# Module 4: Comparative Sequence Analysis

# Aims

- Review the sequences available from different organisms
- Identify putative homologous genes
- Compare genome sequences from different organisms and identify conserved sequences
- Analyse conserved sequences for potential regulatory function

Comparative sequence analysis is a powerful method for aiding human gene identification, inferring function of a gene's product, and identifying novel functional elements such as those involved in transcriptional regulation. This is because biologically important regions of the genome are, generally, under selective constraint. Comparing the genome sequences from a variety of organisms may facilitate the identification of functionally significant units in the human genome.

The information that can be inferred when comparing sequences is dependent on the evolutionary distance between the two organisms. Organisms that are closely related are more likely to share a higher degree of sequence similarity. Distantly related organisms such as yeast and worm share less sequence similarity and are likely to show sequence conservation in coding regions alone. This may also be true for distantly related vertebrates such as fish. More closely related organisms, such as mouse, are likely to be conserved in coding regions, and other functional elements such as regulatory sequences. However, the closer the evolutionary relationship with human, the more 'sequence noise' is likely to arise where non-functional sequence appears similar because insufficient time has elapsed for the two sequences to diverge.

99

#### **Evolutionarily Related Gene Sequences**

Homologous genes are derived from a common ancestor and may either a similar sequence or function. In general, homologous genes can be divided into two classes:

**Orthologues** are genes that often perform the same function in different organisms. They are defined as being homologous genes in different organisms derived from the same gene during speciation. In general, their sequence similarity reflects amount of time since they diverged from a common ancestor i.e., the less time that has elapsed since divergence, the greater the sequence similarity between the two genes. These are genes 1 and 2 in figure 4.1.

**Paralogues** are genes families that are present with in a single species. Often they arise by duplication. These genes are not under the same pressure to maintain their function so that one copy may acquire a novel function. These are genes 2 and 3 in figure 4.1.



Figure 6.1 Homologous Gene Sequences

#### Identifying evolutionarily related gene sequences – where to start...

Most searches for orthologous genes begin with blast searches (for more details refer to module 1 of this manual). The type of search that you should perform depends upon what information you have at your disposal. Protein sequence based queries generally find more distantly related matches because of the redundancy in the genetic code (i.e. some amino acids are encoded by more than one codon). Nucleotide searches using the discontiguous megablast parameters are also very useful. We recommend that you survey a number of different databases using a number of different search parameters to obtain the most informative results.

#### BUT be warned: uncertain Orthologues

You may encounter gene sequences that appear to be orthologous and may be derived from the same ancestor but no longer perform the same biological function (for example genes 1 and 3 in figure 4.1). If you choose to analyse such sequences the sensitivity and specificity of your search will be reduced and it may not yield any informative results.

For example, the gene for bone morphogenetic protein 8 *(BMP8)* was duplicated in a common ancestor of human and mouse giving rise to *BMP8a* and *BMP8b* (see Figure 4.2). BLAST analysis of these four sequences yields quite confusing results. Human and mouse *BMP8a* are reciprocal best alignments using both nucleotide and protein sequences to search. In contrast, both the nucleotide and protein sequences of mouse *Bmp8b* align best to their human *BMP8a* counterparts. Human *BMP8b* mRNA aligns best to mouse *Bmp8b* mRNA, but human BMP8b protein aligns best to mouse Bmp8b mRNA, but human BMP8b aligns best to human protein *BMP8a* (Nardone *et al.,* 2004, figure 4.2).

101



# Figure 4.2 Uncertain BMP8a and BMP8b orthologues.

Therefore, we recommend that you perform the following steps to confirm the true functionality and relatedness of your gene sequences. Much of this can be information can be obtained from genome browsers such as Ensembl, NB KI NCBI or the UCSC.

- Identify any other paralogues that may affect your analysis. This can be achieved by performing a BLAST search your sequence against it source genome or using the self-chain track at the UCSC genome browser.
- Confirm the percentage identity (similarity) at both the nucleotide and protein level between paralogous and orthologous sequences to ensure that you are analysing the most closely related sequences.
- Perform evolutionary analysis of nucleotide/protein sequence (phylogeny). In contrast to similarity-based methods such as BLAST, phylogenetic methods can better take into account the effects of repeated substitutions at one site and variable rates of evolution

among sequences. Multiple genes are placed in an evolutionary tree representing genealogical relationship

- 4) Compare the exon/intron structure of your orthologous genes.Evolutionarily related genes often share a similar gene structure
- 5) Examine the chromosomal context of the two orthologous genes. Closely related species, such as human and mouse often have large conserved segments (for definition see later section Genome sequence analysis) and therefore neighbouring genes are also shared between the two species.

# **Comparative Genome Analysis**

Comparing the DNA sequences of different species is a powerful method for decoding genomic information. This is because functional sequences tend to evolve at a slower rate than non-functional sequences. By comparing the genomic sequences of several species at different evolutionary distances it is possible to identify coding sequences, conserved non-coding sequences and those sequences that are unique to humans.

Advances in bacterial clone mapping and sequencing that evolved during the construction of the sequence-ready maps for particularly the worm and the human genome are now being applied to other organisms. The sequence of many genomes has or is being generated using a combination of the cloneby-clone method (adopted for generating the human genome sequence by the public effort) and whole genome shotgun (WGS - used by Celera to generated the sequence of the fruit fly and their version of the human genome sequence). Finished and unfinished genome assemblies are currently available for 44 vertebrates. The front page of the genome browser Ensembl (http://www.ensembl.org) is shown on the following page (figure 4.3). It displays all species whose genomes have been sequence coverage and participating research institutes can be found by following the appropriate hyperlinks on this page. Similar links can also be found at the UCSC genome browser.



Figure 4.3. Ensembl – available species (May 2009)

When two species diverge from a common ancestor those sequences that maintain their original function are likely to remain conserved in both species throughout their subsequent independent evolution. Therefore comparing sequences in different species is a powerful tool for increasing the confidence of a predicted functional unit, or identifying novel functional units (e.g. human, mouse and zebrafish).

In general, greater evolutionarily distance between the species is reflected by more divergent sequences and fewer shared functional units. Comparing sequences that diverged from a common ancestor approximately 450 million years ago (mya) e.g. human and fish aids the identification of coding sequences. Conserved non-coding regions are generally not identified. If the evolutionary distance between the two species is reduced to approximately 60 mya, e.g. human and mouse, both non-coding and coding units are commonly conserved. A large number of features are conserved between recently evolved species such as human and chimp. The inclusion of a closely related species in a comparative analysis makes it possible to identify coding and non-coding sequences but also those genomic sequences that may be responsible for traits that are unique to the reference species.

