



Identification and Characterization of the *YGHL1* Gene from Yellowtail fish (*Seriola quinqueradiata*) and the evolution of the *YGHL1/HIG1* family in Vertebrates

Abe Shunnosuke*, Ravi Nagganatha Suganthan, Sahi Vaidurya Pratap and Morita Eugene Hayato

Laboratory of Molecular Cell Physiology, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama 790-8566, JAPAN

Available online at: www.isca.in

Received 3rd June 2013, revised 15th June 2013, accepted 21st July 2013

Abstract

A 1529bp full length cDNA of the yellowtail growth hormone like-1 (*YGHL1*) from *Seriola quinqueradiata* was cloned, and its structure, genomic organization and expression were analyzed. The yellowtail *YGHL1* gene is composed of three coding exons and one 5' non-coding exon, with putative transcription factor binding sites present in upstream of the transcription site. The *YGHL1* in yellowtail fish was highly expressed in brain, gill, heart, and kidney, while no appreciable expression was observed in liver and skeletal muscle. This study also shows the phylogeny and evolution of the *YGHL1/HIG1* orthologs in mammals and other vertebrates, by comparing their sequences and syntenic context. According to the deduced peptide sequence alignment, the "YGHL1/HIG1 exon 3 domain" peptide seems to be well conserved in the *YGHL1/HIG1* gene family. Our data also suggest that an ancestral locus similar to *YGHL1* in *Ciona intestinalis* underwent duplications to create orthologous loci of the *YGHL1/HIG1* family in the vertebrates. The regions encoding the *YGHL1/HIG1* paralogs in human and mouse were close to the regions where some homeostatically important genes are clustered. These clusters are prominent in human 3p25-22 and 17q11-12 and their orthologous region in mouse 6D1-6E3 and 11D, showing a concerted evolution between them.

Keywords: *YGHL1/HIG1*, exon 3 domain, yellowtail, evolution, vertebrate lineage.

Introduction

The generation of multiple copies of diverse eukaryotic genes has been attributed to polyploidization in ancestral vertebrates. Following such duplication, a multitude of mechanisms have rearranged the relative positions of different genes and genomic regions, causing diversification of the metabolic pathways that ultimately affect the morphology and function of the organism¹⁻⁶. Taylor *et al.*⁷ reported twenty seven groups of orthologous duplicated fish genes in their study of genome duplication of ray finned fishes which included genes from human, mouse, chicken, frog and zebrafish; and phylogenetic analysis of these genes suggested that genomic duplications took place during the evolution of fish, 300 to 450 million years ago. Genomic studies involving a novel RING zinc finger protein, belonging to the *MKRN* family^{8,9}, have also provided an insight into the functional linkage and evolution of human chromosome region 3p21-25, in which apparently unrelated orthologous loci in fish, *YGHL2/MKRN2*, *RAF1*, and the *HIGDIA*, are present¹⁰. These works also illustrate that syntenic analysis of flanking genes is valuable for elucidation of the concerted evolution of gene families.

The yellowtail growth hormone like-1 (*YGHL1*) cDNA (GenBank: D85880) was obtained on screening a pituitary cDNA library of *Seriola quinqueradiata* (Yellowtail fish or Japanese Amberjack) using the antibody for Salmon growth hormone in 1996. Later, a cDNA encoding a 91aa protein

homologous to the yellowtail *YGHL1* was identified in liver tissue of a euryoxic fish, *Gillichthys mirabilis*, as Hypoxia-Inducible Gene 1, *HIG1* (GenBank: AF266206), which has shown to be induced after prolonged hypoxia in liver¹¹. Based on the report of Gracey *et al.*¹¹, many members of the *HIG1* gene family were identified. Based on the genomic organization and extensive conservation of nucleotide and protein sequences orthologous to the yellowtail *YGHL1*, we propose to call the *HIG1* gene family as *YGHL1/HIG1* family.

In the present study, the yellowtail *YGHL1* was cloned and characterized. Its genomic structure including the 5' upstream region and its expression were studied. We compared these sequences with their orthologous loci in vertebrates and invertebrates to understand the evolution of the extensive *YGHL1/HIG1* family. We also analyzed their flanking genes and showed that multiple vertebrate paralogs have evolved from a single ancestral gene through concerted evolution along with their flanking genes.

Material and Methods

Isolation and analysis of genomic DNA: Testis of yellowtail fish (0.5g) was homogenized in 5 volumes of extraction buffer [10mM Tris-HCl (pH.8), 25mM EDTA (pH.8), 100mM NaCl, 0.25% SDS, 100 µg/ml of Protease-K] and incubated at 55°C for 90min, the mixture then was extracted three times with an equal volume of TE saturated phenol, followed by three times

with an equal volume of phenol/chloroform/isoamyl alcohol, 25:24:1 (v/v/v). The aqueous phase was purified at least twice with an equal volume of chloroform/isoamyl alcohol, 24:1 (v/v). The extracted genomic DNA was precipitated from the aqueous phase with 2.5 volumes of ethanol at 20°C. The pellet was collected by centrifugation and dissolved in TE buffer (10mM Tris-HCl [pH7.5], 1mM EDTA [pH7.5]). The exon/intron boundary structure was analyzed by means of the polymerase chain reaction (PCR) using oligonucleotide primers based on the cDNA sequence. The fragments obtained on PCR were subcloned using a pMOSBlue blunt ended cloning kit (GE Healthcare, USA).

