

Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species

Cindy Martens, Klaas Vandepoele, and Yves Van de Peer*

Department of Plant Systems Biology, VIB, and Department of Molecular Genetics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

Communicated by Marc C. E. Van Montagu, Ghent University, Ghent, Belgium, December 27, 2007 (received for review October 31, 2007)

The chromalveolates form a highly diverse and fascinating assemblage of organisms, ranging from obligatory parasites such as *Plasmodium* to free-living ciliates and algae such as kelps, diatoms, and dinoflagellates. Many of the species in this monophyletic grouping are of major medical, ecological, and economical importance. Nevertheless, their genome evolution is much less well studied than that of higher plants, animals, or fungi. In the current study, we have analyzed and compared 12 chromalveolate species for which whole-sequence information is available and provide a detailed picture on gene loss and gene gain in the different lineages. As expected, many gene loss and gain events can be directly correlated with the lifestyle and specific adaptations of the organisms studied. For instance, in the obligate intracellular Apicomplexa we observed massive loss of genes that play a role in general basic processes such as amino acid, carbohydrate, and lipid metabolism, reflecting the transition of a free-living to an obligate intracellular lifestyle. In contrast, many gene families show species-specific expansions, such as those in the plant pathogen oomycete *Phytophthora* that are involved in degrading the plant cell wall polysaccharides to facilitate the pathogen invasion process. In general, chromalveolates show a tremendous difference in genome structure and evolution and in the number of genes they have lost or gained either through duplication or horizontal gene transfer.

gene gain | gene loss

The chromalveolates (1) account for about half of the recognized species of protists (e.g., apicomplexans and ciliates) and algae (e.g., kelps, diatoms, and dinoflagellates), and are of great ecological and economical importance. They are further subdivided into six diverse groups (2, 3). The Alveolata contain the parasitic apicomplexans, the plastid-less ciliates, and the dinoflagellate algae, whereas the Chromista include the cryptophyte, haptophyte, and stramenopile algae.

Despite general shared features, the alveolates have greatly diverged in many respects, including host specificities, tissue tropisms, and the requirement of multiple hosts (4). The obligate intracellular apicomplexans contain several protozoan pathogens that provoke severe diseases in mammals, including humans. Infections by *Plasmodium falciparum*, which causes human malaria, and *Theileria parva* and *T. annulata*, which are responsible for great economic losses in cattle in Africa, have profound medical, social, and economic effects (5, 6). Others such as *Toxoplasma gondii*, *Cryptosporidium parvum*, and *C. hominis* are primarily health threats in HIV+/AIDS and immunosuppressed populations (7). Ciliates are unique among unicellular organisms in that they separate germline and somatic functions (8).

The haptophytes and stramenopiles share characteristics such as tubular mitochondrial cristae, similar storage products, and fucoxanthin that suggest a specific relationship between these taxa (9). Stramenopiles include autotrophic and heterotrophic species that may differ enormously in their morphology and mode of life. Members of this group occupy key ecological niches in marine environments. For example, the diatoms constitute the most abundant component of marine plankton. Stramenopiles

have also succeeded in occupying terrestrial environments as plant pathogens. For example, *Phytophthora* plant pathogens attack a wide range of agriculturally and ornamentally important plants (10). *Phytophthora sojae* costs the soybean industry millions of dollars each year (11). In California and Oregon, a newly emerged *Phytophthora* species, *P. ramorum*, is responsible for a disease called sudden oak death that affects not only the live oaks that are the keystone species of the ecosystem but also a large variety of woody shrubs that inhabit the oak ecosystems, such as bay laurel and viburnum (12).

All extant species of the chromist and alveolate groups are considered to have evolved from an ancestor that contained a red algal endosymbiont (the “chromalveolate” hypothesis), which later gave rise to the plastids found in most of the chromalveolates. Although there is ample molecular evidence that suggests that the alveolates and stramenopiles form a monophyletic group (13–17), the phylogenetic position of the cryptophytes and haptophytes, and their inclusion in the chromalveolate “superclade” is still debated. Recently, however, Hackett and colleagues (18) found additional support for the monophyly of the cryptophytes and haptophytes and their sister relationship to the chromalveolates based on a multigene dataset of nuclear genes.

