Appendices

Appendix A

Table Appendix A.1: Possible false positive Pfam assignments in the TreeFam proteins. A list of Pfam domains that only a single gene in a gene family is annotated with, and that are not predicted with a high E-value nor cover a high fraction of a domain model.

TreeFam family	Pfam domain	Domain name	Fraction of a model covered	E-value
TF101021	PF01154	HMG_CoA_synt_N	0.14	0.0011
TF101181	PF01576	Myosin_tail_1	0.08	0.00018
TF101220	PF00621	RhoGEF	0.26	2.60E-05
TF102023	PF08092	Toxin_22	0.3	0.0081
TF105126	PF08609	Fes1	0.17	0.00086
TF105285	PF00021	UPAR_LY6	0.15	0.00022
TF105388	PF00580	UvrD-helicase	0.08	0.00011
TF105664	PF07602	DUF1565	0.07	1.30E-05
TF105993	PF08624	CRC_subunit	0.25	0.0001
TF106336	PF05693	Glycogen_syn	0.07	0.0004
TF106337	PF03488	Ins_beta	0.27	1.30E-05
TF300142	PF01370	Epimerase	0.27	0.00018
TF300253	PF00125	Histone	0.28	0.0002
TF300491	PF00128	Alpha-amylase	0.16	0.00017
TF300506	PF08001	CMV_US	0.12	0.00031
TF300523	PF02689	Herpes_Helicase	0.04	3.00E-05
TF300533	PF02672	CP12	0.21	5.30E-05
TF300647	PF08764	Coagulase	0.05	0.02
TF300805	PF05585	DUF1758	0.06	0.00043
TF312998	PF00398	RrnaAD	0.13	0.00046
TF313187	PF02790	COX2_TM	0.2	0.017
TF313234	PF01757	Acyl_transf_3	0.28	2.10E-05
TF313377	PF07732	Cu-oxidase_3	0.27	0.017
TF313568	PF08634	Pet127	0.05	0.0052
TF313594	PF08443	RimK	0.19	5.00E-05
TF313654	PF02932	Neur_chan_memb	0.14	4.10E-05
TF313802	PF06807	Clp1	0.1	0.0013
TF313930	PF04258	Peptidase_A22B	0.08	1.70E-05
TF313947	PF01271	Granin	0.07	0.011
TF314126	PF05904	DUF863	0.02	3.50E-05
TF314165	PF00136	DNA_pol_B	0.09	4.50E-05
TF314440	PF08529	NusA_N	0.18	0.00091
TF314441	PF06127	DUF962	0.01	6.80E-05
TF314495	PF08401	DUF1738	0.15	0.0001
TF314521	PF00650	CRAL_TRIO	0.15	0.0054
TF314774	PF01595	DUF21	0.13	2.10E-05
TF315186	PF06282	DUF1036	0.18	3.60E-05

TF315189	PF00851	Peptidase_C6	0.05	0.00029
TF315227	PF05511	ATP-synt_F6	0.29	5.10E-05
TF315263	PF00600	Flu_NS1	0.12	0.0098
TF315272	PF08320	PIG-X	0.08	0.00088
TF315302	PF08637	NCA2	0.08	0.0026
TF315363	PF00636	Ribonuclease_3	0.14	0.00011
TF315367	PF02752	Arrestin_C	0.26	1.90E-05
TF315472	PF01546	Peptidase_M20	0.29	0.00044
TF315592	PF00775	Dioxygenase_C	0.01	0.0011
TF315712	PF03564	DUF1759	0.26	3.10E-05
TF315897	PF01537	Herpes_glycop_D	0.11	2.40E-05
TF316533	PF07462	MSP1_C	0.04	0.00028
TF316780	PF03238	ESAG1	0.06	0.0053
TF316929	PF02932	Neur_chan_memb	0.13	0.00026
TF317006	PF06650	DUF1162	0.12	1.90E-05
TF317757	PF02093	Gag_p30	0.15	0.0043
TF317925	PF00878	CIMR	0.12	0.00022
TF318379	PF01030	Recep_L_domain	0.26	4.90E-05
TF318668	PF00001	7tm_1	0.01	2.80E-05
TF318706	PF08719	DUF1768	0.15	1.20E-05
TF319588	PF00650	CRAL_TRIO	0.15	0.0054
TF319633	PF07714	Pkinase_Tyr	0.01	2.20E-05
TF319951	PF01461	7tm_4	0.18	7.00E-05
TF321275	PF02682	AHS1	0.08	0.00019
TF321359	PF00106	adh_short	0.29	0.00044
TF321457	PF00443	UCH	0.25	1.20E-05
TF321796	PF05473	Herpes_UL45	0.17	2.00E-05
TF322230	PF02752	Arrestin_C	0.26	1.90E-05
TF323518	PF07933	DUF1681	0.13	0.00012
TF323731	PF00650	CRAL_TRIO	0.17	0.002
TF323819	PF01237	Oxysterol_BP	0.18	0.00014
TF323965	PF00094	VWD	0.15	9.60E-05
TF323987	PF00147	Fibrinogen_C	0.16	0.00011
TF324336	PF08389	Xpo1	0.26	2.50E-05
TF324755	PF09409	PUB	0.26	0.00067
TF324880	PF00261	Tropomyosin	0.07	0.00018
TF325457	PF03052	Adeno_52K	0.16	0.001
TF325523	PF08583	UPF0287	0.21	0.00053
TF326264	PF00168	C2	0.23	0.00015
TF326378	PF02355	SecD_SecF	0.14	0.00029
TF326897	PF00122	E1-E2_ATPase	0.23	2.20E-05
TF328040	PF01613	Flavin_Reduct	0.14	0.00012
TF329290	PF07933	DUF1681	0.13	0.001
TF329430	PF00168	C2	0.25	8.10E-05
TF329606	PF00447	HSF_DNA-bind	0.22	1.30E-05
TF329606	PF01579	DUF19	0.12	0.021
TF329710	PF02250	Orthopox_35kD	0.12	0.00046
TF330156	PF07732	Cu-oxidase_3	0.27	2.90E-05
TF330183	PF05806	Noggin	0.08	4.60E-05
TF330319	PF03401	Bug	0.18	0.00035
TF330845	PF00012	HSP70	0.11	0.00038
TF331115	PF00555	Endotoxin_M	0.29	0.0033
TF331158	PF05642	Sporozoite_P67	0.04	7.20E-05
TF331282	PF00386	C1q	0.26	9.10E-05

