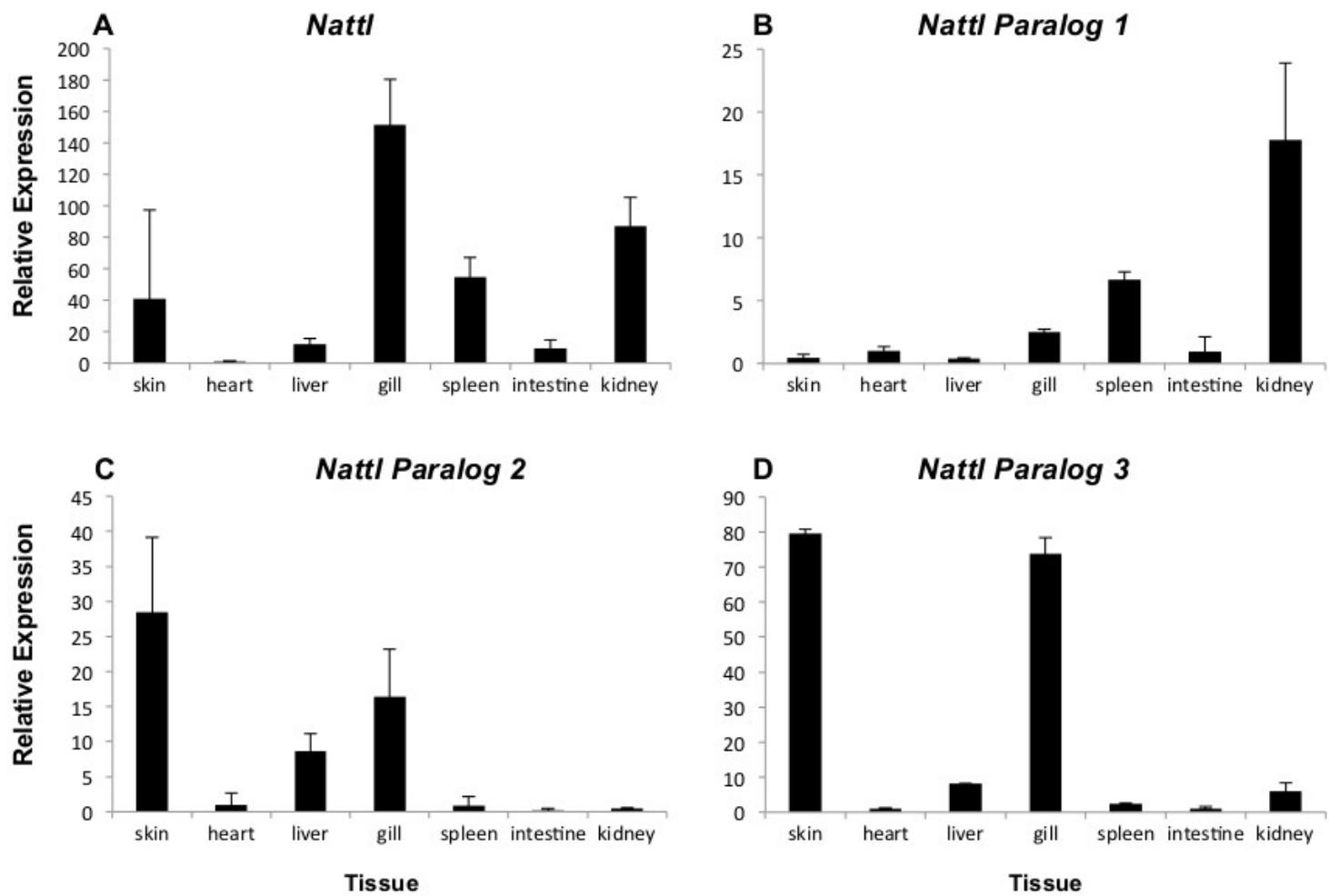


**The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs**  
Jóhannes Gudbrandsson\*, Ehsan P. Ahi\*, Kalina. H. Kapralova\*,  
Sigríður R. Franzdottir\*, Bjarni K. Kristjánsson<sup>§</sup>, S. Sophie Steinhaeuser\*,  
Ísak M. Jóhannesson\*, Valerie H. Maier\*, Sigurdur S. Snorrason\*,  
Zophonías O. Jónsson\* and Arnar Pálsson\*

