

# **R4 regulators of G protein signaling (RGS) identify a conserved genomic region that contains MHC-related markers**

**Jaanus Suurväli<sup>1</sup>, Jacques Robert<sup>2</sup>, Pierre Boudinot<sup>3</sup> and Sirje Rüütel Boudinot<sup>1</sup>**

<sup>1</sup>Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

<sup>2</sup> Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY 14642

<sup>3</sup> INRA, Molecular Virology and Immunology, Domaine de Vilvert, 78350 Jouy en Josas, France

Corresponding author:

Jaanus Suurväli

Tallinn University of Technology

Department of Gene Technology

Akadeemia tee 15, 12618 Tallinn, Estonia

phone: +372 620 4447

fax: +372 620 4401

e-mail: jaanus.suurvali@ttu.ee

### Online Resource 3.

This file contains the evidence supporting orthology relationships between the vertebrate RGS1/RGS16 region genes and their putative amphioxus orthologs as shown on Figure 3.

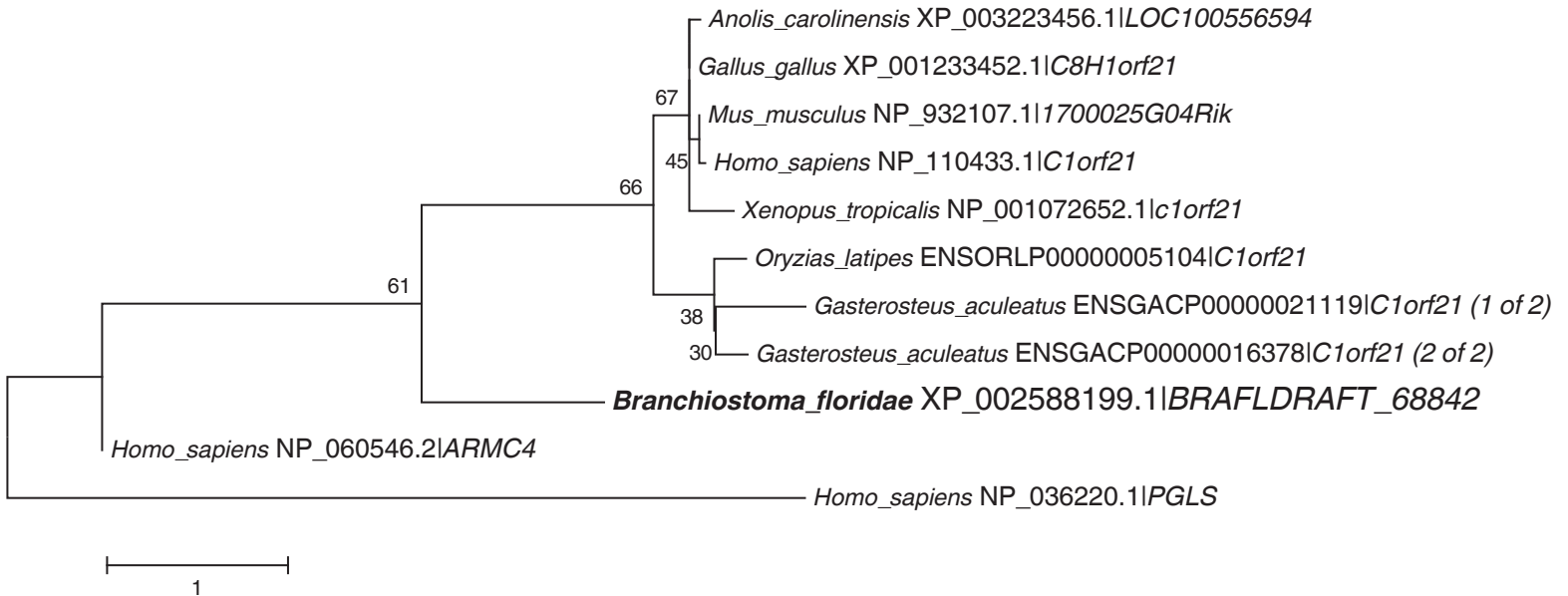
**Online Resource 3A** – Table with an overview of the amphioxus orthologs – such as scaffold they are located on, their EntrezGene ID, E values obtained when blasting these sequences against human and links to online phylogenetic trees in the Phylome Database <http://phylomedb.org/>.

**Online Resource 3B – 3I** contain phylogenetic evidence for RGS1/RGS16 region orthologs for which a phylogeny was not available in the Phylome Database. All of these are Maximum Likelihood trees from 1000 bootstrap replicates (as described in Materials and Methods). Sequences to build the tree were initially retrieved from Genbank (NCBI) using blastp, the dataset was completed with sequences from species such as medaka and stickleback from Ensembl as these were not available in the NCBI database. Amphioxus sequences have been marked with bold on the figures. In several cases, the amphioxus orthologs appear to be in fact co-orthologs of multiple proteins.