Isolation of total RNA, determination of cDNA extremities, and cDNA analysis: A two to three years old yellowtail was obtained from a local fishery and total RNA was immediately extracted from various organs using an RNA Extraction Kit (GE Healthcare, USA). The 5'-cDNA extremity was determined by the rapid amplification of cDNA ends (RACE) method. Poly (A⁺) RNA was purified from total RNA (100 µg) using an mRNA Purification kit (GE Healthcare, USA). After annealing 1 µg of poly (A⁺) RNA with 10 pmoles of a gene specific reverse primer (5'-gtggagataggcacagttgat-3') derived from the partial cDNA sequence, the first strand cDNA was synthesized by incubation with ReverTraAce reverse transcriptase (Toyobo, Japan) for 1 hour at 42°C. Then, the blunt ended second strand was synthesized by simultaneous inclusion of DNA polymerase I (9U/µl), E. coli DNA ligase (6U/µl, Takara, Japan), and T4 DNA polymerase (9U/µl) in the presence of RNase H (1U/µl) in a reaction buffer composed of 33.4mM Tris-HCl (pH 8.5), 4.6mM MgCl₂, 10mM (NH₄)₂SO₄, 6.6mM 2-mercaptoethanol, 0.06mM EDTA, 0.005% BSA, 100mM KCl, and 0.15mM NAD at 16°C for 90min., and the double-stranded cDNA obtained (5µl) was ligated with 12.5 pmoles of an adaptor consisting of complementary oligo DNA, (5'-gtaatacagactcactatagggcgcgcgtgctgcagcggccgggctgtgt-3') and (5'-accagccggg-3'), using 4 U/µl T4 DNA ligase at 16°C overnight. T4 DNA ligase, DNA polymerase I, and RNase H were obtained from Toyobo, Japan. The ligated cDNA fragments were 5'-RACE amplified by using an adapter primer (5'-gtaatacagactcactatagggc-3') with gene specific primers in 25 µl volume reaction mixtures using XL-PCR kit (PE Applied Biosystems), and then subcloned into the pT7blue vector. We also identified and sequenced EST clones homologous to *YGHL1* from *Danio rerio* and *Xenopus leavis*. Putative transcription factor binding sites were identified manually based on the consensus sequences reported. The CRE sequences were found manually and confirmed with CREB target gene database (<http://natural.salk.edu/CREB/>)

DNA sequencing and analysis: Subcloned DNA was isolated and purified using a FlexiPrep kit (GE Healthcare, USA), and sequenced using a BigDye Terminator Cycle Sequencing Ready Reaction Kit (Invitrogen) and an ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems).

Northern blotting analysis: Total RNA was isolated from various freshly harvested yellowtail organs: heart, kidney, brain, muscle, liver, and gill, using a RNA Extraction Kit (GE Healthcare, USA). Each sample (20µg) was loaded on to a 1% agarose gel containing 5% formaldehyde and then capillary-transferred to a nylon membrane, Hybond-XL, (GE Healthcare, USA). The membrane was hybridized with a [α -³²P]-labelled yellowtail *YGHL1* cDNA probe.

Sequence data analysis: Nucleotide sequences were compared with those in the NCBI database using the basic local alignment search tool (BLAST) algorithm (National Center for Biotechnology Information). Amino acid sequences, deduced from cDNA sequences, were analyzed by means of the CLUSTALW multiple sequence alignment tool¹² and tree generation was done using neighbour-joining algorithm¹³ in MEGA 5.1¹⁴. The bootstrap analysis was done with 1000 replications.

Results and Discussion

Structure, genomic organization and expression analysis of the yellowtail *YGHL1*: Screening of the pituitary gland cDNA library of yellowtail fish for growth hormone related sequences resulted in a cDNA of yellowtail growth hormone like-1 in 1996 (GenBank: D85880). The 5' and 3' RACE experiments yielded the full length cDNA of *YGHL1* (GenBank: AB073641), which encoded a short peptide of 92aa (figure 1. A). The 1529bp cDNA has a coding sequence of 279bp, preceded by a 127bp5' UTR and followed by a 1123bp 3'UTR (figure 1. B). The genomic organization of the Yellowtail *YGHL1* was analyzed by comparing the cDNA sequence with the 6658bp genomic sequence of the yellowtail *YGHL1*, (GenBank: AB094665). The gene consists of 4 exons and 3 introns with exon/intron boundaries having the canonical GT/AG splice sites. The first exon is a non-coding exon, whereas the remaining exons share the coding sequence.

The total RNA samples isolated from various tissues of yellowtail (figure 2. A) were hybridized with a probe derived from the ORF of the yellowtail *YGHL1* cDNA (figure 2. B). The northern blot showed a discrete transcript of 1.8kb in gill, brain, heart, and kidney. The highest level of expression was detected in heart, while no transcript was detected in muscle or liver. There seems to be a differential expression of the *YGHL1/HIG1* orthologs in liver of fishes^{11,15}, which can be attributed to the euryoxic ability of the fish species.

Gracey *et al.* proposed that *HIG1* may be regulated by the HIF-1 α mediated transcriptional mechanism, requiring the presence of hypoxia response element (HRE) in the promoter region. We found 5 HRE sites in the 2066 bp upstream sequence of the yellowtail *YGHL1* based on the HRE consensus sequence^{16, 17}. Olive flounder *YGHL1* is reported to have HRE sites in its promoter region¹⁵. Based on the consensus recognition sequence search, we also found putative binding sites for some specific

transcription factors in the yellowtail *YGHL1* (figure 3). The putative sites are TPA Response Elements (TRE) for Activator Protein-1, binding sites for CCAT/Enhancer Binding Proteins and cAMP Response Elements (CRE) for cAMP Binding Proteins¹⁸⁻²⁰.

Presence of these transcription factor binding sites may be a key for understanding the expression and physiological roles of the yellowtail *YGHL1*. The stabilized and activated, HIF heterodimer binds to the consensus HRE in the HIF target genes to promote survival in low-oxygen conditions. HIF-1 also interacts with other transcription factors, for instance, HIF-1 cooperates with ATF-1 and CREB-1 to activate the lactate dehydrogenase. A gene^{21, 22} or with AP-1 binding factors to activate the vascular endothelial growth factor gene²³. C/EBP family transcription factors together with other transcription factors are known to have pivotal roles in numerous cellular responses, including the control of cellular growth and differentiation, immune and inflammatory processes and various diseases²⁴. However, presence of both CRE and HRE sites in the promoter region of yellowtail *YGHL1* might allow the gene to express in both aerobic and anaerobic conditions. Generally, under normoxic conditions, the CREB protein family is principal mediators of gene expression in response to cAMP following phosphorylation by cAMP-dependent protein kinase (PKA) which is necessary for many living organisms and life processes.

