Research article

Comparative analysis of somitogenesis related genes of the *hairy/Enhancer of split* class in *Fugu* and zebrafish

Martin Gajewski* and Chris Voolstra*

Address: Universitaet zu Koeln, Institut fuer Genetik, Weyertal 121, 50931 Koeln, Germany

E-mail: Martin Gajewski* - martin.gajewski@uni-koeln.de; Chris Voolstra* - chr.voolstra@uni-koeln.de

*Corresponding authors

Abstract

Background: Members of a class of bHLH transcription factors, namely the *hairy* (*h*), *Enhancer of split* (*E(spl)*) and *hairy*-related with YRPW motif (*hey*) (*h/E(spl)/hey*) genes are involved in vertebrate somitogenesis and some of them show cycling expression. By sequence comparison, identified orthologues of cycling somitogenesis genes from higher vertebrates do not show an appropriate expression pattern in zebrafish. The zebrafish genomic sequence is not available yet but the genome of *Fugu rubripes* was recently published. To allow comparative analysis, the currently known Her proteins from zebrafish were used to screen the genomic sequence database of *Fugu rubripes*.

Results: 20 *h/E(spl)/hey*-related genes were identified in *Fugu*, which is twice the number of corresponding zebrafish genes known so far. A novel class of c-Hairy proteins was identified in the genomes of *Fugu* and *Tetraodon*. A screen of the human genome database with the *Fugu* proteins yielded 10 *h/E(spl)/hey*-related genes. By analysing the upstream sequences of the *c-hairy* class genes in zebrafish, *Fugu* and *Tetraodon* highly similar sequence stretches were identified that harbour Suppressor of hairless paired binding sites (SPS). This motif was also discovered in the upstream sequences of the *her1* gene in the examined fish species. Here, the Su(h) sites are separated by longer intervening sequences.

Conclusions: Our study indicates that not all *her* homologues in zebrafish have been isolated. Comparison to the human genome suggests a selective duplication of *h/E(spl)* genes in pufferfish or loss of members of these genes during evolution to the human lineage.

Background

Somitogenesis is the phase in embryonic development of vertebrates during which the somites, that is segmented mesodermal tissue along each side of the notochord, are successively generated. Besides other models [1], the "clock-and-wavefront" model [2] is able to explain somite formation. This model proposes the existence of a cellular clock that interacts with a slowly progressing wave (wavefront) moving along the presomitic mesoderm (PSM) from anterior to posterior. Genes of the *h/E(spl)/hey* family have been identified in chicken and mouse that show a cycling expression pattern, namely *c-hairy1*, *c-hairy2*, *c-hey2*, *mhes1* and *mhes7*[3–9]. These bHLH proteins are thought to obey the proposed molecular clock.

Analysis of the family of *h/E(spl)*-related genes in zebrafish has yielded 9 *her*-genes so far (*her1-6*: [10,11], *her7* and *her8* unpublished own work, accession numbers: AF240772, AY007990, *her9*: [12]) and one member of the *hey* class, called *gridlock*[13]. *her1*, *her4*, *her6* and *her7* are thought to play a role in somite formation [14–16]. *her4* and *her6* are expressed in the developing nervous system and in the anterior PSM of zebrafish embryos but they do not cycle [10,17]. The *her1*[15,18] and the *her7* gene (unpublished) are both cyclically expressed in the PSM during somitogenesis. Both genes are only distantly related to the cycling *c-hairy* type genes from higher vertebrates. In zebrafish, *her9*, the apparent orthologue of the cycling genes from chicken and mouse does not show any expression at all in the PSM [12]. Therefore, it has been suggested that the zebrafish *her1* gene is the functional homologue to the cycling *c-hairy* type genes and thus an element of the molecular clock. Alternatively, more than two *c-hairy* type genes (*her6* and *her9*) exist in zebrafish. Since whole genome sequence information for zebrafish is still missing, we chose to exploit the publicly available draft sequence of the *Fugu* genome [19] to identify all the H/E(spl)/Hey-related proteins of this organism. The evolutionary divergence time of both fish species is about 180 myr compared to 430 myr to tetrapods [20]. Hence, the emerging picture might give us a clearer view on the situation of the *h/E(spl)/hey*-related gene family in zebrafish than from comparison to higher vertebrates. Due to a remarkable conservation of exon-intron structure of the *h/E(spl)*-related genes from insects up to man, it is possible to annotate the proteins of this family on the level of genomic sequence for *Fugu rubripes*. Since the automatically performed annotation was not able to correctly predict the protein sequences (the N- and/or C-termini often were missing), it became necessary for comparative analysis to re-compile the sequences on the basis of an exon-intron dataset of known *h/E(spl)* genes.

It has been shown, that the pufferfish *Fugu rubripes* possesses a very small genome with a size of only 400 Mb [21]. Compared to zebrafish or man the genome size in pufferfish is 4 to 8 times reduced. This reduction refers mostly to the intergenic regions and repetitive elements, while *Fugu* seems to possess a similar gene set with the same structure than mammalian organisms [22]. Comparison of promoter sequences from homologous genes in both fish species should facilitate the identification of candidate transcription factor binding sites.

Results and Discussion
In the genome of Fugu rubripes significantly more h/E(spl)/hey-related genes were identified than in Danio rerio are currently known

Screening of the genomic sequence database of *Fugu rubripes* with the so far known zebrafish "Hairy and Enhanc-er of split"-related (Her) proteins by the TBLASTN program [23] revealed 20 *h/E(spl)/hey*-related genes in *Fugu* (table 1, see additional file 1, additional file 2 and additional file 3. A BLASTN search with the compiled genes against the available cDNA database [ftp://ftp.hgmp.mrc.ac.uk/pub/fugu/fugu_cdna.zip] gave no matches, indicating that the dataset (3142 entries) is too small. All assembled sequences showed conserved exon-intron boundaries and no unexpected stop-codons, suggesting that the genes are functional. In general, the *Fugu* proteins were named according to their overall similarity to the zebrafish Her proteins. To reveal the relationships within the bHLH proteins of *Fugu rubripes* the sequences were aligned (fig. 1) and a corresponding tree was calculated (fig. 2A). The *Fugu* proteins could be clearly divided into the three classes Enhancer of split type (3 members, fig. 2A: orange box), Hey type (5 members, fig. 2A: light red box) and Hairy type (12 members, fig. 2A: bluegreen, light green, dark green and violet boxes).

In more detail: Deduced from the intronless genomic structure, which is indicative for *E(spl)* genes in *Drosophila melanogaster*[24], FrHer2 (T002603/1) and both Her4 proteins (T002603/2, T002289) were identified as Enhancer of split type proteins. *Fugu her2* and one of the *her4* genes are located close to each other in a head to tail orientation on contig T002603 (see additional file 2). However, FrHer2 (T002603/1) does not cluster with the other 2 Enhancer of split type proteins in the tree.

