

Supplementary Material

**Saltatory evolution of the *ectodermal neural cortex (ENC)*
gene family at the vertebrate origin**

Nathalie Feiner, Yasunori Murakami, Lisa Breithut, Sylvie Mazan, Axel Meyer and
Shigehiro Kuraku

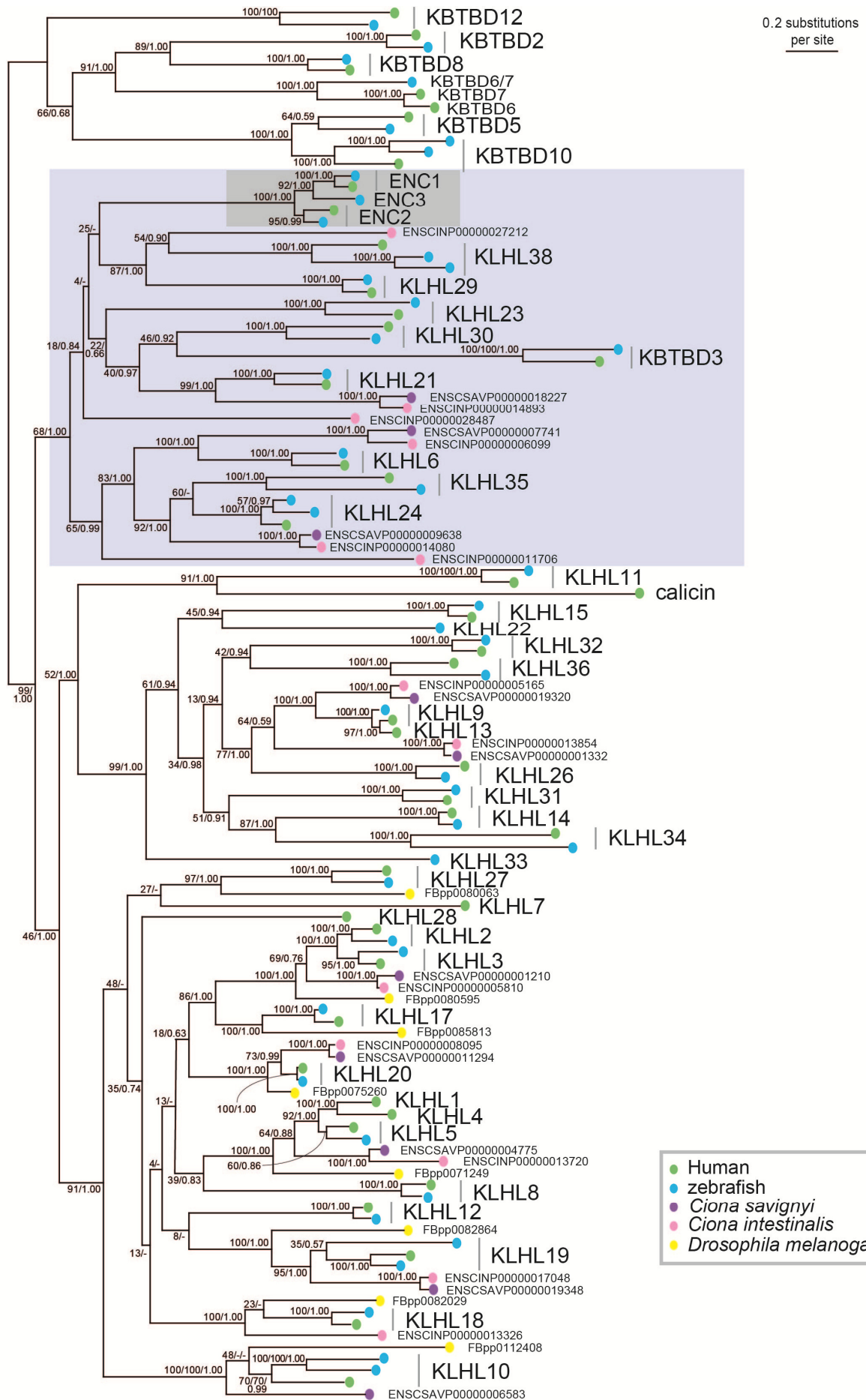


Fig. S1. Phylogenetic tree including distant members of the kelch repeat gene family. Phylogenetic tree of representatives of the kelch repeat superfamily of human, zebrafish, *Drosophila melanogaster*, *Ciona intestinalis* and *C. savignyi* (see color code on the lower right) based on the substitution model LG+I+F+ Γ_4 (shape parameter for gamma distribution $\alpha = 1.68$). A total of 120 amino acid sequences were used, and the alignment consisted of 403 amino acid residues. Support values at nodes are shown in order, bootstrap probabilities in the ML analysis, and Bayesian posterior probabilities. Based on this analysis, the dataset for the outgroup analysis was chosen (blue box): the group of proteins clustering with a bootstrap support of 68 (ML analysis) with the *ENC* family (grey box).

Table S1. Sequences included in molecular phylogenetic tree in Figure S1.

Species	Gene name	Gene name alias	Accession ID
<i>Homo sapiens</i>	KLHL1		NP_065917.1
<i>Homo sapiens</i>	KLHL2		ENSP00000226725
<i>Danio rerio</i>	klhl2		ENSDARP00000082978
<i>Homo sapiens</i>	KLHL3		ENSP00000312397
<i>Danio rerio</i>	klhl3		ENSDARP00000073620
