

TABLE S1. Sources of the proteome files used for PLD biocomputational HHMER profiling.

Species/Strain/Genome version	Internal Species Code ^a	Proteome File Source	Reference
<i>Albugo laibachii</i>	Alai	ftp://ftp.ensemblgenomes.org/pub/protists/release-22/fasta/albugo_laibachii/pep/Albugo_laibachii.ENA1.20.pep.all.fa	(Kemen et al. 2011)
<i>Alexandrium tamarense</i>	Ata	http://hmmer.janelia.org/search/hmmersearch	(Hackett et al. 2013)
<i>Aplanochytrium kerguelense</i> PBS07	Aker	http://genome.jgi.doe.gov/Aplke1/download/Aplke1_GeneCatalog_proteins_20121220_aa.fasta.gz	N.A. ^b
<i>Arabidopsis thaliana</i>	Atha	ftp://ftp.arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/TAIR10_pep_20110103_representative_gene_model	(Arabidopsis Genome Initiative 2000)
<i>Aspergillus niger</i> ATCC 1015 v4.0	Anig	http://genome.jgi.doe.gov/Aspni7/download/Aspni7_GeneCatalog_proteins_20131226_aa.fasta.gz	N.A. ^b
<i>Aurantiochytrium limacinum</i> ATCC MYA-1381	Alim	http://genome.jgi.doe.gov/Aurli1/download/Aurli1_GeneCatalog_proteins_20120618_aa.fasta.gz	N.A. ^b
<i>Aureococcus anophagefferens</i> CCMP1984	Aano	ftp://ftp.jgi-psf.org/pub/JGI_data/Aureococcus_anophagefferens/annotation/v1.0/proteins.Auran1_FilteredModels3.fasta.gz	(Gobler et al. 2011)
<i>Babesia bovis</i> T2Bo strain	Bbov	http://www.vetmed.wsu.edu/research_vmp/babesia-bovis/	(Brayton et al. 2007)
<i>Bathycoccus prasinos</i>	Bpra	https://bioinformatics.psb.ugent.be/gdb/bathycoccus/RELEASE_15jul2011/bathy_PROT.fasta	(Moreau et al. 2012)
<i>Bigeloviella natans</i> CCMP2755	Bnat	http://genome.jgi.doe.gov/Bigna1/download/Bigna1_filtered_proteins.fasta.gz	(Curtis et al. 2012)
<i>Bos taurus</i>	Btau	ftp://ftp.ncbi.nlm.nih.gov/genomes/Bos_taurus/protein/protein.fa.gz	Bovine Genome Seq. and Analysis Consortium 2009)
<i>Branchiostoma floridae</i> v1.0	Bflo	ftp://ftp.jgi-psf.org/pub/JGI_data/Branchiostoma_floridae/v1.0/proteins.Brafl1.fasta.gz	(Putnam et al. 2008)
<i>Chlamydomonas reinhardtii</i> CC-503 cw92 mt+	Crei	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Creinhardtii/annotation/Creinhardtii_236_protein.fa.gz	(Merchant et al. 2007)
<i>Chlorella variabilis</i> NC64A	Cvar	ftp://ftp.jgi-psf.org/pub/JGI_data/Chlorella_NC64A/annotation/v1.0/Chlorella_NC64A_best_proteins.fasta.gz	(Blanc et al. 2010)
<i>Chondrus crispus</i>	Che	http://hmmer.janelia.org/search/hmmersearch	(Collen et al. 2013)
<i>Coccomyxa subellipsoidea</i> .C-169	Csub	http://genome.jgi.doe.gov/Coc_C169_1/download/Coccomyxa_C169_v2_filtered_proteins.fasta.gz	(Blanc et al. 2012)
<i>Cryptosporidium parvum</i>	Cpar	http://cryptodb.org/common/downloads/Current_Release/Cparumlowall/fasta/data/CryptoDB-5.1_Cparumlowall_AnnotatedProteins.fasta	(Abrahamsen et al. 2004, Puiu et al. 2004)
<i>Cryptosporidium merolae</i>	Cmer	http://merolae.biol.s.u-tokyo.ac.jp/download/cds.fasta	(Nozaki et al. 2007)
<i>Cyanophora paradoxa</i>	Cypa	http://cyanophora.rutgers.edu/cyanophora/Cyanophora_paradoxa_MAKER_gene_predictions-022111-aa.fasta	(Price et al. 2012)
<i>Danio rerio</i>	Drer	ftp://ftp.ncbi.nlm.nih.gov/genomes/Danio_rerio/protein/protein.fa.gz	(Howe et al. 2013)
<i>Daphnia pulex</i> v1.0	Dpul	http://genome.jgi.doe.gov/Dappu1/download/FrozenGeneCatalog20110204.proteins.fasta.gz	(Coldbourne et al. 2011)