Five *Fugu* proteins were classified as Hey (4) or Hey-related (1) compared to only one currently known zebrafish member of that class, called Gridlock [13]. Homologues for each of the *hey* or *hey*-related genes in *Fugu* have already been isolated in higher vertebrates [5,25,26] or were previously annotated in the genbank database (accession number: XM_068025 for "bHLH protein similar to *hhesr1/hey1*"). The C-terminal consensus sequence of the Hey proteins is KP**YRPW**GTE(I[Hey1]/V[Hey2])GAF. Focussing on this consensus, FrHey1.1 (contig T004546), FrHey1.2 (contig T000531) and FrHey2 (contig T014948) could be determined for the pufferfish (table 1 and fig. 1). Interestingly, all 3 proteins show at least one exchange to the C-terminal consensus sequence. According to the similarity in the bHLH-domain the two other proteins in contig T003476 and T005657 belong also to the Hey type proteins (fig. 1). However, like mouse HRT3 ([27], accession number: AF172288) FrHey3 (contig T005657) possesses the TEIGAF motif but deviates to a higher degree from the YRPW-motif. The other protein (in contig T003476) does not possess a KP**YRPW**GTE(I[Hey1]/V[Hey2])GAF motif at the C-terminal end. Instead a completely diverged amino acid sequence is present. Furthermore, the basic domain is somehow truncated (fig. 1, see additional file 4). Interestingly, this protein consists of 4 exons. This is in contrast to
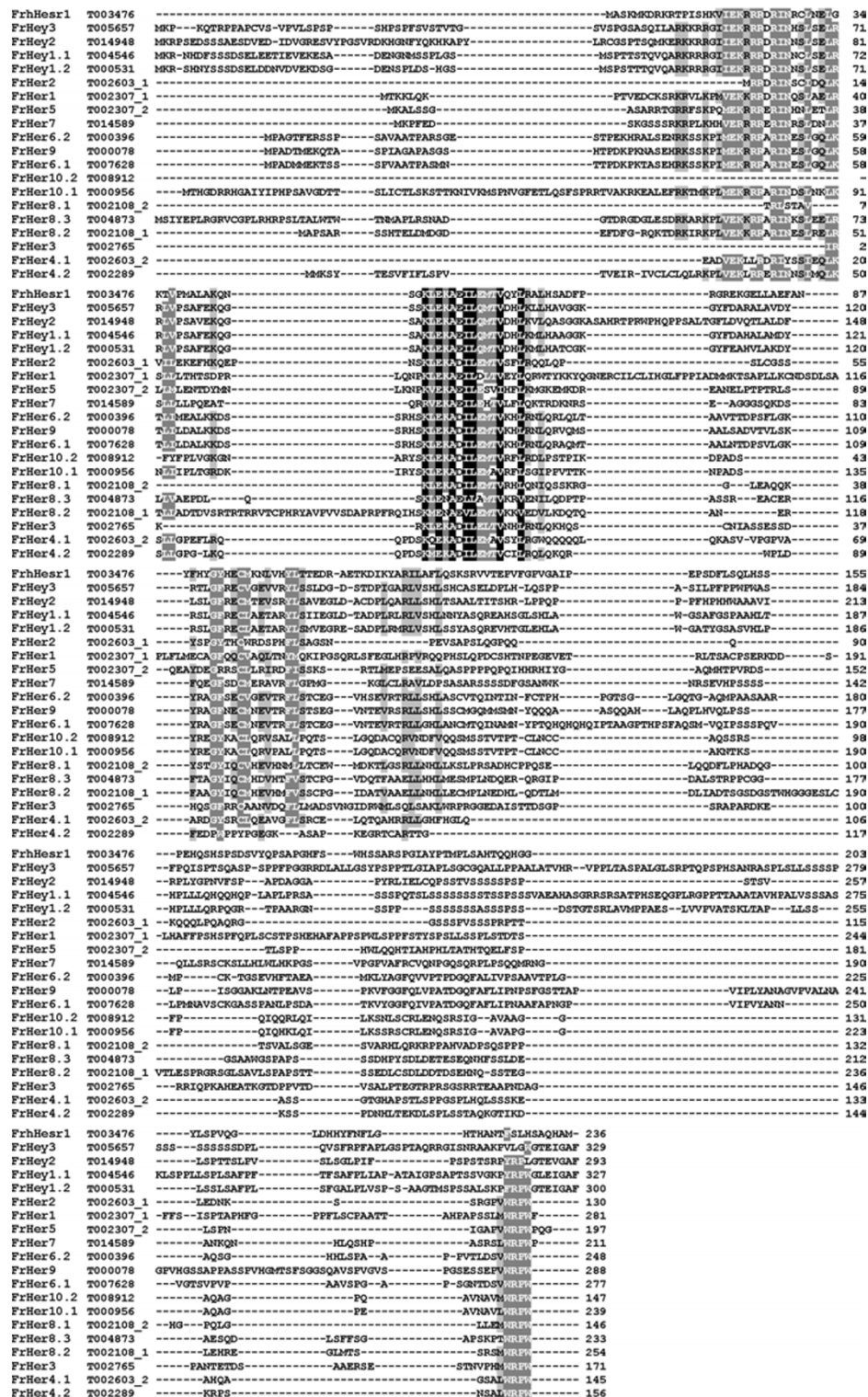
Figure 1
Sequence alignment of H/E(spl)/Hey-related proteins from *Fugu rubripes* Conservation levels: 100% identical residues are indicated in black, 80% or more conserved residues are marked in dark grey, 60% or more conserved residues are marked in lighter grey and less than 60% conserved residues are indicated in very light grey. Naming of the *Fugu* proteins according to the zebrafish homologues, numbering corresponds to the contig in which they were identified (table 1).
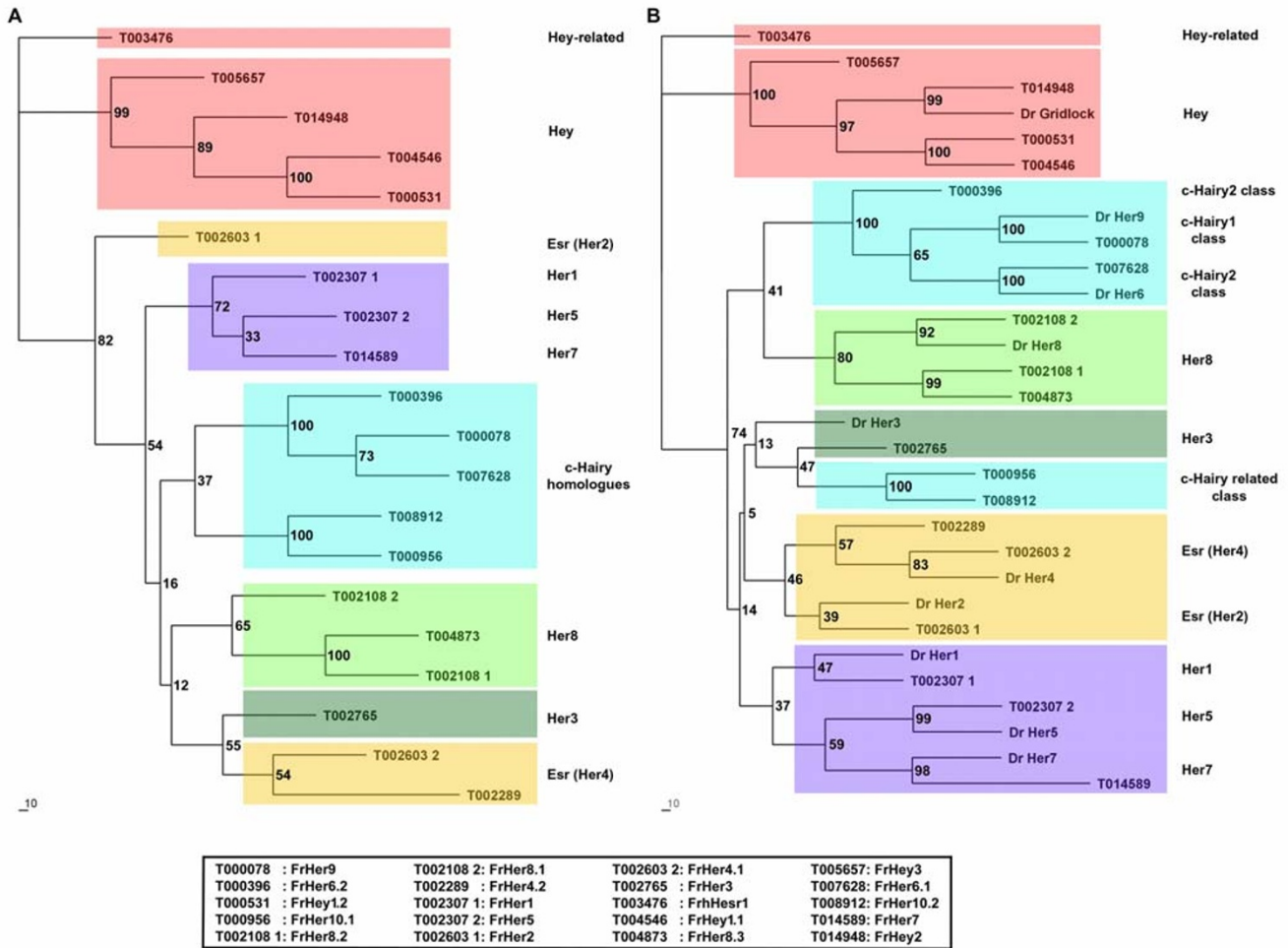
Figure 2
Relationship of H/E(spl)/Hey-related proteins (A) Similarity tree of the different H/E(spl)/Hey-related proteins in *Fugu*. (B) Similarity tree of the different proteins from *Fugu* and zebrafish. Identical colours in (A) and (B) indicate similar proteins. Blue green: c-Hairy type, dark green: Her3 similar, green: Her8 similar, light red: Hey similar, orange: Enhancer of split type, violet: Her1/5/7 similar. The conceptually translated proteins of *Fugu* were named according to the zebrafish homologues. c-Hairy1 class, proteins belonging to the c-Hairy1 similar proteins; c-Hairy2 class, proteins belonging to the c-Hairy2 similar proteins; c-Hairy-related class, proteins similar to c-Hairy1 and 2 building up a novel class of c-Hairy like proteins; Esr, Enhancer of split-related; Her, Hairy and Enhancer of split-related; Hey, Hairy and Enhancer of split-related with YRPW motif; Hey-related, similar to the Hey proteins. Numbering of the *Fugu* proteins according to the contig in which they were identified (see table 1). Trees were computed making use of the Neighbor-Joining method with a bootstrap support of 100 replicates.

the other identified Hey-genes, which have 5 exons in human and *Fugu* (see table 1). Nevertheless, the human orthologue, defined as "bHLH protein similar to Hesr-1/Hey1" (accession number: XM_068025, table 1, additional file 4) shows equal aberrant features from the Hey-type genes. Therefore, we will refer to the *hey*-related gene in *Fugu* as Fr*hhesr1*-related.

12 identified pufferfish proteins belong to the Hairy class of bHLH transcription factors that can be further subdivided into 3 branches. 5 proteins belong to the *c-hairy* type class compared to two genes known so far from zebrafish

(fig. 2: bluegreen box). At a first glance, in the second branch at least two of three proteins seem to be the zebrafish orthologues Her1 and Her7, which are involved in somitogenesis (fig. 2: violet box). A detailed analysis of the c-Hairy type class as well as of the Her1/5/7 is given in the following sections. The third branch consists of three proteins that are related to the zebrafish Her8 (table 1, fig. 2: light green box).