<i>Homo sapiens</i>	KLHL4		ENSP00000362211
<i>Homo sapiens</i>	KLHL5		ENSP00000352716
<i>Danio rerio</i>	klhl5		ENSDARP00000083875
<i>Homo sapiens</i>	KLHL6		ENSP00000433734
<i>Danio rerio</i>	klhl6		ENSDARP00000028820
<i>Homo sapiens</i>	KLHL7		ENSP00000343273
<i>Homo sapiens</i>	KLHL8		ENSP00000273963
<i>Danio rerio</i>	klhl8		ENSDARP00000109421
<i>Homo sapiens</i>	KLHL9		ENSP00000351933
<i>Danio rerio</i>	klhl9		ENSDARP00000112236
<i>Homo sapiens</i>	KLHL10		ENSP00000293303
<i>Danio rerio</i>	klhl10		ENSDARP00000016945
<i>Danio rerio</i>	klhl10		ENSDARP00000067578
<i>Homo sapiens</i>	KLHL11		NP_060613.1
<i>Danio rerio</i>	klhl11		ENSDARP00000066367
<i>Homo sapiens</i>	KLHL12		ENSP00000416886
<i>Danio rerio</i>	klhl12		ENSDARP00000043814
<i>Homo sapiens</i>	KLHL13		ENSP00000360949
<i>Homo sapiens</i>	KLHL14		ENSP00000352314
<i>Danio rerio</i>	klhl14		ENSDARP00000066586
<i>Homo sapiens</i>	KLHL15		ENSP00000332791
<i>Danio rerio</i>	klhl15		ENSDARP00000025618
<i>Homo sapiens</i>	KLHL17		ENSP00000343930
<i>Danio rerio</i>	klhl17		ENSDARP00000119330
<i>Homo sapiens</i>	KLHL18		ENSP00000232766
<i>Danio rerio</i>	klhl18		ENSDARP00000025821
<i>Homo sapiens</i>	KLHL19	KEAP1	ENSP00000171111
<i>Danio rerio</i>	klhl19	keap1a	ENSDARP00000045762
<i>Danio rerio</i>	klhl19	keap1b	ENSDARP00000124228
<i>Homo sapiens</i>	KLHL20		ENSP00000209884
<i>Danio rerio</i>	klhl20		ENSDARP00000056665
<i>Homo sapiens</i>	KLHL21		ENSP00000366886