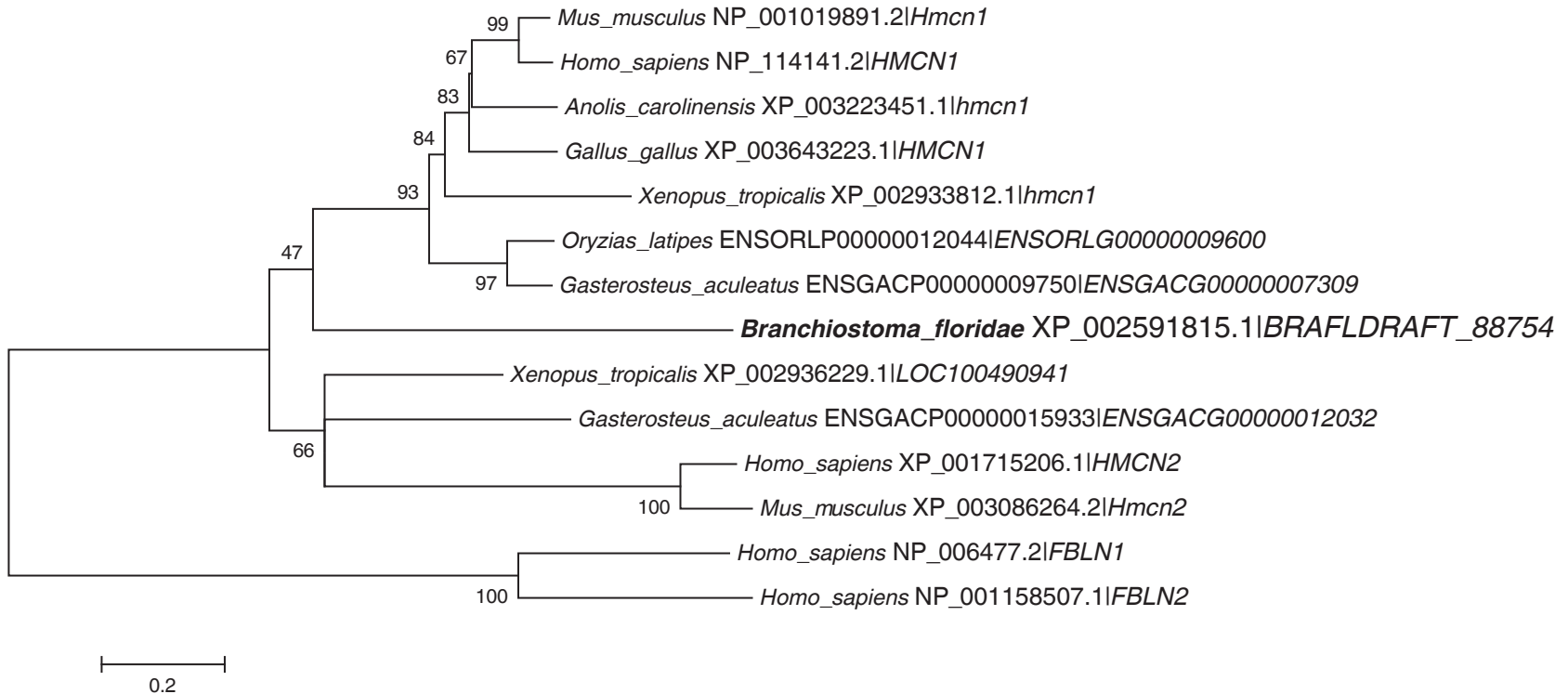
**Online Resource 3J-3K** – Some sequences in amphioxus had a specific domain structure only found in certain proteins – LAP1C domain in vertebrate TOR1AIP proteins or IER domain found in the MHC paralogon markers IER2, IER5 and IER5L. However, the identity of these IER-like and TOR1AIP-like sequences could not be confirmed by phylogenetic analysis. Nevertheless, they were tentatively kept for Figure 3. Online Resource 3J and 3K simply provide the domain structure of these proteins, as analyzed by InterProScan.

**Online Resource 3A.** Amphioxus orthologs of the RGS1/RGS16 region are often located on the same scaffolds as previously identified MHC marker genes (yellow) (Holland et al. 2001; Abi-Rached et al. 2002), and are co-orthologs of several ohnologs. Phylogenetic evidence is shown in the table by either a direct link to the Phylome database <http://phylomedb.org/> or can be found in this file here (Online Resource 3B-3K)