Here, we have analyzed and compared 12 diverse chromalveolate species for which whole-sequence information is available. The aims of this analysis were to study the genome evolution for these organisms and to document gene loss and gene gain along the different chromalveolate lineages. Furthermore, we investigated the extent to which these events can be correlated with the differences in lifestyle between the various species.

Results and Discussion

A Parsimonious Scenario of Gene Loss and Gene Gain in the Chromalveolates. To study gene family evolution in this eukaryotic group of organisms, we assembled the publicly available protein sequences of 12 chromalveolate and eight outgroup species. This resulted in a dataset of 306,696 proteins that were grouped, based on sequence similarity and Markov clustering (see *Methods*), into 32,887 different multigene families and 58,331 orphans—i.e., genes without homology to others in the dataset. For all gene families, phylogenetic profiles were constructed that reflect the presence or absence of a particular gene family in a particular species. To make sure that an apparent gene family absence is not due to errors in annotation, or to increased evolutionary rates hiding “clear” homology, additional analyses

Author contributions: C.M., K.V., and Y.V.d.P. designed research; C.M. and K.V. performed research; C.M. and K.V. analyzed data; and C.M. and Y.V.d.P. wrote the paper.

The authors declare no conflict of interest.

*To whom correspondence should be addressed. E-mail: yves.vandeppeer@psb.ugent.be.

This article contains supporting information online at www.pnas.org/cgi/content/full/0712248105/DC1.