Human vs.	Chimpanzee	Mouse	Opossum	Fish
Size (Gbp)	3.0	2.5	3.5	0.4
Time since divergence	~5 MYA	~65MYA	~150 MYA	~450 MYA
Sequence conservation (in coding regions)	>99%	~80%	~70%	~65%
Aids identification of	Recently changed sequences and genomic rearrangements	Both coding and non- coding sequences	Both coding and non- coding sequences	Primary coding sequences
Background noise	HIGH	MODERATE	LOW	VERY LOW

Table 4.1	: Selection	of Species	for DNA	Comparisons
14810 111		0. 000000		• • • • • • • • • • • • • • • • • • •

Today we will be showing you some of our favourite web-sites where homologous gene sequences can be identified and analysed. These are listed in table 4.2. As with most types of web-based genomic analysis there are a vast number of sites that can be used for this type of analysis. These sites require different amounts of information from the user; some require an input sequences while others contain pre-calculated information in a database that can be interrogated easily.

# Table 4.2 Different steps and web programmes that can be used to identify evolutionarily conserved regions

Step	Using whole genome assemblies	Interrogating and incorporating sequences from databases
Identifying	UCSC – Self Chain	NCBI - blastn
paralogues	Ensembl – Protein family	
Identifying	Ensembl – Orthologue	Discontiguous megablast
orthologues	Prediction	BLAST-n, -p
	UCSC - BLAT	BLink
Confirming true	Ensembl - MultiContigView	Clustalw
orthology	- SyntenyView	
	- AlignSliceView	
Aligning and	ECR Browser	zPicture
identifying	Genome Vista	Vista
conserved	PhastCons	LAGAN
sequences		

#### WORKED EXAMPLES:

#### 1) Identifying paralogous and orthologous genes

The easiest and most common method used to identify homologous sequences exploits the sequence conservation between related genes. This can be identified using sequence similarity searches i.e., BLAST searches against nucleotide or protein databases (see module 1). However, in many cases the genome browsers such as NCBI, Ensembl and the UCSC genome browsers have already done the hard work for you. For example orthologous gene predictions from whole genome assemblies can be accessed from either **HomolGene** (NCBI) or **Orthologue Prediction** (Ensembl); while paralogous genes can be identified by following the **Protein Family** link found in Ensembl GeneView.

#### a) Useful sites at the NCBI

If you have the sequence for your gene of interest in one species and wish to find orthologues BLAST is a good starting point

BLAST	Basic Local Alignment Search Tool	My NCBI P	
Home	t Results Saved Strategies Help	[Sign In] [Register]	BLAST Species
CBI/ BLAST Home		News	BEACT OPECIES
BLAST finds regio	ons of similarity between biological sequences. <u>more</u>		specific databases
Learn more about h	how to use the new BLAST design	BLAST 2.2.18 now available	
	blad Conomoo	2.2.18 are now available at	
DLAST ASSem	bled Genomes	our F1 site	
Choose a species g	enome to search, or list all genomic BLAST databases.	More BLAST news	
Human	B Oniza sativa		
Mouse	Bos taurus     Pan troulodutes		
Rat	Danio rerio Microbes		
Arabidopsis the	aliana Drosophila melanogaster Drosophila melanogaster	Tip of the Day	
		Use Genomic BLAST to	
		see the genomic	he general of
Basic BLAST		context IT U	ine genome of [
Choose a BLAST on	ogram to run	If you are interested in the	ur opening of
onooce a bb tor pr	ogram vo rom.	evolution of a particular <b>yO</b>	ur species of [
		gene or gene family it is	oract is not
<u>nucleotide blast</u>	Search a nucleotide database using a nucleotide query Algorithms: blastn megablast discontiguous megablast	the intro-exon structure	
	i i i i i i i i i i i i i i i i i i i	even across species.	mploto vou con
protein blast	Search protein database using a protein query	Often, the only data	inplete you can
	Algonithms: blastp, psi-blast, phi-blast	sequence from a cDNA or a	AST traco
<u>blastx</u>	Search protein database using a translated nucleotide query	curated database such as	
<u>tblastn</u>	Search translated nucleotide database using a protein query	however, to see how the	chicves for
tblastx	Search translated nucleotide database using a translated nucleotide query	sequence using BLAST and thus arrive at an idea of its	ecific organisms
Specialized BL	AST	structure.	
Choose a type of sp	enialized search (or database name in prontheses )	More tips	nt-use
Search trace are	rchives	l dis	scontiquous
Find conserved	domains in your sequence (cds)		guode
Find sequences	with similar <u>conserved domain architecture</u> (cdart)	l me	egablast
Search sequence	ces that have <u>gene expression profiles</u> (GEO)		3
<ul> <li>Search Immune</li> <li>Search for SNP</li> </ul>	s (con)		
Screen sequence	e for vector contamination (vecscreen)		
Align two seque	ences using BLAST (bl2seq)		
Search protein	or nucleotide targets in PubChem BioAssay		

However NCBI has pre-computed orthologues which can be found in the HomoloGene page. However because this page is precomputed this means that it does not contain the most recently assembled genomes.

#### 1: We will start by searching for our gene of interest MITF at NCBI



r

Protein Alignments Protein multiple alignment, pairwise similarit	y scores and	Conserved Do <u>maine</u> Conserved Domains from CDD found in protein sequences		3: Select Show Multiple Alignment
evolutionary distances.		by rpsblast searching.		
Show Multiple Alignment		S HLH (cd00083)		
Show Pairwise Alignment Scores Pairwise alignments generated using BL	AST	Helix-loop-helix domain, found in specific DNA- binding proteins that act as transcription factors; 60- 100 amino acids long.		
Regenerate Alignments				
XP_001138775.1 (Pan troglodytes) XP_855594.1 (Canis lupus familiaris) BLAST	•	Related Homology Resources Links to curated and computed homology information found in other databases MGI:104554 Orthology group for M.musculus Mtf includes H.sapiens		
		MITF and R.norvegicus Mitf.		
Phenotypes Phenotypic information for the genes in this from model organism databases.	entry imported	UniGene Links to groups of transcribed sequences established by tblastn searching of UniGene.		Further orthologues
Homo sapiens		Bt.29882. Bos taurus		Identified by
MIM:103470 Waardenburg syndrome/ocular albinism MIM:103500 Tietz oendeme	, digenic.	Microphthalmia-associated transcription factor Bt.66572, Bos taurus Transcribed locus, strongly similar to NP_032627.1		searching UniGene
MIM: 193510 Waardenburg syndrome, type IIA.		microphthalmia-associated Bt.98153, Bos taurus Transcribed locus, strongly similar to NP_032627.1 microphthalmia-associated		there is the horse
Mus musculus		Cfa.2682, Canis lupus familiaris		orthologue
MP:0001186 Pigmentation phenotype.		Microphthalmia-associated transcription factor		
MP:0003631 Nerrous system phenotype		Microphthalmia-associated transcription actor a	L	
MP:0005371		Dr.83675, Danio rerio Microphthalmia-associated transpotion factor b		
Limbs/digits/tail phenotype.		🔟 Eca 12983, Equus caballus		
Lethality-postnatal.		Microphthalmia transcription factor		
NP_937802.1         1           XP_001138775.1         1           XP_00101150.1         NP<02627.1	MQSESGIV MQSESGIV	PDPEVGEEPHEEPKTYYELKSQPLKSSSSAEHPGAS PDPEVGEEPHEEPKTYYELKSQPLKSSSSAEHPGAS	44 44	Multiple Alignment of orthologous protein sequences
XP_001065702.1 1	MQSESGIV	ADFEVGEEFHEEPKTYYELKSQPLKSSSSAEHSGAS	44	
<u>NP_990360.1</u> <u>NP_571922.1</u> 1	MQSESGIV	PDFEVGDDFHEEPKTYYELKSQPLQNSNPSEQQ	41	
NP         937802.1         45           XP         001138775.1         45           XP         855594.1         1           NP         001001150.1         1           XP         001065702.1         45           NP         990360.1         1           NP         571922.1         42	KPPISSS KPPISSS  	MTSRILLRQQLMREQMQBQBRREQQQKLQAAQFMQQ MTSRILLRQQLMREQMQBQBRREQQQKLQAAQFMQQ MLEMLEYN	88 88 8 8 8 8 8 8 8 8 7 7	
NP         937802.1         89           XP         001138775.1         89           XP         855594.1         9           NP         001001150.1         9           NP         032627.1         9	RVPVSQTP RVPVSQTP 	AINVSVPTTLPSATQVPMEVLKVQTHLENPTKYHIQ AINVSVPTTLPSATQVPMEVLKVQTHLENPTKYHIQ 	132 132 25 25 25	
<u>XP 001065702.1</u> 89	RVAVSQTP	AINVSVPTTLPSATQVPMEVLKVQTHLENPTKYHIQ	132	
<u>NP_990360.1</u> 37 <u>NP_571922.1</u> 78	RVPVSQTP -ISITHSP	ALNVSVPASLPPATQVPMEVLKVQTHLENPTKYHIQ AINVSHPCGPPSAAQVPMEVLKVQTHLENPTKYHIQ	8U 120	

