

# Table S1: Sources of genomic sequences, statistics on proteins, relative gene orientation, and clustering using OrthoMCL and Pfam

NOTE: Download date is given where no version could be established. When position information was available from protein sequences or from a correspondent feature table no nucleotide sequences were used to map the proteins. pp stands for protein prediction set release date.

ORGANISM	ABB.		VERSION		Total No. Proteins	Reduced No. Proteins	Relative gene orientation			Clustered by PFAM	Clustered by ORTHOMCL
			P	A			→→	→←	←→		
<i>PEZIZOMYCOTINA</i>											
<i>Aspergillus nidulans</i>	Aspnid	Broad	Release 4		10 701	10 699	4 689	2 899	2 879	6 703	9 019
<i>Aspergillus terreus</i>	Aspter	Broad	pp 06-05-23		10 406	10 406	4 610	2 886	2 891	6 696	9 044
<i>Botrytis cinerea</i>	Botcin	Broad	pp 05-04-10		16 448	16 448	7 375	4 242	4 324	5 874	9 957
<i>Chaetomium globosum</i>	Chaglo	Broad	pp 05-06-27		11 124	11 124	4 760	3 165	3 172	5 739	8 808
<i>Coccidioides immitis</i>	Cocimm	Broad	pp 06-25-01		10 459	10 457	4 765	2 841	2 844	4 895	8 008
<i>Fusarium oxysporum</i>	Fusoxy	Broad	FO2 07-03-27		17 609	17 606	8 176	4 661	4 673	8 906	15 543
<i>Fusarium verticillioides</i>	Fusver	Broad	FV3 06/12		14 241	14 175	6 482	3 827	3 830	7 614	12 402
<i>Gibberella zeae</i>	Gibzea	Broad	FG1 07-03-27		13 321	13 320	5 938	3 681	3 684	7 291	11 155
<i>Histoplasma capsulatum</i>	Hiscap	Broad	pp 06-24-005		9 350	9 349	4 047	2 556	2 557	4 626	7 763
<i>Magnaporthe grisea</i>	Maggr	Broad	v5.0		11 109	11 107	3 865	2 739	2 587	6 012	8 710
<i>Neurospora crassa</i>	Neucra	Broad	v3.0		9 865	9 825	4 287	2 716	2 699	5 306	7 590
<i>Sclerotinia sclerotiorum</i>	Scslscl	Broad	pp 05-10-01		14 522	14 522	6 523	3 980	3 983	5 834	9 732
<i>Stagonospora nodorum</i>	Stanod	Broad	SN15 08-05-15		16 597	16 597	8 228	4 148	4 159	7 096	10 007
<i>Uncinocarpus reesii</i>	Uncree	Broad	pp 06-05-26		7 798	7 798	3 371	2 203	2 201	4 522	6 869
<i>SACCHAROMYCOTINA</i>											
<i>Claviceps lusitanae</i>	Clalus	Broad	pp 06-01-25		5 941	5 823	2 777	1 518	1 520	3 871	5 035
<i>Candida tropicalis</i>	Cantr	Broad	pp 06-06-12		6 258	6 210	3 048	1 571	1 570	4 222	5 593
<i>Lodderomyces elongisporus</i>	Lodelo	Broad	2007-01-15		5 798	5 743	2 858	1 431	1 428	3 911	5 226
<i>Pichia guilliermondii</i>	Picgui	Broad	pp 06-15-03		5 920	5 832	2 936	1 442	1 445	4 170	5 315
<i>Saccharomyces cerevisiae</i>	Saccer	SGD	2005-02-17		6 652	6 504	2 864	1 814	1 810	4 378	4 853
<i>TAPHRINOMYCOTINA</i>											
<i>Schizosaccharomyces pombe</i>	Schpom	Broad	2008-04-21		4 995	4 987	2 280	1 350	1 351	4 042	4 174
<i>BASIDIOMYCOTA</i>											
<i>Coprinus cinereus</i>	Copcin	Broad	v1.0		13 544	13 544	6 421	3 345	3 351	5 800	9 897
<i>Cryptococcus neoformans</i>	Cryneo	Broad	v3.0		7 299	7 297	2 613	2 232	2 251	4 506	5 642
<i>Puccinia graminis</i>	Pucgra	Broad	pp 07-07-13		20 567	20 566	9 996	5 097	5 107	5 031	15 691

