesp regulatory sequences in zebrafish are yet to be mapped, it is not known if these fall within the short HCNEs that were found upstream of zebrafish mesp-aa/ab. Intriguingly though, mesp expression in the anterior PSM is very similar in mouse, Xenopus, zebrafish and medaka which suggests that the enhancers regions for the mesp genes in all vertebrates may be conserved. In support of this, enhancer analysis has shown that a 350 bp region upstream of mouse Mesp2, a 2.8 kb region upstream of medaka mesp-b, a 2.9 kb region upstream of fugu mesp-b and a 3.5 kb region upstream of Xenopus laevis mesp-b are able to drive expression of a reporter in anterior PSM in a pattern that recapitulates endogenous Mesp2/mesp expression (Haraguchi et al., 2001, Moreno et al., 2008, Moreno and Kintner, 2004, Terasaki et al., 2006). In addition, the medaka 2.8 kb mesp-b enhancer can substitute for the mouse Mesp2 350 bp enhancer, suggesting the mesp-b anterior PSM enhancer is conserved in vertebrates despite differences in size (Yasuhiro et al., 2008). In Ciona T box binding sites found upstream of mesp are critical for expression of a reporter in the heart/muscle progenitor cells (Davidson et al., 2005), whilst in medaka and mouse T-box sites upstream of mesp-b/Mesp2 are also required for reporter expression in stripes at the anterior PSM (Terasaki et al., 2006, Yasuhiro et al., 2008). Our binding site analysis indicates that clusters of T-box sites are present upstream of all mesp-bs, and that in medaka, fugu, tetraodon and stickleback these sites fall within or close to regions of high conservation. Consistent with this we found that mesp-b expression was dependent on Tbx24 activity. RJBP-k sites upstream of medaka mesp-b and mouse Mesp2 have been shown to act together with T-box sites to enhance reporter gene expression but are not required for activation (Terasaki et al., 2006, Yasuhiro et al., 2008); again this is consistent with our observations which indicate mesp-ba/bb expression in zebrafish is not dependent on Notch signalling. The regulatory regions that govern mesp-a expression have not been investigated, but our segmentation mutant analysis suggests both Notch signalling and Tbx24 are required in a non-redundant manner for normal expression in the anterior PSM. Interestingly clusters of T-box and RJBPK sites are found upstream of mesp-a genes, except in stickleback, and may suggest that these sites are also active in regulating expression from the mesp locus.

We also note that the expression levels of mesp-ab mRNA are higher than mesp-aa during gastrulation and somitogenesis, and similarly mesp-bb expression is higher than mesp-ba during somitogenesis (Fig. 2A). While it is tempting to speculate that mesp-ab and mesp-bb could therefore play a greater role in development than their counterparts on chromosome 7, until protein levels are determined and functional studies performed on these two new genes this can not be determined.

Although the spatio-temporal regulation of mesp gene expression is highly similar the protein sequences are quite dissimilar, particularly in the region outside the central bHLH domain (Supple-
mentary Table S2). The bHLH domain is involved in DNA binding but the function of the N- and C-terminal domains of these proteins is less well understood. Recently however, the N-terminal domain of Mesp2 has been shown to regulate Notch signalling through interacting with and destabilizing the Mastermind-like 1 protein (Sasaki et al., 2011). Despite differences in the sequence of Mesp proteins they seem to have similar functions in the embryo since different Mesp proteins are able to compensate for each other. For instance, although zebrafish mesp-ba and mouse Mesp2 are only 24.8% identical overall (but 73.5% over the bHLH domain) mesp-ba is able to rescue a Mesp2 knockout mouse. In addition, Mesp1 is able to rescue the Mesp2 knockout in mouse, suggesting that these proteins also function in a similar fashion (Saga, 1998). Despite this there are some differences in function between different Mesp proteins as illustrated by a study which showed mouse Mesp2 and Xl-mesp-b are able to activate a lunatic fringe promoter reporter construct more strongly than either mouse Mesp1 or Xl-mesp-a (Hitachi et al., 2009).