Comparing both fish species, the number of *h/E(spl)/hey*-related genes in pufferfish is twice the number of these genes currently known from zebrafish. Evidence from

Table 1: *hairy*, *Enhancer of split* and *hey*-related genes in the genome of *Fugu rubripes*

similarity to D. r.	P value	contig in F. r.	exon positions	examined upstream region	similarity to other vertebrates	P value
Her-1 (AF292032; X97329)	1e-020 3e-020	T002307/1 Frher1	8482-8429 8159-8064 7975-7882 7830-7232	10482-8482	mHes7 (AB050104); hHes7 (AB049064)	2e-014; 8e-014
Her-5 (X95301); Her-1 (AF292032); Her-7 (AF240772)	6e-021 2e-009 2e-009	T002307/2 Frher5	10889-10936 11014-11109 11175-11624	8889-10889	x-esr5 (AF137072); mHes7 (NM_033041); hHes7 (AB049064)	1e-012; 1e-012; 1e-012
Her7 (AF292032); Her-2 (X97330)	4e-013 0.041	T014589 Frher7	2619-2575 2436-2350 2260-1760	4619-2619	no significant similarity found	
Her-6 (X97333)	3e-082	T000396 Frher6.2	71531-71421 71319-71224 71152-71065 70947-70499	73531-71531	x-hairy1 (U36194); hHes1 (NM_005524); mHes1 (NM_008235)	6e-078; 3e-076; 4e-075
Her-6 (X97333)	e-106	T007628 Frher6.1	7030-6923 6803-6708 6594-6507 6345-5867	9030-7030	x-hairy1 (U36194); mHes1 (NM_008235); hHes1 (NM_005524)	7e-098; 9e-098; 4e-097
Her-3 (X97331); Her9 (AF301264)	8e-005 1e-004	T008912 Frher10.2	8182-8296 8390-8715	n.d.	hHes2 (AL031848); mHes2 (NM_008236); x-hairy2b/a (AF022798; AF383159)	2e-012; 2e-011; 2e-006; 2e-005
Her-9 (AF301264)	e-114	T000078 Frher9	38437-38544 38711-38806 38888-38975 39095-39665	36437-38437	x-hairy2b/a (AF356000; AF383159); c-hairy1 (AF032966)	4e-096; 1e-090; 2e-089
Her-9 (AF301264)	3e-016	T000956 Frher10.1	3397-3603 3725-3820 3934-4021 4114-4439	1397-3397	hHes2 (AL031848); mHes2 (NM_008236); x-hairy2a/b (AF383159; AF383160)	2e-023; 6e-022; 6e-017; 1e-016
Her-2 (X97330)	2e-012	T002603/1 Frher2	5189-4800	7189-5189	x-esr1/7 (AF383157; AF146088); mHes5 (NM_010419)	6e-017; 7e-016; 2e-014
Her-4 (X97332)	6e-026	T002603/2 Frher4.1	10235-9801	n.d.	mHes5 (NM_010419); x-esr7/2/1 (AF146088; AF383158; AF383157)	9e-020; 1e-018; 3e-016; 9e-015
Her-4 (X97332)	7e-016	T002289 Frher4.2	13834-14300	11834-13834	x-esr1/7/2 (AF383157, AF146088, AF383158); mHes5 (NM_010419)	4e-12 – 2e-9; 2e-9
Her-3 (X97331)	1e-009	T002765 Frher3	29260-29341 29448-29878	n.d.	x-hairy1 (U36194); mHes1 (NM_008235)	2e-008; 3e-008
Her8b/a (AY007991; AY007990); Her9 (AF301264)	3e-021 6e-021 2e-016	T002108/1 Frher8.2	22094-22008 21903-21648 21508-21090	24094-22094	mHes6 (NM_019479); hHes6 (NM_018645; BC007939; AF260237; XM_059427)	4e-041; 6e-041; 8e-041; 5e-040; 3e-024
Her8b/a (AY007991; AY007990); Her9 (AF301264)	2e-024 5e-024 2e-010	T002108/2 Frher8.1	25843-25748 25499-25156	n.d.	hHes6 (XM_043579; AF260237; NM_018645; BC007939; NM_019479; XM_039925)	3e-013; 3e-013; 3e-013; 3e-013; 1e-012; 1e-010
Her8b/a (AY007991; AY007990); Her9 (AF301264)	5e-024 5e-024 5e-015	T004873 Frher8.3	3983-4135 4229-4315 4393-4474 4571-4946	1983-3983	hHes6 (NM_018645; BC007939); mHes6 (NM_019479); hHes6 (AF260237)	2e-035; 2e-035; 1e-034; 2e-033
Gridlock (AF237948)	8e-014	T003476 Frhhesr1	11666-11640 11101-10997 10845-10749 10524-10052	13666-11666	hHesr-1/Hey1 (XM_068025)	8e-075
Gridlock (AF237948)	4e-049	T004546 Frheyl.1	4462-4377 4295-4220 4137-4054 3964-3883 3667-3039	5955-4462	Cf-Hesr-1/Hey1 (AJ388551); mHesr-1/Hey1 (AF172286); hHerp2/hHesr-1 (AF232239; AF151522)	3e-074; 4e-073; 7e-073; 7e-073
Gridlock (AF237948)	9e-044	T000531 Frheyl.2	54444-54529 54640-54712 54801-54884 54968-55049 55151-55725	52444-54444	hHERP2 (AF232239); Cf-Hesr-1/Hey1 (AJ388551); hHesr-1 (AF311883); mHesr-1 (AF172286);	5e-060; 8e-060; 2e-059; 9e-058
Gridlock (AF237948)	3e-039	T014948 Frhey2	1023-1149 1358-1419 1491-1574 1679-1814 1891-2360	1-1023	mHesr-2 (AF172287); hHesr-2 (XM_028903); Cf-Hesr-1 (AJ388551)	1e-036; 2e-036; 3e-031
Gridlock (AF237948)	2e-040	T005657 Frhey3	15382-15409 15518-15649 15736-15819 15893-15974 16130-16791	13382-15382	hHesr (XM_010727); hHesr-3 (AF311885) Cf-Hesr-1 (AJ388551)	2e-043; 2e-043; 4e-042

c chicken (*Gallus domesticus*), Cf dog (*Canis familiaris*), D.r. (*Danio rerio*), F.r. (*Fugu rubripes*), h human (*Homo sapiens*), m mouse (*Mus musculus*), x frog (*Xenopus laevis*); n.d. not definable

Hox cluster analysis suggests that the extent of duplication in *Fugu*, if any, is far less extensive than in zebrafish [28]. Therefore, it is likely that more *h/E(spl)/hey*-related genes in zebrafish exist. Deduced from the comparative tree it can be estimated that at least for *hey* type genes and in particular for genes of the *c-hairy* type class further zebrafish homologues have to be expected.

A screen with the 20 conceptually translated *Fugu* proteins against the human genomic sequence database yielded 10 *h/E(spl)/hey*-related genes (for accession numbers see Materials and Methods). These 10 genes can be further divided into 6 *h/E(spl)* genes, 3 *hey* genes (*hey1*, *hey2*, *heyL*, [26]) and 1 *hey*-related gene (*hhesr1*, accession number: XM_068025). In Fugu, 15 *h/E(spl)* genes, 4 *hey*-genes and 1 *hey*-related gene were identified. Deduced from the sequence alignment of the Hey type proteins (see additional file 4), FrHey1.1 and FrHey1.2 are similar to human Hey1, FrHey2 is similar to human Hey2 and FrHey3 is similar to human HeyL. One copy of the hey-related gene exists in both genomes (*hhesr1*, accession number: XM_068025, and Fr*hhesr1*). While the set of *Hey*- and *Hey*-related genes is quite comparable in both organisms, the number of *h/E(spl)* genes in *Fugu* (15) is 2.5 times higher than in the human genome (6). This indicates a duplication of those genes in *Fugu* or loss of members of the *h/E(spl)* genes in human. It is not known how these genes are expressed in *Fugu* but interestingly the majority of the zebrafish *her* genes is expressed in subsets of differentiating neuroblasts during nervous system development ([10,12,29]; *her*8 unpublished). Since fish possess a unique set of neurons and sensory organs, compared to higher vertebrates, it might be that selective duplication of the *hairy* genes was one source that led to neuronal/sensory cell diversification.

A novel class of c-Hairy proteins exist in the genomes of two pufferfish species

The 5 c-Hairy type proteins of the Hairy class of bHLH transcription factors in *Fugu rubripes* can be divided into three distinct c-Hairy subclasses, namely the c-Hairy1 class and c-Hairy2 class [4] and a third novel c-Hairy class, which we will refer to as the c-Hairy-related class since no apparent orthologue from chicken is known so far. Members of the three subclasses can be distinguished by length of the amino acid sequence between the orange domain and the WRPW motif (fig. 3). FrHer9 (in contig T000078) possesses the longest amino acid sequence stretch between the two domains and therefore belongs to the c-Hairy1 class, like zebrafish Her9 [12].