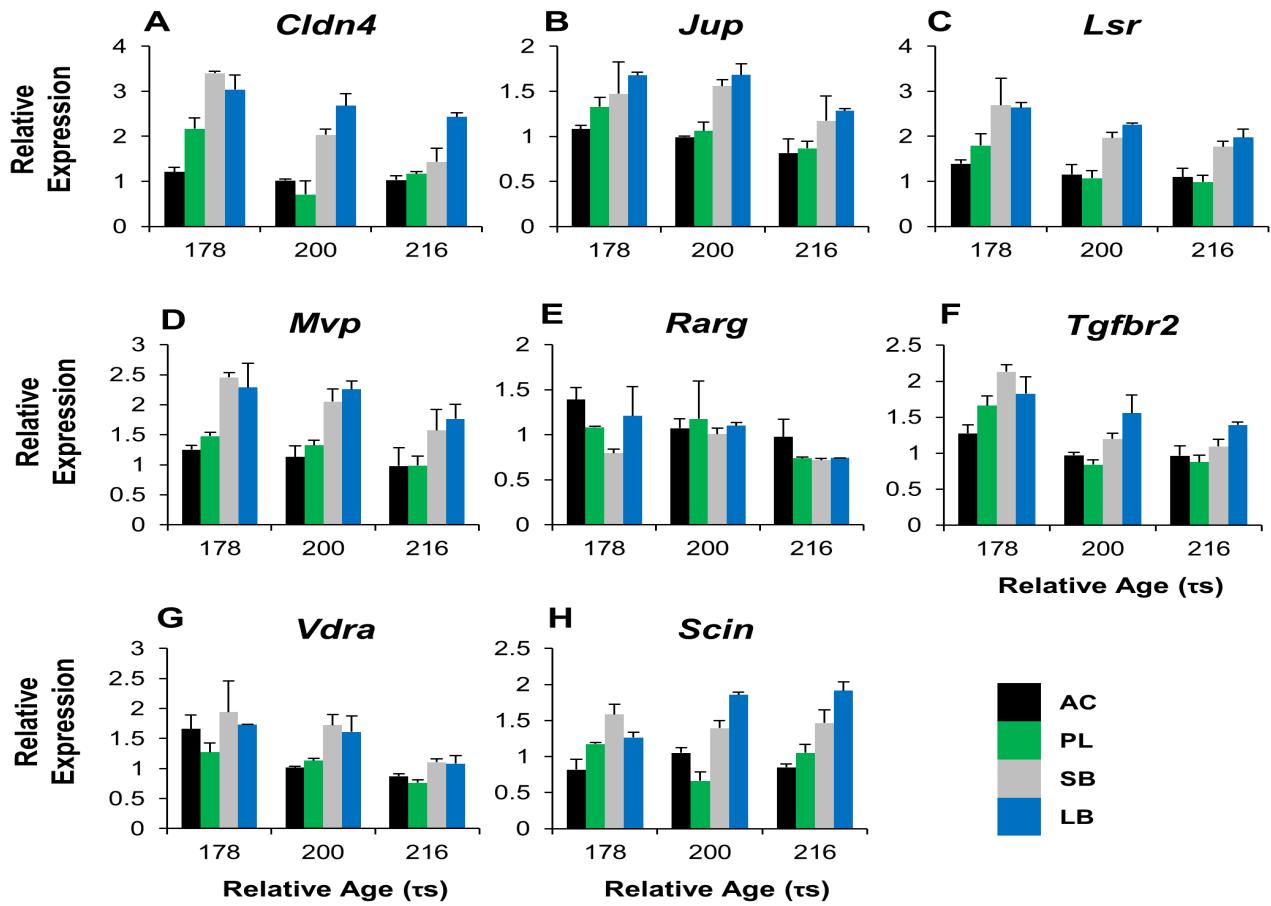
\*Institute of Life and Environmental Sciences  
University of Iceland  
101 Reykjavik, Iceland

<sup>§</sup>Holar University College  
551 Sauðarkrokur, Iceland

Corresponding author:  
Arnar Palsson  
Associate Professor in Bioinformatics  
Institute of Life and Environmental Sciences  
University of Iceland  
Sturlugata 7  
101 Reykjavik, Iceland  
Tel: +(354) 525 4265  
Fax: +(354) 525 4632  
E-mail: apalsson@hi.is