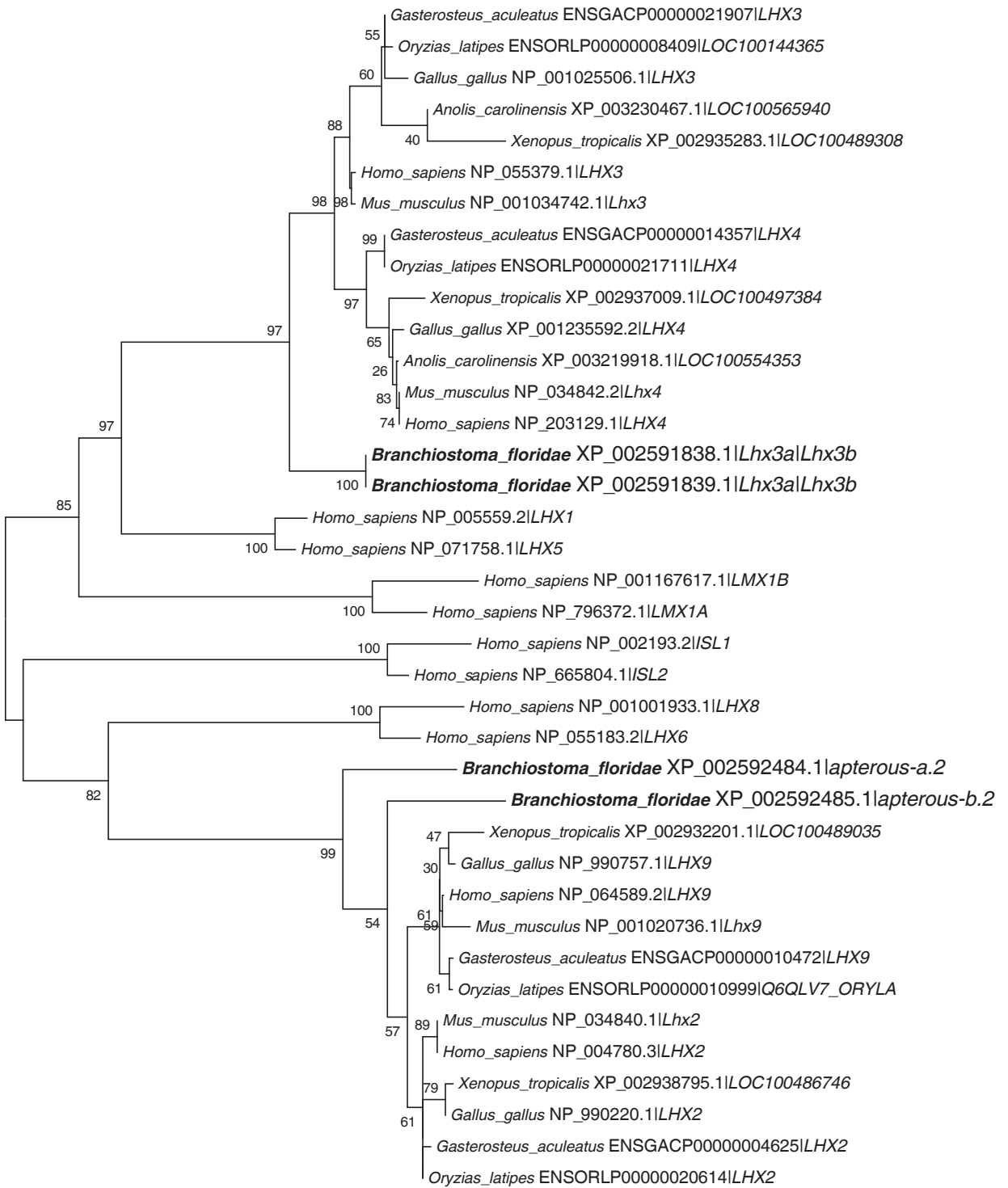
Gene	Amphioxus protein	Gene ID	Scaffold #	Scaffold id	Hits with lowest E values upon running blastp against human	Phylogeny
HMCN1	XP_002591815.1	7246318	209	NW_003101396.1	7e-170 for HMCN1, 4e-101 for HMCN2	Online Resource 3C
ASPM	XP_002591834.1	7246315	209	NW_003101396.1	E = 0 for ASPM	<a href="#">Phy000WXDI</a>
LHX4	XP_002591838.1	7247591	209	NW_003101396.1	1e-144 for LHX3, 2e-136 for LHX4	Online Resource 3D
LHX4	XP_002591839.1	7247591	209	NW_003101396.1	5e-124 for LHX3, 2e-122 for LHX4	Online Resource 3D
QSOX1	XP_002591840.1	7246644	209	NW_003101396.1	5e-135 for QSOX1, 2e-133 for QSOX2	<a href="#">Phy000WXDL</a>
LHX9	XP_002592484.1	7223702	200	NW_003101405.1	1e-117 for LHX9, 8e-115 for LHX2	Online Resource 3D
LHX9	XP_002592485.1	7223703	200	NW_003101405.1	3e-132 for LHX2, 1e-131 for LHX9	Online Resource 3D
CDC73	XP_002592488.1	7221437	200	NW_003101405.1	E = 0 for CDC73	<a href="#">Phy000X62F</a>
TROVE2	XP_002592489.1	7223960	200	NW_003101405.1	E = 0 for TROVE2	Online Resource 3H
UCHL5	XP_002592491.1	7223961	200	NW_003101405.1	3e-148 for UCHL5	<a href="#">Phy000X7E5</a>
ACBD6	XP_002592536.1	7221447	200	NW_003101405.1	1e-85 for ACBD6	<a href="#">Phy000WPXC</a>
STX6	XP_002592538.1	7223711	200	NW_003101405.1	2e-82 for STX6, 3e-74 for STX10	<a href="#">Phy000X62K</a>
ARPC5	XP_002596309.1	7231377	166	NW_003101439.1	7e-57 for ARPC5, 2e-51 for ARPC5L	<a href="#">Phy000X66H</a>
PSMB7, 10	XP_002596413.1	7206715	166	NW_003101439.1	3e-148 for PSMB7, 6e-117 for PSMB10	<a href="#">Phy000X64H</a>
NEK7	XP_002596414.1	7206693	166	NW_003101439.1	E = 0 for NEK7, 5e-176 for NEK6	<a href="#">Phy000X86N</a>
TRMT1L	XP_002593682.1	7227130	192	NW_003101413.1	3e-81 for TRMT1L	<a href="#">Phy000XCB0</a>
GLT25D2	XP_002593693.1	7227132	192	NW_003101413.1	2e-139 for GLT25D1, 2e-135 for CERCAM, 5e-135 for GLT25D2	<a href="#">Phy000X3U2</a>