TF331344	PF00822	PMP22_Claudin	0.12	2.90E-05
TF331400	PF02790	COX2_TM	0.21	0.0077
TF331842	PF01429	MBD	0.25	0.00033
TF332204	PF05253	UPF0224	0.19	2.60E-05
TF332241	PF00059	Lectin_C	0.29	2.00E-05
TF332364	PF02373	JmjC	0.27	2.60E-05
TF332426	PF07798	DUF1640	0.24	1.80E-05
TF332497	PF08124	Lyase_8_N	0.06	0.0006
TF332538	PF05579	Peptidase_S32	0.05	7.90E-05
TF332659	PF08562	Crisp	0.21	0.0011
TF332845	PF04266	ASCH	0.18	0.00011
TF333186	PF09451	ATG27	0.09	5.20E-05
TF333434	PF00836	Stathmin	0.24	0.0017
TF333463	PF00081	Sod_Fe_N	0.21	0.00082
TF333601	PF03255	ACCA	0.2	0.001
TF335097	PF01271	Granin	0.05	0.0082
TF335573	PF02622	DUF179	0.12	0.0003
TF335835	PF01068	DNA_ligase_A_M	0.24	0.00025
TF338389	PF02825	WWE	0.3	0.0016
TF338479	PF00100	Zona_pellucida	0.23	0.0003
TF339541	PF06039	Mqo	0.05	0.00014
TF339848	PF01579	DUF19	0.15	0.012
TF340612	PF03154	Atrophin-1	0.04	0.00079
TF341730	PF01347	Vitellogenin_N	0.14	0.0002

Appendix B

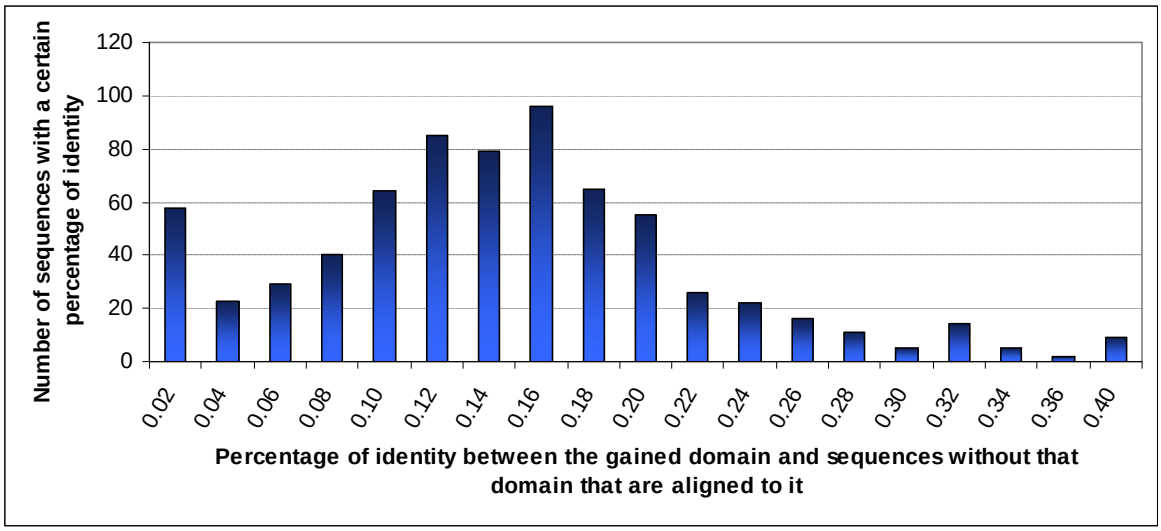

Figure Appendix B.1: Distribution of the percentages of identity between the inferred gained domain and the most similar sequence in the same gene family that does not have that domain assigned. The set of domain gains that is shown in the graph was filtered to include only internal gains and gains that have a descendant with the gained domain in at least one genome of a better quality. Two sequences in the same gene family are aligned either because of the shared ancestry, or because multiple alignment algorithms (MUSCLE in this case, http://www.drive5.com/muscle) over-align similar regions in proteins, even when they are not evolutionarily related. Both of these instances are likely to be present in the regions where inferred gained domains are aligned to sequences of other proteins in the same gene family. The peak at 0.16 could be explained with addition of values from these two scenarios.