*Fig*  
**ure S1.** Relative expression of *nattl* and *nattl* 1-3 in tissues of adult AC-charr. Relative expression of Natterin (A) & Natterin paralogs 1-3 (B-D) within different tissues (skin, heart, liver, gill, spleen, intestine & kidney) of adult aquacultured Arctic charr (RT-qPCR); expression plotted for different tissues, relative to heart tissue (lowest expression levels)



**Figure S2. Relative expression of selected craniofacial candidate genes.** Relative expression of 12 candidate genes with characterized craniofacial expression during zebrafish development (ZFIN website) in the head of SB, LB, PL and AC at three close relative-ages. In the transcriptome data all of the genes had shown higher expression in SB at 200 ts. The expression is normalized to the geometric means of two craniofacial reference genes (ACTB and IF5A1). Expression is relative to a replicate of AC morph at 200 (ts), set to one. Error bars represent standard deviation calculated from two biological replicates and each biological replicate contains homogenate of six heads.

### **Supplementary file S1**

**Parameters and multiple testing corrected p-values for expression analysis.** The file is tab-delimited and the columns are:

"Unigene.Description": the annotation for that gene / paralog group.

"NR.contigs": number of contigs with this annotation.

"logCPM": count per million, log-scale

"logFC.morph": Mean fold change between the morphs, log-scale.

"logFC.T163", "logFC.T200", "logFC.T433": Mean fold change for each timepoints compared to timepoint 141, log-scale

"FDR.morph": P-value for morph difference, mulitple testing corrected

"FDR.time": P-value for time differences, mulitple testing corrected

"Contigs": SalmonDB id for the contigs with the specific annotation.

**Supplemental Table S1. qPCR primers**

Gene	Description	Primer Sequence (5'- 3')	Product Size (bp)	PCR Efficiency	Melting Temperature (°C)	Exon Boundary
<i>Actb</i>	Beta Cytoskeletal Actin	F-GAAGATCAAGATCATGCC R-CAGACTCGTGTACTCCTGCT F-ACAGCATACCTCTGTGGGG	122	1.95	80.5 ± 0.7	Yes
<i>Alp</i>	Alkaline phosphatase	R-GGTGGCATGGTCACACG F-GTGCTGTGC CATCCCAAG	177	1.90	85.12 ± 0.5	Yes
<i>Cldn4</i>	Claudin-4	R-CACCACACAGGTACATCCACA	100	1.98	80.4 ± 0.6	Yes
	Chondroitin beta-1,4-N-acetylgalactosaminyltransferase 2	F-GAGAGCCACTTACTGAGGGG				
<i>Cgat2</i>	Cytochrome c oxidase subunit VIb isoform 1	R-GAATGGACGGAAAAGAGTAACG F-GAGGGTCTACAAATCACTGTGC	120	1.98	81.86 ± 0.3	Yes
<i>Cox6b1</i>	Eukaryotic Translation Elongation Factor 1 Alpha	R-CCTGGAGTCCTACTCATACAAACAT F-GAAGATCGGCTATAACCCCTGC	147	1.93	82.22 ± 0.7	Yes
<i>Ef1α</i>	Eukaryotic Translation Initiation Factor 5A	R-ACCTTCCATCCCTTAGCGTG F-GGCTTCGTGGTGCTGAAG	111	1.94	81.36 ± 0.4	Yes
<i>If5a1</i>	Junction plakoglobin	R-CCATGTGGACCTTAGCGTG F-CACAGCAGACATACCAGGATG G	91	1.91	80.76 ± 0.6	Yes
<i>Jup</i>		R-CTGGCGATCTCTCCCCTGTT F-GCGGGACATCTACACTGCTTA	109	1.97	81.0 ± 0.3	Yes
<i>Krtap4-3</i>	Keratin-associated protein 4-3	R-AGAAGGCTAAAGTCTTAGTGACTATC	151	1.89	81.88 ± 0.6	Yes
	Lipolysis-stimulated lipoprotein receptor	F-TGCTGTCACTCTGGCGA				
<i>Lsr</i>		R-CCGTCTGGCAAGGTTCA G F-TTCCAGATCAACAGCCGCTA	80	1.91	80.77 ± 0.5	Yes
<i>Lyz</i>	Lysozyme	R-GATGCCACTGTGATGTCAT F-ACCAACTCCCAGGAGGCT	111	1.94	81.87 ± 0.7	Yes
<i>Mvp</i>	Major vault protein	R-CCTCTCCAGACGACCACG F-GTAAAGTCACCTGCATGAATG	75	1.97	78.93 ± 0.3	Yes
<i>Nattl</i>	Natterin-like protein	R-CATCTCTCCTTGATGATACCC F-AATCCGTGTCCTACCACAATGA	104	1.98	78.81 ± 0.8	No
<i>Nattl-1</i>	Natterin-like protein paralog-1	R-GGTGTGCGGTCAAAGCA F-TGAAATVTGTCATCACAAAC	135	1.77	78.03 ± 0.1	No
<i>Nattl-2</i>	Natterin-like protein paralog-1	R-GGATCTGGTCGAGGTGGC F-GTACATCCGTTCTCACCAAG	163	1.72	80.50 ± 0.2	No
<i>Nattl-3</i>	Natterin-like protein paralog-1	R-GATGTGCGGTCAAAGCG F-TGGTGGAGTGTGCGCTT	138	1.77	79.12 ± 0.2	No
<i>Ndub6</i>	NADH dehydrogenase 1 beta subcomplex subunit 6	R-CTCTCTGGGAGGTCTGGAA F-CCGTATGAATACCGTCCACAGG	171	1.89	82.40 ± 0.3	Yes
<i>Parp6</i>	Poly (ADP-Ribose) Polymerase Family, Member 6	F-ACCCAGATGTTGCCGTGCTT	147	1.93	81.87 ± 0.7	Yes

