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) (hi/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 hi/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 hi/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 Enhancer 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 hhes1/hey1”). The C-terminal consensus sequence of the Hey proteins is KPYRPWGTE(VHey1/NHey2)GAF. Focussing on this consensus, FrHey1.1 (contig T004546), FrHey1.2 (contig T005531) 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 KPYRPWGTE(VHey1/NHey2)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).
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 Frhhesr1-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

---

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

| similarity to | P value | config in F. | exom positions | examined upstream region | similarity to other vertebrates | P value |
|---------------|---------|--------------|----------------|--------------------------|--------------------------------|---------|
| Her-1 (AF229032) | 3e-020 | T002307/1 | 8402-8492 | 10482-8492 | m.Hes7 (XM050104); h.Hes7 (AF490964) | 2e-014; 8e-014 |
| Her-1 (AF229032) | 0e-009 | Frh.5 | 8159-8064 | 7975-7852 | x-Hes7 (AF313727); m.Hes7 | 1e-012; 1e-012; 1e-012 |
| Her-5 (XM938501) | 6a-013 | T002307/2 | 10889-10889 | 8889-10889 | m.Hes7 (XM050104); h.Hes7 (AF490964) | 1e-012; 1e-012 |
| Her-7 (AF229032) | 4e-013 | T034259 | 6192-2575 | 4619-2619 | no significant similarity found |
| Her-6 (XM7388) | 3e-017 | T002306 | T0862-71531 | 7851-71531 | x-hairy (U31614); m.hairy | 6e-078; 6e-078; 6e-075 |
| Her-6 (XM7388) | 3e-009 | T002628 | T0104-70298 | 7030-70298 | x-hairy (U31614); m.hairy | 7e-099; 9e-099; 6e-009 |
| Her-5 (XM73831) | 5a-005 | T002912 | T0862-71531 | 7030-70298 | x-hairy (U31614); m.hairy | 7e-099; 9e-099; 6e-009 |
| Her-6 (XM73831) | 8e-005 | T002306 | T0104-70298 | 7030-70298 | x-hairy (U31614); m.hairy | 7e-099; 9e-099; 6e-009 |
| Her-9 (AF301364) | 5e-007 | T002306 | T0104-70298 | 7030-70298 | x-hairy (U31614); m.hairy | 7e-099; 9e-099; 6e-009 |
| Her-5 (XM73839) | 2e-012 | T002303/1 | T0104-70298 | 7030-70298 | x-Hes7 (AF383157); h.Hes2 | 1e-023; 6e-022; 6e-017; 1e-016 |
| Her-4 (XM73831) | 2e-026 | T002623/2 | T0104-70298 | 7030-70298 | x-Hes7 (AF383157); h.Hes2 | 1e-023; 6e-022; 6e-017; 1e-016 |
| Her-4 (XM2395) | 7a-016 | T002359 | T0104-70298 | 7030-70298 | x-Hes7 (AF383157); h.Hes2 | 1e-023; 6e-022; 6e-017; 1e-016 |
| Her-3 (XM73831) | 1e-009 | T002306 | T0104-70298 | 7030-70298 | x-Hes7 (AF383157); h.Hes2 | 1e-023; 6e-022; 6e-017; 1e-016 |

**c** chicken (Gallus domesticus), **Df** dog (Canis familiaris), **Dz** (Danio rerio), **F.x** (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 *Frhhesr1*). 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]; *her8* 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, *Frher1* (in contig T002307) is oriented in the same manner to *Frher5*. *Frher7* 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/Her5* 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.
Sequence alignment of the Her1/5/7 proteins from 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.

**Figure 4**

Sequence alignment of the Her1/5/7 proteins from *Fugu rubripes*, *Tetraodon nigroviridis* and *Danio rerio*.

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 Frher6.1 (T007628) and Frher9 (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 *Drosophila 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). Deducing 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-
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 *Ffhesr1*-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.

![Figure 5](image-url)
er, it cannot be excluded that so far uncharacterized miRNAs (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-hairy1 and c-hairy2 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/] 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: X97333, her6: X97334, her7: X97335, her8: X97336, her9: X97337.

### Table 2: Alignment of the identified SPS sequences.

|     | Dm E(spl)m3 | TGTGGGAACACACGACAAATCT---gtTTCCACAGGctcTgtaAAATG  |
|-----|-------------|---------------------------------------------------|
|     | Dm E(spl)m7 | TGTGGGAACCTGAGGGGAAAGT---gtTTCCACAGGctcTgtaAAATG  |
|     | Dm E(spl)m8 | TGTGGGAACCTGAGGGGAAAGT---gtTTCCACAGGctcTgtaAAATG  |
|     | Dm E(spl)m9 | TGTGGGAACCTGAGGGGAAAGT---gtTTCCACAGGctcTgtaAAATG  |
| X1  | hairy2a     | CGTGGGAACAGAAGAAG---gtgAA---gtTTCCACACG             |
| Hs  | hes4        | TGGTGGAACAGAAGAAGAAG---ggTTCACAAGAAG---gtTTCCACACG |
| Dr/Fr/Tn | her9     | TGGTGGAACAGAAGAAGAAG---ggTTCACAAGAAG---gtTTCCACACG |
| Mm/Hs | her1        | TGTGGGAACAGAAGAAG---gtTTCCACACG                     |
| Dr/Fr/Tn | her6     | TGTGGGAACAGAAGAAG---gtTTCCACACG                     |
| Fr  | her1/5      | TGGTGGAACACTGATCGAGAAGA---gtTTCCACACG               |
| Tn  | her1        | GGGTGGAACACTGATCGAGAAGA---gtTTCCACACG               |
| Dr  | her1/7      | GGGTGGAACACTGATCGAGAAGA---gtTTCCACACG               |
|     |             | RTGRGAR                                            |