<i>Danio rerio</i>	klhl21		ENSDARP00000115156
<i>Homo sapiens</i>	KLHL22		ENSP00000331682
<i>Danio rerio</i>	klhl22		ENSDARP00000121854
<i>Homo sapiens</i>	KLHL23		ENSP00000272797
<i>Danio rerio</i>	klhl23		ENSDARP00000099374
<i>Homo sapiens</i>	KLHL24		ENSP00000242810
<i>Danio rerio</i>	klhl24		ENSDARP00000115013
<i>Danio rerio</i>	klhl24		ENSDARP00000088904
<i>Homo sapiens</i>	KLHL25	ENC2	AAV51405.1
<i>Danio rerio</i>	klhl25	ENC2	ENSDARP00000067876
<i>Homo sapiens</i>	KLHL26		ENSP00000300976
<i>Danio rerio</i>	klhl26		ENSDARP00000070477
<i>Homo sapiens</i>	KLHL27	IPP	ENSP00000379739
<i>Danio rerio</i>	klhl27	IPP	NP_001107093.1
<i>Homo sapiens</i>	KLHL28		ENSP00000379434
<i>Homo sapiens</i>	KLHL29	KBTBD9	ENSP00000420659
<i>Danio rerio</i>	klhl29		ENSDARP00000030003
<i>Homo sapiens</i>	KLHL30		ENSP00000386389
<i>Danio rerio</i>	klhl30		ENSDARP00000101071
<i>Homo sapiens</i>	KLHL31	KBTBD1	EAX04428.1
<i>Danio rerio</i>	klhl31		ENSDARP00000110003
<i>Homo sapiens</i>	KLHL32	KIAA1900	ENSP00000358265
<i>Danio rerio</i>	klhl32		ENSDARP00000023457
<i>Danio rerio</i>	klhl33		ENSDARP00000072406
<i>Homo sapiens</i>	KLHL34		ENSP00000368813
<i>Danio rerio</i>	klhl34		ENSDARP00000099083
<i>Homo sapiens</i>	KLHL35		ENSP00000438526
<i>Danio rerio</i>	klhl35		ENSDARP00000087460
<i>Homo sapiens</i>	KLHL36		ENSP00000317442
<i>Danio rerio</i>	klhl36		ENSDARP00000039584
<i>Homo sapiens</i>	KLHL37	ENC1	ENSP00000306356
<i>Danio rerio</i>	klhl37	ENC1	ENSDARP00000118742
<i>Homo sapiens</i>	KLHL38		ENSP00000321475
<i>Danio rerio</i>	klhl38		ENSDARP00000052077
<i>Danio rerio</i>	klhl38		ENSDARP00000058940
<i>Danio rerio</i>	klhl38	ivns1abpa	ENSDARP00000061658
<i>Danio rerio</i>	klhl38	ivns1abpb	ENSDARP00000002476
<i>Homo sapiens</i>	KBTBD2		NP_056298.2
<i>Danio rerio</i>	Kbtbd2		NP_001070628.1
<i>Homo sapiens</i>	KBTBD3		NP_940841.1
<i>Danio rerio</i>	kbtbd3		XP_001342878.1

