

Supplementary File

Putative extremely high rate of proteome innovation in lancelets might be explained by high rate of gene prediction errors.

László Bánya and László Patthy

Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, P.O. Box 286, H-1519, Hungary,

Corresponding author: László Patthy,

POBox address: Budapest, P.O. Box 286, H-1519, Hungary,

Street address: Budapest, Magyar tudósok körútja 2, H-1117, Hungary

Tel: (361) 382 6751

Supplementary Table S1. Comparison of the domain architectures of orthologous proteins of *B. belcheri* and *B. floridæ*. One hundred proteins of *B. belcheri*, containing at least two Pfam-A domains, were selected and their *B. floridæ* orthologs (if any) were identified by the reciprocal best-hit method using a cut off value of 60% protein sequence identity. ¹Amino acid sequence identity over the aligned region. ²Domain architecture is defined as the linear sequence of Pfam-A domains. Domains that differ in the domain architectures of orthologs from the two amphioxus species are highlighted in bold underline. The table compares the domain architectures of proteins that have orthologs in the high quality Swiss-Prot database. Entries whose domain architecture is identical with those of orthologs in the Swiss-Prot database are marked with an asterisk*.

Sequence ID <i>B. belcheri</i>		Protein sequence identity (%) ¹	Domain architecture ² <i>B. belcheri</i>		<i>B. floridæ</i>
001440_PFF0*	XP_002598316.1*	87	P4Ha_N 2OG-FeII_Oxy_3		P4Ha_N 2OG-FeII_Oxy_3
002260_PFF0*	XP_002598117.1*	86	Activin_recp Pkinase_Tyr		Activin_recp Pkinase_Tyr
003500_PFF0*	XP_002598733.1*	82	CLCA VWA_2		CLCA VWA_2
003850_PFF0*	XP_002597871.1	92	Pep_M12B_propep Reprolysin_5 Disintegrin		Pep_M12B_propep Reprolysin_5 Pep_M12B_propep Disintegrin
012350_PFF0*	XP_002604729.1*	80	Kazal_2 SPARC_Ca_bdg Thyroglobulin_1		Kazal_2 SPARC_Ca_bdg Thyroglobulin_1
017590_PFF0*	XP_002598432.1*	77	Cu-oxidase_3 Cu-oxidase Cu-oxidase_2		Cu-oxidase_3 Cu-oxidase Cu-oxidase_2
052490_PFF0*	XP_002608045.1*	93	Fz NTR		Fz NTR
058270_PFF0	XP_002586055.1	72	I-set Fz Kringle		I-set Fz Kringle Fz Kringle I-set Pkinase_Tyr I-set Pkinase_Tyr
078830_PFF0*	XP_002596432.1*	93	CPSase_L_chain CPSase_L_D2 Biotin_carb_C HMGL-like PYC_OADA Biotin_lipoyl		CPSase_L_chain CPSase_L_D2 Biotin_carb_C HMGL-like PYC_OADA Biotin_lipoyl
080250_PFF0	XP_002598506.1	94	LRR_8 I-set Wnt		Wnt

080250_PFF0	XP_002598508.1	94	LRR_8 I-set Wnt	Wnt
091850_PFF0	XP_002591393.1	75	Gal_Lectin OLF HRM GPS GAIN 7tm_2	Gal_Lectin OLF HRM GAIN 7tm_2
102470_PFF0*	XP_002608336.1	98	RNA_pol_Rpb2_1 RNA_pol_Rpb2_2 RNA_pol_Rpb2_3 RNA_pol_Rpb2_4 RNA_pol_Rpb2_5 RNA_pol_Rpb2_6 RNA pol Rpb2 7	RNA_pol_Rpb2_1 RNA_pol_Rpb2_2 RNA_pol_Rpb2_3 RNA_pol_Rpb2_4 RNA_pol_Rpb2_5 RNA_pol_Rpb2_6
103400_PFF0	XP_002593834.1*	91	Ank_3 Ank_2 Ank_4 Ank_2 Arm Arm	Ank_2 Arm
114280_PFF0	XP_002608942.1	89	C2 C2 FerI C2 FerB C2 C2 C2 C2 C2 Ferlin_C	C2 C2 FerI C2 FerB C2 C2 C2 C2 Ferlin_C
117350_PFF0*	XP_002599461.1*	89	TGFb_propeptide TGF_beta	TGFb_propeptide TGF_beta
118940_PFF0*	XP_002587581.1*	86	HAD_2 APH Acyl-CoA_dh_N Acyl-CoA_dh_M Acyl-CoA_dh_1	HAD_2 APH Acyl-CoA_dh_N Acyl-CoA_dh_M Acyl-CoA_dh_1
123650_PFF0*	XP_002603715.1	95	CAP_GLY Dynactin	CAP_GLY
123650_PFF0*	XP_002603716.1	93	CAP_GLY Dynactin	Dynactin
127590_PFF0	XP_002614044.1*	86	Kazal_2 Kazal_2 Kazal_2 SEA Kazal_2 Kazal_2 Kazal_2 Kazal_2 Kazal_1	Kazal_2 Kazal_2 Kazal_2 SEA Kazal_2 Kazal_2 Kazal_2 Kazal_2 Kazal_1