KCNT2	XP_002593744.1	7242106	192	NW_003101413.1	E = 0 for KCNT1, E = 0 for KCNT2	<a href="#">Phy000WW06</a>
PBX1, 2, 3, 4	XP_002593745.1	7245490	192	NW_003101413.1	2e-106 for PBX3, 4e-103 for PBX1, 1e-98 for PBX2, 7e-94 for PBX4	<a href="#">Phy000X9HH</a>
SWT1	XP_002609338.1	7255400	34	NW_003101539.1	2e-55 for SWT1	Online Resource 3G
IER5	XP_002609351.1	7252638	34	NW_003101539.1	3e-13 for IER5L, 1e-12 for IER5, 2e-08 for IER2	Online Resource 3K
RXRA, B, G	XP_002609360.1	7253104	34	NW_003101539.1	E = 0 for RXRA, E = 0 for RXRG, E = 0 for RXRB	<a href="#">Phy000XKGQ</a>
TOR1AIP1	XP_002609408.1	7245599	34	NW_003101539.1	2e-19 for TOR1AIP2, 5e-19 for TOR1AIP1	Online Resource 3J
DHX9	XP_002609434.1	7242207	34	NW_003101539.1	E = 0 for DHX9, 3e-177 for DHX36, 1e-152 for DHX57	<a href="#">Phy000X9KI</a>
LAMC1	XP_002609437.1	7241772	34	NW_003101539.1	E = 0 for LAMC1, E = 0 for LAMC3, E = 0 for LAMA1	<a href="#">Phy000XJ7L</a>
DDX39A, B	XP_002600114.1	7216951	136	NW_003101469.1	E = 0 for DDX39B, E = 0 for DDX39A	<a href="#">Phy000X70J</a>
TOR1AIP1	XP_002600196.1	7220812	136	NW_003101469.1	2e-21 for TOR1AIP1, 7e-18 for TOR1AIP2	Online Resource 3J
TOR1AIP1	XP_002600202.1	7219896	136	NW_003101469.1	2e-19 for TOR1AIP1, 5e-18 for TOR1AIP2	Online Resource 3J
RGL1	XP_002600257.1	7218548	136	NW_003101469.1	3e-136 for RGL1, 1e-111 for RGL3, 2e-93 for RALGDS	Online Resource 3F
NOTCH1,2,3,4	CAC19873.1	Y12539.2	136	NW_003101469.1	E = 0 for NOTCH1, 2, 3, 1e-143 for NOTCH4	Online Resource 3I
NCF2	XP_002600343.1	7226532	136	NW_003101469.1	1e-109 for NCF2, 5e-45 for NOX1	Online Resource 3E
RGS R4	XP_002588151.1	7220373	261	NW_003101344.1	6e-40 for RGS4, 4e-37 for RGS3, 4e-37 for RGS8	<a href="#">Phy000X7D8</a>
C1orf21	XP_002588199.1	7223664	261	NW_003101344.1	4e-10 for C1orf21	Online Resource 3B