ORGANISM	ABB.		VERSION		Total No. Proteins	Reduced No. Proteins	Relative gene orientation			Clustered by PFAM	Clustered by ORTHOMCL
			P	A			→ →	→ ←	← →		
<b>ZYGOMYCOTA</b>											
<i>Rhizopus oryzae</i>	Rhiory	Broad	RO3 04-09-09		17 413	17 413	8 832	4 254	4 246	8 376	13 837
<b>CHYTRIDIOMYCOTA</b>											
<i>Batrachomyces dendrobatidis</i>	Batden	Broad	pp 07-01-03		8 863	8 794	4 318	2 198	2 213	4 378	6 618
<b>MICROSPORIDIA</b>											
		Antonospora locustae									
<i>Antonospora locustae</i>	Antloc	DB	2008-04-17	2008-04-17	2 606	2 470	708	430	334	881	1 923
<i>Encephalitozoon cuniculi</i>	Encunc	NCBI	2008-04-18	2008-04-18	1 996	1 945	941	497	496	1 141	1 454
<b>MAMMALS</b>											
<i>Homo sapiens</i>	Homsap	Ensembl	48.36j		46 610	22 481	10 694	5 849	5 856	16 632	19 365
<i>Mus musculus</i>	Musmus	Ensembl	48.37a		39 458	23 193	11 517	5 766	5 774	17 953	20 319
<b>AVES</b>											
<i>Gallus gallus</i>	Galgal	Ensembl	48.2f		22 195	16 597	7 817	4 364	4 364	11 900	13 772
<b>FISHES</b>											
<i>Danio rerio</i>	Danrer	NCBI	2008-02-26		32 191	31 206	16 305	7 440	7 436	22 665	25 944
<i>Fugu rubripes</i>	Fugrub	Ensembl	48.7b		21 974	21 775	9 907	4 560	4 518	16 289	19 985
<i>Gasterosteus aculeatus</i>	Gasacu	Ensembl	46.1d		27 581	20 730	10 356	4 915	4 899	16 456	19 028
<i>Oryzias latipes</i>	Orylat	Ensembl	48.1d		25 107	19 974	9 830	4 617	4 612	15 459	18 313
<i>Tetraodon nigroviridis</i>	Tetnig	Ensembl	48.1j		28 002	27 932	15 426	6 239	6 240	16 453	19 426
<b>AMPHIBIA</b>											
<i>Xenopus tropicalis</i>	Xentro	JGI	v4.1		29 843	18 811	7 831	4 124	4 055	14 674	16 403
<b>UROCHORDATA</b>											
<i>Ciona intestinalis</i>	Cioint	Ensembl	46.2g		19 858	13 896	6 819	2 720	2 729	7 797	9 890
<b>NEMATODA</b>											
<i>Brugia malayi</i>	Brumal	NCBI	2007-08-31	2007-08-31	379	278	24	5	3	196	249
<i>Caenorhabditis elegans</i>	Caeele	NCBI	2008-02-25		16 429	16 303	8 474	3 912	3 911	9 538	11 768