			Laminin_EGF	Laminin_EGF
			Laminin_EGF	<u>Kazal_2</u>
			EGF	<u>EGF</u>
			<u>Laminin_G_1</u>	<u>Laminin_G_1</u>
			EGF	<u>EGF</u>
			<u>Laminin_G_1</u>	<u>Laminin_G_1</u>
127600_PFF0	XP_002614044.1*	90		<u>Kazal_2</u>
				<u>Kazal_2</u>
				<u>Kazal_2</u>
				SEA
				<u>Kazal_2</u>
				<u>Kazal_2</u>
				<u>Kazal_2</u>
				<u>Kazal_1</u>
				<u>Kazal_1</u>
				<u>Laminin_EGF</u>
				<u>Laminin_EGF</u>
			Kazal_2	Kazal_2
			EGF	EGF
			Laminin_G_1	Laminin_G_1
			Laminin_G_1	Laminin_G_1
			EGF	EGF
			Laminin_G_1	Laminin_G_1
128050_PFF0	XP_002603923.1	71	TSP_1	TSP_1
			<u>TSP_1</u>	<u>TSP_1</u>
			Ldl_recept_a	Ldl_recept_a
			MACPF	MACPF
			TSP_1	TSP_1
			GCC2_GCC3	GCC2_GCC3
133320_PFF0	XP_002610402.1	91	SRCR	SRCR
			Kringle	Kringle
			SRCR	SRCR
			Trypsin	Trypsin
135460_PFF0*	XP_002605157.1*	85	Fz	Fz
			Frizzled	Frizzled
144220_PFF0	XP_002591459.1	92	RICTOR_N	RICTOR_N
			RICTOR_M	RICTOR_M
			RasGEF_N_2	RasGEF_N_2
			RICTOR_V	RICTOR_V
152200_PFF0*	-		FHA	
			CEP170_C	
153460_PFF0*	-		Ank_2	
			Ank_2	
			Ank_2	
			Ion_trans	
157010_PFF0*	XP_002611039.1	87	EF-hand_7	EF-hand_7
			Ferric_reduct	Ferric_reduct
			FAD binding_8	
			NAD_binding_6	NAD_binding_6

157070_PFF0*	XP_002611055.1*	95	ANAPC10 HECT	ANAPC10 HECT
157720_PFF0	XP_002612291.1	90	HYR GCC2_GCC3	Tropomin HYR GCC2_GCC3
158710_PFF0	XP_002600118.1	95	tRNA-synt_1 Anticodon_1 Val_tRNA-synt_C	tRNA-synt_1 Anticodon_1 Val_tRNA-synt_C
158830_PFF0*	XP_002600149.1*	83	IGFBP VWC TSP_1	IGFBP VWC TSP_1
160400_PFF0	XP_002601973.1	93	Integrin_beta Integrin_B_tail Integrin_b_cyt	Integrin_beta Integrin_B_tail Integrin_b_cyt
161590_PFF0	XP_002592408.1	98	FERM_M IRS Talin_middle I_LWEQ VBS VBS I_LWEQ	VBS I_LWEQ
161590_PFF0	XP_002592411.1	92	FERM_M IRS Talin_middle I_LWEQ VBS VBS I_LWEQ	FERM_M IRS Talin_middle
161590_PFF0	XP_002592410.1	92	FERM_M IRS Talin_middle I_LWEQ VBS VBS I_LWEQ	I_LWEQ VBS
172520_PRF0*	XP_002611318.1*	97	Fz NTR	Fz NTR
185920_PRF0	XP_002606276.1*	92	An_peroxidase EGF_CA	An_peroxidase Sushi EGF_CA
186040_PRF0*	XP_002606276.1*	91	An_peroxidase Sushi EGF_CA	An_peroxidase Sushi EGF_CA
189350_PRF0*	XP_002596820.1	99	FbpA DUF814 DUF3441	FbpA

189350_PRF0* XP_002596821.1	80	FbpA DUF814 DUF3441	DUF814 DUF3441
192660_PRF0* XP_002586029*	95	ThiF E1_FCCH E1_4HB ThiF UBA_e1_thiolCys E1_UFD	ThiF E1_FCCH E1_4HB ThiF UBA_e1_thiolCys E1_UFD
196330_PRF0 XP_002594642.1*	86	Ion_trans Ion_trans Ion_trans Ion_trans	Ion_trans Ion_trans
207610_PRF0* XP_002610947.1*	88	TGFb_propeptide TGF_beta	TGFb_propeptide TGF_beta
208940_PRF0* XP_002607923.1	95	ANF_receptor Pkinase_Tyr Guanylate_cyc	Pkinase_Tyr Guanylate_cyc
209230_PRF0 XP_002610589.1*	88	Neur_chan_LBD Neur_chan_memb Neur chan LBD Neur chan memb	Neur_chan_LBD Neur_chan_memb
211270_PRF0* XP_002591619.1*	88	BAR_3 PH ArfGap Ank_2	BAR_3 PH ArfGap Ank_2
215230_PRF0* XP_002589217.1*	88	PMT MIR PMT_4TMC	PMT MIR PMT_4TMC
215230_PRF0* XP_002589229.1*	90	PMT MIR PMT_4TMC	PMT MIR PMT_4TMC
222490_PRF0* XP_002598311.1	97	DMAP_binding DNMT1-RFD zf-CXXC BAH BAH DNA_methylase	DNMT1-RFD zf-CXXC BAH BAH DNA_methylase
223320_PRF0 XP_002601301.1	79	Pep_M12B_propep Reprolysin TSP_1 ADAM Spacer1 TSP_1 TSP_1 TSP_1 TSP_1 TSP 1 TSP 1 TSP 1	Pep_M12B_propep Reprolysin TSP_1 ADAM Spacer1 TSP_1 TSP_1 TSP_1 TSP_1 TSP 1 TSP 1 TSP 1
227800_PRF0 XP_002608349.1	63	SRCR	SRCR

			SRCR	SRCR
			SRCR	Lysyl_oxidase
			SRCR	Lysyl oxidase
			Lysyl oxidase	Lysyl oxidase
237250_PRF0	XP_002612866.1	91		
			A2M_recep	ANATO
			NTR	A2M
				Thiol-ester_cl
				A2M_comp
237260_PRF0	XP_002612866.1	90	A2M_N	A2M_recep
			A2M_N_2	NTR
			ANATO	
			A2M	
			Thiol-ester_cl	ANATO
			A2M_comp	A2M
			A2M_recep	Thiol-ester_cl
			NTR	A2M_comp
237260_PRF0	AF391289_5	92	A2M_N	A2M_N
			A2M_N_2	A2M_N_2
			ANATO	ANATO
			A2M	A2M
			Thiol-ester_cl	
			A2M_comp	
246010_PRF0*	XP_002613969.1*	72	PG_binding_1	PG_binding_1
			Peptidase_M10	Peptidase_M10
			Hemopexin	Hemopexin
247370_PRF0*	-		WIF	
			Pkinase_Tyr	
248080_PRF0	XP_002591824.1*	92	EGF_3	
			Sushi	Sushi
			Sushi	Sushi
			VWA	VWA
			Trypsin	Trypsin
253950_PRF0*	XP_002610496.1*	97	Histone	Histone
			RhoGEF	RhoGEF
			PH	PH
			RasGEF_N	RasGEF_N
			RasGEF	RasGEF
253960_PRF0*	XP_002610495.1*	96	zf-Sec23_Sec24	zf-Sec23_Sec24
			Sec23_trunk	Sec23_trunk
			Sec23_BS	Sec23_BS
			Sec23_helical	Sec23_helical
			Gelsolin	Gelsolin
254170_PRF0*	XP_002595526.1*	87	ABC_membrane	ABC_membrane
			ABC_tran	ABC_tran
			ABC_membrane	ABC_membrane