Amino acid alignment and Phylogenetic tree construction of the *YGHL1/HIG1* homologs: The *YGHL1/HIG1* orthologous to the yellowtail *YGHL1* are present in other fishes, amphibians, birds and mammals. The deduced amino acid sequence of the yellowtail *YGHL1* (GenBank: BAB91154.1) was aligned with other *YGHL1* and *HIG1* polypeptides using CLUSTALW program. The alignment result showed high conservation between the amino acid sequences (figure 4). The yellowtail *YGHL1* protein showed 72-82% identity with other fish sequences; 63% identity with human *HIGD1A* and 56% with mouse *Higd1a*. The alignment also showed a highly similar 46aa polypeptide present in all the sequences at the N-terminus, starting and ending with a glycine. The polypeptide is encoded by the 135bp third exon of the Yellowtail *YGHL1* which is also well conserved in other *YGHL1/HIG1* orthologs. Therefore, we propose a nomenclature for the conserved domain as “*YGHL1/HIG1* Exon 3 domain” which is also called as *HIG-1-N* domain. The well-conserved structure of the *YGHL1/HIG1* exon 3 domain suggests that it is either quite stable or sensitive to alterations that would impair its function. In mouse, a chimeric transcript is formed between *UbiE2* and *Yghl1-4*, a non-homologous but tandem gene, by the intergenic splicing event (Genbank: AB110950). The chimeric transcript retains the *YGHL1/HIG1* exon 3 domain, which strongly indicates that the exon 3 domain is functionally important for the *YGHL1/HIG1* family.

A phylogenetic tree was constructed based on amino acid sequences of *YGHL1* and *HIG1* genes from various organisms

using neighbour joining method (figure 5). The tree showed that the yellowtail *YGHL1* is more closely related to the orthologous sequences in fishes. The sequences of human, mouse and other organisms form a separate group from the fishes but still related altogether. The *Ciona intestinalis* *YGHL1/HIG1* seems to be the ancestral source and very closely related to the human *HIGD2A* and mouse *Higd2a*.

Evolution of the *YGHL1/HIG1* gene family by genomic duplication: The phylogenetic and the genomic analyses show the presence of the *YGHL1/HIG1* genes in both the ray finned fishes and the mammals. This fact shows that the *YGHL1/HIG1* gene family might have evolved 450 mya ago by events of genomic duplication taken place between the emergence of vertebrates and the divergence between tetrapods and ray-finned fish as that of *MKRN* gene family^{8,9,25}. An *YGHL1/HIG1* homolog was also found in the *Ciona intestinalis*, which shows that an ancestral *YGHL1/HIG1* was present in Chordata before urochordates and euarchordates diverged, approximately 500 mya. This follows a trend, which is that gene families with several vertebrate paralogs often have only one homolog in the urochordate *Ciona intestinalis*^{26, 27}. Altogether, we propose a schema depicting the evolution of the *YGHL1/HIG1* paralogs in human and mouse (figure 6). Though we consider *Ciona intestinalis* to be the origin for the *YGHL1/HIG1* gene family, there is a possibility of a more primitive eukaryote to have hosted the *YGHL1/HIG1*. The torafugu *YGHL1* represents the *YGHL1/HIG1* homologs in the fish genome from which the paralogs in human and mouse are derived. In human, *HIGD1A*, *HIGD2A*, *HIGD1C* and *CLST11240* were mapped to 3p22, 5q35, 12q13, and 17q21 respectively. In mouse, *Higd1a*, *Higd2a*, *Yghl1-4* and *Higd1b* were mapped to 9F4, 13B1, 15F1, and 11D, respectively. The human *HIGD1A* (3p22), *HIGD2A* (5q35) and mouse *higd2a* (13B1) seems to be derived from the original source after the genomic duplications. The present chromosomal position of the other *YGHL1/HIG1* paralogs appears to be the result of further duplications during the mammalian lineage.

Duplication of a chromosomal segment leads to groups of paralogs. All the neighboring genes in such segment are duplicated at the same event. The number of duplicate genes depends on the size of the segment. Duplication of large chromosomal segments is often called block or segmental duplication, and the resulting duplicate segments with several linked paralogs are often denoted as a paralogon or paralogous region²⁸. According to our genomic analysis performed previously for 3p25 synteny in human chromosome 3¹⁰, the *HIGD1A* seems to be clustered with some significant members of the nuclear receptor (*NR*) superfamily, the *WNT* and *STAT* families. Paralogs of these clustered genes are distributed together with the other paralogous loci of the *YGHL1/HIG1*. The human *GH1*, *GH2* and mouse *Gh* genomic regions are also adjacent to the *CLST11240* in 17q24 (Human) and *Higd1b* in 11D (Mouse). The synteny associated with *YGHL1/HIG1* paralogs may pose strong evidence that this gene family

underwent concerted evolution with these homeostatically and developmentally important genes in vertebrates, especially mammals (figure 6). This theorizes that the *YGHL1/HIG1* gene family might function or be under control of hormonal and signal pathways, in which these genes are involved^{29,34}.



Figure-1

Sequence and structural analysis of yellowtail *YGHL1*

[A] Nucleotide sequence and deduced amino acid sequence of yellowtail *YGHL1*. The initiation codon (atg) is underlined and the stop codon (tga) is indicated by an asterisk. The shaded box indicates the Kozak initiation motif (annatgg)³⁵. The putative polyadenylation signal (AATAAA) is shown in bold and underlined. The conserved *YGHL1/HIG1* exon 3 domain sequence is shown in bold. [B] Genomic structure of yellowtail *YGHL1*. The exons are shown by boxes, where shaded region show the region encoding the ORF and unshaded region indicate untranslated regions. Introns are shown as lines