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.
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/L)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 ru-bripes, are depicted in table 1. Sequence LGS99222 was used for identification of the FrHer1 C-terminus (PAAVSPGAPSNTDSVWR). Sequence LGS286123 was used for identification of the FrHey 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/Resource.html] and the trace archive [http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?F=HsBlast.html&&ORG=Hs] at NCBI using megBLAST. 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?] at NCBI using megaBLAST. The sequences with the accession numbers were assembled: 89970374, 89981571, 90016949, 90052606, 90066642, 90222617, 90222618, 90275465, 90275466, 90351844, 90421676, 90430225, 95723915, 95738908, 95996399, 97329806, 97457947, 97625306, 99082743, 99085687, 99124118, 99344199, 99379652, 99390392, 99545674 and 10019329 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: COAH037AD02SP1, COAG530DE07SP1, COAG873DH06LP1, COAG1038AA11SP1, C0AG1120DA10SP1 and 90098825, 90098826, 90176607, 90411020, 90411021, 95688576, 95703158, 96052770, 97373175, 97658773, 99123371, 99159494, 99248382, 99288485.
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 Frhesr1-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 at NCBI. 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 hhesr1-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 Frhesr1-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 Frher9 (T000078) was identified at position 38130–38161. The SPS sequence in the promoter region of Frher6 (T007628) was localized at position 7320–7289 (reverse complement). The Frher1/5 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 comparison, 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 (Consoritum 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] |
Additional file 4

Additional file 4 - Sequence alignment of the different Hey type proteins. For accession numbers of the compared genes see Materials and Methods and table 1. 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. Fr, Fugu rubripes; Hs, Homo sapiens.

Click here for file
[http://www.biomedcentral.com/content supplementary/1471-2164-3-21-S4.jpg]

Additional file 5

Additional file 5 - Comparison of the exon-intron boundaries of diverse hairy genes. Exon-intron boundaries of Drosophila hairy were compared with zebrafish her1 and her7 and with respective sequences of other vertebrate hes genes. Exonic sequence parts in capital letters, intronic sequence parts in small letters. Conserved amino acid residues are boldly typed. Dm, Drosophila melanogaster; Dr, Danio rerio; Hs, Homo sapiens; Mm, Mus musculus.

Click here for file
[http://www.biomedcentral.com/content supplementary/1471-2164-3-21-S5.jpg]

Additional file 6

Additional file 6 - assembled sequences from Danio rerio and Tetraodon nigroviridis. In order of appearance: Drherr6 upstream sequence, Drherr9 upstream sequence, Tnher1 genomic region, Tnher5 genomic region, Tnher6.1 genomic region, Tnher6.2 genomic region, Tnher7 genomic region, Tnher9 genomic region, Tnher10.1 genomic region, Tnher10.2 genomic region. Position of potential regulatory elements and of the exons is given in the header of each sequence.

Click here for file
[http://www.biomedcentral.com/content supplementary/1471-2164-3-21-S6.fas]

Acknowledgements

The authors wish to thank Diethard Tautz for giving helpful advice and critical comments to this work. Thanks also to Michael Krohler for critically reading the manuscript. The Fugu sequence data has been provided freely by the Fugu Genome Consortium [http://fugu.hgmp.mrc.ac.uk()]. Genoscope and the Whitehead Institute for Genome Research [http://www.genoscope.cns.fr/externe/Tetraodon/] made the sequence data of Tetraodon nigroviridis publicly available. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 572).

References

1. Meinhardt H. Models of segmentation. In: Somites in developing embryos (Edited by: Bellairs R, Edc 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, Pourquié 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, Pourquié O. Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. Development 2000, 127:421-429
5. Leimeister C, Dale K, Fischer A, Klamt B, Hrabe de Angelis M, Radde F, McGrew MJ, Pourquié O, Gossler 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, Shioita 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, Pourquié O: A molecular clock involved in somite segmentation. Curr Top Dev Biol 2001, 51:221-248
10. Weizsäcker E. Molekulargenetische Untersuchungen an sechs Zebrabfisch-Genen mit Homologie zur Enhancer of split Gen-Familie von Drosophila. Universität zu Köln, Inaugural-Dissertation 1994
11. Muller M, v. Weizsäcker E, Campos-Ortega JA: Expression domain 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-hairy (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, Yamasa K, Kuroiwa A, Saga Y, Takeda H. Zebrafsh 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 HS, Meek S, Smith S, Warner S, Edwards YJK, Bouchier A, Cortage A, Yeoh GSH, Urquaza 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 OA: Evolutionary genetics of fish. Adv Genet 1991, 20:119-228
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, Venkatase B, Aparicio S: Characterization of the pufferfish (Fugu) genome as a compact pact model vertebrate genome. Nature 1993, 364:263-268
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403-410
24. Klambt C, Knust E, Tietze K, Campos-Ortega JA: Closely related transcripts encoded by the neurogenic gene complex enhancer of split of Drosophila melanogaster. Embio J 1989, 8:203-210
25. Leimeister C, Externbrink A, Klamt B, Gossler 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, Gossler M: Characterization of the human and mouse HEY1, HEY2, and HETL genes: cloning, mapping, and mutation screening of a new bHLH family gene. 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, Dormeijer P v, Weitzsäcker 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, Kirschnner 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, 213:33-33
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. Nucleic 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, Glattling K, Ernst P, Hotz-Wagenblatt A, Suhai S: 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 visualization 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