In conclusion, the expression of mesp-ab and mesp-bb appears conserved with their homologues in fish and other vertebrates during early mesoderm formation and somitogenesis, suggesting they may also function in a conserved way in mesoderm and somite patterning.

Materials and Methods

Zebrafish lines
AB wild type and mutant (fsis/tbx2
+/-, bea/deltaC
+/-, ael/deltaD
+/-, and wit/mib
2-2) embryos were staged according to Kimmel et al., 1995.

Isolation of mesp-ab and mesp-bb
PCR primers were designed against reported cDNA sequence for mesp-ab (previously known as mesp2) and coding genomic sequence for mesp-bb and used to amplify cDNA from a mixed stage sample consisting of 75%, bud, 4-6 somite stage and 9-11 somite stage. Products were cloned into pGEM-T Easy vector (Promega) and verified by sequencing. Primer sequences are reported in Supplementary Table S2. 5' and 3' RACE was then used to identify full-length cDNA at 4-6 somite stage. Briefly for 5' RACE, total RNA was reverse transcribed with a gene specific primer and GoScript reverse transcriptase (Promega). A homopolymeric dC tail was added to the 5' end. For 3' RACE, total RNA was primed with 3'RACE adaptor primer and reverse transcribed as above. UTR sequences were obtained by PCR with primers gene specific and adaptor primers, cloned into pGEM-T Easy vector and verified by sequencing. Full-length sequences are provided in Supplementary Table S1 and have been submitted to Genbank with accession numbers JN084105 (mesp-ab) and JN084106 (mesp-bb).

Sequence retrieval and phylogenetic analysis
Where possible full-length cDNA and protein sequences were downloaded from Ensembl or Genbank. For zebrafish mesp-ab and mesp-bb we used the full-length sequences identified in this paper. For medaka mesp-a and mesp-b we used sequences from Terasaki et al., 2006. For fugu mesp-a, fugu mesp-b and tetraodon mesp-a sequence analysis suggested the sequences available in the databases were not full length and so we used additional genomic sequence, alignment with known full-length transcripts and conserved exon/intron boundaries to add 5' and 3' sequence to the putative coding transcript. For stickleback and tetraodon mesp-b sequences, which have not been annotated, we searched the genomic sequence between mesp-a and sv2b and found evidence for a mesp-b-like bHLH coding sequence. The alignment of this sequence with other fish mesp-bs and conserved exon/intron structure allowed us to identify a putative coding sequence. Searches of databases at NCBI did not uncover any additional sequences for these fish. For salmon and minnow searches of EST, genome and trace archive sequences were performed. All sequences used here are reported in Supplementary Table S1.

Sequence alignment and phylogenetic tree analysis was performed in MEGA5 (Tamura et al., 2011). MUSCLE was used to align sequences and then improved manually; analyses used alignments of the bHLH domain plus some flanking amino acid sequence. Both Neighbour Joining and Maximum Likelihood analyses were performed and gave similar results. 1000 bootstrap replicates were performed. Evolutionary distances were computed using the JTT matrix-based method.

Syntenic analysis
We analysed the region around mesp-aa/mesp-ba and mesp-ab/mesp-bb in zebrafish and in other species manually and using Genomicus (Muffato et al., 2010). The region from 11646289 to 11178241 on chr.25 (Zv9) is shown in Fig. 1, but it should be noted that annotations for another mesp-related gene, sv2b and anpep can also be found in the interval 11129393-11051080 on chr.25. However the quality of sequence and assembly over this region of chr.25 is poor and this apparent duplication could be due to an assembly error (see pgpviewer.ensembl.org).