			ABC_tran	ABC_tran
256530_PRF0	B6MUN4.1	64	<u>Myb_DNA-binding</u> SWIRM <u>JAB</u>	SWIRM <u>JAB</u>
256520_PRF0	B6MUN4.1	76	JAB	<u>SWIRM</u> JAB
257610_PRF0*	XP_002587133.1*	84	EST1 EST1_DNA_bind EST1_DNA_bind PIN_4	EST1 EST1_DNA_bind EST1_DNA_bind PIN_4
258810_PRF0	XP_002605408.1*	60	Glyco_transf_29 <u>Glyco transf 29</u>	Glyco_transf_29
258870_PRF0*	XP_002595259.1	98	<u>STAT_int</u> STAT_alpha STAT_bind SH2	STAT_alpha STAT_bind SH2
258870_PRF0*	XP_002585719.1	92	STAT_int <u>STAT_alpha</u> <u>STAT_bind</u> <u>SH2</u>	STAT_int
258870_PRF0*	XP_002595260.1	100	<u>STAT_int</u> <u>STAT_alpha</u> STAT_bind <u>SH2</u>	STAT_bind
258870_PRF0*	XP_002595256.1	96	<u>STAT_int</u> <u>STAT_alpha</u> STAT_bind SH2	STAT_bind SH2
259020_PRF0*	XP_002600999.1*	97	GRAM GRAM RabGAP-TBC	GRAM GRAM RabGAP-TBC
272770_PRF0*	XP_002601715.1*	90	AAA_11 AAA_11 AAA_12	AAA_11 AAA_11 AAA_12
283240_PRF0	CAA06854.1	95	Reeler Spond_N fn3 TSP_1 fn3 TSP_1 TSP_1	Reeler Spond_N fn3 TSP_1 fn3 TSP_1 TSP_1
293630_PRF0*	XP_002613483.1*	92	Chorein_N ATG_C	Chorein_N ATG_C
293740_PRF0	XP_002613924.1	92	<u>VPS13</u> <u>VPS13_mid_rpt</u> <u>UBA</u> <u>SHR-BD</u>	<u>VPS13</u> <u>VPS13_mid_rpt</u> <u>UBA</u> <u>SHR-BD</u>

			<u>VPS13_C</u>	<u>VPS13_C</u>
209920_PRM0	XP_002613924.1	89	UBA	<u>VPS13</u> <u>VPS13_mid_rpt</u> UBA <u>SHR-BD</u> <u>VPS13_C</u>
209930_PRF0	XP_002613924.1	81		<u>VPS13</u> <u>VPS13_mid_rpt</u> UBA <u>SHR-BD</u> <u>VPS13_C</u>
209940_PRF0	XP_002613924.1	86	<u>Chorein_N</u> VPS13 VPS13_mid_rpt	VPS13 VPS13_mid_rpt <u>UBA</u> <u>SHR-BD</u> <u>VPS13_C</u>
209940_PRF0	XP_002613923.1	94	Chorein_N <u>VPS13</u> <u>VPS13_mid_rpt</u>	Chorein_N
321850_PRF0*	XP_002585883.1*	85	ANF_receptor Lig_chan-Glu_bd Lig_chan	ANF_receptor Lig_chan-Glu_bd Lig_chan
319230_PRF0*	XP_002600703.1	77	cNMP_binding DEP cNMP_binding <u>RasGEF_N</u> RasGEF	cNMP_binding DEP cNMP_binding
319230_PRF0*	XP_002600702.1	95	<u>cNMP_binding</u> DEP <u>cNMP_binding</u> <u>RasGEF_N</u> RasGEF	<u>cNMP_binding</u> DEP <u>cNMP_binding</u> <u>RasGEF_N</u> RasGEF
325460_PRF0	XP_002608790.1*	88	FG-GAP FG-GAP Integrin_alpha2	FG-GAP FG-GAP Integrin_alpha2
320370_PRF0	XP_002594515.1	72	<u>zf-RING_5</u> zf-B_box fn3 <u>SPRY</u>	<u>zf-C3HC4_3</u> zf-B_box fn3
325780_PRF0*	XP_002602117.191*	91	Glycos_transf_2 Ricin_B_lectin	Glycos_transf_2 Ricin_B_lectin
328450_PRF0	XP_002607180.1*	79	Kunitz_BPTI NTR	<u>WAP</u> <u>Kazal_2</u> <u>Kunitz_BPTI</u> Kunitz_BPTI NTR

328460_PRF0	XP_002607180.1*	85	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI	WAP Kazal_2 Kunitz_BPTI NTR
328470_PRF0	XP_002607185.1	96	ABC_membrane ABC_tran ABC_membrane ABC_tran ABC_membrane ABC_tran ABC_membrane ABC_tran	ABC_membrane ABC_tran ABC_membrane ABC_membrane ABC_membrane ABC_membrane ABC_membrane ABC_tran
328810_PRF0	XP_002592166.1	91	zf-HC5HC2H FYRN FYRC SET	zf-HC5HC2H FYRN FYRC SET
328940_PRF0	XP_002612925.1	90	dsrm dsrm A_deamin A_deamin	A_deamin
328940_PRF0	XP_002612924.1	99	dsrm dsrm A_deamin A_deamin	dsrm dsrm
329040_PRF0*	XP_002612907.1*	82	Flavodoxin_1 FAD_binding_1 NAD_binding_1	Flavodoxin_1 FAD_binding_1 NAD_binding_1
329370_PRF0*	XP_002611195.1	70	PLAT Lipoxygenase	Catalase PLAT Lipoxygenase
329970_PRF0	XP_002606233.1	92	DAG_kinase_N C1_1 C1_1 DAGK_cat DAGK_acc	DAG_kinase_N C1_1 C1_1 DAGK_cat DAGK_acc
330040_PRF0*	-		Activin_recp Pkinase	
330210_PRF0*	XP_002602990.1	98	EF-hand Mito_carr Mito_carr Mito_carr	EF-hand Mito_carr Mito_carr Mito_carr Peptidase M2