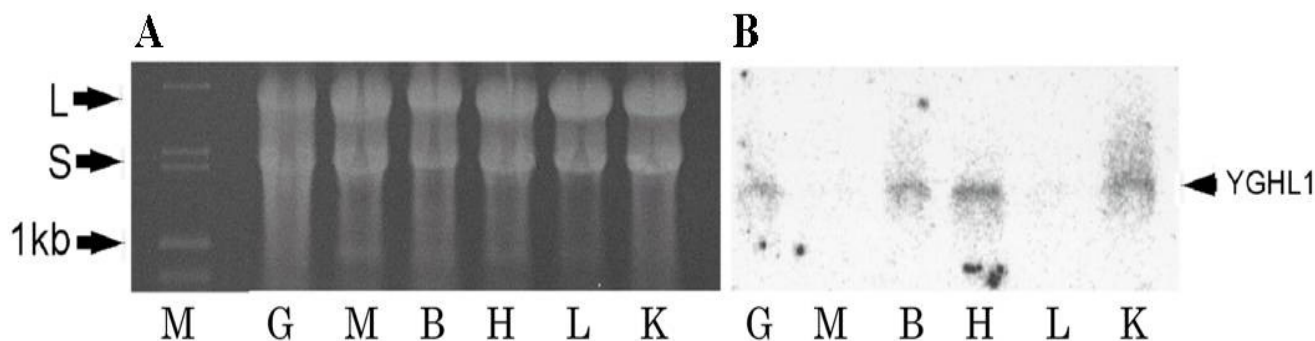


Figure-2

Northern blot analysis of yellowtail *YGHL1* in *Seriola quinqueradiata* organs

[A] Total RNA was isolated and electrophoresed on a 1% agarose/formaldehyde gel from the gill (G), muscle (M), brain (B), heart (H), liver (L) and kidney (K). *Hind* III digest of λ DNA was used as marker (M). L and S with arrows to the left indicate ribosomal large and small RNAs, respectively. [B] A northern blot of the gel from panel A was hybridized with a probe derived from yellowtail *YGHL1* cDNA. Lanes are as mentioned in panel A. The position of the yellowtail *YGHL1* transcript (approximately 1800 bp) is indicated by an arrow mentioning *YGHL1* to the right.

-2066 cgacggccggggtgggtaccagccagtaatacgaactcactatagg**GCACGCGTGGTGG**acggccggggtgggtctg**TTAGGTAAG**taaaagtctgccggtgtaacctttagagaaccttt
HRE C/EBP

-1946 agtgaacattccccagggttacctgtttgctgggtagtgtacacttgcactttggcatctttacaggtaccatctttatgagtgggtgattaacatattcaccactggatggaggtgtt

-1826 tcttcttttgagaaggccaggttccatttaaaaagtcctttgaatttcagtaactatttttgcagcgcacaaggcaaaagtgaat**TTGTGCAAT**aaaataagtaaaagtattagaaaat
C/EBP

-1706 attaataattagttagtaacaatattgacttcaccaggttaacagagactagtatagtacagtacagtataaggatgacactacactatacaacaacaaaaatgtgtatttttaattat

-1586 aggggtggtcactaatagtcataattacataaaggaaatactacaggtcattagagcttcaagctatgggtatgttagttgatgttttgaggatgaaccatagacttttatatgaaa

-1466 agggatgaacaaggcagacgctctcacaacaaatgtgaaaagttgctattgtaacaaacaaatccttagatttaagtggaacgcttgggttgagtgtatcacatcaggcttaagtagaa

-1346 ttatttgcaagaagtacaactactca**TGAATCA**tgcaaaatttctacttctaattag**CGTCA**caaatatttctatgacaacaaaactctcttagtgaag**TGATTCA**gggcagcattaaa
AP-1 CRE AP-1

-1226 agtcgttttaaggacgagtttgtgtaacaagaatgtgtcaatatacacacagacaatgaatgtcaatactgttaatgatagtagccagggcactacaactaatattgttaggacctgatt

-1106 gattagctctattattgatgataattcctaattattttacaacaatatttctatgactaataataattacactaatataattattatcatttgccaaaaaaatattataattacagtaata

-986 aagaataagcattttcttatgcaaatattaattctaaaatcattgctaaaagtattacacaaagtcatttaagagtggtgtaaacatgtctttacgaccactaataataatccg**CGTC**
CRE

-866 **A**ttggccctcgtattattattacaagtaactgtaatatttaatatatttaatttttaaacatgatctgaattagatgcagcagctgtg**AGAAACGTGCCAG**gagcctcttgacatgag
HRE

-746 gccactgttacatcatctcagagat**TAACACGTGAAGC**aaga**TGATTCA**cctccaactacttccaactttgtttaataccggtaactcatattgtggcattctacagtaagatgtttaa
HRE AP-1

-626 ac**TATTACGTGGGCA**atagtgtatggatagatattagttgaagtgaagaggcaggttcaatttcttgagaataactgggttgcttaacattaaccattaaatagttgttagtttcccttg
HRE

-506 tttccagattagtagtcttccacaatcttcgattttacaagtagagttaccattgatgtgctgcttcaggacttctcggattattacaacatcaaatgttgactgtaacgatgaaag

-386 agtatgttatgtgcttagagtaagattatgtgatgtcatgtttaccattttgattaatacatagctctca**CGTCA**accactaggacttcttatagggttaagagctgtgtcaacaacca
CRE

-266 catccaacagacagcgcacaaaagtcattgtttatactatattgaaggacgtgccataggtacaaattttatttttttctacttatttgagggcttttttttagaacagactaat

-146 ttgttttaacataaattcataactggcaagcgttcatttaaaacaacacaaattttcaaggcacagacattattaccattttctcttgtaactttaacgcagcacactcgtgagaaactg

-26 c**TGACGTC****AGAACTGCGTGCGAT**ggt
CRE HRE

Figure-3

Nucleotide sequence of the 5' upstream region of the yellowtail YGHL1 gene

Putative binding sites of the transcription factors are shown underlined and bold. A line is drawn to separate the CRE and HRE regions at end of the sequence

Conclusion

We cloned, sequenced and characterized the yellowtail *YGHL1* gene and its promoter region in *Seriola quinqueradiata*. The well conserved *YGHL1/HIG1* Exon 3 domain was identified by multiple polypeptide analysis of sequences orthologous to yellowtail *YGHL1*. The ancestral homolog of *YGHL1/HIG1* has been the source of this extensive *YGHL1/HIG1* family. The

synteny of *YGHL1/HIG1* paralogs along with linked flanking genes in human and mouse are results of genomic duplication events. Further empirical studies are required to reveal the function of yellowtail *YGHL1* and to determine the role of transcription factor binding sites in yellowtail *YGHL1* and the *YGHL1/HIG1* family.