In FrHer6.1 (in contig T007628) and FrHer6.2 (in contig T000396) a sequence stretch of around 30–40 amino acids just N-terminal of the WRPW-motif is missing compared to c-Hairy1 class members (fig. 3). This feature is

indicative for c-Hairy2 class proteins [4]. Thus, FrHer6.1 and FrHer6.2 belong to the c-Hairy2 class and can be considered as Her6 homologues.

Deduced from the high conservation in the bHLH domain to the other c-Hairy proteins, FrHer10.1 (in contig T000956) and FrHer10.2 (in contig T008912) constitute a third to our knowledge novel c-Hairy class. In both proteins a sequence stretch of around 40 amino acids is missing, like in the c-Hairy2 class. In addition, a second stretch just C-terminal to the orange domain is absent compared to the already known two classes (fig. 3). Proteins of this c-Hairy type are so far not known in other vertebrates. However, these two *Fugu* proteins show similarity to the Hes2 proteins from mouse and human (fig. 3). But in the Hes2 proteins the number of amino acids between the orange domain and the WRPW motif is even more reduced in comparison to the c-Hairy-related proteins. To elucidate whether the class of c-Hairy-related proteins is unique in *Fugu*, the shotgun sequences of *Tetraodon nigroviridis* were screened with the 5 *Fugu c-hairy* type nucleotide sequences. Like in *Fugu*, the two c-Hairy-related class members (TnHer10.1 and TnHer10.2, TnHer10_1 and TnHer10_2 in fig. 3) and also the second member of the c-Hairy2 class were found (TnHer6.2, TnHer6_2 in fig. 3) supporting that respective proteins have to be expected in zebrafish.

Is the zebrafish her1/7 gene complex conserved in pufferfish?

One branch of the tree consists of the *Fugu* homologues of zebrafish *her1*, *her5* and *her7* (see fig. 2B violet box and table 1). The *her1* and *her7* genes in zebrafish have been shown to be cyclically expressed and are thought to play an important role in somitogenesis. Besides, they are arranged in a head to head orientation (accession number: AF292032). In contrast, Fr*her1* (in contig T002307) is orientated in the same manner to Fr*her5*. Fr*her7* was identified on another contig (T014589). But sequence similarity does not strictly mean functional similarity. It has been shown that the specificity of the biological action *in vivo* of proteins of the H/E(spl) family is mediated by the orange domain [30]. Sequence comparison of the orange domain of FrHer5 with zebrafish Her7 and Her5 shows that the domain from the *Fugu* protein is equally similar to that of the zebrafish proteins (fig. 4). Besides, FrHer5 possesses a prolin residue C-terminal of the WRPW motif, which is indicative for Her7/Hes7 proteins of zebrafish, mouse and human. Moreover, in the FrHer7 basic domain the prolin residue, which is necessary for the repressive function by binding to N-boxes [31], is missing. Instead, a histidin residue is present at this position (fig. 4). Additionally, the loop region of FrHer7 is shorter in comparison to zebrafish Her7. Length differences in the loop region are known to be responsible for functional specifi-
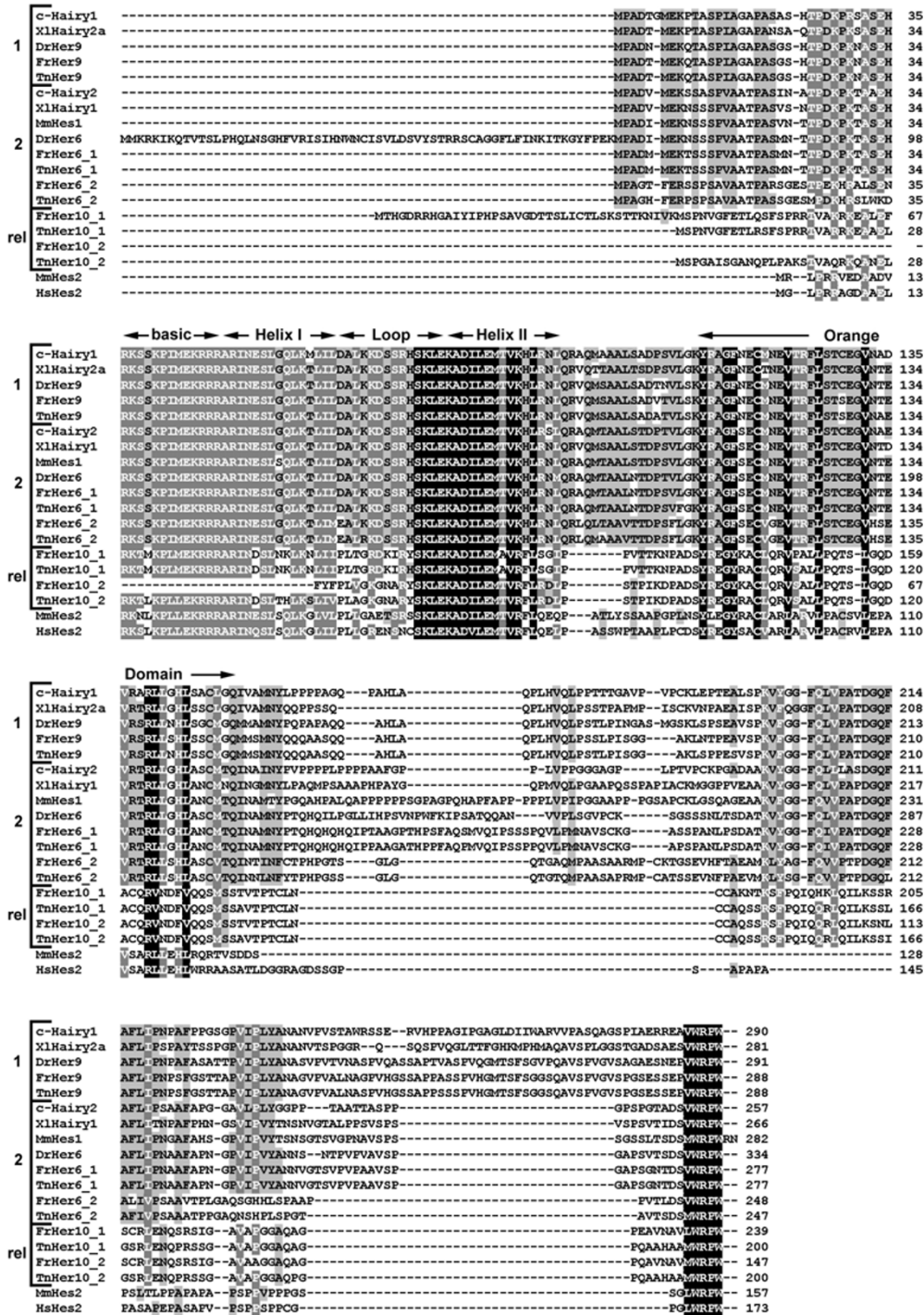
Figure 3
Sequence alignment of the different c-Hairy type proteins The Her6, Her9 and Her10 proteins from pufferfish *Fugu rubripes* and *Tetraodon nigroviridis* were aligned with *Danio rerio* and other corresponding vertebrate proteins. Conservation levels: 100% identical residues are indicated in black, 80% or more conserved residues are marked in dark grey, 60% or more conserved residues are marked in lighter grey and less than 60% conserved residues are indicated in very light grey. 1, proteins belonging to the c-Hairy1 class; 2, proteins belonging to the c-Hairy2 class; rel, proteins belonging to the novel c-Hairy-related class. c, *Gallus domesticus*; Dr, *Danio rerio*; Fr, *Fugu rubripes*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Tr, *Tetraodon nigroviridis*; Xl, *Xenopus laevis*.
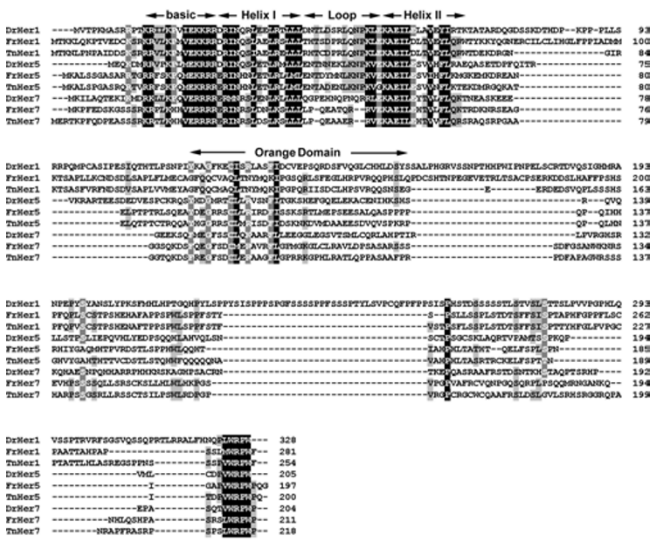
Figure 4
Sequence alignment of the Her1/5/7 proteins from *Fugu rubripes, Tetraodon nigroviridis* and *Danio rerio* Conservation levels: 100% identical residues are indicated in black, 80% or more conserved residues are marked in dark grey, 60% or more conserved residues are marked in lighter grey and less than 60% conserved residues are indicated in very light grey. Dr, *Danio rerio*; Fr, *Fugu rubripes*; Tr, *Tetraodon nigroviridis*.

city [32]. All observed differences indicate that FrHer7 is not the functional equivalent of zebrafish Her7. Furthermore, the Her1, Her5 and Her7 proteins in *Tetraodon* show identical characteristics as the *Fugu* proteins (fig. 4). Conclusively, evidence from sequence analysis implies that the *Fugu her1/5* gene complex is equivalent to the zebrafish *her1/7* gene complex.