Table Appendix B.2: High confidence domain gain events. Information about descendants of the gain event is shown only for the gains in the human lineage.

TreeFam family	Pfam domain	Representative transcript	Descendants
TF340491	PF02518	ENST00000275580	Primates
TF331377	PF04680	ENST00000290291	Primates
TF352220	PF05033,PF00856	ENST00000307483	Primates
TF331083	CL0074	ENST00000338965	Primates
TF342157	PF04698	ENST00000354668	Primates
TF328297	CL0219,PF02023	ENST00000357581	Primates
TF340395	CL0159	ENST00000359050	Primates
TF314793	PF00271,CL0008	ENST00000370424	Primates
TF351422	PF10409	ENST00000381866	Primates
TF105356	CL0023	ENST00000194097	Mammals
TF335271	CL0041	ENST00000254691	Mammals
TF328011	PF02023	ENST00000259883	Mammals
TF337552	PF00096	ENST00000262637	Mammals
TF328424	PF05386	ENST00000262715	Mammals
TF337951	PF00147	ENST00000301455	Mammals
TF300253	PF03002	ENST00000320498	Mammals
TF350810	PF01352	ENST00000338637	Mammals
TF338854	PF01352	ENST00000344099	Mammals
TF331962	PF00612	ENST00000366709	Mammals
TF338165	PF04711	ENST00000367990	Mammals
TF336000	PF08065	ENST00000368654	Mammals
TF330114	CL0175	ENST00000373330	Mammals
TF333425	PF04593	ENST00000388827	Mammals
TF325887	PF10522	ENST00000394516	Mammals
TF105660	PF08062	ENST00000399466	Mammals
TF330855	PF03523	ENST00000262101	Mammals
TF334740	CL0006,PF00621	ENST00000296794	Mammals
TF329807	PF06049	ENST00000367797	Mammals
TF324004	PF02008	ENST00000373644	Mammals
TF317779	PF09307	ENST00000009530	Vertebrates
TF326567	CL0003	ENST00000046794	Vertebrates
TF325130	PF00023	ENST00000160373	Vertebrates
TF106374	CL0172	ENST00000199447	Vertebrates
TF105392	CL0159,PF03160	ENST00000200181	Vertebrates
TF325426	PF00632	ENST00000206595	Vertebrates
TF319848	PF01033	ENST00000229003	Vertebrates
TF106352	PF00023	ENST00000230792	Vertebrates
TF313285	PF01759,PF01821	ENST00000245907	Vertebrates
TF329176	CL0081	ENST00000249910	Vertebrates
TF330078	PF10393	ENST00000255132	Vertebrates
TF320327	PF04812	ENST00000257555	Vertebrates
TF320327	PF04813	ENST00000257555	Vertebrates
TF330114	PF05485	ENST00000260045	Vertebrates
TF331062	CL0072	ENST00000260283	Vertebrates
TF313938	CL0154	ENST00000260983	Vertebrates
TF316484	PF01463,CL0022,PF01822	ENST00000262304	Vertebrates
TF316148	PF03815	ENST00000262424	Vertebrates
TF312824	PF00612	ENST00000262457	Vertebrates
TF316113	CL0003	ENST00000262878	Vertebrates
TF317402	CL0159	ENST00000263798	Vertebrates