Seriola quinqueradiata_YGHL1	-----MT--AYDENE-----SKLMRKVKENPFVP	22
Paralichthys olivaceus_YGHL1	-----MS--AYTENE-----SKFMRKAKENPFVP	22
Takifugu rubripes_YGHL1	-----MS--GIEENE-----SKFLRKAKENPFVP	22
Gillichthys mirabilis_HIG1	-----MSL--DYEEHQ-----SKFMRKAKENPFVP	23
Danio rerio_YGHL1	-----MSDIGYEDNE-----SKLMRKIKQNPFPV	24
Xenopus laevis_YGHL1	-----MAQPGD-VLPTYDMSDS-----QTSKLIRKSKEPFVP	32
Gallus gallus_HIG1	-----MSSGHEPVYPEYETETS-----QTSKLIRKFKETPFVP	33
Homo sapiens_HIGD1A	-----MSTDTGVSLPSYEED-----QGSKLIRKAKEAPFVP	31
Mus musculus_Higd1a	-----MSTNTDLSSLSSYDEG-----QGSKFIRKAKETPFVP	31
Homo sapiens_HIGD1C	-----MSSDNQWSADEDEGQL-----SRLIRKSRSDFPFVP	30
Mus musculus_Yghl1-4	-----MSSD--EWSAAEDEGQL-----SRLIRKSRSDFPFVP	29
Homo sapiens_CLST11240	-----MSANRRWVVPDDEDC-----VSEKLIRKTRRESPLVP	32
Mus musculus_Higd1b	-----MSANKGWVVPPEGEDN-----LSKKFLRKTRRESPLVP	32
Homo sapiens_HIGD2A	MATPGVPIPEVPFEPSKPPVIEGLSPVYRNPESEFKEKFRKTRNPVVP	50
Mus musculus_Higd2a	MAAPGPVSPAPPFDPSPVIEGFSPTVYSNPEGFKEKFRKTRNPVVP	50
Ciona intestinalis_YGHL1	-----MTKEEILK-----FLQHCKDNPFVP	20
	:::: :: *.**	
Seriola quinqueradiata_YGHL1	VGIAGFFAI VGYRLMKMKNRGDTKMSVHLIHMVAAQGFVVGAMTVG-VL	71
Paralichthys olivaceus_YGHL1	VGLAGFFAI VGYRLMKMKTRGDTKMSVHLIHMVAAQGFVVGAMTVG-VL	71
Takifugu rubripes_YGHL1	VGMAGFLAI VGYRLMKMRSRGDTKMSVHLIHMVAAQGFVVGAMTVG-VL	71
Gillichthys mirabilis_HIG1	VGITGELGIVAYQLMKMKHRGNVMSVHLIHMVAAQGFVVGAMTVG-VI	72
Danio rerio_YGHL1	AGMAGFFAI VAYRLFKLSRGDTKMSVHLIHMVAAQGFVVGAMTVG-VI	73
Xenopus laevis_YGHL1	IGMAGFAAVVAFGLFKLSRGNTKMSVHLIHMVAAQGFVVGAMTCG-VL	81
Gallus gallus_HIG1	IGMAGFAVVVGYGGLYKLRGNTKMSLHLIHMVAAQGFVVGAVTTCG-VL	82
Homo sapiens_HIGD1A	VGIAGFAAI VAYGLYKLSRGNTKMSIHLIHMVAAQGFVVGAMTVG-MG	80
Mus musculus_Higd1a	IGMAGFAAI VAYGLYKLSRGNTKMSIHLIHMVAAQGFVVGAMTVG-MG	80
Homo sapiens_HIGD1C	IGIAGFVTTVSCGLYKLYRRDQKMSIHLIHMVAAQGFVVGAVTTCG-VL	79
Mus musculus_Yghl1-4	VGMAGFVAVLSYGLYKLSRRQKMSLHLIHMVAAQGFVVGAVTTCG-VL	78
Homo sapiens_CLST11240	IGLGGCLVVAAYRIYRLRSRGSTKMSIHLIHMVAAQACAVGAIMLG-AV	81
Mus musculus_Higd1b	IGVAGCLVIAAYRIYRLKARGSTKLSIHLIHMVAAQACAVGAIMLG-AM	81
Homo sapiens_HIGD2A	IGCLATAAALTGYLYSFH-RGNSQRSQIMMRTRIAAQGFVVAAILLG-LA	98
Mus musculus_Higd2a	IGCLGTAAALTGYLYCFH-RGQSHRSQIMMRTRIAAQGFVVAAILLG-LA	98
Ciona intestinalis_YGHL1	LGLSATVLFLENGLRHMR-LGNRQQSQVMMRGRVMAQGFVIAITTCGAVY	69
	* . : . . : * :: * : * . : * :	
Seriola quinqueradiata_YGHL1	YSMYRDFIVKPREEQKSMQNK-----	92
Paralichthys olivaceus_YGHL1	YTMRYDVIKPRE-LKAVENK-----	91
Takifugu rubripes_YGHL1	FSMYKDYFAKPKENHK-----	87
Gillichthys mirabilis_HIG1	YSMYRDYIAHPSEKKDE-----	90
Danio rerio_YGHL1	YSMYNEYILKPAEAQKALEQKALEKK	99
Xenopus laevis_YGHL1	YSMYKEYLAKPSEH-----	95
Gallus gallus_HIG1	YSMFREHVIVKPE-----	95
Homo sapiens_HIGD1A	YSMYREFWAKPKP-----	93
Mus musculus_Higd1a	YSMYQEFWANPKPKP-----	95
Homo sapiens_HIGD1C	YSMYKDYIRPRFFSESKE-----	97
Mus musculus_Yghl1-4	YSMYKDYIRPRFFNVPK-----	96
Homo sapiens_CLST11240	YTMYSYVVKRMAQDAGEK-----	99
Mus musculus_Higd1b	YTMRYDYIKRVSEDA-EK-----	98
Homo sapiens_HIGD2A	VTAMKSRP-----	106
Mus musculus_Higd2a	ASAMKSQA-----	106
Ciona intestinalis_YGHL1	QSAFKSKKENVV-----	81
	: .	