Promoters of c-hairy1 and c-hairy2 type genes can be classified into two distinct groups

In the upstream sequences of the *c-hairy1* and *c-hairy2* class genes Fr*her6.1* (T007628) and Fr*her9* (T000078) a sequence stretch of approximately 100 nucleotides matches with the proximal part in the promoters of *x-hairy2a* and *hhes1*. The corresponding promoter parts of the orthologous genes in *Tetraodon* and zebrafish show identical composition, organisation and localization of the regulatory elements with respect to the 5' end of the corresponding gene (fig. 5). The fish *her6* and *her9* upstream sequences consist of a CCAAT-box, a TATA box and a SPS motif. It has been shown that this site is a crucial element in the regulation of the *x-hairy2a* gene, which is responsible for the (striped) expression in the PSM in *Xenopus*[33]. The SPS motif is a bipartite binding site for the Suppressor of hairless protein. The binding sites are separated by 30 or 29 nucleotides in the promoters of *E(spl)* genes of *Dro-*

sophila melanogaster and higher vertebrates, respectively [33,34]. One of the binding sites occurs in a reverse orientation to the other. Furthermore, a hexamer motif, which lies between or within the motifs, has a functional aspect.

Although the promoters consist of identical regulatory elements, on the level of nucleotide sequence they can be divided into two groups that correspond to the c-Hairy1 and c-Hairy2 protein classes (fig. 5). In particular, this classification is even reflected in the intervening regions of the two Su(h) sites of the identified SPS motifs. Thus, the promoters of the *c-hairy* class genes show more similarities within their classes in different species than among themselves in the same species.

Identification of potential regulatory motifs in zebrafish her1

The zebrafish *her1/7* gene complex was compared with the *Fugu her1/5* gene complex. The intergenic region of zebrafish is 11.4 kb large (accession number: AF292032), the region of *Fugu* consists of 2.4 kb. Alignment of the intergenic regions of the two clusters did not show large sequence stretches of significant similarity. Since conservation of promoter sequences most of the time show functional regions the reverse must not be true [35]. Therefore, we comparatively analysed the zebrafish *her1/7* promoter with the upstream sequences of the *Fugu her1/5* gene complex and *Tetraodon her1* by MatInspector [36]. First, we focused on potential Su(h) binding sites in the pufferfish promoters since *Fugu* and *Tetraodon* are closely related. Re-examination focussing on Su(h) sites that are able to constitute a SPS motif yielded one site in the intergenic region of the *Fugu her1/5* gene complex and in the upstream sequence of *Tetraodon her1* (table 2). The SPS sequences in the *her1* promoters of both pufferfish deviate from the other sites identified so far. In both Su(h) sites of *Tetraodon her1* one nucleotide exchange is present at the same position. Both SPS sequences in *Fugu her1/5* and *Tetraodon her1* show a distance of 32 nucleotides between the two Su(h) binding motifs (table 2). Deduced from this data we suggest that: 1. The distance between the two Su(h) sites has to be expected in the range of 29 – 32 nucleotides for fish. 2. Single exchanges to the GTGGGAA consensus might exist. Taking these parameters into account examination of the zebrafish *her1/7* promoter yielded one SPS candidate sequence (table 2). The two potential Su(h) sites have a distance of 31 nucleotides and a hexamer motif. Both sites deviate from the GTGGGAA consensus. Nevertheless, for the Su(h) protein of *Drosophila* it has been shown, that it binds to oligonucleotide sequences with the consensus RTGRGAR [34]. At least one binding site matches to this consensus. The second Su(h) site deviates in one position from the identified motifs in pufferfish. However, it has to be experimentally verified, whether the identified SPS sequences in this study are targets for tran-
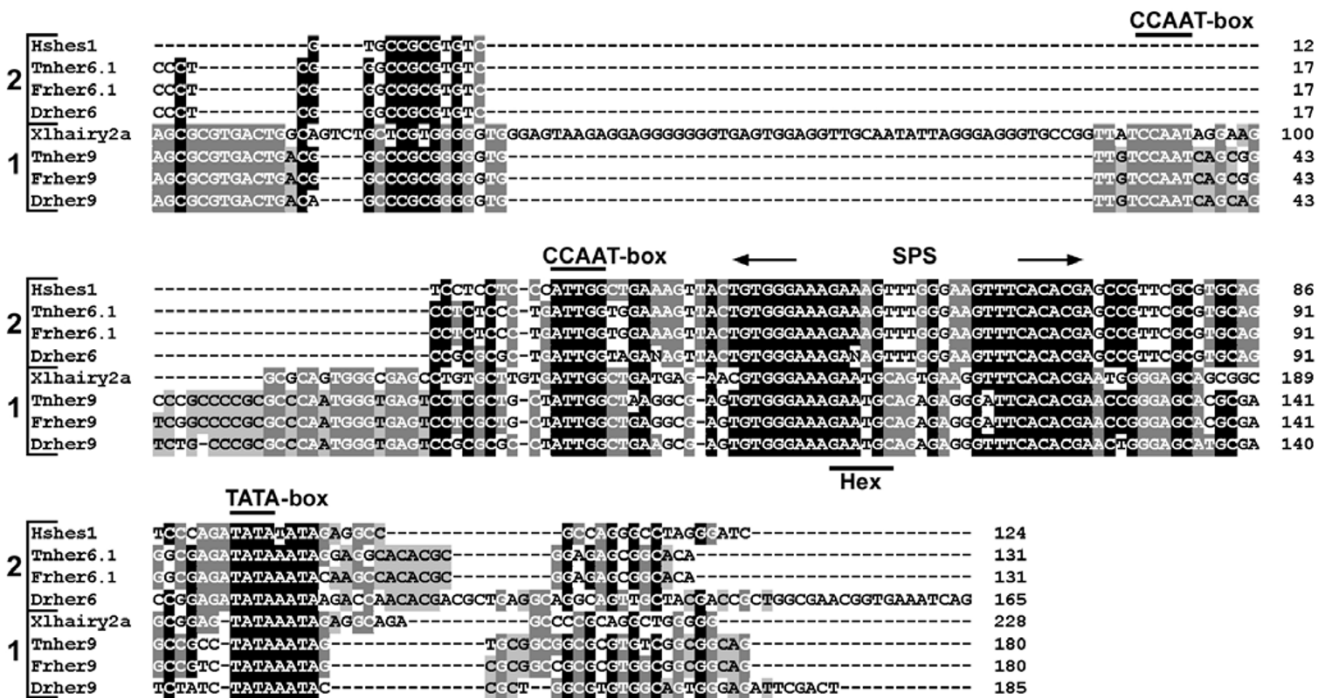
Figure 5
Alignment of the upstream sequences of *c-hairy* class genes The upstream sequences of the *c-hairy1/2* class genes *her*6 and *her*9 from *Danio*, *Fugu* and *Tetraodon* were aligned with the promoters of *x-hairy2a* and human *hes1*. Conservation levels: 75% or more conserved nucleotides are marked in black, 50% or more conserved nucleotides are marked in dark grey and nucleotides conserved to more than 25% are indicated in light grey. Boundaries of SPS, the potential Suppressor of hairless paired site, are indicated by the arrows. Potential CCAAT-boxes, hexamer sequences and TATA-boxes are indicated by a bold bar. 1, upstream sequences belonging to the *c-hairy1* class; 2, upstream sequences belonging to the *c-hairy2* class; Hex, hexamer sequence; SPS, Suppressor of hairless paired binding site; Dr, *Danio rerio*; Fr, *Fugu rubripes*; Hs, *Homo sapiens*; Tr, *Tetraodon nigroviridis*; Xl, *Xenopus laevis*.

scriptional regulation. Screening with a model according to these findings gave no further SPS motifs in the upstream sequences of the other *Fugu* genes than already identified (see Materials and Methods).

In *Xenopus*, it has been shown that the SPS regulatory motif and a specific 3'UTR element are sufficient to drive mesoderm expression [33]. Since the zebrafish *her1* gene is exclusively expressed in the PSM, we assume that SPS sequences are specific mesoderm regulator elements in fish and presumably in vertebrates.