4: return to HomoloGene page and select Show Pairwise Alignment Scores.

Display Alignment Scores	🖌 Show 🕻	20 💌 Send to	*					
All: 1 Fungi: 0 Mammal	Is: 0 🛠							
1: HomoloGene:4892. Gene	conserved in Euteleo	stomi						Linko to poinvico
\lignment Scores								nrotoin PLAST
Gene		Identity (%) Substitution Rates <sup>1</sup>				//	protein BLAST	
Species	Symbol	Protein	DNA	d	$d_{\rm H}/d_{\rm S}$	$d_{\rm HR}/d_{\rm HC}$		Alignment
Homo sapiens	MITE						▶ //	
vs. Pan troglodytes	MITE	99.6	99.6	0.005	0.111	0.000	Blast	
vs. Canis lupus familiaris	MITE	97.3	93.5	0.068	0.050	0.695	Blast	
vs. Bos taurus	MITE	97.0	93.1	0.073	0.042	0.296	Blast	
vs. Mus musculus	Mitf	93.8	88.1	0.130	0.051	0.834	Blast	
vs. Rattus norvegicus	Mitf	94.0	89.5	0.113	0.063	0.632	Blast	
vs. Gallus gallus	MITE	91.9	83.8	0.182	0.031	0.399	Blast	
vs. Danio rerio	mitfb	77.0	73.0	0.335	undef	0.452	Blast	
Pan troglodytes	MITE							
vs. Homo sapiens	MITE	99.6	99.6	0.005	0.111	0.000	Blast	
vs. Canis lupus familiaris	MITE	96.8	93.6	0.067	0.060	0.496	Blast	
vs. Bos taurus	MITE	97.0	93.1	0.072	0.042	0.297	Blast	
vs. Mus musculus	Mitf	93.4	87.9	0.132	0.054	0.740	Blast	
vs. Rattus norvegicus	Mitf	93.7	89.2	0.116	0.064	0.579	Blast	
vs. Gallus gallus	MITE	91.5	83.8	0.183	0.034	0.376	Blast	
vs. Danio rerio	mitfb	76.6	72.4	0.344	undef	0.501	Blast	
Canis lupus familiaris	MITE							
vs. Homo sapiens	MITE	97.3	93.5	0.068	0.050	0.695	Blast	
vs. Pan troglodytes	MITE	96.8	93.6	0.067	0.060	0.496	Blast	
vs. Bos taurus	MITF	97.8	93.4	0.069	0.028	0.426	Blast	
vs. Mus musculus	Mitf	93.1	87.6	0.136	0.054	0.642	Blast	
vs. Rattus norvegicus	Mitf	92.3	86.8	0.146	0.059	0.731	Blast	
vs. Gallus gallus	MITE	91.6	83.1	0.191	0.032	0.617	Blast	
vs. Danio rerio	mitfb	77.9	71.7	0.355	undef	0.521	Blast	

- **d**: the number of nucleotide substitutions per site, corrected for multiple substitutions using the method of Jukes and Cantor (1969).
- d<sub>N</sub>/d<sub>s</sub>: the ratio of the rate of nonsynonymous substitutions (d<sub>N</sub>) to the rate of synonymous substitutions(d<sub>s</sub>), calculated using the method of <u>Nei and Gojobori</u> (<u>1986</u>). A high value of this metric indicates adaptive selection, whereas a low value indicates purifying selection.
- d<sub>NR</sub>/d<sub>NC</sub>: the ratio of radical nonsynonymous substitutions (d<sub>NR</sub>) to conservative nonsynonymous substitutions (d<sub>NC</sub>), calculated using the method of <u>Hughes et al.</u> (<u>1990</u>). This metric is analogous to d<sub>N</sub>/d<sub>S</sub>, but it has the advantage of being useful for studying the evolution of sequences that diverged in the distant past.

#### b) The Ensembl genome browser

- 1. From the Ensembl homepage select the **Zebrafish** genome browser.
- 2. Search *e*! zebrafish gene: **eng2b** and press go.

#### 3. Select Ensembl Gene: ENSDAR00000038868

This should bring you to the GeneView page for the eng2b gene:

#### Zebrafish Workshop

Done

Done

000		Ensembl genom	e browser 54: D.reri	o - Gene summary	/ – Gene: eng2b (EN	ISDARG00000038	868)	$\bigcirc$
	$\times$ $\bigcirc$ (	Attp://www.ensembl.	org/Danio_rerio/Gene	/Summary?g=ENSD	ARG00000038868	<b>N</b> 1	🗘 🔻 🕞 🕻 Google	٩)
Most Visited - Gettin	g Started Lates	t Headlines ৯ Apple - A	mazon eBay Yahoo	o! News ≠				
<i>C</i> Ensemb	d -						<b>2</b> .	Q
Home > Zebrafish Location: 2:27,938,253-27,	941,786 Gene: eng	g2b Transcript: eng2b					Account   Logout   BLAST	7/BLAT   BioMart   Docs & FAQs
Gene: eng2b	Gene: eng2b (E	NSDARG00000038868)						
Gene summary     Supporting evidence     Sequence     External references (0)     Regulation     Comparative Genomics	Homeobox protein of Location Transcripts	engrailed-2b (Zf-En-1)(Eng3) <u>Source</u> <u>Chromosome 2: 27,938,253-27,1</u> There is 1 transcript in this gene: Name Trans eng2b <u>ENSDARTO</u>	E UniProtK8/Swiss-Prot P31533 341,786 forward strand. hide transcripts cript ID Prote 2000056748 ENSDARPOO	ein ID Descri 000056747 protein_	ption			
- Genomic alignments (				Gene s	ummary help			Splice variants »
Gene Tree (alignm)     Orthologues (12)     Paralogues (7)     Protein families (1)     Genetic Variation     Variation Table     Variation Table     Variation mage     External Data	Name Synonyms Gene type Prediction Method	eng2b (ZFIN) en3, eng3, engrailed 3 [To view all Known protein coding Transcripts were annotated by th Transcripts	Ensembl genes linked to the name e Ensembl g <u>enebuild</u> .	click here.]				
ID History     Gene bistory		۲.			23.53 Kb		Forward stra	nd 💻
Gene history     Serie history     Serie history     Serie history     Personal annotation		Ensembl gene	27,930,000	27,934,000	27,938,000 eng2b >	27,942,000	27,946,000 27,9	950,000
<ul> <li>Configure this page</li> <li>Add custom data to page</li> <li>Export data</li> </ul>			Contigs Ensembl gene			< EX 51036 0.8	oding Ensembligene	< cnpy1 Known protein
<ul> <li>Bookmark this page</li> </ul>		-	27,930,000 Reverse strand	27,934,000	27,938,000 23.53 Kb	27,942,000	27,946,000 27,9	950,000
							Export image	
		Tip: use the "Configure t	play his page" link on the left t	o show additional data ir	this region.			
	Ensembl release 54 - Permanent link - View	May 2009 © <u>WTSI</u> / <u>EBI</u> v in archive site						About Ensembl I Contact Us I Help

# 4. Click on the Protein Family hyperlink on the left side of the page:

000	Ensemb	I genome browser 54: D.rerio – Protein I	ramilies – Gene: eng2b (ENSDARG000000	138868)	0
	X n Attp://www.	ensembl.org/Danio_rerio/Gene/Family?g=E	NSDARG00000038868		ioogle Q
Most Visited - Getting	g Started Latest Headlines ని Ap	ple - Amazon eBay Yahoo! News -			
PEnsemb				2.	Q
Home > Zebrafish	141.786 Gene: eno2b Transcrint: eno2	n		Account   Logou	it   BLAST/BLAT   BioMart   Docs & FAQs
Gene: eng2b	Gene: eng2b (ENSDARG000000	38868)			
Gene summary     Splice variants (1)     Supporting evidence     Sequence     External references (0)     Regulation     Comparative Genomics	Homeobox protein engrailed-2b (Zf-En-1)( Location Chromosome 2: 27,9) Transcripts There is 1 transcript i Name eng2b	Transcript ID         Protein ID           Transcript ID         Protein ID           ENSDART00000056748         ENSDARP00000056747	Description protein_coding		
Gene Tree	« Paralogues	1	Protein families help		Variation Table >
Gene Tree (algom     Orthologues (12)     Paratogues (7)     Protein Families (1)     Variation Table     Variation Table     Variation Table     Variation Table     Genetic Variation     Genetic Variation     Orthologues     ID History     Genetistory     User data     Personal annotation     Configure this page     Export data     Bookmark this page	Eamily ID ENSFM0050000270570 (4 genes) (all proteins in family) Ensemb release 54 - May 2009 © <u>WTS</u> / <u>EB</u> Permanent Ink - View in archive site	Consensus annotation RECNAME: FULL=HOMEOBOX ENGRAILED	Other Zebrafish transcripts in this	family b)	Multiple alignments 39 Ensemb members of this family <u>JulVew</u> 94 members of this family <u>JulVew</u> About Ensemb  ContactUs  Heb

Click on the 4 genes link to bring up family view.



Ensembl '<u>FamilyView</u>' provides a list of closely related Ensembl gene predictions together with a consensus family description and shows the chromosomal location of family members on a karyotype ideogram. It also provides a list of vertebrate <u>UniProt</u> sequences and Ensembl protein predictions from other species that have been used to define the family. It therefore provides a way of exploring orthologues and closely related homologues across a range of animal species.

Putative paralogues are also listed in the hyperlinks on the left side of the page – Paralogue prediction.

Genomic alignments (								
Gene Tree	« Ortholo	gues	Paralogues help					
- Gene Tree (text)	-							
Gene Tree (alignme	The following gene	e(s) have been identified as putative paralogues (within species):						
<ul> <li>Orthologues (12)</li> </ul>	Taxonomy Leve	el Gene identifier						
Paralogues (7)     Protein families (1)     Genetic Variation	Euteleostomi	ENSDARG0000026599 (eng2a) (Mult:ContigView) [Align] Homeobox protein engrailed-2a (ZI-En-2)(Eng2) [Source: UniProtKB/Swiss-Prot (P090) [Target %id: 67; Query %id: 68]	5))					
Variation Table     Variation Image     External Data		ENSDARG0000015073 (eng1b) [MultiContigView] [Align] hypothetical protein LOC541371 [Source:RefSeq peptide;Acc:NP_001013516] [Target %id: 47; Query %id: 49]						
- ID History - Gene history - Liser data		ENSDARG00000074916 (B8A685_DANRE) [MultiContigView] [Align] Novel protein similar to vertebrate engrailed homolog 1 (Zgc:109892, EN1) Fragment [Target %id: 76; Query %id: 31]	Source: UniProtKB/TrEMBL (B8A685)]					
Personal annotation		ENSDARG0000014321 (eng1a) (MultiContiqView) [Align] Homeobox protein engrailed-1a (Eng1) [Source: UniProtKB/Swiss-Prot (Q04896)] [Target %id: 51; Query %id: 45]						
Add custom data to page	Bilateria	ENSDARG0000007891 (meox1) [MultiContigView] [Align] mesenchyme homeobox 1 [Source:RefSeq peptide;Acc:NP_001002450] [Target %id: 16; Query %id: 15]						
Bookmark this page		ENSDARG0000061818 (NP_001038589.1) [MultiContigView] [Align] hypothetical protein L0C566969 [Source:RefSeq peptide;Acc:NP_001038589] [Target %id: 14; Query %id: 16]						
		ENSDARG00000040911 (LOC556898) [MultiContigView] [Align] No description [Target %id: 12; Query %id: 14]						
	View sequence alig	gnments of all homologues.						

The gene orthology and paralogy predictions are generated by a pipeline where maximum likelihood phylogenetic gene trees (generated by PHYML) play a central role. They aim to represent the evolutionary history of gene families, i.e. genes that diverged from a common ancestor. These gene

trees reconciled with their species tree (using RAP) have their internal nodes annotated to distinguish duplication or speciation events.





Click on view fully expanded tree, under view options to see the full image.

Gene TreeView represents the evolutionary history of gene families.