<i>Dictyostelium purpureum</i> QSDP1	Dpur	ftp://ftp.jgi-psf.org/pub/JGI_data/Dictyostelium_purpureum/v1.0/annotation/Dicpu1_best_proteins.fasta.gz	(Sugang et al. 2011)
<i>Ectocarpus siliculosus</i>	Esil	https://bioinformatics.psb.ugent.be/gdb/ectocarpus/Ectsi1_prot_LATEST.tfa.gz	(Cock et al. 2010)
<i>Eimeria tenella</i>	Eten	http://www.sanger.ac.uk/resources/downloads/protozoa/cimeria-tenella.html	(Lim et al. 2012)
<i>Emiliania huxleyi</i> CCMP1516	Ehux	http://genome.jgi.doe.gov/Emihu1/download/Emihu1_best_proteins.fasta.gz	(Read et al. 2013)
<i>Euglena gracilis</i>	Egra	http://hmmer.janelia.org/search/hmmersearch	(Dos Santos Ferreira et al. 2007)
<i>Fragilariopsis cylindrus</i>	Fcyl	http://genome.jgi.doe.gov/Fracy1/download/portalData/Fracy1_GeneModels_FilteredModels2_aa.fasta.gz	N.A. ^b
<i>Giardia intestinalis</i>	Gint	http://giardiadb.org/common/downloads/Current_Release/GintestinalisAssemblageA/fasta/data/GiardiaDB-3.1_GintestinalisAssemblageA_AnnotatedProteins.fasta	(Aurrecochea et al. 2009a, Franzén et al. 2009)
<i>Glycine max</i>	Gmax	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Gmax/annotation/Gmax_189_protein.fa.gz	(Schmutz et al. 2010)
<i>Guillardia theta</i> CCMP2712	Gthe	http://genome.jgi.doe.gov/Guith1/download/Guith1_GeneCatalog_proteins_20101209_aa.fasta.gz	(Curtis et al. 2012)
<i>Homo sapiens</i>	Hsap	ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/protein/protein.fa.gz	(IHGSC 2004)
<i>Ichthyophthirius multifiliis</i>	Imul	http://ich.ciliate.org/system/downloads/img1_0407_aa.fsa	(Coyne et al. 2011)
<i>Leishmania major</i>	Lmaj	ftp://ftp.sanger.ac.uk/pub/pathogens/Leishmania/major/DATASETS/LmjFwholegenome_20070731_V5.2.pep	(Ivens et al. 2005)
<i>Micromonas pusilla</i> CCMP1545 v3.0	Mpus	http://genome.jgi.doe.gov/Micpu3/download/Micpu3_GeneCatalog_proteins_20110615_aa.fasta.gz	(Worden et al. 2009)
<i>Micromonas</i> sp. RCC299	Misp	http://genome.jgi.doe.gov/MicpuN3/download/MicromonasRCC299v3.FrozenGeneCatalog_20090404.proteins.fasta.gz	(Worden et al. 2009)
<i>Naegleria gruberi</i>	Ngru	http://genome.jgi.doe.gov/Naegr1/download/Naegr1_best_proteins.fasta.gz	(Fritz-Laylin et al. 2010)
<i>Nannochloropsis gaditana</i>	Ngad	http://www.nannochloropsis.org/download/naga_prot.faa	(Radakovits et al. 2012)
<i>Nannochloropsis oceanica</i> CCMP1779	Noce	http://www.nannochloropsis.org/download/Noceanica_proteins.fasta	(Vieler et al. 2012)
<i>Nematostella vectensis</i> v1.0	Nvec	ftp://ftp.jgi-psf.org/pub/JGI_data/Nematostella_vectensis/v1.0/annotation/proteins.Nemve1FilteredModels1.fasta.gz	(Putnam et al. 2007)
<i>Neurospora tetrasperma</i> FGSC 2508 mat A v2.	Ntet	http://genome.jgi.doe.gov/Neute_mata2/download/N_tetrasperma_matA_v2_FilteredModels.proteins.fasta.gz	N.A. ^b
<i>Oryza sativa</i>	Osat	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Osativa/annotation/Osativa_204_protein.fa.gz	(Ouyang et al. 2007)
<i>Ostreococcus lucimarinus</i> CCE9901	Oluc	ftp://ftp.jgi-psf.org/pub/JGI_data/Ostreococcus_lucimarinus/O_lucimarinus.FM_aa.fasta.gz	(Palenik et al. 2007)
<i>Ostreococcus tauri</i> iOTH95	Otau	ftp://ftp.jgi-psf.org/pub/JGI_data/Ostreococcus_tauri/O_tauri.FM_aa.fasta.gz	(Derelle et al. 2006, Palenik et al. 2007)
<i>Paramecium tetraurelia</i>	Ptet	http://paramecium.cgm.cnr-s.gif.fr/download/fasta/Ptetraurelia_peptides_cur.fasta	(Arnaiz et al. 2007, Arnaiz & Sperling, 2011)