**Online Resource 3B.** Amphioxus sequence XP\_002588199.1 is orthologous to vertebrate C1orf21. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.

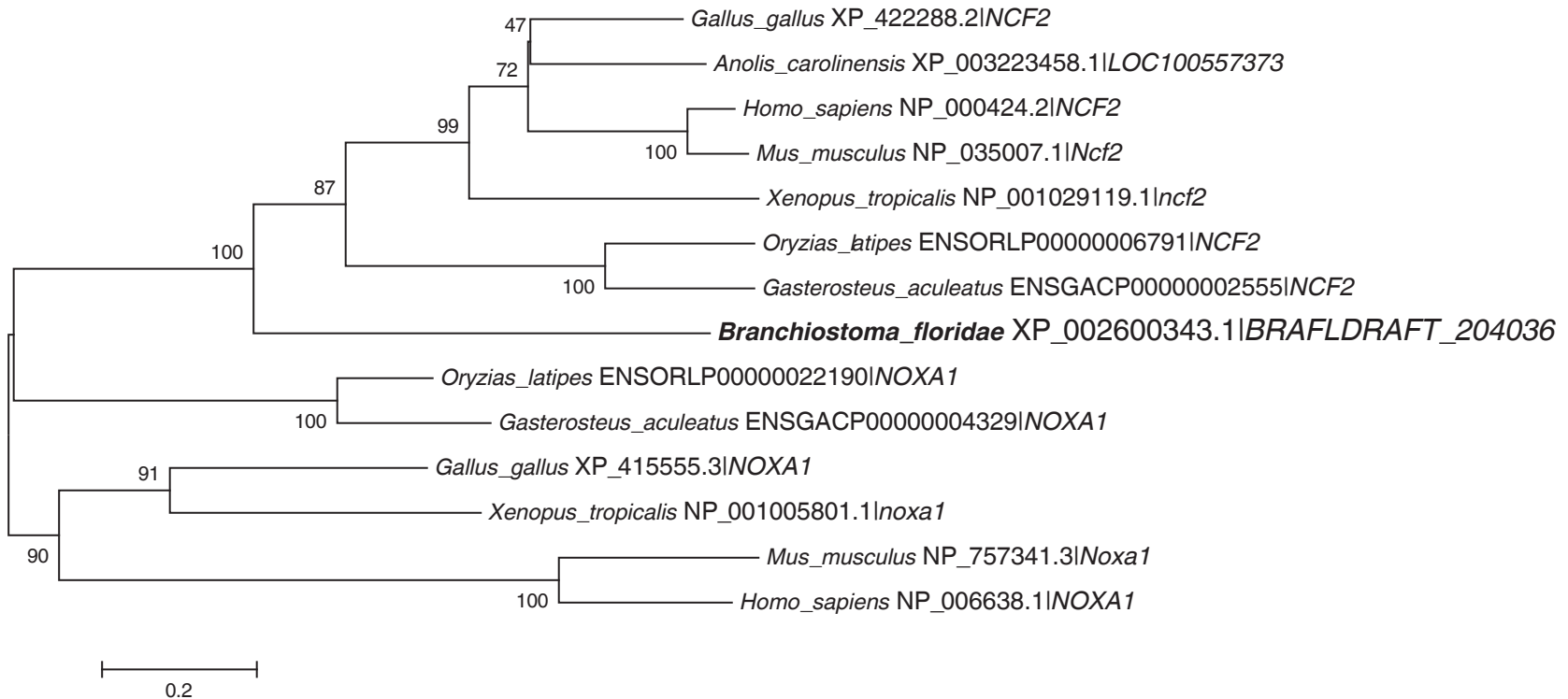


**Online Resource 3C.** Amphioxus sequence XP\_002591815.1 is orthologous to vertebrate HMCN1. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.

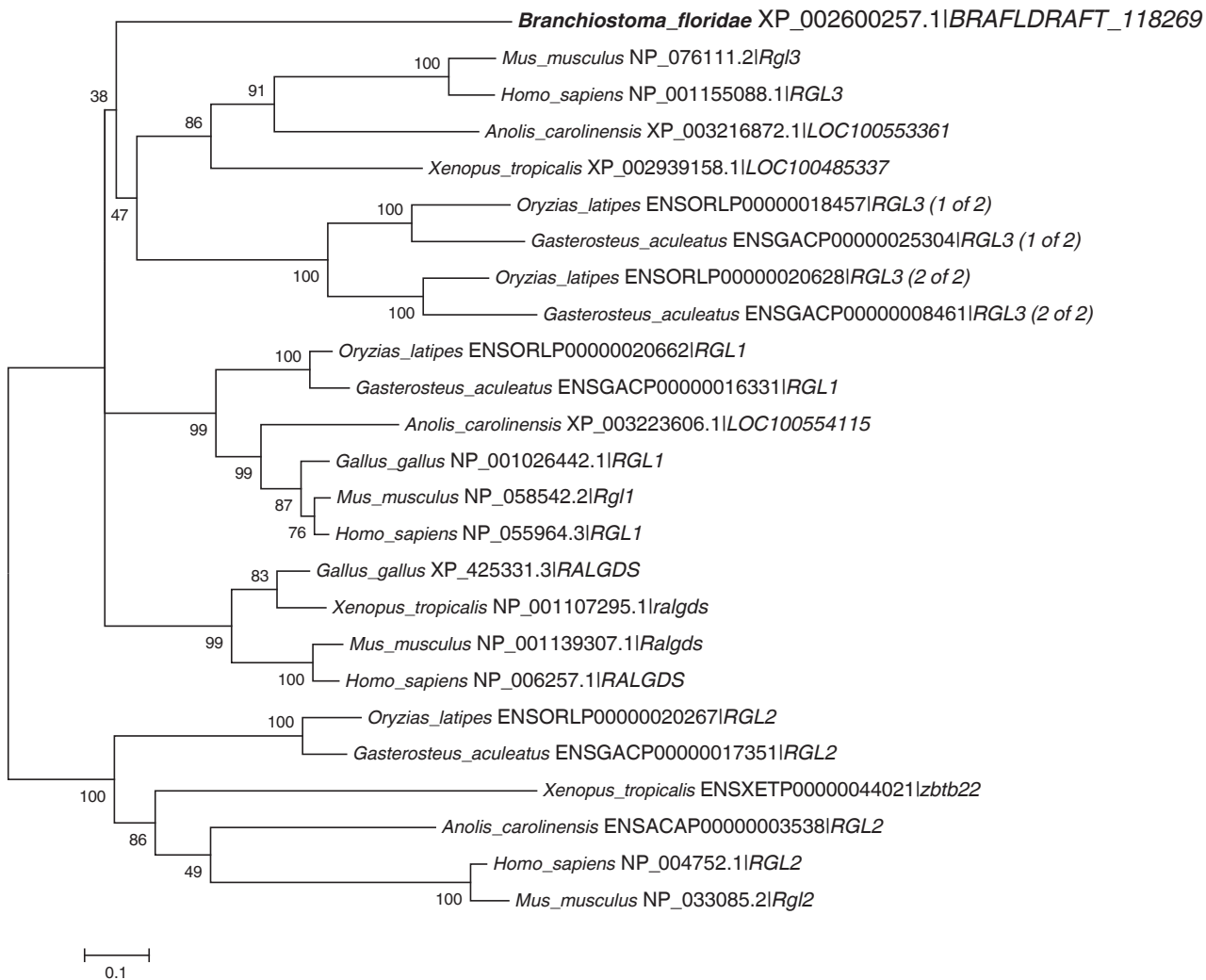


0.1

**Online Resource 3D.** Amphioxus sequences XP\_002591838.1 and XP\_002591839.1 are orthologous to vertebrate LHX4 (and LHX3). Sequences XP\_002592484.1 and XP\_002592484.1 are orthologous to vertebrate LHX9 (and LHX2). This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.