The Gene Tree displays the maximum likelihood phylogenetic tree (using PHYML) representing the evolutionary history of gene families. Red squares show duplication and blue show speciation. The gereen bars to the right are a representation of the multiple alignment of peptides made using MUSCLE. Full boxes indicate matches/mismatches, open boxes indicate gaps in the alignment. Both the image and the alignment can be dumped.

a	Ensembl genome browser 54: D.rerio - Gene Tree - Gene: eng2b (ENSDA	RG0000038868) - Iceape	_ <b>D</b> X
Elle Edit View Go Bookm		00000000000000000000000000000000000000	A a
A Home Bookmarks	Red Hat Network $\$ Training $\$ Documentation $\$ Support $\$ Software $\$ Developers $\$ Embedd	ded & Search & Hardware & Downloads & Shop & News	0
Gene Tree (alignment			<b></b>
Paralogues (7) Protein families (1)	C REOK 1 Onyzias latipes		
Genetic Variation     Variation Table	MEOX1 Takfugu rubriges		
Variation Image     External Data	a meax 1 Danio rerio		
E- ID History	ENSTGUG0000002388 Taeniopygia guttata		
User data     Personal annotation	C ENSIGUIUUUU 15769 Taenopygia guttata		
Configure this page	C MEOX1 Anolis carolinensis		
Export data     Bookmark this page	MEOXI Sorex araneus		
	MEOX1 Tursiops truncatus		
	D NP_001030453.1 Bos taurus		
	□ XP_001491616.1 Equus caballus		
	MEOX1 Canis familiaris		
	r MEOX 1 Peris Caus		
	HEOK1 Ochotona princeps		
	MEOX1 Cavia porcellus		
	MEOX1 Homo sapiens		
	C MEOX1_PANTR Pan troglodytes		
	MEDX1_PONP' Pongo pygmaeus		
	b MEOX1 Macaca mulatta		
	MEUX1 Utotemur gametti MEUX1 Spermophilus tridecemlineatus		
	Meax1 Mus musculus		
	□ ENSOCUG0000007860 Oryctolagus cuniculus ■ D MEOX1 Loxodonta africana		
	MEOX1 Procavia capensis		
	D MEOX 1 Pteropus vampyrus		
	P MEOX1 Directors europeeus		
	MEOX1 Tarsius syrichta		
	CI MEOX1 Tupaia belangeri		Ê
	MEDX1 Monodelphis domestica		
	MEOX2 Anolis carolinensis	18 8 1	
	MEOX2 Tak flugu rubripes     MEOX2 Tetraodon nigroviridis		
	NP_001038589.1 Danio rerio		
	D MEOX2 Gasterosteus a culeatus		
	LOC556898 Danio rerio		
	CI MEDX2 Tarsius syrichta		
	MEOX2 Proceedings that must     MEOX2 Dasypus novem cinctus		
	MEDX2 Procavia capensis		
	D MEDX2 Tupala belangeri D MP_0010915141 Bos taurus		
	MEOX2 Vicugna pacos		
	HEOK2 Tupaia belangeri <sup>503</sup>		
	t MEOX2 Canis familiaris		
	D MEOX2 Pheropus vampyrus		
	MEOK2 Pan troglodytes		
	MEOX2 Homo sapiens		
	11 MEDX2 Pongo pygmaeus 15 MEDX2 Macaca mulatta		
	MEOX2 Erinaceus europaeus		
	Meox 2 Mus musculus		
	Meax 2 Dipodo mys ordii		
	D MEOX2 Spermophilus tride camlineatus		
	G MEOX2 Ochotona princeps		
	MEOX2 Monodelphis domestica		
	HEOX2 Taeniopygia guttata		
	MEOX2 Xenopus tropicalis     Discretification of the statement of the		
	Q 4H363_CIOIN Ciona intestinalis		
	Dibth Drosophila melanogaster		
	AGAP00 3406 Anopheles gambiae     AGAP00 3406 Anopheles gambiae     AGAP00 3406 Anopheles gambiae		
	CG15696 Drosophila melanogaster		
	l ם R06F6.6 Caenorhabditis elegans רים ENSTNIG000000 18144 Tetraodon nigroviridis		
	G ENSTRUG00000011240 Tak Éugu rubripes		
	CI ENSORLG0000008839 Oryzlas latipes     CI eng2b Danio rerio		
			-11-6



#### Orthologue prediction at Ensembl

This is one of the easiest and most accurate sites for identifying orthologous genes. Information is only available for species that have a sequenced and assembled genome, so if you want to identify sequences from other species you will need to use Blast.

Search for fgf8a:

PEnsomh	1			2 -	Q
Home > Zebrafish	*: 			Account   Logout   BLAST/BLA	T   BioMart   Docs & FAQs
Location: 13:33,446,378-33	,452,845 Gene: fgf8a	Transcript: fgf8a			
Gene: fgf8a	Gene: fgf8a (ENSD	ARG0000003399			
Gene summary     Splice variants (1)     Supporting evidence     Sequence     External references (0)     Regulation     Comparative Genomics	fibroblast growth factor i Location <u>Chro</u> Transcripts Ther	8 a [Source:RefSeq peptie processories 13: 33,446,376 re is 1 transcript in this gr Name Tra fgf8a ENSDAR	Acc:NP_S71356] 31.452.865 reverse strand. e: hide transcripts script ID Protein ID Description 000000255883 ENSDARP00000018653 protein_coding		
- Genomic alignments (			Gene summary he!p		Splice variants »
Gene Tree (alignm     Orthologues (29)     Paralogues (4)     Protein families (1)     Genetic Variation     Variation Table     Variation Image     External Data	Name fof8 Synonyms ace, Gene type Know Prediction Tran Method Tran	ia (ZFIN) acerebellar, cb110, etID wn protein coding iscripts were annotated b nscripts	9886.13, fgf-8, fgf8, iddibd5031, wurfb73a06 [To view al Ensembligenes linked to the name <u>click her</u> the Ensembli <u>genebuild</u> .	٤)	
Hollstory     Gene history     Gene history     User data     Personal annotation     Configure this page     Add custom data to     page     Export data     Bookmark this page		Contigs Ensembl gene	33.44 Mb         33.45 Mb         33.45 Mb         33.45 Mb           33.44 Mb         33.45 Mb         33.45 Mb         33.45 Mb           33.44 Mb         33.45 Mb         33.45 Mb         33.45 Mb           33.45 Mb         33.45 Mb         33.45 Mb         33.45 Mb           33.45 Mb         33.45 Mb         33.45 Mb         33.45 Mb	Forward strand provided	
		Tip: use the "Configu	this page" link on the left to show additional data in this region.		
	Ensembl release 54 - May 2 Permanent link - View in an	2009 © <u>WTSI</u> / <u>EBI</u> chive site		Ab	out Ensembl I <u>Contact Us</u> I <u>Helo</u>

Click on the orthologues link in the left hand menu.