TF317511	PF00017	ENST00000263915	Vertebrates
TF331945	PF07525	ENST00000264607	Vertebrates
TF316876	PF00093	ENST00000264895	Vertebrates
TF316876	PF03160	ENST00000264895	Vertebrates
TF324610	PF02732	ENST00000267430	Vertebrates
TF351678	PF01392	ENST00000273857	Vertebrates
TF314731	PF10565	ENST00000279593	Vertebrates
TF313240	PF10606	ENST00000282753	Vertebrates
TF316380	CL0011	ENST00000283296	Vertebrates
TF316380	PF01390	ENST00000283296	Vertebrates
TF351678	CL0202	ENST00000284885	Vertebrates
TF329158	CL0023	ENST00000285928	Vertebrates
TF314232	PF00569,CL0220	ENST00000288642	Vertebrates
TF316484	PF02010	ENST00000289672	Vertebrates
TF332664	PF07776	ENST00000289816	Vertebrates
TF313285	CL0005	ENST00000291440	Vertebrates
TF336193	PF01342,PF03172	ENST00000291582	Vertebrates
TF323966	CL0214	ENST00000294383	Vertebrates
TF317402	PF01403	ENST00000296474	Vertebrates
TF329295	CL0124	ENST00000296498	Vertebrates
TF329059	CL0001	ENST00000296575	Vertebrates
TF331157	CL0041	ENST00000297350	Vertebrates
TF312852	CL0219	ENST00000298139	Vertebrates
TF323475	CL0003	ENST00000298229	Vertebrates
TF323480	CL0005	ENST00000302495	Vertebrates
TF331319	PF01822,CL0164	ENST00000303746	Vertebrates
TF106506	PF00023	ENST00000303941	Vertebrates
TF106401	PF00249	ENST00000310806	Vertebrates
TF327329	PF00051,PF09396	ENST00000311907	Vertebrates
TF314204	PF02816	ENST00000313478	Vertebrates
TF324155	PF00023	ENST00000313581	Vertebrates
TF315996	CL0006	ENST00000314276	Vertebrates
TF316105	CL0188	ENST00000317133	Vertebrates
TF317614	PF06959	ENST00000317905	Vertebrates
TF315956	PF05485	ENST00000321679	Vertebrates
TF317659	PF01391	ENST00000322313	Vertebrates
TF329915	PF00040	ENST00000323926	Vertebrates
TF106510	PF02161	ENST00000325455	Vertebrates
TF313103	PF07941	ENST00000328224	Vertebrates
TF318980	PF02165	ENST00000332351	Vertebrates
TF333138	PF01391	ENST00000333570	Vertebrates
TF329287	PF00642	ENST00000333834	Vertebrates
TF317921	PF00023	ENST00000340022	Vertebrates
TF316214	PF04621	ENST00000343495	Vertebrates
TF105669	PF00458	ENST00000344102	Vertebrates
TF318080	CL0016	ENST00000344204	Vertebrates
TF315606	CL0041	ENST00000344227	Vertebrates
TF329345	CL0010	ENST00000344936	Vertebrates
TF321873	CL0056	ENST00000355044	Vertebrates
TF326161	PF01284	ENST00000355237	Vertebrates
TF300189	PF10574	ENST00000357484	Vertebrates
TF330032	PF01033	ENST00000357639	Vertebrates
TF300851	PF00642	ENST00000357720	Vertebrates
TF328589	PF09303	ENST00000358316	Vertebrates
TF323607	PF06462	ENST00000359520	Vertebrates
TF323475	PF00017	ENST00000359570	Vertebrates
TF315592	PF01392	ENST00000360986	Vertebrates
TF331681	PF00057	ENST00000361205	Vertebrates
TF326495	PF06663	ENST00000367213	Vertebrates
TF315841	PF02205	ENST00000367288	Vertebrates

TF334159	PF05177	ENST00000367856	Vertebrates
TF315806	CL0123	ENST00000368474	Vertebrates
TF314133	CL0003	ENST00000369075	Vertebrates
TF329606	PF03509	ENST00000369235	Vertebrates
TF316297	PF06839	ENST00000369466	Vertebrates
TF316833	PF06484	ENST00000371130	Vertebrates
TF313103	PF03521	ENST00000371741	Vertebrates
TF101106	PF10487	ENST00000372577	Vertebrates
TF331727	PF05604	ENST00000372970	Vertebrates
TF312900	CL0202	ENST00000373187	Vertebrates
TF330498	CL0154	ENST00000373209	Vertebrates
TF330345	CL0011	ENST00000373401	Vertebrates
TF320194	CL0196	ENST00000373638	Vertebrates
TF300648	CL0172	ENST00000375663	Vertebrates
TF300648	PF00043	ENST00000375663	Vertebrates
TF313965	PF00084	ENST00000377034	Vertebrates
TF331310	CL0033	ENST00000377674	Vertebrates
TF106001	PF02344,PF01056	ENST00000377970	Vertebrates
TF313698	PF03700	ENST00000380285	Vertebrates
TF315592	CL0202	ENST00000380605	Vertebrates
TF324293	CL0154	ENST00000380868	Vertebrates
TF316876	CL0056	ENST00000380881	Vertebrates
TF332820	PF08365	ENST00000381389	Vertebrates
TF329720	CL0084,PF00533	ENST00000381989	Vertebrates
TF323983	CL0179	ENST00000383733	Vertebrates
TF105391	CL0128	ENST00000389202	Vertebrates
TF317067	CL0266	ENST00000389247	Vertebrates
TF316056	PF09004	ENST00000389568	Vertebrates
TF318198	CL0188	ENST00000389821	Vertebrates
TF106341	PF00010	ENST00000389936	Vertebrates
TF330156	PF03815	ENST00000392504	Vertebrates
TF331707	CL0219,PF09091	ENST00000392723	Vertebrates
TF331055	CL0010	ENST00000393398	Vertebrates
TF336041	CL0001	ENST00000394980	Vertebrates
TF337303	PF00435	ENST00000395209	Vertebrates
TF314963	CL0208	ENST00000396197	Vertebrates
TF317532	CL0011	ENST00000396906	Vertebrates
TF317402	PF01833,PF01437	ENST00000397752	Vertebrates
TF106276	PF08959	ENST00000398892	Vertebrates
TF331207	PF00014	ENST00000399429	Vertebrates
TF106451	PF07452	ENST00000204604	Bilateralia
TF314081	CL0033	ENST00000215739	Bilateralia
TF313754	PF00805	ENST00000221200	Bilateralia
TF331485	PF00988,CL0014	ENST00000233072	Bilateralia
TF323999	PF00773	ENST00000252889	Bilateralia
TF323502	PF02185	ENST00000254260	Bilateralia
TF324918	PF00057,CL0186	ENST00000260197	Bilateralia
TF313551	PF08912	ENST00000261535	Bilateralia
TF313326	CL0190	ENST00000261875	Bilateralia
TF323159	CL0020	ENST00000263635	Bilateralia
TF351276	CL0072	ENST00000264042	Bilateralia
TF354308	CL0221	ENST00000278279	Bilateralia
TF315363	PF00611	ENST00000281092	Bilateralia
TF324744	PF00642	ENST00000295373	Bilateralia
TF315892	CL0010	ENST00000295713	Bilateralia
TF318935	PF02218	ENST00000301843	Bilateralia
TF315897	PF03765	ENST00000306726	Bilateralia
TF323280	PF00630	ENST00000323468	Bilateralia
TF318014	PF00412	ENST00000336180	Bilateralia
TF101179	PF09465	ENST00000338179	Bilateralia