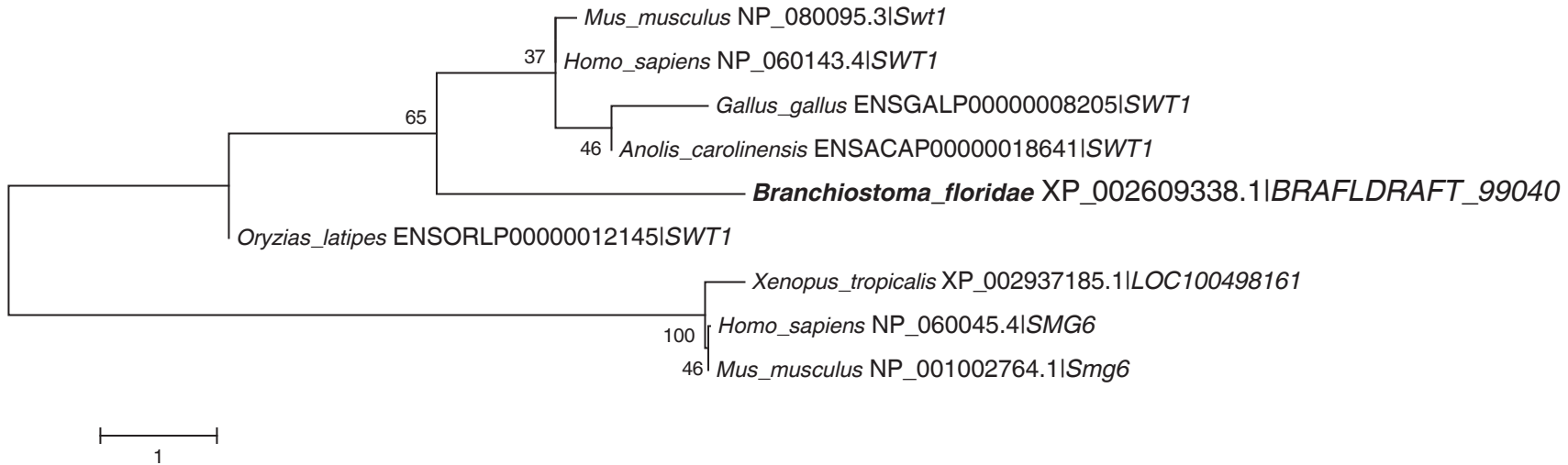


**Online Resource 3E.** Amphioxus sequence XP\_002600343.1 is orthologous to vertebrate NCF2. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.

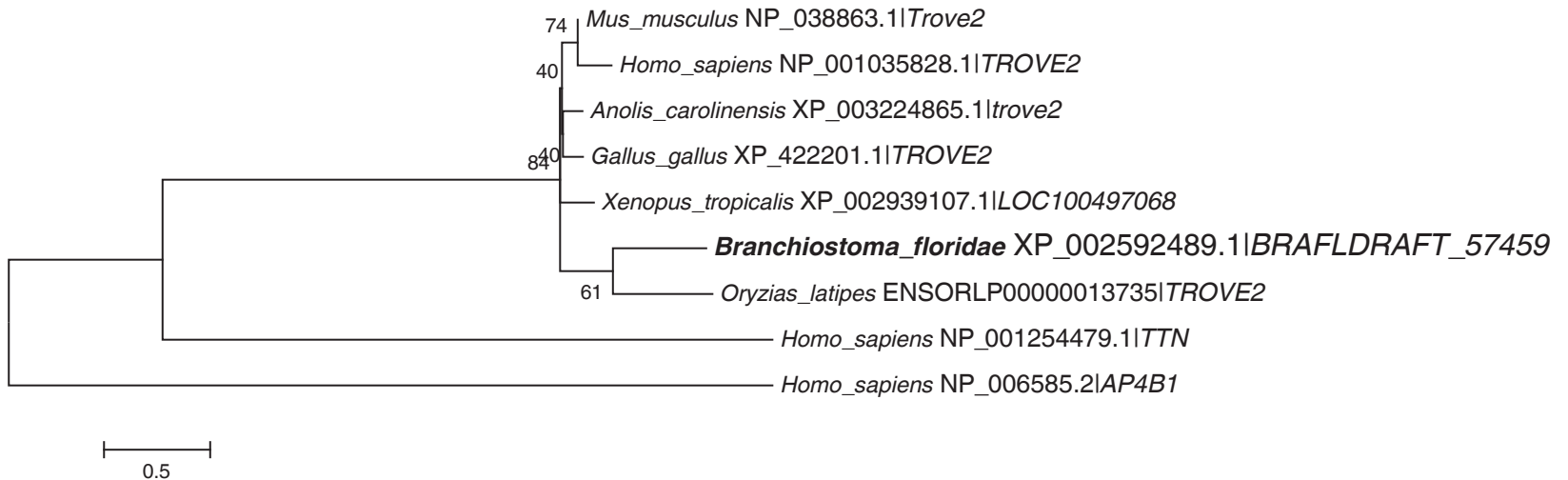


**Online Resource 3F.** Amphioxus sequence XP\_002600257.1 is orthologous to vertebrate RGL1,2,3 and RALGDS proteins. The sequence appears most similar to RGL3 and the tree has expected structure. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.