<ul> <li>Genomic alignments (0)</li> </ul>					
Gene Tree     Gene Tree (text)	« Gene Tree				Orthologues help Para
<ul> <li>Gene Tree (alignment</li> <li>Orthologues (29)</li> </ul>	The following gene(s) have been identified as put (N.B. If you don't find a homologue here, it may I	ative orthologue	is: pecies par	alogue". Please view the gene tre	e info or export between-species paraloques with BioMart to see more.)
- Paralogues (4)	Species	Туре	dN/dS	Ensembl identifier	External ref.
Protein families (1)     Genetic Variation	Anole Lizard (Anolis carolinensis)	1-to-many		ENSACAG00000011928 Target %id: 0.3; Query %id: 00 (Alian)	FGF8 Plorobiast growth factor 8 Precessor (PS R 8)(Hepatin blinding growth factor 8)(HBG R 8)(Androgen indeced growth factor)(AIG P) (Source: UniPro1KB)Sniss-Pro1; acc: <u>05507.9</u>
Variation Image	Cow (Bos taurus)	1-to-many		ENSBTAG0000001530 Target Nid: 67; Query Nid: 69 (Alian)	Q863D7_BOVIN Abrobiat growth factor & Pragment (Source: UniProt KB(trEMBL; acc <u>Q863D7</u> )
ID History	Dog (Canis familiaris)	1-to-many		ENSCAFG0000009918 Target Nid: 68; Query Nid: 71 (Align)	FGF8_CANFA Hbrobilist growth factor 8 (P6 F-8) (H60 F-8) (Androgen-Indeced growth factor) (Al6P) (Pregnent), (Source: UniProf/K8);Soiss-Prof, acc: <u>062683</u>
User data	Guinea Pig (Cavía porcollus)	1-to-many		ENSCP 0 G00000021533 Target Nid: 72; Query Nid: 63 (Align)	FGF8 Mbrobiast growth factor & Precusor (PDF &)Oreparin binding growth factor &)(HBDF &)(Androgen-indeced growth factor)(ADP) (Source: UniProtKB)Swiss-Prof, acc: PSS07.9
Configure this page	Ciona intestinalis	1-to-many		ENSCING000000302 Target Nid: 10; Query Nid: 15 (Align)	NP_001027648.1 Revolusing rowth Sector 0/17/10 (Source: RefSeq_pepfide; acc: <u>NP_001027648</u>
Add custom data to page Export data	Ciona savignyi	1-to-many		ENSCSAVG00000011185 Target Skid: 14; Query Skid: 27 (Align)	Q86GG2_CIOSA Mbroblast growth factor 017710 isoform 1. (Source: UniProfikButrEMB1; acc: <u>0957940</u>
Bookmark this page	Horse (Equus caballus)	1-to-many		ENSECAGODODOD10125 Target Skid: 47; Query Skid: 46 (Align)	FGF8 Fibroblast growth factor 0 Precessor (F0 F 0)01eparity binding growth factor 0)(H00 F 0)(Androgen induced growth factor)(AIOT) (Source: UniProtK0)Source Prot, acc. <u>PSS07.9</u>
	Cat (Felis catus)	1-to-many		ENSECAG0000007177 Target Skid: 55; Query Skid: 53 (Align)	FGF8 Rbroblast growth factor 8 Precessor (F0F8)(Heparin-binding growth factor 8)(H80F8)(Androgen-induced growth factor)(A0F) (Source: UniProtK8)(Source Prot, acc. <u>PS5073</u>
	Chicken (Gallus gallus)	1-to-many		ENSGAL G0000007706 Target Sid: 76; Query Sid: 61 (Align)	FGP8_CHICK Ribroblist growth factor 8 precessor (F0F-8) (H80F-8). (Source: UniProfK8)Sniss-Prot, acc. <u>090722</u> )
	Stickleback (Gasterosteus aculeatus)	1-to-1		ENSGAC G0000003803 Target Sid: 61; Query Sid: 63 (Align)	FGP8 (1 of 2) Ribroblast growth factor 8 Precessor (FGF 8)(Heparin binding growth factor 8)(HBGF 8)(Androgen indeced growth factor)(AIGP) (Source: UniProfKB)Swiss Prof, acc: <u>P5507.3</u>
	Gorilla (Gorilla gorilla)	1-to-many		ENSGG0G0000015108 Target Sid: 25; Query Sid: 16 (Align)	FGF8 Rbroblast growth factor 8 Precessor (F6F 8)(Heparin blinding growth factor 8)(H86F 8)(Androgen indeced growth factor)(AIGF) (Source: UniProtK8)Swiss Prot, acc. <u>P5507.9</u>
	Human (Homo sapiens)	1-to-many		ENSCO0000107831 Target %id: 65; Query %id: 75 (Alian)	FGP8 Ribrobiast growth factor 8 Precessor (PGF 8)(Heparin binding growth factor 8)(HBGF 8)(Androgen indeced growth factor)(AIGP) (Source: UniProtKB(Sixiss: Prot) acc. <u>P5507.9</u>
	Macaque (Macaca mulatta)	1-to-many		ENSMMUG00000015771 Target %id: 73; Query %id: 46 (Alian)	FGF8 Hibblish growth factor 8 Precessor (F6F8)(Heparin binding growth factor 8)(H86F8)(Androgen indeced growth factor)(AI6P) (Source: UniProfit B(Swiss-Prof; acc: <u>P55079</u>
	Mouse Lemur (Microcobus murinus)	1-to-many		ENSMIC G00000011733 Target Nid: 65; Query Nid: 75 (Alian)	FGF8 Hbrobist growth factor 8 Precessor (FSF 8)(Hepath-binding growth factor 8)(H86F 8)(Androgen-indeced growth factor)(A06P) (Source: UniProfix 8)Swiss-Prof; acc: <u>P5507.9</u>
	Opossum (Monodelphis domestica)	1-to-many		ENSMODG0000011718 Target Nid: 50, Query Nid: 72 (Align)	FGR8 Hibbits growth factor 8 Precessor (F6 F 8)(Hepath-binding growth factor 8)(H86 F 8)(Androgen-indeced growth factor)(A0 P) (Source: UniProfix 8)Swise Prof. acc: <u>P5507.9</u>
	Mouse (Mas mascans)	1-to-many		Target Nid: 56; Query Nid: 71 (Align)	FgT8 Hbrobiad growth factor 6 Precessor (PD F 8)(Heparin-binding growth factor 8)(H80 F-8)(Androgen-indeced growth factor)(AIOP) (Source: UniProfix B)Swise Prof. acc: P372377
	Microbal (Myotis Ricitigus)	1-to-many		Target Slid: 66; Query Slid: 70 (Align)	FGP8 Hibrobian growth factor 8 Precessor (PD F 8)(0+eparin-binding growth factor 8)(HBO F-8)(Androgen-induced growth factor)(AIO 7) (Source: UniProfix B)(Swise-Prof), acc: P3507.9
	Medaka (Oryzias latipes)	1-to-1		Target Slid: 78; Query Slid: 79 (Align)	FGPB (1 01 2) If broket growth factor & Precursor (TD F 8)0Hepwin-binding growth factor 8)(HBO F 8)(Androgen-induced growth factor)(AIOP) (Source: UniProfxB(Swise-Prof; acc: P5507.9) International State (State State Stat
	Chimpanzee (Pan troglodytes)	1-to-many		Target Sid: 61; Query Sid: 60 (Align)	FGP8 Hibrobiast growth factor & Precessor (FD F 8)(Heparin binding growth factor 8)(HBD F 8)(Androgen induced growth factor)(AID F) (Source: UniProtKB(Swise Prot, acc. <u>P5507.9</u>
	Orangutan (Pongo pygmaeus)	1-to-many		Target Skid: 55; Query Skid: 43 (Align)	FGP8 Ribrobiast growth factor & Precessor (FDF 8)(Heparin binding growth factor 8)(HBDF 8)(Androgen indeced growth factor)(AIDF) (Source: UniProtKB(Swise Prot, acc. <u>P5507.9</u>
	hyrax (Procavia capetisis)	1-to-many		Target Sid: 65, Query Sid: 75 (Align)	FLAFD Planta
	Megabat (Preropus Valapyrus)	I-to-many		Target Sid: 55, Query Sid: 62 (Align)	FLGF0 Florb0
	nai (naicus un vegnuls)	1-to-many		Target Sid: 56; Query Sid: 71 (Align)	rgro Ploobart growth factor 8 Precessor (PGF 8)(Heparin binding growth factor 8)(HBGF 8)(Androgen indeced growth factor)(AIGF) (Source: UniProtKB)Swiss Prot, acc: <u>P37237</u> = con
	Zahra Einch / Zaanionunia auttata	1-to-many		Target Taid: 60; Query Taid: 42 (Align)	Honobia growth factor 8 Precursor (FGF 8)(Heparin binding growth factor 8)(HBGF 8)(Androgen-Indeced growth factor)(AIGP) (Source: UniProfixB)Selss-Prob acc: P55073
	Zebra rinch ( <i>raennopygla gittata</i> )	1-to-many		Target Wid: 01; Query Mid: 70 (Align)	r urro Hibobard growth factor 8 Precessor (FGF 8)(Heparin binding growth factor 8)(HBGF 8)(Androgen indeced growth factor)(AIGF) [Source: UniProtKB(Swiss-Prot) acc: <u>P55079</u> FCCP0. (1 = 4.0.)
	Tetranden (Tetranden niemuisidie)	1-10-1		Target Nid: 7+; Query Nid: 60 (Align)	Handbild growth factor 8 Precessor (PDP 8)(Heparis binding growth factor 8)(HBDP 8)(Androgen-indexed growth factor)(AIDP) (Source: UniProtKB)84655-Pro); acc: <u>P51079</u>
	Delabio (Tursiana transatus)	1-to-man		Target Nid: 05; Query Nid: 07 (Align)	Handbild growth factor & Precessor (PDP b)cheparis-binding growth factor 6)(HBDP 6)(Androgen-indeced growth factor)(AIDP) (Source: UniProtKB)SHISSE Pro); acc: PS107.9
	Vananua francisalia	1-to-many		Target Nid: 67; Query Nid: 70 (Align)	Handblad growth factors Precessor (PDP-6)(Hapwis-binding growth factors)(HBDP-6)(Androgen-indeced growth factor)(ADP) [Source: UniProtKB)Swise Prot, Acc. <u>PS102.9</u>
	Manu exemptes disperants of all here the	r=co-many		Target Nid: 81; Query Nid: 81 (Align)	nyno Pafo-prov protein, (Source: UniProfXBUTEMBU; acc: <u>064133</u>
	were sequence and memory or an noniologues.				