Unusual high degree of conservation in the promoter of the hesr1-related gene of zebrafish, human, mouse and Fugu

To detect conserved regions, the upstream sequences of the *Fugu* genes, for which we could identify the first exon, were compared with the "genbank" database at NCBI by using the BLASTN algorithm under low stringency conditions [23]. Only in the case of the upstream sequence of

Fr*hhesr1*-related (in contig T003476) and the corresponding human, mouse and zebrafish promoters (fig. 6; see Materials and Methods for retrieval of the sequences from the different organisms) an extraordinary high degree of similarity was observed. The sequences are not conserved over the entire length. Highly conserved blocks are disrupted by diverged sequence stretches between the different species. The conserved blocks do not seem to contain any regulatory motif characterized so far. It is unlikely that the conserved regions do code for a protein. By using GENSCAN [http://genes.mit.edu/GENSCAN.html] one potential exon 500 bp upstream of the Start-Methionine in *Fugu* was identified, which is not conserved in mouse and human. However, a BLASTP search for the deduced 31 amino acid long peptide gave no significant similarities. A sorted six-frame translation with the *Fugu* sequence with a minimum ORF size of 50 amino acids and any Start Codon gave 18 ORFs that showed no remarkable similarities in a BLASTP search. Furthermore, a BLASTX search over the entire sequence gave no positive results. Howev-

Table 2: Alignment of the identified SPS sequences.

The SPS sequences from the two pufferfish species *Fugu* and *Tetraodon* and zebrafish *Danio rerio* were aligned with the corresponding sites of *Drosophila E(spl)* genes and with the sites identified in *h/E(spl)* genes of higher vertebrates. *Drosophila* SPS sequences from [34], higher vertebrate SPS sequences according to [33]. For retrieval of fish SPS sequences see Materials and Methods as well as additional file 6. The consensus sequence (RTGRGAR) for binding of Su(h) proteins as experimentally defined by [34] is shown two lines below the SPS sequence of *Danio*. The hexamer motifs are boldly underlined. Single base substitutions to the higher vertebrate Su(h) binding motif found in fish sequences are also underlined. Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Fr, *Fugu rubripes*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Tn, *Tetraodon nigroviridis*; Xl, *Xenopus laevis*.

er, it cannot be excluded that so far uncharacterized miR-NAs (reviewed in [37]) are encoded within the conserved regions. Nevertheless, it would be worthwhile to investigate if this sequence truly represents a promoter and what function the respective protein carries out.

Conclusions

By analysing the genome database of *Fugu* [http://Fugu.hgmp.mrc.ac.uk/] 20 *h/E(spl)/hey* genes in *Fugu* were determined, compared to 10 currently known members of this family in zebrafish. Deduced from this data, at least for *hey* type genes and in particular for genes of the *c-hairy* type class further zebrafish homologues have to be expected. A novel *c-hairy* type class could be identified in *Fugu* and *Tetraodon* that is so far not known in other vertebrates. Since no apparent orthologue from chicken exists, we refer to this class as the *c-hairy*-related class. Indicative for this class is the difference in length of the amino acid sequence between the orange domain and the WRPW motif compared to the already known two classes. Although the upstream sequences of the *c-hairy*1 and *c-hairy*2 class genes show identical composition of regulatory elements

in zebrafish, *Fugu* and *Tetraodon*, the classification coming from protein sequence analysis is reflected in the promoter regions on the level of nucleotide sequence. Furthermore, the identified SPS motifs in the promoters of the analysed *c-hairy1* and *c-hairy2* type genes are highly conserved, shown by comparison with higher vertebrates. For the SPS motif in the *her1* upstream sequences of the analysed fish species longer intervening sequences between the Su(h) sites were observed. Deduced from mesoderm specific expression of the *her1* gene in zebrafish, we suppose that the "Suppressor of hairless paired binding site" is a specific mesoderm regulator in fish and presumably in vertebrates.

Materials and Methods

Gene identification and sequence retrieval

By using the TBLASTN program [23] the *Fugu* database (first release from 25.10.01) at the UK HGMP Resource Centre [http://Fugu.hgmp.mrc.ac.uk/][19] was screened with the 9 different Her proteins of *Danio rerio*. The accession numbers of the used zebrafish bHLH genes are: *her1*: X97329, *her2*: X97330, *her3*: X97331, *her4*: X97332, *her5*:
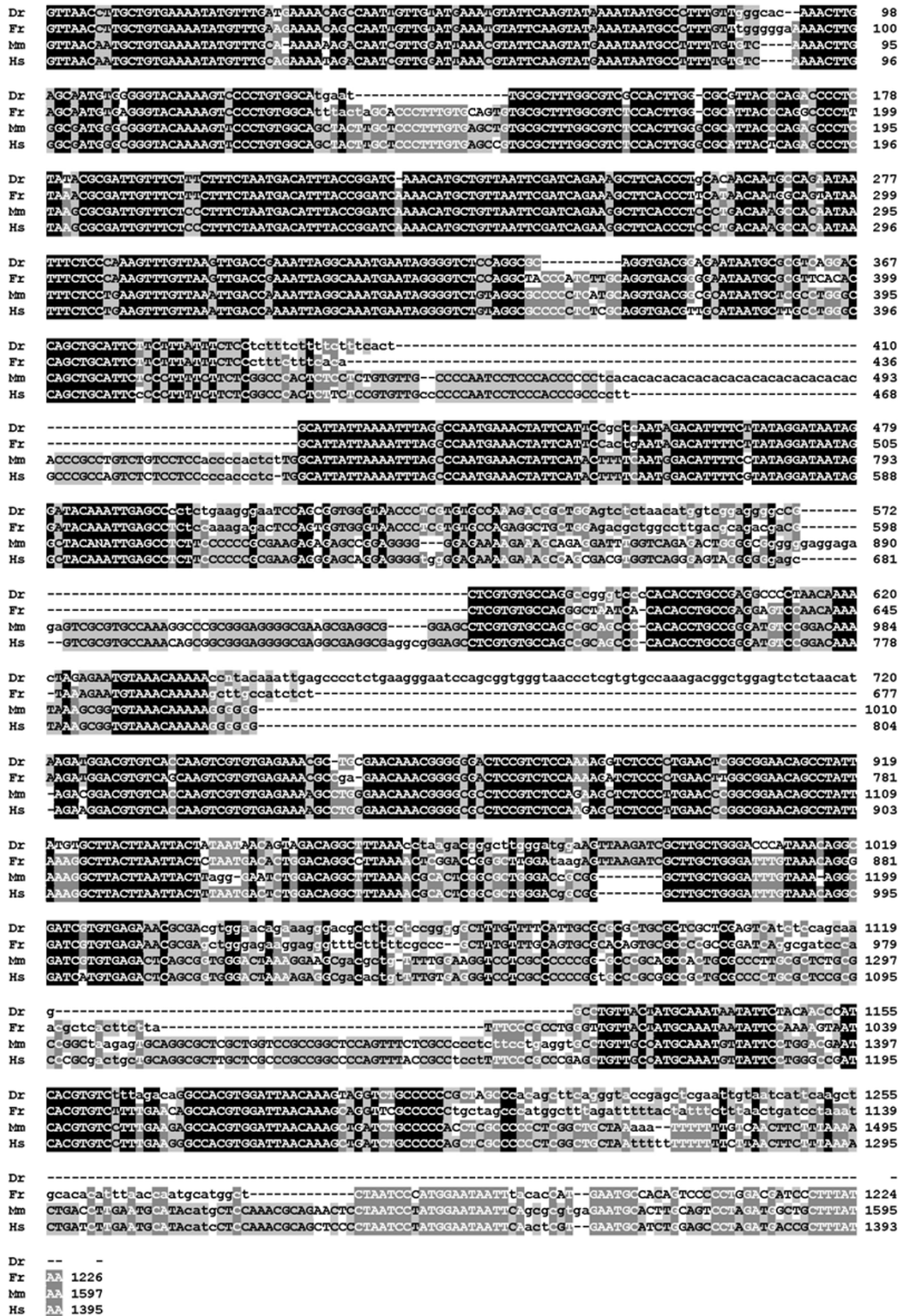
Figure 6
Alignment of the upstream sequences of different vertebrate *hesr1*-related genes For accession numbers of the compared genes see Materials and Methods. Conservation levels: 100% identical residues are indicated in black, 80% or more conserved residues are marked in dark grey, 60% or more conserved residues are marked in lighter grey and less than 60% conserved residues are indicated in very light grey. Dr, *Danio rerio*; Fr, *Fugu rubripes*; Hs, *Homo sapiens*; Mm, *Mus musculus*.

X95301, *her6*: X97333, *her7*: AF240772, *her8*: AY007990, *her9*: AF301264. The TBLASTN results were screened manually for the occurrence of sequence motifs known to be conserved in H/E(spl)-related proteins. These motifs are the basic domain with the consensus sequence KPx(M/V/I)E(K/R)(R/K)R, the highly conserved second loop with the consensus sequence (L/V)EKA(D/E)(I/V)LE and the WRPW motif located at the C-terminal end of the proteins. For further analysis, at least two of these motifs had to occur in a reasonable distance within a contig. Because the N- and/or C-termini often were missing in the automatically annotated proteins [http://www.jgi.doe.gov/fugu/index.html], a training set for identification of the exonic regions was created, which contains the following known *hairy* genes: *Drosophila hairy*: X15904, human *hes1*: L19314, mouse *hes1*: D16464, mouse *hes7*: AB050104, zebrafish *her1* and *her7*: AF292032 (see additional file 5). The regions of coding sequences, which we have conceptually translated into the respective proteins of *Fugu rubripes*, are depicted in table 1. Sequence LGS99222 was used for identification of the FrHer6.1 C-terminus (PAAVSPGAPSGNTDSVWRPW). Sequence LGS286123 was used for identification of the FrHey1 C-terminus (WGLEIGAF). The BLASTX algorithm [38] was used to identify the closest relatives to the *Fugu* Her proteins in zebrafish and higher vertebrates (table 1).