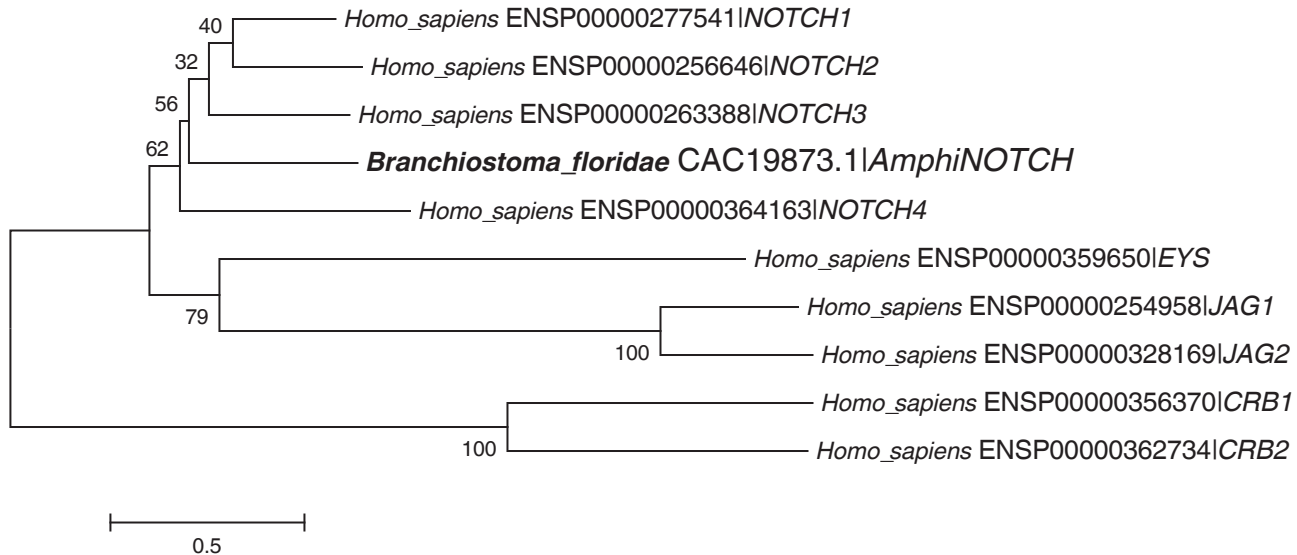




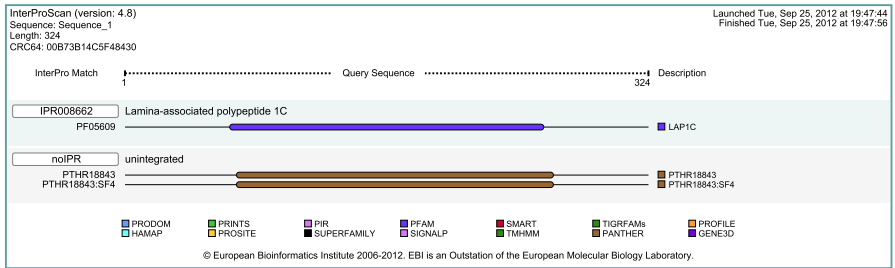
**Online Resource 3G.** Amphioxus sequence XP\_002609338.1 is orthologous to vertebrate SWT1. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.



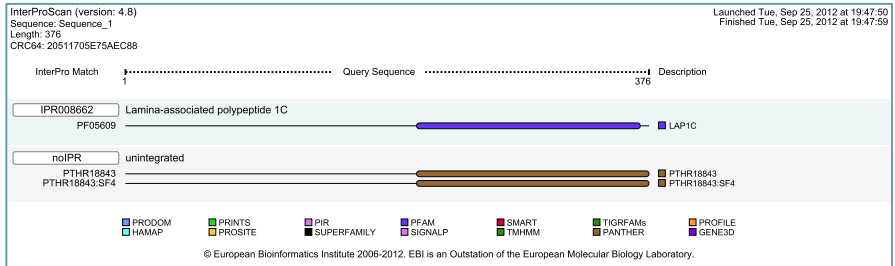
**Online Resource 3H.** Amphioxus sequence XP\_002592489.1 is orthologous to vertebrate TROVE2. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.



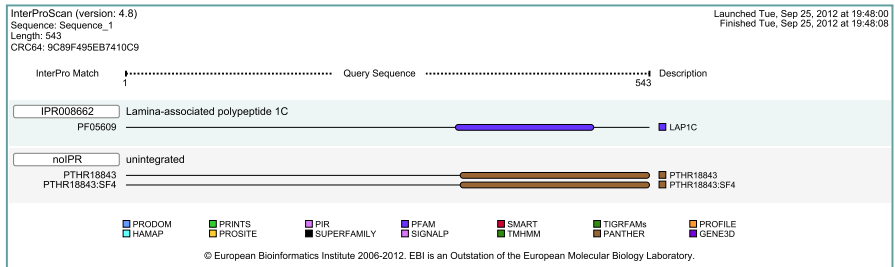
**Online Resource 3I.** Amphioxus sequence CAC19873.1 is related to the human NOTCH proteins. This Maximum Likelihood tree was constructed with 1000 bootstrap replicates.