Ensembl calculates the closest putative orthologues for species pairs. The stable identifiers of putative orthologous genes lead to the corresponding '<u>GeneView</u>' display of that gene within the web site for the other species. The homologous genes represent the best reciprocal BLAST hits for the two species with additional pairs obtained by a combination of BLAST and location information for more closely related species. These homologues may therefore potentially represent orthologues. Types of orthologous gene pairs are described on the webpage; dN/dS ratios (non-synonymous to synonymous substitutions – an indicator of selective pressure) are also displayed.

Click on align under the human orthologue to show the alignment:

Gene: fgf8a	Gene: fgf8a (ENSDARG0000003399)						
Gene summary     Splice variants (1)     Supporting evidence     Sequence     External references (0)     Regulation     Comparative Genomics     Gene Tree     Gene Tree     Gene Tree (text)     Gene Tree (alignment     Orthologues (29)     Paralogues (4)	fibroblast growth factor 8 a [Source:RefSeq peptide;Acc:NP_571356]       Location     Chromosome 13: 33,446,378-33,452,845 reverse strand.       Transcripts     There is 1 transcript in this gene:						
	Transcripto Transcripto	Name Transcript ID Protein ID Description					
		fgf8a ENSDART00000025583	ENSDARP00000018653	protein coding			
	702						
	« Orthologues Ortholog Alignme						
	Ortholog type: 1 to many orthologue						
	Species		Gene ID		Peptide I		
Protein families (1)	Danio rerio	ENSDARG0000	0003399	ENSDARP0000	0018653		
Genetic Variation	Homo sapiens	ENSG00000107	831	ENSP00000321	1797		
- Variation Table - Variation Image - External Data ⊡- ID History - Gene history ⊡- User data - Personal annotation	CLUSTAL W(1.81) multiple sequence alignment						
	ENSDARP00000018653/1-210 ENSP00000321797/1-244	MRL IPSRLSYLFLHLFAFCYYAQ MGSPRSALSCLLLHLLVLCLQAQEGPGRGP * * ** *: ***: * **	AL GREL ASL FRAGREP QGUSQQQ	JT IQSPPN JTVQSSPN **: **. **			
<ul> <li>Configure this page</li> <li>Add custom data to page</li> <li>Export data</li> <li>Bookmark this page</li> </ul>	ENSDARP00000018653/1-210 FTQHVSEQSKVTDEVSRELIRTYQLYSETSGKHVQVLANKKINAMAEDGDVHAKLIVETD ENSP00000321757/1-244 FTQHVSEQSLVTDQLSELIRTYQLYSETSGKHVQVLANKEINAMAEDGDPFALIVETD						
	ENSDARP00000018653/1-210 TFGSRURIKGAETGFYICMIRRGKLIGKKNGLGKDCIFTEIVLENNYTALQNVKYEGWYM ENSP00000321797/1-244 TFGSRURVRGAETGLYICMIRKGKLIAKSNGKGKDCVFTEIVLENNYTALQNAKYEGWYM *******::******						
	ENSDARP0000018653/1-210 AFTRKGRPRKGSKTRQHQREUHIMKRLPKGHQIAEHRPFDFINYPFINRTKRTR ENSP00000321757/1-244 AFTRKGRPRKGSKTRQHQREUHIMKRLPRGHHTT-EQSLRFEFLNYPFFTRSLRGSQRTW ************************************						
	ENSDARP00000018653/1-210 ENSP00000321797/1-244	YSG-ER APEP-R *					

N.B. Orthologue prediction only allows two aligned sequences to be viewed. Aligned eutherian or amniota genome sequences can be viewed within Ensembl using the MLAGAN parameters. Multiple sequence alignment programmes such as ClustalW should be used to align more than two orthologous nucleotide (mRNA) or protein sequences (covered in modules 2 and 3).

# Confirming true orthology:

Other features of orthologue prediction will now be used to assess the exon/intron structures and chromosomal context of putative orthologues. Orthologous gene pairs often have similar transcript structures. These can be compared visually using AlignSliceView at Ensembl. This is not available in the current Ensembl release, so we will have to use an archive site.

Click back to the Gene summary page for *fgf8a*, and click on the view in archive site link at the bottom of the page and chose ensembl 47.