TF315363	CL0266	ENST00000348343	Bilateralia
TF323999	PF07145	ENST00000358691	Bilateralia
TF323312	PF00641	ENST00000359653	Bilateralia
TF324164	CL0223	ENST00000369443	Bilateralia
TF326321	PF01424,CL0196	ENST00000371527	Bilateralia
TF324293	CL0266,PF00621	ENST00000380868	Bilateralia
TF323674	PF02825	ENST00000389044	Bilateralia
TF314351	CL0126	ENST00000061240	AllAnimals
TF330032	CL0263	ENST00000075322	AllAnimals
TF329240	CL0200	ENST00000202017	AllAnimals
TF313988	PF04707	ENST00000251170	AllAnimals
TF314316	PF01463,CL0022	ENST00000252804	AllAnimals
TF335359	PF06009	ENST00000252999	AllAnimals
TF314796	CL0041	ENST00000261600	AllAnimals
TF317296	CL0266	ENST00000261752	AllAnimals
TF320906	PF00787	ENST00000262211	AllAnimals
TF313191	PF08403	ENST00000262461	AllAnimals
TF105399	PF06466	ENST00000263754	AllAnimals
TF313184	PF00595	ENST00000264431	AllAnimals
TF323502	CL0031	ENST00000265562	AllAnimals
TF317067	CL0006	ENST00000268676	AllAnimals
TF314219	PF02809	ENST00000289528	AllAnimals
TF102004	CL0072	ENST00000295797	AllAnimals
TF314470	CL0186	ENST00000298125	AllAnimals
TF316118	PF00439	ENST00000302054	AllAnimals
TF314638	CL0183	ENST00000310298	AllAnimals
TF300359	CL0220	ENST00000310454	AllAnimals
TF312822	CL0271	ENST00000311630	AllAnimals
TF105056	CL0137	ENST00000313698	AllAnimals
TF314677	PF09141	ENST00000314888	AllAnimals
TF314748	CL0154	ENST00000324068	AllAnimals
TF319230	PF00023	ENST00000332509	AllAnimals
TF106173	PF02148	ENST00000334136	AllAnimals
TF312960	CL0010	ENST00000338257	AllAnimals
TF313629	CL0266	ENST00000339416	AllAnimals
TF316643	PF00373	ENST00000340930	AllAnimals
TF318080	CL0011	ENST00000344204	AllAnimals
TF316643	PF03623	ENST00000346049	AllAnimals
TF105282	PF08070	ENST00000348049	AllAnimals
TF314159	PF06311	ENST00000355058	AllAnimals
TF106448	CL0114	ENST00000357008	AllAnimals
TF106151	CL0196	ENST00000358896	AllAnimals
TF313758	PF00880	ENST00000359988	AllAnimals
TF351123	CL0159	ENST00000360304	AllAnimals
TF323658	PF00397	ENST00000361125	AllAnimals
TF105224	CL0186	ENST00000361961	AllAnimals
TF314076	CL0186	ENST00000367097	AllAnimals
TF354311	CL0221	ENST00000367122	AllAnimals
TF300807	PF02225	ENST00000367512	AllAnimals
TF320809	CL0010	ENST00000369405	AllAnimals
TF314566	PF09162	ENST00000372788	AllAnimals
TF317034	PF00620	ENST00000373026	AllAnimals
TF314897	PF01585	ENST00000373451	AllAnimals
TF323767	CL0003	ENST00000373886	AllAnimals
TF319104	PF00880	ENST00000377187	AllAnimals
TF323577	PF00784	ENST00000377307	AllAnimals
TF314263	CL0016	ENST00000378168	AllAnimals
TF324293	CL0010	ENST00000380868	AllAnimals
TF300785	PF07533	ENST00000382194	AllAnimals
TF102004	CL0266	ENST00000392038	AllAnimals