<i>Rarg</i>	Retinoic acid receptor gamma-A	F-AAGGCGAGCCCCCTTCTTC R-TGCTCTGGGTCTCCACCG F-CACCTGATCCCAGACATCCAA	82	1.92	78.62 ± 0.3	Yes
<i>Scin</i>	Scinderin/ Adseverin	R-CCTCACTCAACAACCTCGC F-CTGCTCCGAGGACGAGTG	136	1.90	83.24 ± 0.7	No
<i>Tgfbr2</i>	TGF-beta receptor type-2	R-ACCGACACCACCTGGGAG F-AATAAGGATGATTGAGGTGGTTG	72	1.93	79.02 ± 0.5	Yes
<i>Ubl5</i>	Ubiquitin-like protein 5	R-ATGAGCTTCTTCAGGTCTCC F-CGAGAAGGGACAGGTGTGTC	99	1.95	78.44 ± 0.3	Yes
<i>Ub2l3</i>	Ubiquitin-Conjugating Enzyme E2L 3	R-ACCAACGCAATCAGGGACT F-CGTCACCAAGGCGGGTCA	96	1.93	79.62 ± 0.3	Yes
<i>Vdra</i>	Vitamin D3 receptor A	R-TGGAGCTTG AGTTCTTCAGGC	81	1.93	78.12 ± 0.3	Yes

**Supplemental Table S2: Mapping of Illumina reads to *S.salar* EST data.**

Number of reads aligning to salmon reference for each sample

Alignment per read	SB 141	SB 163	SB 200	SB 433	AC 141	AC 163	AC 200	AC 433
0	33088778	30492314	27175901	25569628	32159386	30051365	31267710	28563169
1	6979368	11791558	11449549	11058555	11599602	11320997	11027195	10650748
2	2742358	4021683	3814418	3734404	4328402	4523686	3959198	3655786
3	2099068	2964994	2748108	2651522	3111277	3332577	2878729	2515303
4	1228292	1777846	1720902	1968251	1977738	2182392	1929818	1980420
5	914704	1317556	1284262	1434314	1471739	1679277	1447604	1426744
6	645264	946579	938290	1087959	1001350	1083025	1045157	1081063
7	425856	595785	578175	726290	657220	750523	690286	735351
8	293065	428003	424426	590100	530040	591332	527821	579860
9	206205	319401	334861	455838	296169	334264	387901	485653
10+	749074	1419362	1761275	3041930	1092980	1189781	1967857	3294222
<b>Total reads</b>	<b>49372032</b>	<b>56075081</b>	<b>52230167</b>	<b>52318791</b>	<b>58225903</b>	<b>57039219</b>	<b>57129276</b>	<b>54968319</b>

**Supplemental Table S3. ANOVA results for 9 genes tested for differential expression between whole Small Benthic and Aquaculture charr embryos, at two developmental timepoints (161 and 200 day degrees).**