<i>Homo sapiens</i>	KBTBD5		NP_689606.2
<i>Danio rerio</i>	kbtbd5		ENSDARP00000057021
<i>Homo sapiens</i>	KBTBD6		NP_690867.3
<i>Homo sapiens</i>	KBTBD7		CAG38589.1
<i>Danio rerio</i>	-		CAP19416.1
<i>Homo sapiens</i>	KBTBD8		NP_115894.2
<i>Danio rerio</i>	kbtbd8		NP_001103500.1
<i>Homo sapiens</i>	KBTBD10		EAX11275.1
<i>Danio rerio</i>	kbtbd10		XP_001922967.1
<i>Danio rerio</i>	kbtbd10b		NP_945330.1
<i>Homo sapiens</i>	KBTBD12		NP_997218.2
<i>Danio rerio</i>	kbtbd12		NP_001029089.1
<i>Homo sapiens</i>	calicin		NP_005884.2
<i>Danio rerio</i>		ENC3	ENSDARP00000075626
<i>Ciona intestinalis</i>			ENSCINP00000014080
<i>Ciona intestinalis</i>			ENSCINP00000006099
<i>Ciona intestinalis</i>			ENSCINP00000028487
<i>Ciona intestinalis</i>			ENSCINP00000005810
<i>Ciona intestinalis</i>			ENSCINP00000008095
<i>Ciona intestinalis</i>			ENSCINP00000027212
<i>Ciona intestinalis</i>			ENSCINP00000017048
<i>Ciona intestinalis</i>			ENSCINP00000011706
<i>Ciona intestinalis</i>			ENSCINP00000014893
<i>Ciona intestinalis</i>			ENSCINP00000013720
<i>Ciona intestinalis</i>			ENSCINP00000005165
<i>Ciona intestinalis</i>			ENSCINP00000013326
<i>Ciona intestinalis</i>			ENSCINP00000013854
<i>Ciona savignyi</i>			ENSCSAVP00000009638
<i>Ciona savignyi</i>			ENSCSAVP00000001210
<i>Ciona savignyi</i>			ENSCSAVP00000011294
<i>Ciona savignyi</i>			ENSCSAVP00000004775
<i>Ciona savignyi</i>			ENSCSAVP00000019348
<i>Ciona savignyi</i>			ENSCSAVP00000007741
<i>Ciona savignyi</i>			ENSCSAVP00000006583
<i>Ciona savignyi</i>			ENSCSAVP00000018227
<i>Ciona savignyi</i>			ENSCSAVP00000019320
<i>Ciona savignyi</i>			ENSCSAVP00000001332
<i>Drosophila melanogaster</i>	diablo		FBpp0075260
<i>Drosophila melanogaster</i>			FBpp0082864
<i>Drosophila melanogaster</i>			FBpp0082029
<i>Drosophila melanogaster</i>			FBpp0080595

Drosophila melanogaster

FBpp0085813

Drosophila melanogaster

FBpp0080063

Drosophila melanogaster

FBpp0071249

Drosophila melanogaster

FBpp0112408

Table S2. Nucleotide and deduced amino acid sequences of *Petromyzon marinus* *ENC-A* (*PmENC-A*).