Ensembl release 47: Danio rerio Gene report for ENSDARG0000003399 - Iceape									
_ Elle Edit View Go Bookmarks Iools Window Help									
	🕟 http://oct2007.archive.ensembl.org/Danio_rerio/geneview?gene=ENSDARG00000003399								
🔟 🏦 Home 🛛 Bookmarks 🛇 Red Hat Network 🛇 Training 🛇 Documentation 🛇 Support 🛇 Software 🛇 Developers 🛇 Embedded 🛇 Search 🗞 Hardware 🛇 Downloads 🛇 Shop 🛇 News									
Archive: Ensembl Zebrafish GeneView									
Ensembl release 47 - Oct 2007				HOME + BLAST + BIOMART + SITEMAP					
Your Archive EnsEMBL	Ensembl Gene Report	Ensembl Gene Report for ENSDARG0000003399							
Login or Register	Gene	fall8 (ZFIN) Synonyms: ace, acerebellar, cb110, et1D309886.13, fgf-8, id3bd5031, wu/tb73a06 to view all Archive ensembligenes linked to the name dick here.							
About User Accounts	Ensembl Gene ID	ENSDARG0000003399							
Gene information	Genomic Location	This gene can be found on Chrom The start of this gene is located in	osome 13 at locatio Contig CR925797	iton 20 <u>586,784-20,573,252</u> . 97.8.					
Genomic sequence     Genomic sequence alignment	Description	floroklast growth factor 8 Source: Refseq_pupide.NP_571356							
Gene splice site image     Gene tree info.	Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWeseExonerate model from a database protein or a set of aligned cDNAs followed by an ORF prediction. GeneWeseExonerate models are further combined with available aligned cDNAs to annotate UTRs (For more information see V Curven et al., Genome Res. 2004 14 942-50)							
<ul> <li>Gene variation info.</li> <li>ID history</li> </ul>	Transcripts	1018 ENSDART00000025583	3	ENSDARP00000018653 [Transcript info] [Exon info] [Peptide info]					
Transcript information     Exon information     Protein information     Export gene data     Chromosome 13		Fe Chr. Len DN/	Features V Chr. 13 Length DNA(contigs)	20.58 Ma 20.57 Ma 20.57 Ma 20.57 Ma 20.57 Ma 20.57 Ma 20.58 Ma 20.57 Ma 20.57 Ma 20.58 Ma 20.57 Ma 20.					
20,566,784 - 20,573,252		E	Ensembl trans.	< fg/8 Ensembl Known Protein Coding					
View of Chromosome 13     Graphical view		ľ	Length	20.58 Mb 20.57 Mb 20.57 Mb 20.57 Mb 20.58 Mb					
Export from region Export Gene info in region Export SNP info in region Ensembl Archive	Orthologue Prediction	The following gene(s) have been id (N.B. If you don't find a homologu Species Bos taurus Canis familiaris	dentified as putative le here, it may be a <b>Type</b> 1-to-many 1-to-many	Ive orthologues: a between-species paralogue. Please view the <u>ourse tree info</u> or export between-species paralogues with BioMart to see more.) <b>Gene Geneticity</b> <b>FIRSTACOMMONITY</b> <b>FIRSTACOMMONITY</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTACTIONALS</b> <b>FIRSTA</b>					
<ul> <li>View previous release of page in Archive!</li> <li>Stable Archive! link for this</li> </ul>	retributs release of page intel Architer link for this Ciona intestinalis Architer link for this Ciona intestinalis 1-to-mary ENSCINC0000003322 (NP_00116276) 1 to-mary		Phenobials around factors of (PF-6) (HBCF-6) (Androgen-indiced growth factor) (ANDP) (Phagnenth, (SourcesUnipro1)(SWISSPR01,Acc:062682) Phage State (SourcesUn						
page		Ciona savignyi	1-to-many	moreousi shortim nachra (na logani kan shorta) papino (mr. 2010/2016) prave tuki na je voji kuki na y ENSCSAV/20000011185 (2086/GG2_CTOSA) pretitionalistvisti plajan					
Sanger емес-ее		Felis catus	1-to-many	Depart Mid: 14 (Sury Mid: 28) 9 ENSECAG00000007172 (FGRB) ( <u>Italificantia Vivia) (Alago</u> ) Physikal growth actore precursor (TO-F6) (HBD-F6) (Androgan: induced growth factory (Alago), [SourceUnipro(ISWISSPROT,Acc.P55073) Physikal growth actore precursor (TO-F6) (HBD-F6) (Androgan: induced growth factory (Alago), [SourceUnipro(ISWISSPROT,Acc.P55073)					
		Gallus gallus	1-to-many	Tevenit Nuk 31; journ Nuk 30) y <u>ENSGAL G00000007706</u> (FGRB_CHICK) <u>Burthconfisitivity (Aligg)</u> Ribroblargrowth actore specurace (TGR-8) (HBGR-8) (Source-Uniport/SWS SPR01,AccQ90722)					
		Gasterosteus aculeatus	1-to-1	Length Hair, I's Very Januar equi ENIS GLA COMD0003883 2 (FGF8 (1 of 2)) <u>(Initianity Height)</u> Physiolast growth factor 8 processor (FGF 8) (HBGF8) (Andregen-induced growth factor) (AGF). [Searce:UniprofitSWISSPR0T;Acc:PS5073] Termot Mid # 24 (Journ Mid # 34)					
		Homo sapiens	1-to-many	Encode and year					
		Macaca mulatta	1-to-many	<u>PINSMMUG0000015771</u> (FGR8) ( <u>tableconfiguley) (Align</u> ) <u>Rhobilist growth factor 5 precessor (FGF-8) (HBGF-8) (Androgen- induced growth factor) (AGF), (SourceUniprot(SWISSPR01;Acc.PS073)             <u>Read Visit 7 SourceVisit 40         </u> </u>					
		Monodelphis domestica	1-to-many	y ENSMOD 600000011718 (FGR8) (with confusitive) (Alex) Physical and a second second and a second					
		Mas muscakis	1-to-many	V ENSINUS G00000025219 (Fgf8) (Inuticontautus) (Alexa)     froncisar provint factors (Source Instant Symbol, Acc: h06:9960-4)     Invest bids 35 pourty bids 70					
		Myotis kıcılagas	1-to-many	y ENSMLUG0000004959 (FGR9) (milliconfolded (king) Physicial another by the prevator (FGF-8) (HBGF-8) (Androgen- induced growth factor) (Aid P), (SourceUniprot(SWISSPR0t;Acc.P55073) Therebilds (Source) wide S9					
		Oryzias latipes	1-to-1	ENSIDEL 00000000319 (FGF8 (1 of 2)) tentfloanfactived (Alian) Phyrobial arowth actors preceiver (FGF8) (HBOF8) (Androyen: Induced growth factor) (Alio PL (SourceUniprofitSWBSSPR0T,AccP55073) (prayshild): 70 count side 30					
		Pan troglodytes	1-to-many	y ENSETTERS000000000000000000000000000000000					

From within orthologue prediction find the human entry and click on multicontigview to view long-range sequence conservation between zebrafish and human.