Conservation and binding site analysis
Genomic sequence of the mesp locus from medaka, fugu, tetraodon, stickleback and zebrafish chr. 7 and chr. 25, including 5kb upstream of mesp-bb and 2kb downstream of mesp-a was analysed using MLAGAN in VISTA (Frazer et al., 2004). The same sequence was then analysed to discover T-box and RBP-Jκ binding sites using RSAT (Thomas-Chollier et al., 2008, Turatsinze et al., 2008). Previously identified position weight matrixes for Brachyury and Tbx16 (Bryne et al., 2008, Garnett et al., 2009) and RBP-Jk (Krejci et al., 2009, Tun et al., 1994) were used. Sites with a p value of <0.0001 were considered binding sites. RSAT provides a graphical output of these sites which was used in Fig. 5 then further annotated with the mesp genes and conserved regions identified in VISTA.

qPCR
Total RNA was extracted using Trizol reagent (Invitrogen) and treated with RQ1 DNase. The total RNA was retrotranscribed with GoScript reverse transcriptase (Promega) using random hexamers and oligo(dT)15. Products were purified using Qiaquick PCR purification kit (Qiagen). Reactions were performed using SYBR Green I Master kit (Roche) on a Stratagene MX3005P and data analyzed using MxPro software. Cycling parameters were as follows, 95°C 10 min for 1 cycle then 95°C 30 sec, 60°C 30 sec, 72°C 30 sec for 30 cycles then 95°C 60 sec, 60°C 30 sec, 95°C 30 sec for all primer sets. Primer sequences used can be found in Supplementary Table S3. For each replicate data were normalized to an 18S rRNA endogenous control. The mean expression level with standard error of the mean of three replicates are reported.

In situ hybridization
mesp-aa and mesp-bab plasmids were a kind gift of Hiroyuki Takeda (Sawada et al., 2000). For mesp-ab and mesp-bb we amplified full-length sequence from cDNA and cloned these into pGEM-T Easy vector. For probes without bHLH domain we used PCR to isolate the 3' end of the transcript. For primers details see Table S3. Whole-mount in situ hybridizations were performed according to standard protocols (Nusslein-Volhard and Dahm, 2002). Whole-mount double fluorescent in situ hybridizations were performed as described previously (Julich et al., 2005) with minor modifications: the red signal was developed with FAST RED (Roche AP substrate) and the green with Tyramide FITC (POD substrate).

Acknowledgements
We thank Dr Roll Roberts for his advice on phylogenetics and members of our labs for discussion. We are grateful to Lara Carvalho and Fábio Valério for excellent fish care and Ana Margarida Cristofoelho for technical support. FCW is supported by a Medical Research Council Career Development Award and Lister Institute Research Prize. R.F. is supported by
References


Further Related Reading, published previously in the *Int. J. Dev. Biol.*

**The Xenopus Bowline/Ripply family proteins negatively regulate the transcriptional activity of T-box transcription factors**

Keisuke Hitachi, Hiroki Danno, Shunsuke Tazumi, Yuko Aihara, Hideho Uchiyama, Koji Okabayashi, Akiko Kondow and Makoto Asashima  
*Int. J. Dev. Biol.* (2009) 53: 631-639

**Dynamic expression pattern of distinct genes in the presomitic and somitic mesoderm during Xenopus development**

Audrey Bourdelas, Hong-Yan Li, Clémence Carron and De-Li Shi  
*Int. J. Dev. Biol.* (2009) 53: 1075-1079

**Xtbx6r, a novel T-box gene expressed in the paraxial mesoderm, has anterior neural-inducing activity**

Shigeharu Yabe, Shunsuke Tazumi, Jun Yokoyama and Hideho Uchiyama  
*Int. J. Dev. Biol.* (2006) 50: 681-689

**Bowline, a novel protein localized to the presomitic mesoderm, interacts with Groucho/TLE in Xenopus**

Akiko Kondow, Keisuke Hitachi, Tempei Ikegame and Makoto Asashima  
*Int. J. Dev. Biol.* (2006) 50: 473-479

**Evolution of the Hox/Parahox gene clusters.**

David E K Ferrier and Carolina Minguillón  
*Int. J. Dev. Biol.* (2003) 47: 605-611

---

**5 yr ISI Impact Factor (2010) = 2.961**