Gene or protein details	Sequence
Nucleotide coding sequence of <i>PmENC-A</i>	<p>ATGTCGGTGTTCGGTGCACGAGAACC GCAAGTCGCGCGGAGCACGGGCTCCATGAACCTGGCG CTCTTCCACAAGCCGGGCCACGCCGACTGCGTGCTGGGGCGCCTCTGCGCCCTGCGCAAGCGC TCGCTCTTCACGGACGTGGTGCTGCGGGCGGGCGAGCGCGACTTCCCCTGCCACCCGCGCCGTG CTGGCCGCGTGCGAGCCGCTACTTCGAGGCGATGTTCCGCCGGCGGGCTGCGCGAGAGCCGTGAC GGCGAGGTGGACTTCCACGAGGCGATCCACCCCGAGGTGCTCGAGCTCCTCCTCGACTTCGCC TACACGGCCAAGGTGATGATCAACGAGGAGAACGCCGAGTCCCTGCTGGAGGCGGGCGACATG CTCGAGTTTCACGACGTTTCGTGACGCCTGCGCCGAGTTTCTCGAGAAGAACCTCCACCCCTCC AACTGCCTGGGCATGATGCTGCTGTGCGGAGGCGCACCAAGTGCCGTGCGCTCCACGACCTCTCG TGGCGCATGTGCCTCGCCAACCTTCGGCGCCCTGGCGCTCGCCGATGACTTCTGTGCTGCGG TGCACAAAGCTCGCCGAGCTGGTCTCCAGCGAGGAGCTGGAGGTGGAGGACGAGCGCGTGGCC TACGAGGCGGCGATGCGCTGGGTGAGCCACGACCCCGGGACCCGGGCGGCCACCTGCCCGC CTGCTGGCGGGCTGCGGCTCGCCATGCTGCCGCTGGTCTTCTCGCCGAGGTGCGCGGCGAC GAGCTGATCCGCGGGCAGGCGCGCAGCCGCGAGATGGTTCGAGGAGGCCATGGGCTGCAAGCTG AAGATCCTGAGGAACGACGGCGTCTGTGACCAGCCCCTGCGCCAGGCCCGCAAGTCCGGCCAC GCGCTGCTCCTGCTCGCCGGGCAGACGTTTATGTGCGACAAGATCTACCTGGTGGACCTCAAG GCGCAGCAGATCCTGCCCAAGGCAGACCTGCCGAGCCCGCGCAAGGAGTTCAGCGCGTGCGCC GTGGGCTGCAAGGTCTACGTGATGGGCGGCCGCGGCTCGGAGAACGGTGTGTCCAAGGACGTG TGGGTGTACGACACGGTGCACGATGAGTGGTCCAAGGCGGCCCCCATGCTCATCGCGGATTC GGCCACGGCTCCGCCGAGCTGGAGCACTGCC</p>
Deduced amino acid sequence of <i>PmENC-A</i>	<p>MSVSVHENRKSRASTGSMNLALFHKPGHADCVLGRRLCALRKRSFLT D VVLRAGERDFPCHRAV LAACSRYFEAMFAGGLRESRDGEVDFHEA IHPEVLELLLDFA Y TAKKVMINEENAESLLEAGD MLEFHDVRDACA E FLEKNLHP SNCLGMMLLSEAHQCRRLHDL SWRMCLANFGALALADDFLSL PCDKLAELVSSEELEVEDERVAYEAAMRWVSHDPGTRAAHL PRLLAGLRLAMLPLVFLAEVAG DELIRGQARSREMVEEAMGCKLKI LRNDGVVTS PCARPRKSGHALLLLAGQTFMCDKIYLVDL KAQQILPKADLP SPRKEFSACAVGCKVYVMGGRGSENGVSKDVVYD TVHDEWSKAAPMLIAR FGHGS AELEHC</p>

These nucleotide and deduced amino acid sequences of a partial *ENC* gene were derived from the *Petromyzon marinus* whole genome assembly. This gene was detected by gene prediction performed on the version 3 assembly (see Materials and Methods). The N-terminus of the ORF seems to be complete, while the C-terminus is incomplete.

Table S3. Sequences included in molecular phylogenetic tree in Figure 1B.