Supplementary Table S2. Comparison of the domain architectures of orthologous proteins of *B. belcheri* and *B. floridae*.
 One hundred proteins of *B. belcheri*, containing at least two Pfam-A domains, were selected and their *B. floridae* orthologs (if any) were identified by the reciprocal best-hit method using a cut off value of 60% protein sequence identity.¹ Amino acid sequence identity over the aligned region. ²Domain architecture is defined as the linear sequence of Pfam-A domains. Domains that differ in the domain architectures of orthologs from the two amphioxus species are highlighted in bold underline. The table compares the domain architectures of proteins that have no orthologs in the high quality Swiss-Prot database.

Sequence ID		Protein sequence identity (%) ¹	Domain architecture ²	
<i>B. belcheri</i>	<i>B. floridae</i>		<i>B. belcheri</i>	<i>B. floridae</i>
000200_PFF0	AAN62850.1	85	V-set CBM_14	V-set CBM_14
001280_PFF0	XP_002608265.1	88	CBM_14 ILEI	CBM_14 ILEI
017700_PFF0	XP_002598460.1	85	Gal_Lectin Gal_Lectin GPS 7tm_2	Gal_Lectin Gal_Lectin GPS 7tm_2
018300_PFF0	XP_002588697.1	68	Lectin_C fn2 fn2 fn2 fn2 fn2	Lectin_C fn2 fn2 fn2 fn2 fn2
040970_PFF0	XP_002596007.1	95	kringle kringle trypsin	kringle kringle trypsin
068220_PFF0	XP_002586877.1	79	fn2 fn2 fn2 fn2 CUB CUB	fn2 fn2 fn2 fn2 CUB CUB
080050_PFF0	XP_002606926.1	92	N-glycanase_N N-glycanase_C	N-glycanase_N N-glycanase_C
124390_PFF0	-		cEGF FXa_inhibition EGF_CA EGF_CA NIDO EGF_CA EGF_CA SEA EGF_CA	
159900_PFF0	XP_002600642.1	82	AAA_8 MT AAA_9 Dynein_heavy	AAA_8 MT AAA_9 Dynein_heavy

167280_PRF0	XP_002611740.1	66	Kringle Somatomedin_B 7tm_2	Lectin_C Kringle Somatomedin_B 7tm_2
173360_PRF0	XP_002593467.1	92	Astacin kringle	Astacin kringle
180420_PRF0	XP_002609978.1	90	Somatomedin_B 7tm_2	5-FTHF cyc-lig Somatomedin_B 7tm_2
184730_PRF0	XP_002597992.1	65	Ldl_recept_a Kringle Kringle 7tm_2	Ldl_recept_a Kringle Kringle 7tm_2
188610_PRF0	XP_002609579.1	86	ASC F5_F8_type_C F5_F8_type_C Kringle ASC	ASC F5_F8_type_C F5_F8_type_C Kringle ASC
196740_PRF0	XP_002601342.1	89	Fibrinogen_C Fibrinogen_C I-set <u>fn3</u> <u>fn3</u> <u>fn3</u> <u>fn3</u> Y_phosphatase Y_phosphatase	Y_phosphatase Y_phosphatase
196740_PRF0	XP_002601345.1	78	Fibrinogen_C Fibrinogen_C I-set fn3 <u>fn3</u> <u>fn3</u> <u>fn3</u> Y phosphatase Y phosphatase	Fibrinogen_C Fibrinogen_C fn3
200810_PRF0	XP_002608508.1	75	Astacin CUB MAM	Astacin CUB MAM
258300_PRF0	XP_002605710.1	90	Ldl_recept_a SRCR Ldl_recept_a Trypsin	Trypsin
258300_PRF0	XP_002605711.1	90	Ldl_recept_a SRCR Ldl_recept_a Trypsin	SRCR
260260_PRF0	-		BACK Kelch_1	
269420_PRF0	XP_002605854.1	73	CUB	CUB

			Ldl_recept_a	Ldl_recept_a
			Fz	Fz
			Kringle	Kringle
			Neur_chan_memb	Neur_chan_memb
304470_PRF0	XP_002587583.1	88	Ldl_recept_a	Ldl_recept_a
			VWA	VWA
			Zona_pellucida	Zona_pellucida
320550_PRF0	-		Kringle	
			Kringle	
			Lectin_C	
			CUB	
			Lectin_C	
			MAM	
			Lectin_C	
			Kringle	
			SRCR	
329690_PRF0	XP_002611153.1	90	BTB	BTB
			BTB	BTB
			BACK	BACK
			Kelch_1	Kelch_1
329890_PRF0	XP_002606025.1	77	COR	COR
			Death	Death

Supplementary Table S3. Comparison of the domain architectures of proteins of *B. belcheri* selected from dataset *Branchiostoma.belcheri_HapV2_proteins.fa* (release 1) with those of their equivalents in the dataset *Branchiostoma.belcheri_v15h11.r2_protein.fa* (release 2) and dataset *Branchiostoma.belcheri_v18h27.r3_ref_protein.fa* (release 3). One hundred proteins of *B. belcheri*, containing at least two Pfam-A domains, were selected from release 1 and their matches were identified in release 2 and release 3. ²Domain architecture (DA) is defined as the linear sequence of Pfam-A domains. Entries whose domain architecture is identical with those of orthologs in the high quality Swiss-Prot database are marked with an asterisk*. Mispredicted entries with DA that deviates from those of orthologous Swiss-Prot entries are highlighted in bold underline. The data are arranged in two categories based on the presence or absence of orthologs in the Swiss-Prot database.