<i>Penicillium canescens</i> ATCC 10419 v1.0	Pcan	http://genome.jgi.doe.gov/Penca1/download/Penca1_GeneCatalog_proteins_20130524_aa.fasta.gz	N.A. ^b
<i>Perkinsus marinus</i> ATCC 50983	Pmar	http://www.ncbi.nlm.nih.gov/genome/proteins/280?project_id=46451	N.A. ^c
<i>Phaeodactylum tricoratum</i> v2.0	Ptri	http://genome.jgi.doe.gov/Phatr2/download/Phatr2_chromosomes_geneModels_FilteredModels2_aa.fasta.gz	(Bowler et al. 2008)
		http://genome.jgi.doe.gov/Phatr2/download/Phatr2_bd_unmapped_GeneModels_FilteredModels1_aa.fasta.gz	
<i>Phycomitrella patens</i> ssp. <i>patens</i> Gransden	Ppat	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Ppatens_v1.6/annotation/Ppatens_152_protein.fa.gz	(Rensing et al. 2008)
<i>Phytophthora infestans</i> T30-4	Pinf	http://www.broadinstitute.org/annotation/genome/phytophthora_infestans/download/?sp=EAProteinsFasta&sp=SPI_T30-4&sp=S.zip	(Haas et al. 2009)
<i>Plasmodium berguei</i>	Pber	http://plasmodb.org/common/downloads/Current_Release/PbergheiANKA/fasta/data/PlasmoDB-10.0_PbergheiANKA_AnnotatedProteins.fasta	(Aurrecochea et al. 2009b)
<i>Plasmodium chabaudi</i>	Pchau	http://plasmodb.org/common/downloads/Current_Release/PchabaudiChabaudi/fasta/data/PlasmoDB-10.0_PchabaudiChabaudi_AnnotatedProteins.fasta	(Aurrecochea et al. 2009b)
<i>Plasmodium falciparum</i>	Pfal	http://plasmodb.org/common/downloads/Current_Release/Pfalciparum3D7/fasta/data/PlasmoDB-10.0_Pfalciparum3D7_AnnotatedProteins.fasta	(Gardner et al. 2002, Aurrecochea et al. 2009b)
<i>Plasmodium knowlesi</i>	Pkno	http://plasmodb.org/common/downloads/Current_Release/PknowlesiH/fasta/data/PlasmoDB-10.0_PknowlesiH_AnnotatedProteins.fasta	(Aurrecochea et al. 2009b)
<i>Plasmodium vivax</i>	Pvix	http://plasmodb.org/common/downloads/Current_Release/PvivaxSal1/fasta/data/PlasmoDB-10.0_PvivaxSal1_AnnotatedProteins.fasta	(Aurrecochea et al. 2009b)
<i>Porphyridium purpureum</i>	Ppur	http://cyanophora.rutgers.edu/porphyridium/Porphyridium_genemodels_UPDATED.fasta	(Bhattacharya et al. 2013)
<i>Pseudo-nitzschia multiseriata</i> CLN-47 v1.0	Pmul	http://genome.jgi.doe.gov/Psemu1/download/Psemu1_GeneCatalog_proteins_20111101_aa.fasta.gz	N.A. ^b
<i>Pythium irregulare</i>	Pirr	ftp://ftp.ensemblgenomes.org/pub/protists/release-22/fasta/pythium_irregulare/pep/Pythium_irregulare.GCA_000387425.2.22.pep.all.fa.gz	(Adhikari et al. 2013)
<i>Rhizopus microsporus</i> var. <i>microsporus</i> v1.0	Rmic	http://genome.jgi.doe.gov/Rhimi1_1/download/Rhimi1_1_GeneCatalog_proteins_20130624_aa.fasta.gz	N.A. ^b
<i>Saccharomyces cerevisiae</i> M3707 v1.0	Scer	http://genome.jgi.doe.gov/SacceM3707_1/download/SacceM3707_1_GeneCatalog_proteins_20120712_aa.fasta.gz	N.A. ^b
<i>Schizochytrium aggregatum</i> ATCC 28209 v1.0	Sagg	http://genome.jgi.doe.gov/Schag1/download/Schag1_GeneCatalog_proteins_20121220_aa.fasta.gz	N.A. ^b
<i>Selaginella moellendorffii</i>	Smoe	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/SmoeIIendorffii/annotation/SmoeIIendorffii_91_protein.fa.gz	(Banks et al. 2011)
<i>Solanum lycopersicum</i>	Slyc	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Slycopersicum/annotation/Slycopersicum_225_protein.fa.gz	(Tomato Genome Consortium 2012)
<i>Symbiodinium minutum</i>	Smin	http://marinegenomics.oist.jp/genomes/download/symbB.v1.2.augustus.prot.fa.gz	(Shoguchi et al. 2013)