Species	Gene	Accession ID
<i>Anolis carolinensis</i>	<i>ENC1</i>	ENSACAP00000009090
<i>Callorhinchus milii</i>	<i>ENC1</i>	AAVX01039520.1
<i>Danio rerio</i>	<i>enc1</i>	ENSDARP000000051292
<i>Gallus gallus</i>	<i>ENC1</i>	ENSGALP000000024036
<i>Gasterosteus aculeatus</i>	<i>enc1</i>	ENSGACP000000020964
<i>Heterodontus francisci</i>	<i>ENC1</i>	HE981762 ^a
<i>Monodelphis domestica</i>	<i>ENC1</i>	ENSMODP000000037487
<i>Ornithorhynchus anatinus</i>	<i>ENC1</i>	ENSOANP000000024873
<i>Scyliorhinus canicula</i>	<i>ENC1</i>	HE981756 ^a
<i>Xenopus tropicalis</i>	<i>ENC1</i>	ENSXETP000000011068
<i>Anolis carolinensis</i>	<i>ENC2</i>	ENSACAP000000014425
<i>Danio rerio</i>	<i>enc2</i>	ENSDARP000000067876
<i>Gallus gallus</i>	<i>ENC2</i>	ENSGALP000000011021
<i>Gasterosteus aculeatus</i>	<i>enc2</i>	ENSGACP000000002621
<i>Lepisosteus platyrhincus</i>	<i>ENC2</i>	HE981757 ^a
<i>Monodelphis domestica</i>	<i>ENC2</i>	ENSMODP000000035157
<i>Ornithorhynchus anatinus</i>	<i>ENC2</i>	ENSOANP000000015376
<i>Xenopus tropicalis</i>	<i>ENC2</i>	ENSXETP000000016578
<i>Danio rerio</i>	<i>enc3</i>	ENSDARP000000075626
<i>Gallus gallus</i>	<i>ENC3</i>	HE981758 ^a
<i>Gasterosteus aculeatus</i>	<i>enc3</i>	ENSGACP000000013800
<i>Heterodontus francisci</i>	<i>ENC3</i>	HE981763 ^a
<i>Monodelphis domestica</i>	<i>ENC3</i>	ENSMODP000000032041
<i>Negaprion brevirostris</i>	<i>ENC3</i>	HE981764 ^a
<i>Ornithorhynchus anatinus</i>	<i>ENC3</i>	ENSOANP000000008247
<i>Scyliorhinus canicula</i>	<i>ENC3</i>	HE981759 ^a
<i>Xenopus tropicalis</i>	<i>ENC3</i>	ENSXETP000000014113
<i>Eptatretus burgeri</i>	<i>EbENC-A</i>	HE981760 ^a
<i>Petromyzon marinus</i>	<i>PmENC-A</i>	^b
<i>Petromyzon marinus</i>	<i>PmENC-B</i>	ENSPMAP00000000631

^acDNA sequences were identified in this study, and their deduced amino acid sequence was used in the phylogenetic analysis.

^bsee Table S2.

Table S4. Correspondence between the amphioxus genomic region harboring a putative *ENC* orthology and the chicken genome.

Amphioxus protein ID	Location on scaffold57 in amphioxus		Chromosome of ortholog in chicken
	start	end	
75371	958045	960817	5
75372	963257	965668	2
75373	967000	969879	2
75374	971556	980507	2
75376	997641	1012762	20
75377	1013446	1036242	11
75378	1037359	1040085	4
75379	1043492	1055498	Z
75381	1063448	1067187	2
75382	1070793	1075392	21
75394	1192422	1197546	Z
75398	1266985	1279861	11
75400	1306023	1308153	10
75405	1345455	1349172	20
75408	1358683	1361086	2
75413	1431698	1435911	13
75415	1449669	1450593	4
75416	1455708	1490465	7
75418	1521022	1544307	11
75421	1575234	1587209	2
75426	1709972	1719358	3
75428	1725398	1728443	11
75432	1763248	1765607	3
75433	1784221	1784366	4
75439	1849697	1851268	10
58277	1923884	1925722	9
75451	2069199	2070927	11
75452	2136349	2140605	4
75459	2289085	2302680	2
75466	2360675	2363653	11
75474	2563284	2568020	2
75476	2601049	2609313	13
75477	2618764	2637782	2
75479	2656393	2659461	11

This table shows the location and protein ID of amphioxus genes which are located on scaffold57 within 1 Mb distance to the *ENC* orthology candidate 'XP_002612442' (shown in red). This information is derived from the genome browser at DOE Joint Genome Institute (<http://genome.jgi.doe.gov/Brafl1/Brafl1.home.html>). Location of the putative ortholog (the sequence identified with the largest similarity by blastp searches) in the chicken genome is shown for each amphioxus protein-coding gene. Highlighted in grey are the amphioxus genes with their putative chicken orthologs located on the *ENC*-containing chromosomes (Z, 8, 10, 25 and 28; see Figure 4). With only few chicken genes identified on those chromosomes, this synteny analysis provided no convincing evidence of orthology between this amphioxus gene and the vertebrate *ENC* genes.