Sequence ID release 1	DA	Sequence ID release 2	DA	Sequence ID release 3	DA
Orthologs are present in the Swiss-Prot database					
001440_PFF0*	P4Ha_N 2OG-FeII_Oxy_3	<u>053450F.t1</u>	P4Ha_N	089550F.t1*	P4Ha_N 2OG-FeII_Oxy_3
002260_PFF0*	Activin_recp Pkinase_Tyr	243600F.t1*	Activin_recp Pkinase_Tyr	297980F.t2*	Activin_recp Pkinase_Tyr
003500_PFF0*	CLCA VWA_2	041250R.t1*	CLCA VWA_2	265240R.t1*	CLCA VWA_2
003850_PFF0*	Pep_M12B_propep Reprolysin_5 Disintegrin	<u>062250R.t1</u>	Reprolysin_5 Disintegrin	095570F.t1*	Pep_M12B_propep Reprolysin_5 Disintegrin
012350_PFF0*	Kazal_2 SPARC_Ca_bdg Thyroglobulin_1	229830F.t1*	Kazal_2 SPARC_Ca_bdg Thyroglobulin_1	197610F.t1*	Kazal_2 SPARC_Ca_bdg Thyroglobulin_1
017590_PFF0*	Cu-oxidase_3 Cu-oxidase Cu-oxidase_2	002580R.t1*	Cu-oxidase_3 Cu-oxidase Cu-oxidase_2	<u>204890R.t2</u>	Cu-oxidase_3 Cu-oxidase Cu-oxidase_2 Cu-oxidase
052490_PFF0*	Fz NTR	017500F.t1*	Fz NTR	127920R.t1*	Fz NTR
<u>058270_PFF0</u>	I-set Fz Kringle	142670R.t2	I-set Fz Kringle I-set Pkinase_Tyr	192300R.t2	I-set Fz Kringle I-set Pkinase_Tyr
078830_PFF0*	CPSase_L_chain CPSase_L_D2 Biotin_carb_C HMGL-like PYC_OADA Biotin_lipoyl	097650F.t2*	CPSase_L_chain CPSase_L_D2 Biotin_carb_C HMGL-like PYC_OADA Biotin_lipoyl	252380R.t2*	CPSase_L_chain CPSase_L_D2 Biotin_carb_C HMGL-like PYC_OADA Biotin_lipoyl
<u>080250_PFF0</u>	LRR_8 I-set Wnt	<u>001910R.t1</u>	Wnt	062470F.t1*	Wnt

Wnt					
091850_PFF0	Gal_Lectin OLF HRM GAIN GPS 7tm_2	<u>324940R.t1</u>	Gal_Lectin OLF HRM GAIN GPS 7tm_2	191840R.t2	Gal_Lectin OLF HRM GAIN GPS 7tm_2
102470_PFF0*	RNA_pol_Rpb2_1 RNA_pol_Rpb2_2 RNA_pol_Rpb2_3 RNA_pol_Rpb2_4 RNA_pol_Rpb2_5 RNA_pol_Rpb2_6 RNA_pol_Rpb2_7	010690R.t1*	A_pol_Rpb2_1 A_pol_Rpb2_2 A_pol_Rpb2_3 A_pol_Rpb2_4 A_pol_Rpb2_5 A_pol_Rpb2_6 A_pol_Rpb2_7	282790F.t1*	A_pol_Rpb2_1 A_pol_Rpb2_2 A_pol_Rpb2_3 A_pol_Rpb2_4 A_pol_Rpb2_5 A_pol_Rpb2_6 A_pol_Rpb2_7
<u>103400 PFF0</u>	Ank_3 Ank_2 Ank_4 Arm Arm	<u>028040F.t1</u>	Ank_4 Ank_2 Arm Arm	<u>043790F.t1</u>	Ank_4 Ank_2 Arm Arm
<u>114280 PFF0</u>	C2 C2 FerI C2 FerB C2 C2 C2 C2 Ferlin_C	261060R.t1*	C2 C2 FerI C2 FerB C2 C2 C2 C2 Ferlin_C	167750F.t1*	C2 C2 FerI C2 FerB C2 C2 C2 C2 Ferlin_C
117350_PFF0*	TGFb_propeptide TGF_beta	099520R.t1*	TGFb_propeptide TGF_beta	195530R.t1*	TGFb_propeptide TGF_beta
118940_PFF0*	HAD_2 APH Acyl-CoA_dh_N Acyl-CoA_dh_M Acyl-CoA_dh_1	092840R.t1*	HAD_2 APH Acyl-CoA_dh_N Acyl-CoA_dh_M Acyl-CoA_dh_1	003000F.t1*	HAD_2 APH Acyl-CoA_dh_N Acyl-CoA_dh_M Acyl-CoA_dh_1
123650_PFF0*	CAP_GLY Dynactin	265490R.t1*	CAP_GLY Dynactin	121780F.t5*	CAP_GLY Dynactin
<u>127590 PFF0</u>	Kazal_2 Kazal_2 Kazal_2 SEA	<u>303040R.t1</u>	Kazal_2 Kazal_2 Kazal_2 SEA Kazal_2 Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_2 Kazal_2 Kazal_2	212090F.t4*	Kazal_2 Kazal_2 Kazal_2 SEA

	Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_1 Laminin_EGF Laminin_EGF	Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_1 Laminin_EGF Laminin_EGF Kazal_2 EGF Laminin_G_1 Laminin_G_1 EGF Laminin_G_1	Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_1 Laminin_EGF Laminin_EGF Kazal_2 EGF Laminin_G_1 Laminin_G_1 EGF Laminin_G_1	
<u>127600_PFF0</u>		<u>303040R.t1</u>	Kazal_2 Kazal_2 Kazal_2 SEA Kazal_2 Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_2 Kazal_2 Kazal_2 SEA Kazal_2 Kazal_2 Kazal_2 Kazal_1 Kazal_1 Laminin_EGF Laminin_EGF Kazal_2 EGF Laminin_G_1 Laminin_G_1 EGF Laminin_G_1	212090F.t4*
128050_PFF0	TSP_1 Ldl_recept_a MACPF TSP_1 GCC2_GCC3	013570R.t2	TSP_1 TSP_1 Ldl_recept_a MACPF TSP_1 GCC2_GCC3	174680F.t1 Kazal_2 EGF Laminin_G_1 Laminin_G_1 EGF Laminin_G_1
133320_PFF0	SRCR Kringle SRCR Trypsin	250080R.t1	SRCR Kringle SRCR Trypsin	228670R.t1 SRCR Kringle SRCR Trypsin
135460_PFF0*	Fz Frizzled	222910F.t1*	Fz Frizzled	<u>019690R.t1</u> Frizzled
<u>144220_PFF0</u>		<u>045000R.t1</u>	RCC1 RCC1 RCC1 RCC1 RICTOR_N RICTOR_M	<u>081090R.t3</u> RCC1 RCC1 RCC1 RCC1 RICTOR_N RICTOR_M

	RasGEF_N_2 RICTOR_V	RasGEF_N_2 RICTOR_V	RasGEF_N_2 RICTOR_V		
152200_PFF0*	FHA CEP170_C	<u>098310R.t1</u>	251720F.t1*	FHA CEP170_C	
153460_PFF0*	Ank_2 Ank_2 Ank_2 Ion_trans	227850R.t1*	Ank_2 Ank_2 Ank_2 Ion_trans	Ank_2 Ank_2 Ank_2 Ion_trans	
157010_PFF0*	EF-hand_7 Ferric_reduct FAD_binding_8 NAD_binding_6	030190R.t2*	EF-hand_7 Ferric_reduct FAD_binding_8 NAD_binding_6	EF-hand_7 Ferric_reduct FAD_binding_8 NAD_binding_6	
157070_PFF0*	ANAPC10 HECT	<u>030090R.t1</u>	ANAPC10 HECT ANAPC10 HECT	276460F.t1*	ANAPC10 HECT
157720_PFF0	HYR GCC2_GCC3	124730F.t1	HYR GCC2_GCC3	<u>261920R.t2</u>	Troponin HYR GCC2_GCC3
<u>158710_PFF0</u>	tRNA-synt_1 Anticodon_1 Val_tRNA-synt_C	<u>189810F.t1</u>	tRNA-synt_1 Anticodon_1	<u>259960R.t1</u>	tRNA-synt_1 Anticodon_1
158830_PFF0*	IGFBP VWC TSP_1	190070F.t1*	IGFBP VWC TSP_1	259710R.t1*	IGFBP VWC TSP_1
160400_PFF0	Integrin_beta Integrin_B_tail Integrin_b_cyt	100620R.t1	Integrin_beta Integrin_B_tail Integrin_b_cyt	196660R.t1	Integrin_beta Integrin_B_tail Integrin_b_cyt
<u>161590_PFF0</u>	FERM_M IRS Talin_middle I_LWEQ VBS VBS I_LWEQ	<u>181770F.t1</u>	FERM_M Talin_middle I_LWEQ I_- VBS VBS I_LWEQ	<u>085620R.t1</u>	FERM_M IRS Talin_middle LWEQ
<u>161590_PFF0</u>	FERM_M IRS Talin_middle I_LWEQ VBS VBS I_LWEQ	<u>181780F.t1</u>		<u>085620R.t3</u>	VBS VBS
172520_PRF0*	Fz NTR	078910F.t1*	Fz NTR	190570F.t1*	Fz NTR
<u>185920_PRF0</u>	An_peroxidase EGF_CA	<u>262810R.t1</u>	An_peroxidase EGF_CA	222960R.t1*	An_peroxidase Sushi EGF_CA

<u>227800 PRF0</u>	SRCR SRCR SRCR Lysyl_oxidase	<u>010500R.t1</u>	SRCR SRCR SRCR Lysyl_oxidase	<u>283050F.t3</u>	SRCR SRCR SRCR Lysyl_oxidase
<u>237250 PRF0</u>		111050F.t1*	A2M_N A2M_N_2 ANATO A2M Thiol-ester_cl A2M_comp A2M_recep NTR	171620R.t1*	A2M_N A2M_N_2 ANATO A2M Thiol-ester_cl A2M_comp A2M_recep NTR
<u>237260 PRF0</u>	A2M_N A2M_N_2 ANATO A2M Thiol-ester_cl A2M_comp	111050F.t1*	A2M_N A2M_N_2 ANATO A2M Thiol-ester_cl A2M_comp A2M_recep NTR	171620R.t1*	A2M_N A2M_N_2 ANATO A2M Thiol-ester_cl A2M_comp A2M_recep NTR
246010_PRF0*	PG_binding_1 Peptidase_M10 Hemopexin Hemopexin Hemopexin Hemopexin	098400F.t1*	PG_binding_1 Peptidase_M10 Hemopexin Hemopexin Hemopexin Hemopexin	<u>217220R.t1</u>	PG_binding_1 Peptidase_M10 Hemopexin Hemopexin Hemopexin Hemopexin MFS_1
247370_PRF0*	WIF Pkinase_Tyr	113360F.t1*	WIF Pkinase_Tyr	045780R.t1*	WIF Pkinase_Tyr
248080_PRF0	EGF_3 Sushi Sushi VWA Trypsin	073650R.t1	EGF_3 Sushi Sushi VWA Trypsin	246540R.t1	EGF_3 Sushi Sushi VWA Trypsin
253950_PRF0*		<u>290630R.t1</u>	Histone RhoGEF Histone RhoGEF PH RasGEF_N RasGEF	135030R.t1*	Histone RhoGEF Histone RhoGEF PH RasGEF_N RasGEF
253960_PRF0*	zf-Sec23_Sec24 Sec23_trunk Sec23_BS Sec23_helical Gelsolin	290640R.t1*	zf-Sec23_Sec24 Sec23_trunk Sec23_BS Sec23_helical Gelsolin	135040R.t1*	zf-Sec23_Sec24 Sec23_trunk Sec23_BS Sec23_helical Gelsolin
254170_PRF0*	ABC_membrane ABC_tran ABC_membrane ABC_tran	154120R.t1*	ABC_membrane ABC_tran ABC_membrane ABC_tran	026920R.t1*	ABC_membrane ABC_tran ABC_membrane ABC_tran
<u>256520 PRF0</u>		044700R.t1*	Myb_DNA-binding SWIRM	105830R.t1*	Myb_DNA-binding SWIRM