<i>Tetrahymena thermophila</i>	Tthe	http://www.ciliate.org/system/downloads/T_thermophila_oct2008_proteins.fasta	(Eisen et al. 2006)
<i>Thalassiosira pseudonana</i> v3.0	Tpse	http://genome.jgi.doe.gov/Thaps3/download/Thaps3_chromosomes_geneModels_FilteredModels2_aa.fasta.gz http://genome.jgi.doe.gov/Thaps3/download/Thaps3_bd_unmapped_GeneModels_FilteredModels1_aa.fasta.gz	(Armbrust et al. 2004)
<i>Theileria annulata</i>	Tann	ftp://ftp.sanger.ac.uk/pub/pathogens/T_annulata/TANN.GeneDB.pep	(Pain et al. 2005)
<i>Theileria parva</i>	Tpar	http://www.ncbi.nlm.nih.gov/genome/proteins/38?project_id=16136	(Gardner et al. 2005, Hayashida et al. 2013)
<i>Toxoplasma gondii</i>	Tgon	http://toxodb.org/common/downloads/Current_Release/TgondiiGT1/fasta/data/ToxoDB-9.0_TgondiiGT1_AnnotatedProteins.fasta	(Kissinger et al. 2003, Bontell et al. 2009)
<i>Trichoderma longibrachiatum</i> ATCC 18648 v3.	Tlon	http://genome.jgi.doe.gov/Trilo3/download/Trilo3_GeneCatalog_proteins_20130918_aa.fasta.gz	N.A. ^b
<i>Trichomonas vaginalis</i>	Tvag	http://trichdb.org/common/downloads/Current_Release/Tvaginalis/fasta/TvaginalisAnnotatedProteins_TrichDB-1.3.fasta	(Carlton et al. 2007, Aurrecochea et al. 2009a)
<i>Trypanosoma brucei</i>	Tbru	ftp://ftp.sanger.ac.uk/pub/pathogens/Trypanosoma/brucei/Latest_Whole_Genome_Sequence/Tb927_WGS_2013_08_01/fasta/Tb927_all_v5.1_annotated_proteins_20130801.fa	(Aslett et al. 2010, Jackson et al. 2010)
<i>Trypanosoma congolense</i>	Tcon	ftp://ftp.sanger.ac.uk/pub/pathogens/Trypanosoma/congolense/GeneDB_Tcongolense_Protein_v1.fas.gz	(Aslett et al. 2010)
<i>Trypanosoma cruzi</i>	Teru	http://tritrypdb.org/common/downloads/Current_Release/TcruziCLBrener/fasta/data/TriTrypDB-6.0_TcruziCLBrener_AnnotatedProteins.fasta	(El-Sayed et al. 2005, Aslett et al. 2010)
<i>Trypanosoma vivax</i>	Tviv	http://tritrypdb.org/common/downloads/Current_Release/TvivaxY486/fasta/data/TriTrypDB-6.0_TvivaxY486_AnnotatedProteins.fasta	(Aslett et al. 2010)
<i>Vitis vinifera</i>	Vvin	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Vvinifera/annotation/Vvinifera_145_protein.fas.gz	(Jaillon et al. 2007)
<i>Volvox carteri fnagariensis</i>	Vcar	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Vcarteri/annotation/Vcarteri_199_protein.fas.gz	(Prochnik et al. 2010)
<i>Xenopus tropicalis</i> v4.1	Xtro	ftp://ftp.jgi-psf.org/pub/JGI_data/Frog/v4.1/proteins.Xentr4.fasta.gz	(Hellsten et al. 2010)
<i>Zea mays</i>	Zmay	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Zmays/annotation/Zmays_181_protein.fas.gz	(Schmable et al. 2009)