Gene	Term	Df	F value	p value	Significance	FDR RNA-seq
<i>Alp</i>	Morph	1	13.4797	0.0214	*	0.0697
	Time	1	14.9526	0.0180	*	0.0012
	M x T	1	3.9519	0.1177	.	
<i>Cgat2</i>	Morph	1	0.0257	0.8804	.	0.0035
	Time	1	1.5141	0.2859	.	0.3312
	M x T	1	0.1866	0.6880	.	
<i>Cox6B1</i>	Morph	1	0.0898	0.7793	.	0.0580
	Time	1	3.8312	0.1219	.	0.6320
	M x T	1	0.7359	0.4393	.	
<i>Krtap4-3</i>	Morph	1	30.0255	0.0054	**	0.0121
	Time	1	0.3902	0.5661	.	0.2784
	M x T	1	4.5225	0.1006	.	
<i>Lyz</i>	Morph	1	64.1566	0.0013	**	0.0406
	Time	1	1.0390	0.3657	.	0.0005
	M x T	1	1.2026	0.3344	.	
<i>Nattl</i>	Morph	1	8.1148	0.0465	*	7.718e-07
	Time	1	14.6659	0.0186	*	6.714e-14
	M x T	1	0.2958	0.6154	.	
<i>Ndub6</i>	Morph	1	0.7447	0.4368	.	0.0982
	Time	1	7.3316	0.0537	.	0.6698
	M x T	1	0.2269	0.6587	.	
<i>Parp6</i>	Morph	1	11.2682	0.0284	*	0.1076
	Time	1	0.7393	0.4384	.	0.3789
	M x T	1	0.2343	0.6537	.	
<i>Ubl5</i>	Morph	1	1.1420	0.3454	.	0.0587

Time	1	0.2434	0.6476	.	0.0025
M x T	1	0.3974	0.5627	.	

**Significance:** p > 0.05; \* p < 0.05; \*\* p < 0.01.

**Supplemental Table S4 ANOVA for relative expression levels of Natterin and Natterin Paralogs 1-3 in Arctic charr whole embryos (among SB, AC and PL morphs) and tissues from adult A-charr**

Gene	Term	Df	F value	p value	Significance
<i>Nattl</i>	Morph	2	11.5515	0.0002	***
	Time	5	8.3202	3.99e-05	***
	M x T	9	4.4758	0.0007	***
<i>Nattl1</i>	Morph	2	19.4070	0.0001	***
	Time	3	5.9346	0.0089	**
	M x T	5	4.5761	0.0126	*
<i>Nattl2</i>	Morph	2	14.2921	0.0005	***
	Time	3	15.0463	0.0001	***
	M x T	5	3.2462	0.0404	*
<i>Nattl3</i>	Morph	2	34.4888	6.33e-06	***
	Time	3	4.4204	0.0238	*
	M x T	5	4.1843	0.0174	*
<i>Nattl</i>	Tissue	6	15.468	1.42e-08	***
<i>Nattl1</i>	Tissue	6	12.022	0.0002	***
<i>Nattl2</i>	Tissue	6	7.6811	0.0011	**
<i>Nattl3</i>	Tissue	6	46.182	8.89e-06	***

**Significance:** p > 0.05; \* p < 0.05; \*\* p < 0.01.

**Supplemental Table S5 Gene Ontology analyses of derived SNPs in SB-charr.**

Category	Observed	In category	TERM	FDR adjusted p-value
GO:0006412	24	189	translation	4.34E-006
GO:0006396	8	32	RNA processing	0.0016
GO:0006414	6	19	translational elongation	0.0038
GO:0006313	5	20	transposition, DNA-mediated	0.0498
GO:0015074	5	21	DNA integration	0.0510
GO:0006260	6	35	DNA replication	0.0679
GO:0055114	20	285	oxidation-reduction process	0.0679

**Supplemental Table S6. Predicted effect of SNP-candidates differing in frequency between charr morphs**

Effect on transcribed region	Uni_SB	Uni_AC	Rep_SB	Rep_AC
5'prime	32	19	35	24
Synonymous	232	179	176	113
Non-synonymous	112	72	81	72
3'prime	147	123	59	74

From RNA-reads that mapped to one (Uni) or more (Rep) *S. salar* ESTs.

The candidate SNPs frequencies differ more than 50% between SB and AC, summarized by which morph with higher frequency of the derived allele.