	JAB		JAB		JAB
<u>256530 PRF0</u>	Myb_DNA-binding SWIRM	044700R.t1*	Myb_DNA-binding SWIRM JAB	105830R.t1*	Myb_DNA-binding SWIRM JAB
257610_PRF0*	EST1 EST1_DNA_bind EST1_DNA_bind PIN_4	174840F.t1*	EST1 EST1_DNA_bind EST1_DNA_bind PIN_4	039650R.t1*	EST1 EST1_DNA_bind EST1_DNA_bind PIN_4
<u>258810 PRF0</u>	Glyco_transf_29 Glyco_transf_29	<u>000220F.t1</u>	Glyco_transf_29 Glyco_transf_29	299600F.t1*	Glyco_transf_29
258870_PRF0*	STAT_int STAT_alpha STAT_bind SH2	049290R.t1*	STAT_int STAT_alpha STAT_bind SH2	154380F.t1*	STAT_int STAT_alpha STAT_bind SH2
259020_PRF0*	GRAM GRAM RabGAP-TBC	011870R.t1*	GRAM GRAM RabGAP-TBC	075250R.t1*	GRAM GRAM RabGAP-TBC
272770_PRF0*	AAA_11 AAA_11 AAA_12	091600R.t1*	AAA_11 AAA_11 AAA_12	182450R.t1*	AAA_11 AAA_11 AAA_12
283240_PRF0	Reeler Spond_N fn3 TSP_1 fn3 TSP_1 TSP_1	241240R.t1	Reeler Spond_N fn3 TSP_1 fn3 TSP_1 TSP_1	179850R.t1	Reeler Spond_N fn3 TSP_1 fn3 TSP_1 TSP_1
293630_PRF0*		<u>312700R.t1</u>	TPR_12 TPR_12 TPR_7 TPR_12 TPR_7 TPR_12 Chorein_N ATG_C	110520R.t2*	
<u>293740 PRF0</u>		<u>151150R.t1</u>	SHR-BD VPS13_C SHR-BD VPS13_C	<u>257130F.t1</u>	
	SHR-BD VPS13_C				SHR-BD VPS13_C
<u>209920 PRM0</u>	UBA	<u>151150R.t1</u>		<u>257120F.t1</u>	UBA UBA
			SHR-BD VPS13_C SHR-BD VPS13_C		
<u>209940 PRF0</u>	Chorein_N VPS13 VPS13_mid_rpt	<u>151170R.t1</u>	VPS13 VPS13_mid_rpt	<u>257100F.t1</u>	VPS13 VPS13_mid_rpt
321850_PRF0*	ANF_receptor Lig_chan-Glu_bd Lig_chan	<u>030780R.t1</u>	Lig_chan-Glu_bd Lig_chan	275750F.t1*	ANF_receptor Lig_chan-Glu_bd Lig_chan

319230_PRF0*		<u>102020R.t1</u>	cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding	158680R.t1*	cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding cNMP_binding
			DEP DEP cNMP_N cNMP_N cNMP_N cNMP_N cNMP_N cNMP_N		DEP DEP cNMP_N cNMP_N cNMP_N cNMP_N cNMP_N cNMP_N
			SEA SEA		SEA SEA
325460_PRF0	FG-GAP FG-GAP Integrin_alpha2	100710F.t1	FG-GAP FG-GAP Integrin_alpha2	196740F.t1	FG-GAP FG-GAP Integrin_alpha2
<u>320370 PRF0</u>	<u>zf-RING_5</u>	<u>145570R.t1</u>	<u>zf-C3HC4_3</u> zf-B_box fn3 SPRY	<u>031850F.t1</u>	<u>zf-RING_5</u> zf-B_box fn3
325780_PRF0*		<u>100260F.t1</u>	Glycos_transf_2 BACK Glycos_transf_2 Ricin_B_lectin Glycos_transf_2 Ricin_B_lectin	<u>196320F.t1</u>	Glycos_transf_2 Ricin_B_lectin Glycos_transf_2 Ricin_B_lectin
			Glycos_transf_2 Ricin_B_lectin		Glycos_transf_2 Ricin_B_lectin Glycos_transf_2 Ricin_B_lectin
<u>328450 PRF0</u>		212930R.t1*	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI NTR	040720F.t1*	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI NTR
<u>328460 PRF0</u>	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI	212930R.t1*	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI NTR	040720F.t1*	WAP Kazal_2 Kunitz_BPTI Kunitz_BPTI NTR
<u>328470 PRF0</u>	ABC_membrane ABC_tran ABC_membrane ABC_tran ABC_membrane ABC_tran ABC_membrane ABC_tran	<u>212960R.t3</u>	ABC_membrane ABC_membrane ABC_tran ABC_membrane ABC_membrane ABC_membrane ABC_membrane ABC_membrane ABC_tran	<u>040670F.t2</u>	ABC_membrane ABC_tran ABC_membrane ABC_membrane ABC_membrane
<u>328810 PRF0</u>		<u>279190F.t1</u>	zf-HC5HC2H zf-HC5HC2H PHD PHD PHD zf-HC5HC2H FYRN FYRC FYRC SET	<u>172210F.t4</u>	zf-HC5HC2H PHD PHD PHD zf-HC5HC2H FYRN FYRC SET
<u>328940 PRF0</u>	dsrm	<u>278970F.t1</u>	dsrm	<u>172050F.t2</u>	

	dsrm A_deamin A_deamin		dsrm A_deamin adh_short		dsrm A_deamin
329040_PRF0*	Flavodoxin_1 FAD_binding_1 NAD_binding_1	278780F.t1*	Flavodoxin_1 FAD_binding_1 NAD_binding_1	171870F.t1*	Flavodoxin_1 FAD_binding_1 NAD_binding_1
329370_PRF0*	PLAT Lipoxygenase	<u>246260F.t1</u>	PLAT Lipoxygenase Lipoxygenase	286940R.t1*	PLAT Lipoxygenase
329970_PRF0	DAG_kinase_N C1_1 C1_1 DAGK_cat DAGK_acc	242250R.t1	DAG_kinase_N C1_1 C1_1 DAGK_cat DAGK_acc	057640F.t1	DAG_kinase_N C1_1 C1_1 DAGK_cat DAGK_acc
330040_PRF0*	Activin_recp Pkinase	242390R.t1*	Activin_recp Pkinase	<u>057490F.t2</u>	Activin_recp Pkinase Activin_recp Pkinase
330210_PRF0*	EF-hand_1 Mito_carr Mito_carr Mito_carr	242710R.t1*	EF-hand_1 Mito_carr Mito_carr Mito_carr	057220F.t1*	EF-hand_1 Mito_carr Mito_carr Mito_carr