^a It refers to the species code that was used in the compiled fasta file (e.g. if a sequence is Zmay000001, it corresponds to protein #000001 from *Zea mays*)

^b These sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community, permission to use proteome data was obtained from the Principal Collaborators.

^c The *Perkinsus marinus* Genome Sequencing Project, joint effort between J. Craig Venter Institute (formerly TIGR) and the Center for Marine Biotechnology (COMB).

Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S. et al. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 304:441-5.

Adhikari, B. N., Hamilton, J. P., Zerillo, M. M., Tisserat, N., Lévesque, C. A. & Buell, C. R. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS One* 8:e75072.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M. et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. 306:79-86.

Arnaiz, O. & Sperling, L. 2011. ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. *Nucleic Acids Res.* 39:D632-6.

Arnaiz, O., Cain, S., Cohen, J. & Sperling, L. 2007. ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. *Nucleic Acids Res.* 35:D439-44.

Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X. et al. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 38:D457-62.

Aurrecochea, C., Brestelli, J., Brunk, B. P., Carlton, J. M., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G. et al. 2009b. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 37:D526-30.

Aurrecochea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S. et al. 2009a. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* 37:D539-43.

Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., Albert, V. A., Aono, N., Aoyama, T., Ambrose, B. A. et al. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science*. 332:960-3.

Bhattacharya, D., Price, D. C., Chan, C. X., Qiu, H., Rose, N., Ball, S., Weber, A. P., Arias, M. C., Henrissat, B., Coutinho, P. M. et al. 2013. Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* 4:1941.

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S. et al. 2012. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* 13:R39.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S.,

- Pangilinan, J., Polle, J. 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*. 22:2943-55.
- Bontell, I. L., Hall, N., Ashelford, K. E., Dubey, J. P., Boyle, J. P., Lindh, J. & Smith, J.E. 2009. Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. *Genome Biol.* 10:R53.
- Bovine Genome Sequencing and Analysis Consortium. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324:522-8.
- Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otilar, R. P. et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*. 456:239-44.
- Brayton, K. A., Lau, A. O. T., Herndon, D. R., Hannick, L., Kappmeyer, L. S., Berens, S. J., Bidwell, S. L., Brown, W. C., Crabtree, J., Fadrosch, D. et al. 2007. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog.* 3:1401-13.
- Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C. M., Besteiro, S. et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. 315:207-12.
- Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J. M., Badger, J. H. et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617-21.
- Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K. et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331:555-61.
- Collén, J., Porcel, B., Carré, W., Ball, S. G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K. et al. 2013. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. U.S.A.* 110:5247-52.
- Coyne, R. S., Hannick, L., Shanmugam, D., Hostetler, J. B., Bami, D., Joardar, V. S., Johnson, J., Radune, D., Singh, I., Badger, J. H. et al. 2011. Comparative genomics of the pathogenic ciliate *Ichthyophthirius multifiliis*, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. *Genome Biol.* 12:R100.
- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M. C., Ball, S. G., Gile, G. H., Hirakawa, Y. et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*. 492:59-65.
- Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R. et al. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U.S.A.* 103:11647-52.
- Dos Santos Ferreira, V., Rocchetta, I., Conforti, V., Bench, S., Feldman, R. & Levin, M.J. 2007. Gene expression patterns in *Euglena gracilis*: insights into the cellular response to environmental stress.