For identification of the *c-hairy* homologues as well as *her1*, *her5* and *her7* in *Tetraodon nigroviridis* the corresponding *Fugu* sequences were used to screen the shotgun sequence database at genoscope by BLASTN [http://www.genoscope.cns.fr/externe/tetraodon/Ressource.html] and the trace archive [http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?] at NCBI using megaBLAST. The Sequences, which are described here, can be retrieved at the Ensembl Trace Server [http://trace.ensembl.org] or at the NCBI Trace archive [http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?]. Sequences with the accession numbers 89970374, 89981571, 90016949, 90052606, 90066642, 90222617, 90222618, 90275465, 90275466, 90351844, 90421676, 90430225, 95723915, 95738908, 95996399, 97329806, 97457947, 97625306, 99082743, 99085687, 99124118, 99344199, 99379652, 99390932, 99545674 and 101019329 were assembled for 1 kb upstream of the first exon up to 5 kb downstream of the third exon of *Tetraodon her1*. The fourth exon of *Tetraodon her1* was identified in sequence COAG1029AA11SP1 (reverse complement) at position 97–614. For *Tetraodon her5* the sequences with the following accession numbers were assembled: C0AH037AD02SP1, C0AG530DE07SP1, C0AG873DH06LP1, C0AG1038AA11SP1, C0AG1120DA10SP1 and 90098825, 90098826, 90176607, 90411020, 90411021, 95688576, 95703158, 96052770, 97373175, 97658773, 99123371, 99159494, 99248382, 99288485

101130421. For *Tetraodon her6.1* the sequences with the following accession numbers were assembled: 89972279, 90058082, 90064995, 90119043, 90135880, 90238776, 90289277, 90444073, 90476831 and C0AG202CG02SP1, C0AG336DF02SP1, C0AG465AB08LP1, C0AG467BE05LP1, C0AG600DE09LP1, C0AG635CC05LP1, C0AG1333AE06SP1. *Tetraodon her6.2* was assembled with sequences of the following accession numbers: 97704800, C0AG515DB08SP1 and C0AG1340BB02SP1. For *Tetraodon her7* the sequences with the accession numbers C0BG101CF02LP1, C0AG256AH01SP1 C0AG325AF02LP1, C0AG542AG07SP1, C0AG575CF01SP1, C0AG735BC08SP1, C0AG793CF06SP1, C0AG799AD01LP1, C0AG900AF01SP1, C0AG1022CB06LP1, C0AG1034AH05SP1, C0AG1243AA03SP1, C0AG1342BE07LP1, C0AG1541DH04LP1, C0AH109BE02LP1 and C0BG045CF02SP1 were used. For *Tetraodon her9* the sequences with the accession numbers C0AA019AG08C1, C0AG201DA11SP1, C0AG302BF10LP1, C0AG622DB06LP1, C0AG997DB09LP1, C0AG1070BF12SP1, C0AG1174AH07SP1, C0AG1255CB09SP1 and C0AG1350CB05LP1 were utilised. For *Tetraodon her10.1* the sequences with the accession numbers C0AG132BG03LP1, C0AG214BF08LP1, C0AG471CC07SP1, C0AG852AF01LP1, C0AG1268AC03SP1, C0BG067CF10LP1, 97442563 and 100725254 were assembled. For *Tetraodon her10.2* the following sequences were assembled and the reverse complement of it was analysed: 101123589, C0AG438BH03LP1, C0AG859BF07LP1, C0AH077AD07SP1, C0BG092BE02LP1 (see additional file 6).

The human genome database was screened with the 20 *Fugu* proteins by TBLASTN [http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs]. The accession numbers (contig id, protein id) in which the human H/E(spl)/Hey-related proteins were identified, are: Hes1/Hry (NT_005571, gi: 20536141, XP_039925), Hes2 (AL031848, gi: 4914512, CAB46198), Hes4 (NT_032954, gi: 20472908, NP_066993), Hes5 (NT_004350, gi: 20535853, not annotated yet, 1. exon 161111–161164 and 2. exon 161264–161803), Hes6 (NT_005120, gi: 20535565, NP_061115), Hes7 (NT_010823, gi: 20560291, NP_115969). Hey1 (NT_008209, gi: 20544348, XP_113553), Hey2 (NT_030710, gi: 20550030, NP_036391), HeyL (NT_004511, gi: 20537057, NP_055386), protein similar to Hesr-1 (NT_006169, gi: 20535686, XP_068025).

Promoter identification and sequence retrieval
For examination of the corresponding promoter regions 2 kb upstream of the Start-Methionine were used in a BLASTN search [23] to detect conserved regulatory elements. Search for potential transcription factor binding sites was done using the MatInspector program [36]. Due to bad sequence reads we were not able to identify the first exon in 4 genes (indicated with n.d. in table 1), which were therefore excluded from upstream sequence analysis.

For identification of the zebrafish *her6* and *her9* upstream sequences the corresponding cDNA sequences were used for a blastn search against the trace archive at NCBI. The sequences with the accession numbers 15682371 and 25608918 were assembled for the *her6* promoter. The first 53 nucleotides of the published *her6* cDNA sequence were found on a different sequence stretch. The sequences with the accession numbers 25425900, 98681634, 15611875, 98577137, 100352443 and 102690745 were assembled for around 1 kb upstream of *her9* (see additional file 6).

For comparison of the promoter region of the Fr*hhesr1*-related gene in contig T003476 (table 1) the corresponding human sequence (accession number: 16931048, nt 18662–20664) was used for a BLASTN search against the trace archive to obtain corresponding sequences of mouse and zebrafish. The mouse upstream region was assembled using trace archive files with the following accession numbers 34431248 nt 1–741 of 741, 20297240 nt 135–751 of 751 (reverse complement), 7193321 nt 1–628 of 705 (reverse complement) and 43361903 nt 272–942 of 942 (reverse complement). The zebrafish promoter was compiled by utilizing the following files (accession numbers): 83152916 nt 1–602 of 612 (reverse complement), 30330598 nt 191–612 of 612 and 42667125 nt 1–612 of 612. All sequences are immediate to the first ATG of the respective h*hesr1*-related coding sequence, except for zebrafish. Here, no sequence information about the 3'end of this promoter as well as the beginning of the coding sequence was available. Potential exonic regions in the upstream sequence of the Fr*hhesr1*-related gene were investigated by GENSCAN [http://genes.mit.edu/GENSCAN.html] and potential ORFs were identified by BioEdit [39].

The SPS sequence in the upstream sequence of Fr*her9* (T000078) was identified at position 38130–38161. The SPS sequence in the promoter region of Fr*her6* (T007628) was localized at position 7320–7289 (reverse complement). The Fr*her1/5* gene complex was found in contig T002307 and a corresponding SPS sequence was identified at position 8677–8643 (reverse complement). Deduced from the created dataset of SPS sites, we used FASTM at Genomatix [http://www.genomatix.de/cgi-bin/fastm2/fastm.pl] to define a model for this motif for fish. The model consists of two binding site elements separated by 20 to 45 nucleotides (distance from start of element 1 to start of element 2). Both elements had to occur on one strand, the elements are defined as: GTGRRAR and WTYMCAC. With this dataset, we re-examined all available upstream sequences of the *Fugu* genes for the occurrence of SPS sites.

Sequence comparisons and phylogeny
Alignments of amino acid sequences were done making use of the CLUSTALW algorithm [40] in the program BioEdit or by using the Pileup program of the GCG software package [41,42]. Phylogenetic trees based on these alignments were computed by the Neighbor-Joining method [43] with a bootstrap support of 100 replicates. For the tree calculations the program PHYLIP [44] was used. Trees were displayed using Treeview [45]. The program GeneDoc was used for displaying the alignments [46]. For the alignment of the *c-hairy* homologues the accession numbers of the compared genes are: *c-hairy1*: AF032966, *c-hairy2*: [4], mouse *hes1*: NM008235, *x-hairy1*: U36194, *x-hairy2A*: AF383159. For comparsion, the mouse *hes2* (NM_008236) and human *hes2* (CAB46198) were used.

For alignment of the upstream sequences the DIALIGN program [47] from Genomatix [http://www.genomatix.de/cgi-bin/dialign/dialign.pl] was used.

Authors' contributions
The authors contributed equally to this work.