Figure-4

Alignment of the deduced amino acid sequences of YGHL1/HIG1 from *Seriola quinqueradiata* and other organisms

The deduced amino acid sequences of the yellowtail YGHL1 is aligned with YGHL1/HIG1 homologs of other organisms using CLUSTALW alignment tool (Thompson et al. 1994). Asterisks mark absolutely conserved residues, and two dots or one dot represent highly or moderately similar amino acids, respectively. Dashes represent gaps inserted by the program to optimize the alignment. The red coloured region shows the highly conserved YGHL1/HIG1 exon 3 domain. GenBank accession numbers of the YGHL1/HIG1 amino acid sequences are as follows: *Seriola quinqueradiata* YGHL1 (GenBank: BAB91154.1), *Gillichthys mirabilis* HIG1 (GenBank: AAG13326.1), *Takifugu rubripes* YGHL1 (GenBank: BAD98705.1), *Danio rerio* YGHL1 (GenBank: BAC82438.1), *Paralichthys olivaceus* YGHL1 (GenBank: AEP34037.1), *Xenopus laevis* YGHL1 (GenBank: BAC75823.1) *Homo sapiens* HIGD1A (GenBank: NP_054775.2), *Homo sapiens* HIGD2A (GenBank: NP_620175.1), *Homo sapiens* CLST11240 (GenBank: NP_057522.1), *Homo sapiens* HIGD1C (GenBank: NP_001103089.1), *Mus musculus* Higd1a (GenBank: BAE92116.1), *Mus musculus* Higd2a (GenBank: NP_080209.1), *Mus musculus* Higd1b (GenBank: AAH05568.1), *Mus musculus* Yghl1-4 (GenBank: BAE92117.1), *Gallus gallus* HIGD1A (GenBank: NP_001026128.1), *Ciona* Yghl1/HIG1 (GenBank: XP_002130406.1)

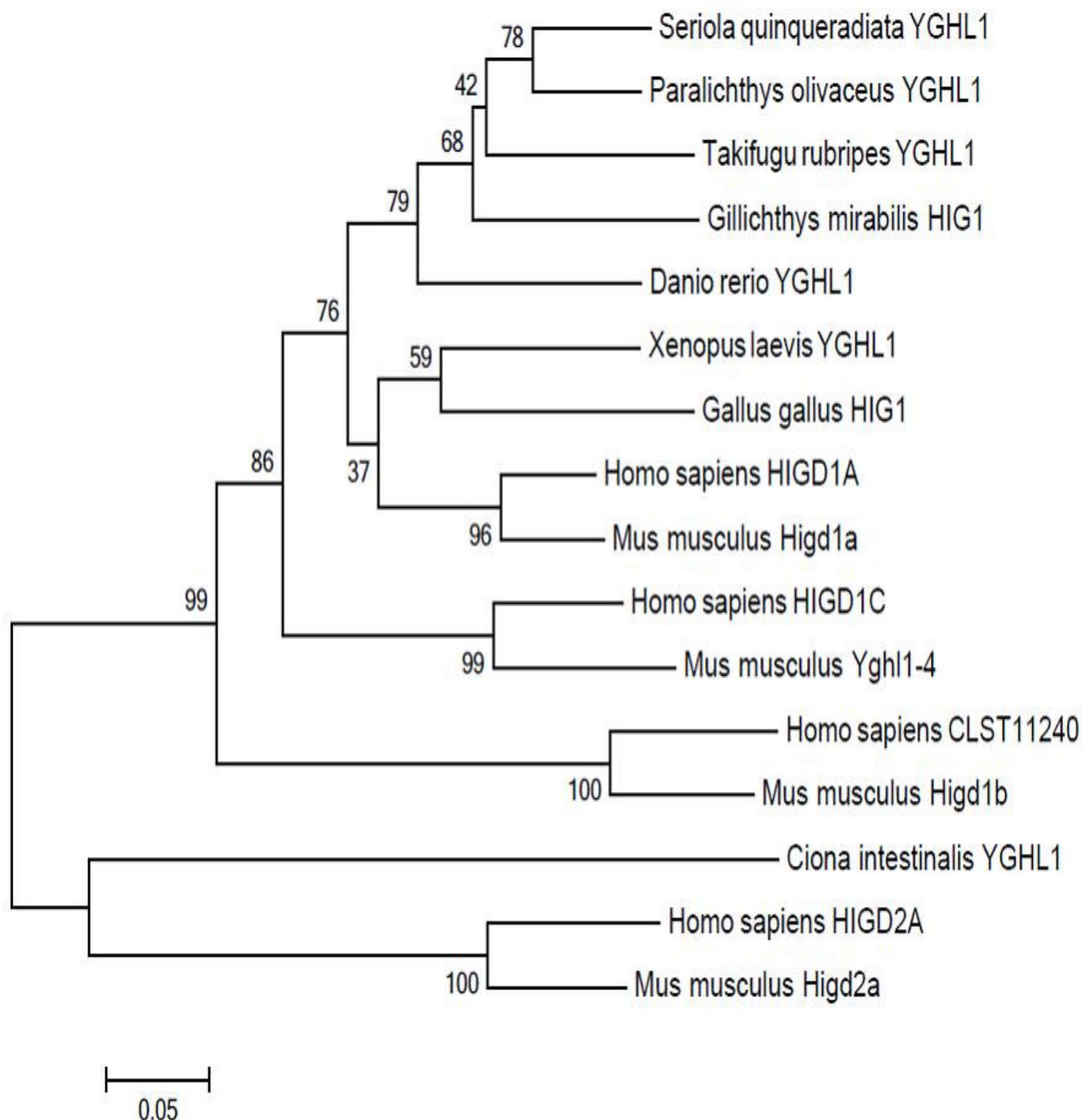


Figure-5

Phylogenetic analysis of the deduced amino acid sequences of YGHL1/HIG1 from *Seriola quinqueradiata* and other organisms

A Phylogenetic tree of YGHL1/HIG1 amino acid sequences from *Seriola quinqueradiata* and other organisms was constructed using the neighbor-joining algorithm in MEGA version 5.1 (Tamura et al., 2011). Bootstrap values supporting the branches were determined with 1000 repetitions and indicated at the corresponding nodes. The tree illustrates the evolutionary relationships among YGHL1/HIG1 in various organisms