ORGANISM	ABB.		VERSION		Total No. Proteins	Reduced No. Proteins	Relative gene orientation			Clustered by PFAM	Clustered by ORTHOMCL
			P	A			→→	→←	←→		
<i>INSECTS</i>											
<i>Aedes aegypti</i>	Aedaeg	Ensembl	44.1a		16 789	15 324	6 091	3 902	3 719	10 573	12 981
<i>Anopheles gambiae</i>	Anogam	Ensembl	46.3i		13 133	12 359	5 936	3 208	3 209	8 658	10 514
<i>Drosophila melanogaster</i>	Dromel	Ensembl	46.43b		19 837	13 622	5 888	3 860	3 862	9 370	10 423
<i>CRUSTACEA</i>											
<i>Daphnia pulex</i>	Dappul	JGI	v1.0	v1.0	22 636	14 776	6 521	3 201	3 204	6 668	12 174
<i>CNIDARIA</i>											
<i>Nematostella vectensis</i>	Nemvec	JGI	v1.0	v1.0	115 414	17 131	5 690	4 040	3 937	1 903	14 514
<i>PLACOZOA</i>											
<i>Trichoplax adhaerens</i>	Triadh	JGI	v1.0	v1.0	2 360	2 283	1 023	626	628	1 645	1 945
<i>CHOANOFLAGELLATA</i>											
<i>Monosiga brevicollis</i>	Monbre	JGI	v1.0	v1.0	24 784	6 191	2 124	1 936	1 961	3 940	4 584
<i>MYCETOZOA</i>											
<i>Dictyostelium discoideum</i>	Dicdis	DictyBase/NCBI	2007-01-23	2008-02-20	13 345	13 343	6 655	3 342	3 340	6 868	9 107
<i>ENTAMOEBIDAE</i>											
<i>Entamoeba histolytica</i>	Enthis	Sanger	2007-08-31		17 203	9 641	4 100	1 977	1 952	4 191	8 013
<i>STREPTHOPHYTA</i>											
<i>Arabidopsis thaliana</i>	Aratha	NCBI	2005-11-16		31 711	26 812	14 060	6 373	6 374	20 040	22 596
<i>Oryza sativa</i>	Orysat	NCBI	2008-03-04		26 777	26 674	13 559	6 552	6 551	17 646	19 028
<i>GREEN ALGAE</i>											
<i>Chlamydomonas reinhardtii</i>	Chlrei	JGI	v3.0	v3.0	9 411	6 419	2 772	1 503	1 491	3 280	4 432
<i>Volvox carteri</i>	Volcar	JGI	v1.0	v1.0	5 605	4 304	1 787	1 113	1 113	1 828	3 594
<i>RED ALGAE</i>											
<i>Cyanidioschyzon merolae</i>	Cyamer	Cyanidioschyzon merolae DB	2008-02-26		5 014	4 953	1 670	1 606	1 601	3 363	3 497

ORGANISM	ABB.		VERSION		Total No. Proteins	Reduced No. Proteins	Relative gene orientation			Clustered by PFAM	Clustered by ORTHOMCL
			P	A			→→	→←	←→		
<b>PELAGOPHYTES</b>											
<i>Aureococcus anophagefferens</i>	Aurano	JGI	v1.0	v1.0	8 292	7 215	2 941	1 793	1 800	4 702	5 635
<b>DIATOMEA</b>											
<i>Thalassiosira pseudonana</i>	Thapse	JGI	v3.0	v3.0	8 542	8 227	3 187	2 505	2 508	4 412	5 782
<b>OOMYCETES</b>											
<i>Phytophthora infestans</i>	Phyinf	Broad	pp 09-09-04		17 898	17 892	7 833	4 682	4 646	7 974	13 563
<b>APICOMPLEXA</b>											
<i>Cryptosporidium parvum</i>	Crypar	NCBI	2008-02-26		3 805	3 798	1 715	1 036	1 039	1 905	2 202
<i>Plasmodium falciparum</i>	Plafal	Plasmodb	Release 5.3		27 917	9 173	4 816	2 170	2 169	3 519	6 239
<i>Theileria parva</i>	Thepar	TIGR	2007-08-30	2007-08-30	3 271	2 608	1 422	587	590	1 286	1 702
<i>Toxoplasma gondii</i>	Toxgon	Toxodb	v4.2	v4.2	4 579	4 514	2 091	1 167	1 168	1 943	2 478
<b>EUGLENOZOA</b>											
<i>Leishmania infantum</i>	Leiinf	Sanger	V5.2	V5.2	6 802	5 906	5 747	53	70	2 993	3 044
<b>HETEROLOBOSEA</b>											
<i>Naegleria gruberi</i>	Naegrü	JGI	v1.0	v1.0	52 330	12 511	5 078	3 446	3 447	6 598	9 304
<b>DIPLOMANDIDA</b>											
<i>Giardia lamblia</i>	Gialam	GiardiaDB	Release 1.1		4 889	4 790	2 275	1 240	1 237	2 188	2 699
<b>PARABASALIDEA</b>											
<i>Trichomonas vaginalis</i>	Trivag	TrichDB	Release 1		59 672	59 637	22 295	10 860	9 475	13 271	48 480