Additional material

Additional file 1
Additional file 1 – contigs of Fugu rubripes used for the analysis (Consortium data – Scaffolds 25.10.01). The sequences are provided in multiple FASTA format. The sequences are ordered according to the contig number. The header includes also the identified gene(s) within a contig.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-21-S1.fas]

Additional file 2
Additional file 2 – assembled cDNA sequences of Fugu rubripes and Tetraodon nigroviridis used in this analysis. The sequences are provided in multiple FASTA format. The sequences are ordered alphabetically.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-21-S2.fas]

Additional file 3
Additional file 3 – protein sequences of the H/E(spl)/Hey class in Fugu rubripes and Tetraodon nigroviridis used in this analysis. The sequences are provided in multiple FASTA format. The sequences are ordered alphabetically. File name: pufferfishherproteins.fas
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-21-S3.fas]

References

1. Meinhardt H: **Models of segmentation.** *In: Somites in developing embryos (Edited by: Bellairs R, Ede DA, Lash JW) Plenum New York* 1986, 179-189
2. Cooke J, Zeeman EC: **A clock and wavefront model for control of the number of repeated structures during animal morphogenesis.** *J Theor Biol* 1976, **58:**455-476
3. Palmeirim I, Henrique D, Ish-Horowicz D, Pourquie O: **Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis.** *Cell* 1997, **91:**639-648
4. Jouve C, Palmeirim I, Henrique D, Beckers J, Gossler A, Ish-Horowicz D, Pourquie O: **Notch signalling is required for cyclic expression of the *hairy*-like gene HES1 in the presomitic mesoderm.** *Development* 2000, **127:**1421-1429
5. Leimeister C, Dale K, Fischer A, Klamt B, Hrabe de Angelis M, Radtke F, McGrew MJ, Pourquie O, Gessler M: **Oscillating expression of *c-Hey2* in the presomitic mesoderm suggests that the segmentation clock may use combinatorial signaling through multiple interacting bHLH factors.** *Dev Biol* 2000, **227:**91-103
6. Bessho Y, Miyoshi G, Sakata R, Kageyama R: **Hes7: a bHLH-type repressor gene regulated by *Notch* and expressed in the presomitic mesoderm.** *Genes Cells* 2001, **6:**175-185
7. Bessho Y, Sakata R, Komatsu S, Shiota K, Yamada S, Kageyama R: **Dynamic expression and essential functions of *Hes7* in somite segmentation.** *Genes Dev* 2001, **15:**2642-2647
8. Saga Y, Takeda H: **The making of the somite: molecular events in vertebrate segmentation.** *Nat Rev Genet* 2001, **2:**835-845
9. Maroto M, Pourquie O: **A molecular clock involved in somite segmentation.** *Curr Top Dev Biol* 2001, **51:**221-248
10. v Weizsäcker E: **Molekulargenetische Untersuchungen an sechs Zebrafisch-Genen mit Homologie zur *Enhancer of split* Gen-Familie von *Drosophila*.** *Universität zu Köln, Inaugural-Dissertation* 1994
11. Müller M, v. Weizsäcker E, Campos-Ortega JA: **Expression domains of a zebrafish homologue of the *Drosophila* pair-rule gene *hairy* correspond to primordia of alternating somites.** *Development* 1996, **122:**2071-2078
12. Leve C, Gajewski M, Rohr KB, Tautz D: **Homologues of *c-hairy1* (*her9*) and *lunatic fringe* in zebrafish are expressed in the developing central nervous system, but not in the presomitic mesoderm.** *Dev Genes Evol* 2001, **211:**493-500
13. Zhong TP, Rosenberg M, Mohideen MA, Weinstein B, Fishman MC: **Gridlock, an HLH gene required for assembly of the aorta in zebrafish.** *Science* 2000, **287:**1820-1824
14. Takke C, Campos-Ortega JA: ***her1*, a zebrafish pair-rule like gene, acts downstream of notch signalling to control somite development.** *Development* 1999, **126:**3005-3014
15. Holley SA, Geisler R, Nüsslein-Volhard C: **Control of *her1* expression during zebrafish somitogenesis by a *Delta*-dependent oscillator and an independent wave-front activity.** *Genes Dev* 2000, **14:**1678-1690
16. Stickney HL, Barresi MJF, Devoto SH: **Somite development in zebrafish.** *Dev Dyn* 2000, **219:**287-303
17. Pasini A, Henrique D, Wilkinson DG: **The zebrafish *Hairy/Enhancer-of-split*-related gene her6 is segmentally expressed during the early development of hindbrain and somites.** *Mech Dev* 2001, **100:**317-321
18. Sawada A, Fritz A, Jiang Y, Yamamoto A, Yamasu K, Kuroiwa A, Saga Y, Takeda H: **Zebrafish *Mesp* family genes, *mesp-a* and *mesp-b* are segmentally expressed in the presomitic mesoderm, and Mesp-b confers the anterior identity to the developing somites.** *Development* 2000, **127:**1691-1702
19. Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJK, Bouchireb N, Cottage A, Yeo GSH, Umrania Y, Williams G, Brenner S: **Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning.** *Genome Res* 1999, **9:**960-971
20. Powers DA: **Evolutionary genetics of fish.** *Adv Genet* 1991, **29:**119-228
21. Hinegardner R: **Evolution of cellular DNA content in Teleost fishes.** *The American Naturalist* 1968, **102:**517-523
22. Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S: **Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome.** *Nature* 1993, **366:**265-268
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410
24. Klämbt C, Knust E, Tietze K, Campos-Ortega JA: **Closely related transcripts encoded by the neurogenic gene complex *enhancer of split* of *Drosophila melanogaster*.** *Embo J* 1989, **8:**203-210
25. Leimeister C, Externbrink A, Klamt B, Gessler M: **Hey genes: a novel subfamily of *hairy*- and *Enhancer of split* related genes specifically expressed during mouse embryogenesis.** *Mech Dev* 1999, **85:**173-177
26. Steidl C, Leimeister C, Klamt B, Maier M, Nanda I, Dixon M, Clarke R, Schmid M, Gessler M: **Characterization of the human and mouse HEY1, HEY2, and HEYL genes: cloning, mapping, and mutation screening of a new bHLH gene family.** *Genomics* 2000, **66:**195-203
27. Nakagawa O, Nakagawa M, Richardson JA, Olson EN, Srivastava D: ***HRT1, HRT2, and HRT3:* a new subclass of bHLH transcription factors marking specific cardiac, somitic, and pharyngeal arch segments.** *Dev Biol* 1999, **216:**72-84

28. Kurosawa G, Yamada K, Ishiguro H, Hori H: ***Hox* gene complexity in *Medaka* fish may be similar to that in pufferfish rather than zebrafish.** *Biochem Biophys Res Commun* 1999, **260:**66-70

29. Takke C, Dornseifer P, v. Weizsäcker E, Campos-Ortega JA: ***her4,* a zebrafish homologue of the *Drosophila* neurogenic gene *E(spl),* is a target of NOTCH signalling.** *Development* 1999, **126:**1811-1821

30. Dawson SR, Turner DL, Weintraub H, Parkhurst SM: **Specificity for the hairy/enhancer of split basic helix-loop-helix (bHLH) proteins maps outside the bHLH domain and suggests two separable modes of transcriptional repression.** *Mol Cell Biol* 1995, **15:**6923-6931

31. Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S: **Two mammalian helix-loop-helix factors structurally related to *Drosophila hairy* and *Enhancer of split.*** *Genes Dev* 1992, **6:**2620-2634

32. Bae S-K, Yasumasa B, Hojo M, Kageyama R: **The bHLH gene *Hes6,* an inhibitor of *Hes1,* promotes neuronal differentiation.** *Development* 2000, **127:**2933-2943

33. Davis RL, Turner DL, Evans LM, Kirschner MW: **Molecular targets of vertebrate segmentation: two mechanisms control segmental expression of *Xenopus hairy2* during somite formation.** *Dev Cell* 2001, **1:**553-565

34. Nellesen DT, Lai EC, Posakony JW: **Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* complex genes to common transcriptional activators.** *Dev Biol* 1999, **213:**33-53

35. Takahashi H, Mitani Y, Satoh G, Satoh N: **Evolutionary alterations of the minimal promoter for notochord-specific *Brachyury* expression in ascidian embryos.** *Development* 1999, **126:**3725-3734

36. Quandt K, Frech K, Karas H, Wingender E, Werner T: **MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.** *Nucleic Acids Res* 1995, **23:**4878-4884

37. Pasquinelli AE: **MicroRNAs: deviants no longer.** *Trends Genet* 2002, **18:**171-173

38. Gish W, States DJ: **Identification of protein coding regions by database similarity search.** *Nat Genet* 1993, **3:**266-272

39. Hall T: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Se* 1999, **41:**95-98

40. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680

41. Devereux J, Haeberli P, Smithies O: **A comprehensive set of sequence analysis programs for the VAX.** *Nucleic Acids Res* 1984, **12:**387-395

42. Senger M, Flores T, Glatting K, Ernst P, Hotz-Wagenblatt A, Suhai S: **W2H: WWW interface to the GCG sequence analysis package.** *Bioinformatics* 1998, **14:**452-457

43. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4:**406-425

44. Felsenstein J: **PHYLIP (Phylogeny Inference Package).** *Seattle, University of Washington (distributed by the author)* 1991

45. Page RDM: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12:**357-358

46. Nicholas KB, Nicholas HB, Deerfield DW II: **GeneDoc: Analysis and viualization of genetic variation.** *EMBNEW NEWS* 1997, **4:**14

47. Morgenstern B, Frech K, Dress A, Werner T: **DIALIGN: finding local similarities by multiple sequence alignment.** *Bioinformatics* 1998, **14:**290-294