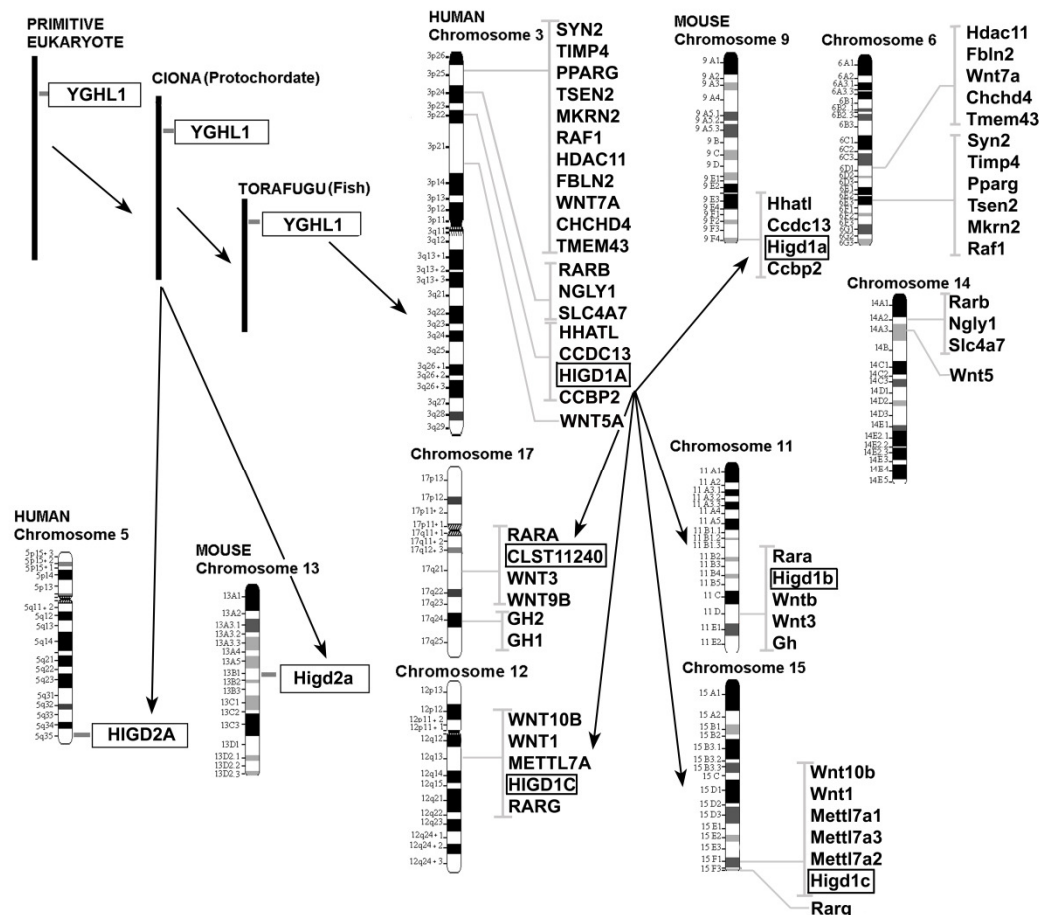


Figure-6

Evolution of YGHL1/HIG1 paralogs in mouse and human

A schema showing the possible origin of primitive YGHL1/HIG1 in *Ciona intestinalis*, its duplication and location in human and mouse chromosomal positions as YGHL1/HIG1 paralogs. The genes flanking the YGHL1/HIG1 paralogs show the syntenic relationships in human and mouse.

References

- Ohno S., Evolution by gene duplication, Springer-Verlag (1970)
- Kessel M. and Gruss P., Murine developmental control genes, *Science*, **249**, 374-379 (1990)
- Nadeau J.H. and Sankoff D., Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution, *Genetics*, **147**, 1259-1266 (1997)
- Gibson T.J. and Spring J., Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins, *Trends Genet.*, **14**, 46-49 (1998)
- Otto S.P., The evolutionary consequences of polyploidy, *Cell*, **131**(3), 452-462 (2007)
- Gerstein A.C. and Otto S.P., Ploidy and the causes of genomic evolution, *J Hered.*, **100**(5), 571-581 (2009)
- Taylor J.S., Van de Peer Y., Braasch I. and Meyer A., Comparative genomics provides evidence for an ancient genome duplication event in fish, *Philos Trans R Soc Lond B Biol Sci.*, **356**, 1661-1679 (2001)
- Gray T.A., Hernandez L., Carey A.H., Schaldach M.A., Smithwick M.J., Rus K., Graves J.A.M., Stewart C.L. and Nicholls R.D., The ancient source of a distinct gene family encoding proteins featuring Ring and C3H zinc-finger motifs with abundant expression in developing brain and nervous system, *Genomics*, **66**, 76-86 (2000)
- Gray T.A., Azama K., Whitmore K., Min A., Abe S. and Nicholls R.D., Phylogenetic conservation of the makorin-2 gene, encoding multiple zinc-finger protein, antisense to the RAF1 proto-oncogene, *Genomics*, **77**, 119-126 (2001)
- Abe S., Chiba S., Mishra N., Minamino Y., Nakasuji H., Doi M. and Gray T.A., Origin and evolution of the genomic region encoding *RAF1*, *MKRN2*, *PPARG*, and *SYN2* in