TF314028	PF00355	ENST00000399167	AllAnimals
TF323674	PF06701	ENST00000399332	AllAnimals
TF332135	PF06046	ENSMUST00000011407	
TF316484	PF02140	ENSMUST00000040422	
TF316155	PF02178	ENSMUST00000040802	
TF328297	PF06747	ENSMUST00000041466	
TF335390	PF00096	ENSMUST00000051869	
TF329295	PF00100	ENSMUST00000084509	
TF344032	CL0016	ENSMUST00000086209	
TF352132	PF02415	ENSMUST00000087258	
TF335097	PF00530	ENSMUST00000090986	
TF343969	CL0072	ENSMUST00000096028	
TF350794	PF01352	ENSMUST00000098508	
TF327726	PF08742	ENSMUST00000098633	
TF106451	PF08742,PF01826,PF00094	ENSMUST00000101614	
TF313537	CL0164	ENSMUST00000102891	
TF331090	CL0188	ENSMUST00000106224	
TF313147	CL0202	ENSMUST00000106949	
TF334740	CL0266	ENSMUST00000109426	
TF332078	PF03172	ENSMUST00000113392	
TF317514	CL0069,CL0011	ENSRNOT00000011676	
TF101514	PF08155	ENSRNOT00000012798	
TF319471	PF00096	ENSRNOT00000034133	
TF337163	PF08384	ENSRNOT00000041557	
TF335163	PF03501	ENSRNOT00000043986	
TF314473	CL0010	ENSXETT00000002556	
TF336376	PF10479	ENSXETT00000010407	
TF313664	PF00628,CL0008	ENSXETT00000017293	
TF352568	PF01759	ENSXETT00000037556	
TF327588	CL0291	ENSXETT00000041950	
TF343001	PF06512	ENSXETT00000045061	
TF343800	CL0266	ENSXETT00000049369	
TF343807	PF02135	ENSXETT00000049701	
TF330284	CL0164	ENSXETT00000055961	
TF316425	CL0004	ENSGALT00000003763	
TF330943	CL0102	ENSGALT00000012528	
TF343232	PF00612	ENSGALT00000017818	
TF331401	CL0066	ENSGALT00000036204	
TF326300	PF02181	ENSDART00000002526	
TF326300	PF02205	ENSDART00000002526	
TF342779	PF00681	ENSDART00000026448	
TF333311	PF08344	ENSDART00000045905	
TF329914	CL0184	ENSDART00000054641	
TF335519	PF08441	ENSDART00000055263	
TF318964	CL0266	ENSDART00000075811	
TF343508	CL0016	ENSDART00000076763	
TF316498	PF05485	ENSDART00000078494	
TF300180	CL0202	ENSDART00000078606	
TF329039	CL0188	ENSDART00000080545	
TF350019	PF00260	ENSDART00000081597	
TF106435	PF00748	ENSDART00000081614	
TF315837	PF02188,CL0072	ENSDART00000084559	
TF330777	PF01049	ENSDART00000086138	
TF315536	CL0272	ENSDART00000087610	
TF315645	CL0001	ENSDART00000097691	
TF332213	CL0229	ENSDART00000098581	
TF351676	PF01033	ENSDART00000104096	
TF343963	PF07500	ENSDART00000036529	
TF335838	PF07776	CG10431-RA	
TF343858	PF05030	CG10555-RA	

TF327367	PF05267	CG10912-RA
TF326895	CL0229	CG10916-RA
TF325393	PF02757	CG11066-RB
TF313668	PF04568	CG11079-RA
TF343304	CL0081	CG13598-RA
TF332191	CL0155	CG13676-RA
TF326889	PF00079	CG14470-RA
TF319090	CL0155	CG14608-RA
TF344100	PF06818	CG15365-RA
TF325916	CL0056	CG15378-RA
TF329913	PF08742,PF00094	CG15671-RA
TF350188	PF02757	CG15731-RA
TF324584	PF00631	CG15844-RA
TF351124	CL0011	CG16974-RA
TF316403	PF01049	CG17941-RA
TF321823	PF00014	CG18296-RA
TF312905	CL0220	CG31216-RA
TF323648	CL0004	CG32226-RA
TF343869	PF03128	CG32580-RA
TF343188	CL0155	CG32656-RA
TF351975	PF00650	CG32697-RA
TF319052	PF05444	CG34040-RA
TF319243	PF00412	CG4656-RA
TF343781	PF01753	CG4877-RA
TF315391	PF07776	CG5034-RA
TF313817	CL0229	CG5071-RB
TF316895	PF00023	CG5424-RB
TF343612	CL0155	CG5756-RA
TF327546	PF05267	CG5765-RA
TF313415	PF10545	CG6279-RA
TF105127	CL0031	CG7042-RA
TF313080	CL0265	CG7067-RA
TF336220	PF00569,CL0220	CG8529-RC
TF317532	PF01753,CL0049	CG8569-RA
TF322044	PF00628	CG8677-RA
TF316872	CL0004	CG9138-RA
TF314883	PF07773	CG9227-RA
TF317819	PF09607	CG9653-RA
TF105292	CL0220	CG9847-RA
TF326676	CL0126	CG9850-RB