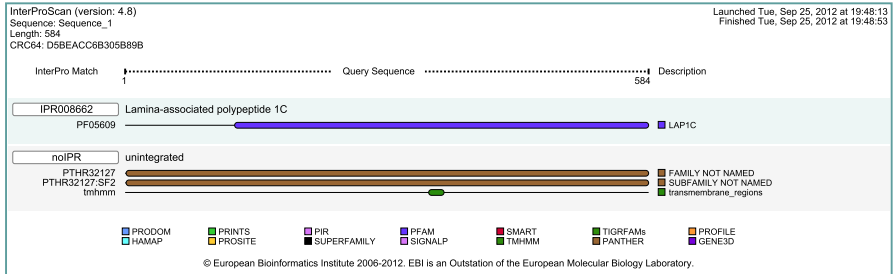
*B. floridae*  
 XP\_002600196.1  
 BRAFLDRAFT\_66701



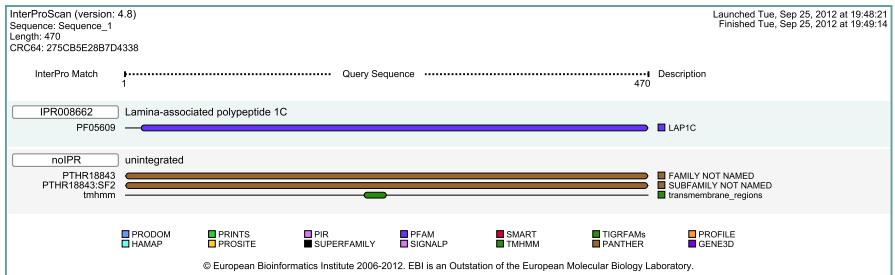
*B. floridae*  
 XP\_002600202.1  
 BRAFLDRAFT\_66707



*B. floridae*  
 XP\_002609408.1  
 BRAFLDRAFT\_86510



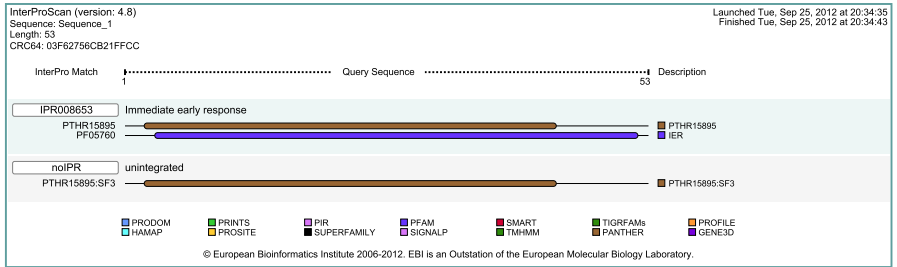
*Homo sapiens*  
 NP\_001254507.1  
 TOR1AIP1



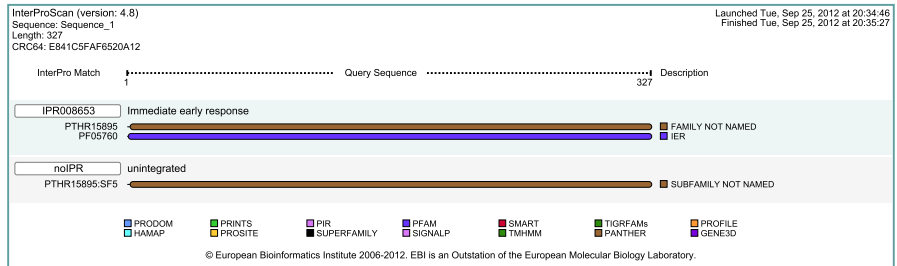
*Homo sapiens*  
 NP\_659471.1  
 TOR1AIP2

**Online Resource 3J.** Amphioxus sequences XP\_002600196.1, XP\_002600202.1 and XP\_002609408.1 share the conserved unique domain structure characteristic to the TOR1AIP proteins of *Homo sapiens* and other vertebrates, although this is not supported by phylogenetic evidence.

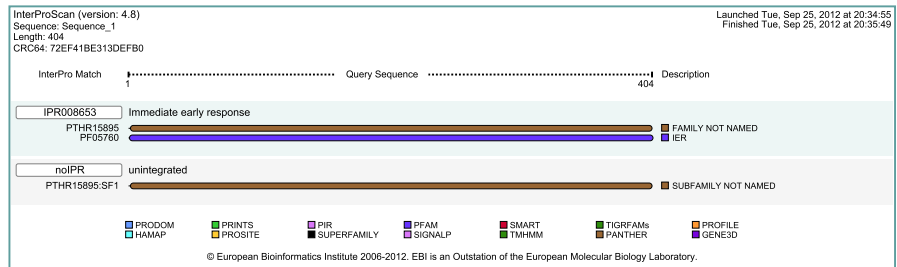
*B. floridae*  
 XP\_002609351.1  
 BRAFLDRAFT\_236223



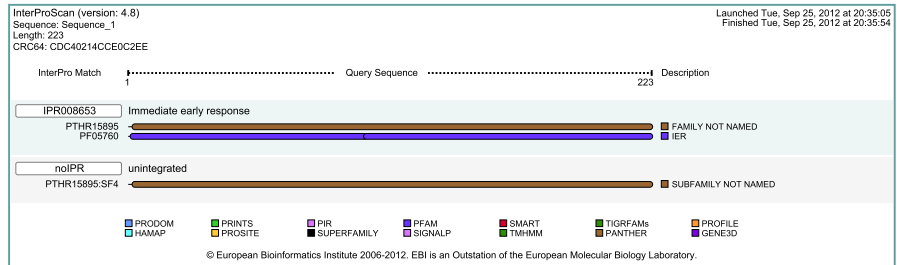
*Homo sapiens*  
 NP\_057629.2  
 IER5



*Homo sapiens*  
 NP\_982258.2  
 IER5L



*Homo sapiens*  
 NP\_004898.2  
 IER2



**Online Resource 3K.** Amphioxus sequence XP\_002609351.1 shares the conserved unique domain structure characteristic to the IER5, IER5L and IER2 MHC paralogon marker proteins in human and other vertebrates, although this is not supported by phylogenetic evidence.