Orthologs are absent in the Swiss-Prot database

000200_PFF0	V-set CBM_14	094410F.t1	V-set CBM_14	144580F.t1	V-set CBM_14
001280_PFF0	CBM_14 ILEI	259500R.t1	CBM_14 ILEI	169170F.t1	CBM_14 ILEI
017700_PFF0	Gal_Lectin Gal_Lectin GPS 7tm_2	002350R.t1	zf-B_box Gal_Lectin Gal_Lectin GPS 7tm_2	204650R.t1	zf-B_box Gal_Lectin Gal_Lectin GPS 7tm_2
018300_PFF0	Lectin_C fn2 fn2 fn2 fn2 fn2	261420F.t1		222240F.t1	Lectin_C fn2 fn2 fn2 fn2 fn2
040970_PFF0	kringle kringle trypsin	236680F.t1	kringle kringle trypsin	076590R.t1	kringle kringle trypsin
068220_PFF0	fn2 fn2 fn2 fn2 CUB CUB	288190R.t1	fn2 fn2 fn2 fn2 CUB CUB	100050R.t1	fn2 fn2 fn2 fn2 CUB CUB
080050_PFF0	N-glycanase_N N-glycanase_C	072890R.t1	N-glycanase_N N-glycanase_C	251010F.t1	N-glycanase_N N-glycanase_C
124390_PFF0	cEGF FXa_inhibition EGF_CA EGF_CA NIDÖ ⁰ EGF_CA EGF_CA SEA EGF_CA	140980F.t1	cEGF FXa_inhibition EGF_CA EGF_CA NIDÖ ⁰ EGF_CA EGF_CA cEGF SEA EGF_CA	141260F.t1	cEGF FXa_inhibition EGF_CA EGF_CA NIDÖ ⁰ EGF_CA EGF_CA cEGF SEA EGF_CA
159900_PFF0	AAA_8 MT AAA_9 Dynein_heavy	101480R.t1	AAA_8 MT AAA_9 AAA_9 Dynein_heavy	158030R.t1	AAA_8 MT AAA_9 AAA_9 Dynein_heavy
167280_PRF0	Kringle Somatomedin_B 7tm_2	007220R.t1	Kringle Somatomedin_B 7tm_2	118250R.t1	Kringle Somatomedin_B 7tm_2
173360_PRF0	Astacin kringle	299040F.t1	Astacin kringle	053640F.t1	Astacin kringle

180420_PRF0	Somatomedin_B 7tm_2	050300R.t1	Somatomedin_B 7tm_2	069800F.t1	Somatomedin_B 7tm_2
184730_PRF0	Ldl_recept_a Kringle Kringle 7tm_2	002500F.t1	Ldl_recept_a Kringle Kringle 7tm_2	176740F.t1	Ldl_recept_a Ldl_recept_a Kringle Kringle Somatomedin_B 7tm_2
188610_PRF0	ASC F5_F8_type_C F5_F8_type_C Kringle ASC	021920R.t1	ASC F5_F8_type_C F5_F8_type_C Kringle ASC	155310R.t1	ASC F5_F8_type_C Kringle ASC
196740_PRF0	Fibrinogen_C Fibrinogen_C I-set fn3 fn3 fn3 fn3 Y_phosphatase Y_phosphatase	024030F.t2	Fibrinogen_C Fibrinogen_C I-set fn3 fn3 fn3 fn3 Y_phosphatase Y_phosphatase	086240F.t1	Fibrinogen_C Fibrinogen_C I-set fn3 fn3 fn3 fn3 Y_phosphatase Y_phosphatase
200810_PRF0	Astacin CUB MAM	016310R.t1	Astacin CUB MAM	104370F.t1	Astacin CUB MAM
258300_PRF0		060660R.t1	SRCR Kringle Kringle Ldl_recept_a PAN_1 Ldl_recept_a SRCR Ldl_recept_a Trypsin	006090R.t1	SRCR Kringle Kringle Ldl_recept_a PAN_1 Ldl_recept_a SRCR Ldl_recept_a Trypsin
260260_PRF0	BACK Kelch_1	162090R.t1	BACK Kelch_1	152800F.t1	BACK Kelch_1
269420_PRF0		303590R.t1	Death CUB CUB Ldl_recept_a Fz Kringle Neur_chan_memb	009080F.t1	Death CUB CUB Ldl_recept_a Fz Kringle Neur_chan_memb
304470_PRF0	Ldl_recept_a VWA Zona_pellucida	092860F.t1	Ldl_recept_a VWA Zona_pellucida	002980R.t1	Ldl_recept_a VWA Zona_pellucida
320550_PRF0	Kringle Kringle Lectin_C CUB Lectin_C MAM	145290F.t1		009580R.t1	Kringle Kringle Lectin_C CUB Lectin_C MAM

	Lectin_C Kringle SRCR		Lectin_C Kringle SRCR		Lectin_C Kringle SRCR
329690_PRF0	BTB BTB BACK Kelch_1 Kelch_1 Kelch_1 Kelch_1	245850F.t1	BTB BTB BACK Kelch_1 Kelch_1 Kelch_1 Kelch_1		287390R.t1
329890_PRF0	COR Death	047410F.t1	COR Death		173700F.t1 Death