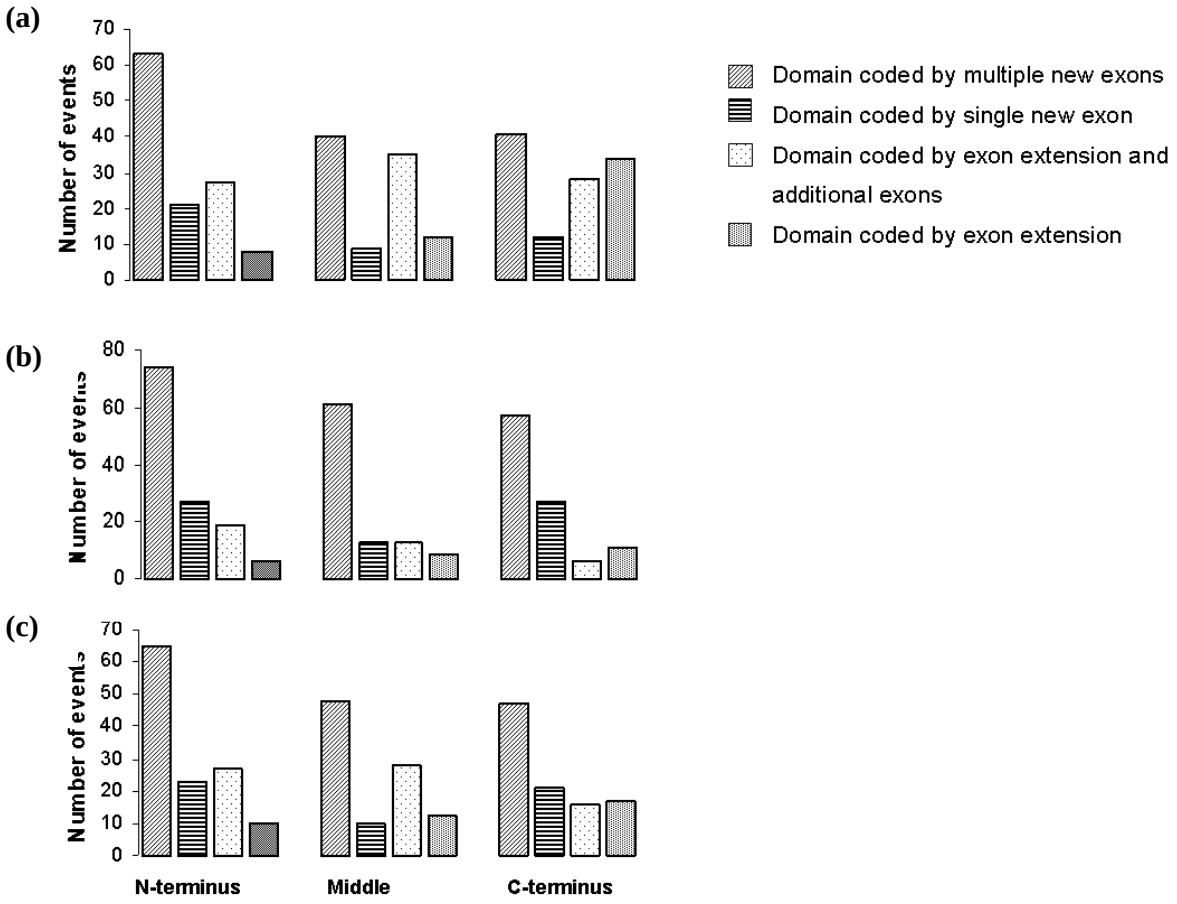

Figure Appendix B.3: Distribution of high confidence domain gain events according to the position of domain insertion and number of exons gained according to three different classification methods. (a) Method 1 was used for classification of high confidence domain gain events in Chapter 3. (b) Method 2 classifies a domain gain as an exon extension if there are at least 20 amino acids towards either of the exon borders and at least 30% of these are identical to a sequence in the alignment that does not contain the gained domain. (c) Method 3 classifies a domain gain as exon extension if there are at least 15 amino acids towards either of the exon borders and at least 25% of these are identical to a sequence in the alignment that does not contain the gained domain. Methods 2 and 3 classify each domain gain as a gain at the termini if towards the termini there are at least 80% unaligned residues or there are less than 10% identical residues in any of the sequences without the gained domain. There were seven and six domain gain events with ambiguous positions obtained by Methods 2 and 3, respectively which are not included in the Figure.

Table Appendix B.4: Domains that are gained by insertion of new exons(s) into the introns of ancestral genes. Phases of introns that surround the exons coding for the gained domains are shown for each gain event. In two cases (marked with * next to domain name) introns surrounding domains did not have symmetrical phases, however additional exons appeared to have been gained together with the one(s) coding for these domains and phases of introns surrounding all inserted exons were symmetrical. It is also noted whether the gained domain(s) is/are coded by single or multiple new exons.