- human chromosome 3p25, *Mar. Biotechnol.*, **6**, S404-S412 (2004)
11. Gracey A.Y., Troll J.V. and Somero G.N., Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*, *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 1993-1998 (2001)
 12. Thompson J.D., Higgins D.G. and Gibson T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice *Nucleic Acids Res.*, **22**, 4673-4680 (1994)
 13. Saitou N. and Nei M., The neighbor-joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, **4**, 406-425 (1987)
 14. Tamura K., Peterson D., Peterson N., Steker G., Nei M. and Kumar S., MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, **28**, 2731-2739 (2011)
 15. Kim Y.O., Park E.M., Moon J.Y., Kong H.J., Nam B.H., Kim W.J., Lee J.H., Kim K.K., Lee S.J., Molecular characterization and transcriptional analysis of the olive flounder (*Paralichthys olivaceus*) YGHL1 gene in response to hypoxia and infection, *Mol. Cell. Biochem.*, **357**, 305-312 (2011)
 16. Wenger R.H., Stiehl D.P., Camenisch G., Integration of oxygen signaling at the consensus HRE. *Sci STKE*, **re12** (2005)
 17. Ferreira T.C., Hertzberg L., Gassmann M., Campos E.G., The yeast genome may harbor hypoxia response elements (HRE), *Comp Biochem Physiol C Toxicol Pharmacol.*, **146**, 255-263 (2007)
 18. Akira S., Isshiki H., Sugita T., Tanabe O., Kinoshita S., Nishio Y., Nakajima T., Hirano T., Kishimoto T., A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family, *EMBO J.*, **9**(6), 1897-907 (1990)
 19. Zhang X., Odom D.T., Koo S.H., Conkright M.D., Canettieri G., Best J., Chen H., Jenner R., Herbolsheimer E., Jacobsen E., Kadam S., Ecker J.R., Emerson B., Hogenesch J.B., Unterman T., Young R.A., Montminy M., Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues, *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 4459-4464 (2005)
 20. Angel P., Hess J., The Multi-Gene Family of Transcription Factor AP-1. In: Bradshaw RA, Dennis EA (ed) Regulation of Organelle and Cell Compartment Signaling: Cell Signaling Collection, Academic Press, 53-62 (2011)
 21. Firth J.D., Ebert B.L. and Ratcliffe P.J., Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements, *J. Biol. Chem.*, **270**, 21021-21027 (1995)
 22. Ebert B.L. and Bunn H.F., Regulation of transcription by hypoxia requires a multiprotein complex that includes hypoxia-inducible factor 1, an adjacent transcription factor, and p300/CREB binding protein, *Mol. Cell. Biol.*, **18**, 4089-4096 (1998)
 23. Damert A., Ikeda E. and Risau W., Activator-protein-1 binding potentiates the hypoxia-inducible factor-1-mediated hypoxia-induced transcriptional activation of vascular-endothelial growth factor expression in C6 glioma cells, *Biochem. J.*, **327**, 419-423 (1997)
 24. Ramji D.P. and Foka P., CCAAT/enhancer-binding proteins: structure, function and regulation, *Biochem. J.*, **365**, 561-575 (2002)
 25. Böhne A., Darras A., D'Cotta H., Baroiller J-F., Galiana-Arnoux D. and Volff J-N., The vertebrate makorin ubiquitin ligase gene family has been shaped by large-scale duplication and retroposition from an ancestral gonad-specific, maternal-effect gene, *BMC Genomics*, **11**, 721 (2010)
 26. Dehal P., Satou Y., Campbell R.K., Chapman J., Degnan B., De Tomaso A., Davidson B., Di Gregorio A., Gelpke M., Goodstein D.M., Harafuji N., Hastings K.E., Ho I., Hotta K., Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen I.A., Nacula S., Nonaka M., Putnam N., Rash S., Saiga H., Satake M., Terry A., Yamada L., Wang HG., Awazu S., Azumi K., Boore J., Branno M., Chin-Bow S., DeSantis R., Doyle S., Francino P., Keys DN., Haga S., Hayashi H., Hino K., Imai K.S., Inaba K., Kano S., Kobayashi K., Kobayashi M., Lee B.I., Makabe K.W., Manohar C., Matassi G., Medina M., Mochizuki Y., Mount S., Morishita T., Miura S., Nakayama A., Nishizaka S., Nomoto H., Ohta F., Oishi K., Rigoutsos I., Sano M., Sasaki A., Sasakura Y., Shoguchi E., Shin-i T., Spagnuolo A., Stainier D., Suzuki M.M., Tassy O., Takatori N., Tokuoka M., Yagi K., Yoshizaki F., Wada S., Zhang C., Hyatt P.D., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T., Richardson P., Lucas S., Kohara Y., Levine M., Satoh N., Rokhsar D.S., The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins, *Science*, **298**, 2157-2167 (2002)
 27. Guo D., Holmlund C., Henriksson R. and Hedman H., The LRIG gene family has three vertebrate paralogs widely expressed in human and mouse tissues, and a homolg in Ascidiacea, *Genomics*, **84**, 157-165 (2004)
 28. Coulier F., Popovici C., Villet R. and Birnbaum D., MetaHox gene clusters, *J. Exp. Zool.* **288**, 345-351 (2000)
 29. Caron H., van Schaik B., van der Mee M., Baas F., Riggins G., van Sluis P., Hermus M.C., van Asperen R., Boon K., Voute P.A., Heisterkamp S., van Kampen A., Versteeg

- R., The human transcriptome map: Clustering of highly expressed genes in chromosomal domains, *Science*, **291**, 1289 (2001)
30. Hurst L.D., Pal C. and Lercher M.J., The evolutionary dynamics of eukaryotic gene order, *Nat. Rev. Genet.*, **5**, 299-310 (2004)
31. Khaitovich P., Muetzel B., She X.W., Lachmann M., Hellmann I., Dietzsch J., Steigele S., Do H.H., Weiss G., Enard W., Heissig F., Arendt T., Nieselt-Struwe K., Eichler E.E. and Paabo S., Regional patterns of gene expression in human and chimpanzee brains, *Genome Res.*, **14**, 1462-1473 (2004)
32. Liu C., Ghosh S., Searls D.B., Saunders A.M., Cossman J., Roses A.D., Clusters of adjacent and similarly expressed genes across normal human tissues complicate comparative transcriptomic discovery, *Omics*, **9**, 351-363 (2005)
33. Singer G.A., Lloyd A.T., Huminiecki L.B. and Wolfe K.H., Clusters of co-expressed genes in mammalian genomes are conserved by natural selection, *Mol. Biol. Evol.*, **22**, 767-775 (2005)
34. Purmann A., Toedling J., Schueler M., Carninci P., Lehrach H., Hayashizaki Y., Huber W. and Sperling S., Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality, *Genomics*, **89**, 580-587 (2007)
35. Kozak M., Initiation of translation in prokaryotes and eukaryotes, *Gene*, **234**, 187-208 (1999)