**Supplemental table S7 A. Data on candidate polymorphisms, primers and melting temperatures.**

Sequence	Genome contig	Gene name	Position	Ref	Var	Freq_AC	Freq_SB	FreqP_PL	FreqP_SB	FreqP_LB
NC_000861.1	n.a.	12S ribosomal RNA	1829	G	A	0 / 53	77 / 81	0 / 6	3 / 4	1 / 8
NC_000861.1	n.a.	16S ribosomal RNA	3119	A	T	46 / 87	18 / 28	0 / 8	0 / 8	0 / 8
NC_000861.1	n.a.	16S ribosomal RNA	3411	C	T	0 / 119	26 / 33	0 / 8	5 / 8	1 / 8
NC_000861.1	n.a.	tRNA-Lys	8876	C	A	73 / 779	74 / 352	0 / 4	0 / 4	n.a.
NC_000861.1	n.a.	NADH dehydrogenase 6	15240	G	A	2 / 3608	137 / 2702	2 / 4	0 / 4	n.a.
SS2U034121	AGKD01052493.1	Eukaryotic translation initiation factor 4 gamma 2	233	C	T	0 / 95	22 / 40	2 / 4	2 / 2	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1037	G	A	28 / 33	0 / 56	n.a.	n.a.	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1188	C	T	0 / 53	19 / 25	4 / 4	4 / 4	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1283	G	A	4 / 60	12 / 15	4 / 4	4 / 4	n.a.
SS2U055283	AGKD01013777.1	DNA2-like helicase	1822	G	A	1 / 65	25 / 50	3 / 4	n.a.	n.a.
SS2U055923	AGKD01022586.1	Bystin	615	G	A	106 / 109	7 / 190	0 / 4	n.a.	n.a.
SS2U058906	AGKD01005918.1	Mid1-interacting protein 1-like	350	G	T	0 / 49	67 / 68	4 / 4	4 / 4	n.a.

Sequence: name of the genebank sequence or EST-contig used as reference for mapped reads

Genome contig: name of salmon genome (ICSASG\_v1) contig with best sequence match to the respective EST-contig

Ref: Reference variant

Var: The derived variant

Freq\_AC and Freq\_SB: Frequency of variant reads as fraction of total numbers of reads mapped in Aquaculture (AC) or Small benthic (SB)

FreqP: The frequency of variant in genotyping by PCR and direct sequencing, as a fraction of total number of chromosomes sequenced

**Supplemental table S7 B.**

Sequence	Position	Forward primer	Reverse primer	Tm forward	Tm reverse	Paralogs
NC_000861.1	1829	GTGCCTCAGACCCACCTAGA	TCTGTCGCCCGTACTAAGGT	60.26	59.76	No
NC_000861.1	3119	GGCCAGAGTAAACACCGAGA	CCTGGATTACTCCGGTCTGA	60.25	60.07	No
NC_000861.1	3411	GGCCAGAGTAAACACCGAGA	CCTGGATTACTCCGGTCTGA	60.25	60.07	No
NC_000861.1	8876	GACGTCCTTCACTCCTGAGC	GGGCTCATAAACTGGTCGAA	59.99	60.07	No
NC_000861.1	15240	ACCCTAAAACCGAACGATCC	TGGCTAGGAAGAGTCCGGTA	60.19	59.83	No
SS2U034121	233	CTCAACGTGTTGACCAGTG	CCCTTACCCCTCCAGGATCTC	60.5	59.89	Yes
SS2U054644	1037	AAGGACGGCCACTATGGTCT	GGGGCATAGAGTGCACAGG	60.9	61.65	Yes
SS2U054644	1188	TCAGAGATAGTGAAGAAGATGCTG	CGTACTTGATAAGACCTGTCGGTA	57.92	59.62	No
SS2U054644	1283	TCAGAGATAGTGAAGAAGATGCTG	CGTACTTGATAAGACCTGTCGGTA	57.92	59.62	No
SS2U055283	1822	TGTGTGAGGTGGTTGAGGAG	GGGTCATTGCTCCCTACAGA	59.7	60.07	No
SS2U055923	615	GTGGACCCAGAGGATGAGAA	AGAACCTGCTCCCAGTTGA	60.05	59.84	No
SS2U058906	350	GCCAAAACCTCCACAATGAT	AACTGGCCTTCCAGATCAGA	59.8	59.8	Yes/No

Paralogs: indicates whether the PCR and sequencing yielded mixed products, indicative of paralogous genes