TreeFam family	Domain gained	Phase of 5' intron	Phase of 3' intron	Is single exon coding for the gained domain
TF336041	CL0001	1	1	Yes
TF335097	PF00530	1	1	Yes
TF331962	PF00612*	1	1	Yes
TF313965	PF00084	1	1	Yes
TF351678	CL0202	1	1	No
TF330156	PF03815	1	1	No
TF329915	PF00040	1	1	No
TF325130	PF00023	1	1	No
TF324293	CL0010	1	1	No
TF323674	PF06701	1	1	No
TF321873	CL0056	1	1	No
TF318080	CL0011	1	1	No
TF317532	CL0011	1	1	No
TF317402	CL0159	1	1	No
TF316484	PF02140	1	1	No
TF316380	CL0011	1	1	No
TF315592	PF01392	1	1	No
TF315592	CL0202	1	1	No
TF313537	CL0164	1	1	No
TF105391	CL0128	1	1	No
TF331319	PF01822,CL0164	1	1	No
TF324293	CL0266,PF00621	1	1	No
TF314677	PF09141	0	0	No
TF314133	CL0003	0	0	No
TF314081	CL0033*	0	0	No
TF313551	PF08912	0	0	No
TF300785	PF07533	0	0	No
TF106435	PF00748	0	0	No
TF325887	PF10522	0	1	Yes
TF324610	PF02732	0	1	Yes
TF323999	PF00773	1	2	Yes
TF322044	PF00628	0	1	Yes
TF315892	CL0010	1	2	Yes
TF354311	CL0221	2	1	No
TF350794	PF01352	0	2	No
TF335359	PF06009	1	0	No
TF331062	CL0072	2	1	No
TF329158	CL0023	2	1	No
TF319848	PF01033	1	2	No
TF319230	PF00023	2	1	No
TF317921	PF00023	2	0	No
TF317614	PF06959	0	2	No
TF316118	PF00439	0	1	No

TF314638	CL0183	2	0	No
TF313938	CL0154	0	2	No
TF313629	CL0266	0	1	No
TF312900	CL0202	1	0	No
TF106448	CL0114	0	1	No
TF327329	PF00051,PF09396	1	2	No

Appendix C

Table Appendix C.1: Predicted binding sites and annotated PTM sites in the set of tissue-specific exons compared to the sets of cassette and constitutive exons. Column headed 'N_+' shows the number of exons with the examined characteristic, 'N_-'of those without it, and column headed 'Fraction$_+$' shows a fraction of exons with the examined characteristic. PTM sites were taken from the UniProtKB/Swiss-Prot database. The P-value shows the results of the comparison with the set of Tissue-specific exons, and is obtained with the Chi-square test.

Analysis	Set of exons	N_+	N_-	N_{total}	Fraction$_+$	P-value
ANCHOR	Tissue-specific	410	1,016	1,426	0.288	N/A
	Cassette	8,821	40,203	49,024	0.180	$P<2.2x10^{-16}$
	Constitutive	27,374	122,564	149,938	0.183	$P<2.2x10^{-16}$
PTM sites	Tissue-specific	119	798	917	0.130	N/A
	Cassette	1,521	20,272	21,793	0.070	$P=9.9x10^{-12}$
	Constitutive	7,671	85,360	93,031	0.082	$P=3.2x10^{-7}$

Table Appendix C.2: All BioCarta pathways, and clusters of BioCarta pathways that a set of genes with tissue-specific isoforms is enriched in. EASE P-values represent modified Fisher exact P-values. Column 'Benjamini' shows P-values after applying the Benjamini correction for multiple tests.

Pathway	EASE P-value	Benjamini P-value
Enriched individual pathways:		
PDZ pathway: Synaptic Proteins at the Synaptic Junction	2.3×10^{-5}	7×10^{-3}
IntegrinPathway: Integrin Signaling Pathway	0.09	0.89
MapkPathway: MAPKinase Signaling Pathway	0.06	0.90
HifPathway: Hypoxia-Inducible Factor in the Cardiovascular System	0.09	0.90
Pitx2Pathway: Multi-step Regulation of Transcription by Pitx2	0.09	0.90
p35alzheimersPathway: Deregulation of CDK5 in Alzheimers Disease	0.08	0.90
BiopeptidesPathway: Bioactive Peptide Induced Signaling Pathway	0.06	0.92
VegfPathway: VEGF, Hypoxia, and Angiogenesis	0.08	0.92
Her2Pathway: Role of ERBB2 in Signal Transduction and Oncology	0.05	0.93
CaCaMPathway: Ca++/ Calmodulin-dependent Protein Kinase Activation	0.06	0.93
NdkDynaminPathway: Endocytotic role of NDK, Phosphins and Dynamin	0.04	0.94
At1rPathway: Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	0.04	0.97
RhoPathway: Rho cell motility signaling pathway	0.02	0.98

ArapPathway: ADP-Ribosylation Factor	0.04	0.98
Enriched clusters of pathways with similar gene members:		
Cluster I		
MapkPathway: MAPKinase Signaling Pathway	0.06	0.90
p38mapkPathway: p38 MAPK Signaling Pathway	0.26	0.98
ErkPathway: Erk1/Erk2 Mapk Signaling pathway	0.35	0.99
Cluster II		
At1rPathway: Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	0.04	0.97
BiopeptidesPathway: Bioactive Peptide Induced Signaling Pathway	0.06	0.92
IntegrinPathway: Integrin Signaling Pathway	0.09	0.89
pyk2Pathway: Links between Pyk2 and Map Kinases	0.11	0.91
Fcer1Pathway: Fc Epsilon Receptor I Signaling in Mast Cells	0.18	0.97
Cxcr4Pathway: CXCR4 Signaling Pathway	0.18	0.96
EcmPathway: Erk and PI-3 Kinase Are Necessary for Collagen Binding in Corneal Epithelia	0.18	0.96
BcrPathway: BCR Signaling Pathway	0.60	1.00
MetPathway: Signaling of Hepatocyte Growth Factor Receptor	0.69	1.00
TcrPathway: T Cell Receptor Signaling Pathway	0.82	1.00