Gene. 389:136-45.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M. et al. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* 4:e286.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G. et al. 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 309:409-15.

Franzén, O., Jerlström-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D. S., Palm, D., Andersson, J. O., Andersson, B. & Svärd, S.G. 2009. Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species?. *PLoS Pathog.* 5:e1000560.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredes, A., Chapman, J., Pham, J. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell*. 140:631-42.

Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M. et al. 2005. Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science*. 309:134-7.

Gobler, C. J., Berry, D. L., Dyhrman, S. T., Wilhelm, S. W., Salamov, A., Lobanov, A. V., Zhang, Y., Collier, J. L., Wurch, L. L., Kustka, A. B. et al. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl. Acad. Sci. U.S.A.* 108:4352-7.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T. et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*. 461:393-8.

Hackett, J. D., Wisecaver, J. H., Brosnahan, M. L., Kulis, D. M., Anderson, D. M., Bhattacharya, D., Plumley, F. G. & Erdner, D.L. 2013. Evolution of saxitoxin synthesis in cyanobacteria and dinoflagellates. *Mol. Biol. Evol.* 30:70-8.

Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D. & Sugimoto, C. 2013. Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. *DNA Res.* 20:209-20.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L. et al. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science*. 328:633-6.

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren K., Matthews, L. et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498-503.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature*. 431:931-45.

- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M., Adlem, E., Aert, R. et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 309:436-42.
- Jackson, A. P., Sanders, M., Berry, A., McQuillan, J., Aslett, M. A., Quail, M. A., Chukualim, B., Capewell, P., MacLeod, A., Melville, S. E. et al. 2010. The genome sequence of *Trypanosoma brucei* gambiense, causative agent of chronic human african trypanosomiasis. *PLoS Negl Trop Dis*. 4:e658.
- Jaillon, O., Aury, J., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C. et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 449:463-7.
- Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A. C., Balmuth, A. L., Robert-Seilaniantz, A., Bailey, K., Holub, E., Studholme, D. J., Maclean, D. & Jones, J. D. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol*. 9:e1001094.
- Kissinger, J. C., Gajria, B., Li, L., Paulsen, I. T. & Roos, D.S. 2003. ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res*. 31:234-6.
- Lim, L., Tay, Y., Alias, H., Wan, K. & Dear, P.H. 2012. Insights into the genome structure and copy-number variation of *Eimeria tenella*. *BMC Genomics*. 13:389.
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L. et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. 318:245-50.
- Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F. et al. 2012. Gene functionalities and genome structure in *Bathycoccus prasinus* reflect cellular specializations at the base of the green lineage. *Genome Biol*. 13R74.
- Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T. et al. 2007. A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biol*. 5:28.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L. et al. 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res*. 35:D883-7.
- Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C. et al. 2005. Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science*. 309:131-3.
- Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S. et al. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U.S.A.* 104:7705-10.

- Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A P, Schwacke R, Gross J, Blouin NA, Lane C et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science*. 335:843-7.
- Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L. K. et al. 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*. 329:223-6.
- Puiu, D., Enomoto, S., Buck, G. A., Abrahamsen, M. S. & Kissinger, J.C. 2004. CryptoDB: the *Cryptosporidium* genome resource. *Nucleic Acids Res*. 32:D329-31.
- Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K. et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-71.
- Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V. et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86-94.
- Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L. & Posewitz, M.C. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat Commun*. 3:686.
- Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A. et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature*. 499:209-13.
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P., Lindquist, E. A., Kamisugi, Y. et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*. 319:64-9.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J. et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature*. 463:178-83.
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A. et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 326:1112-5.
- Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M. et al. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr Biol*. 23:1399-408.
- Sucgang, R., Kuo, A., Tian, X., Salerno, W., Parikh, A., Feasley, C. L., Dalin, E., Tu, H., Huang, E., Barry, K. et al. 2011. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome Biol*. 12:R20.
- Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635-41.

Vieler, A., Wu, G., Tsai, C. H., Bullard, B., Cornish, A. J., Harvey, C., Reza, I. B., Thornburg, C., Achawanantakun, R., Buehl, C. J. et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* 8:e1003064.

Worden, A. Z., Lee, J., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V. et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science.* 324:268-72.