© 2008 by The National Academy of Sciences of the USA

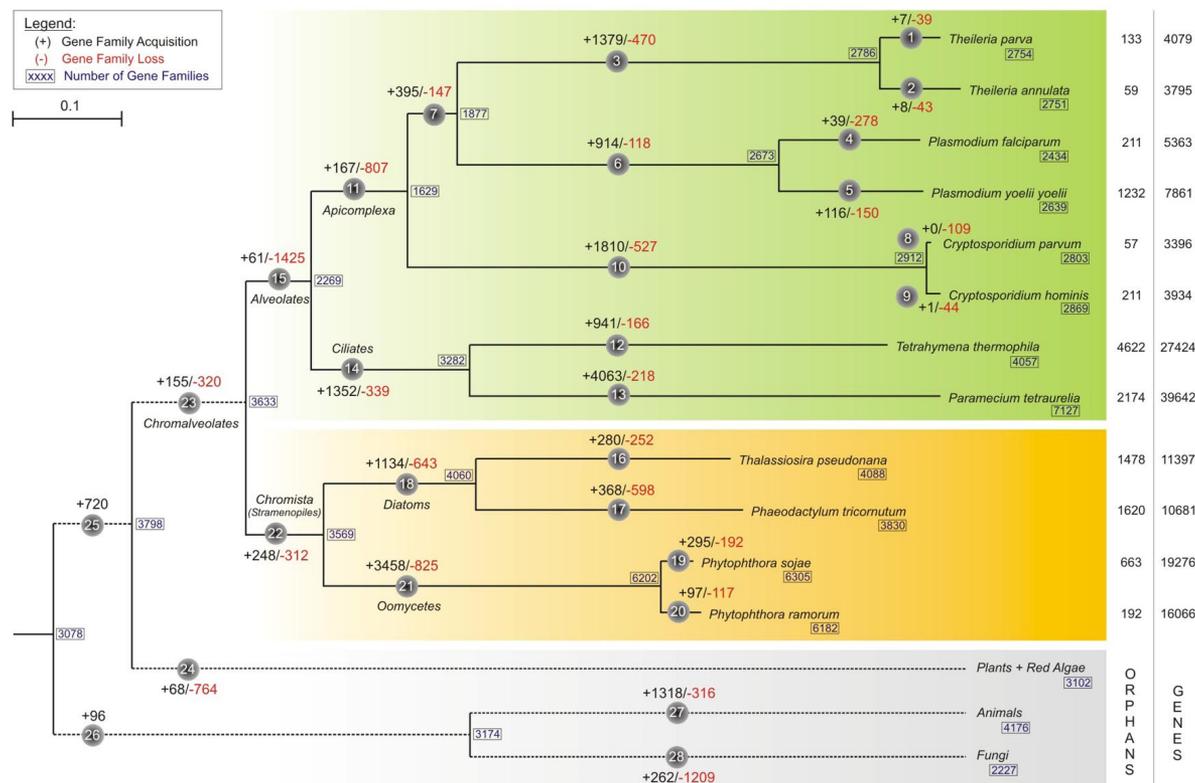


Fig. 1. A parsimonious scenario of gene loss and gene gain in the chromalveolates. At every time point (gray circles) in the chromalveolate tree (distance method, Poisson correction, NJ, all nodes 100% BS-supported), gene loss (red) and gene acquisition (black) was counted based on the Dollo parsimony principle. The inferred ancestral gene sets are shown in boxes. Next to every species, the number of orphans and number of predicted genes is indicated. The part of the tree drawn in solid black lines is drawn to scale; the part drawn in dashed lines is not. Detailed information on which genes and gene families have been gained or lost at certain points in chromalveolate evolution can be obtained from a clickable map of this figure at our web site (<http://bioinformatics.psb.ugent.be>).

were performed to see whether these two phenomena could have seriously biased our results [see [supporting information \(SI Methods\)](#)]. Based on these analyses, it seems that both annotation errors and increased evolutionary rates probably only had a very minor effect on our results ([SI Fig. 4](#) and [SI Table 1](#)). To be able to study gene family evolution by means of phylogenetic profiles, information about the phylogenetic relationships between the different organisms in the dataset is required. Therefore, we have used a concatenated set of single copy core genes (20 families) to resolve the phylogenetic relationships between the 12 chromalveolates. After extracting these 20 families, multiple alignments were created and manually improved, and finally concatenated into one large multiple alignment of 5,360 aa (see [SI Methods](#)). The inferred phylogenetic tree is shown in Fig. 1.

Combining both the species tree and phylogenetic profiles, we were able to reconstruct a parsimonious scenario of chromalveolate genome evolution by mapping different gene loss and gain events onto the branches of the phylogenetic tree. Furthermore, to delineate the minimal gene set for the different ancestral nodes, the Dollo parsimony approach was applied (19), which assumes gene loss to be irreversible. In other words, a gene (family) can be lost independently in several evolutionary lineages but cannot re-evolve (except in a noticeably different form). By applying this methodology to all phylogenetic profiles (which were treated as strings of 0 and 1; i.e., absence resp. presence of the given gene family in the given species), we could associate every branch of the chromalveolate tree (defined as time points or TPs) with both gene family loss and gene family acquisition (Fig. 1). For more information regarding the performance of the Dollo parsimony principle on our data, see [SI Methods](#) and [SI Fig. 5](#). To link the different evolutionary events

(i.e., loss and gain) with information about the molecular function or biological process the corresponding gene families are involved in, the Gene Ontology (GO) vocabulary was used (20). Detailed information on which genes and gene families have been gained or lost during chromalveolate evolution can be found at <http://bioinformatics.psb.ugent.be>.

One of the most notable observations is the occurrence of massive gene loss in the branches leading to the obligate intracellular Apicomplexa [i.e., TP11 and TP15 (Fig. 1)]. This massive gene loss can be linked to the transition of a free-living to an obligate intracellular lifestyle (21). In addition, *Phytophthora* species seem to have lost a very large number of gene families (825 families, TP21). Although this observation seems surprising given their large number of genes [$>16,000$ (Fig. 1)], this massive loss is again explained by the lifestyle of these successful plant pathogens. When we consider the functions of these families in both lineages ([SI Table 2](#), TP11, TP15, and TP21), we indeed observe that these organisms have lost many gene families involved in general processes such as amino acid metabolism, carbohydrate metabolism, and lipid metabolism.

Interestingly, the continuous loss of gene families (i.e., loss at different, successive TPs) in the Apicomplexa lineage is compensated by a similarly continuous acquisition of novel gene families in this lineage. As could be expected from the lifestyle of the Apicomplexa, we observed a significant gain (q -value <0.05) of gene families involved in pathogenesis in the ancestor of the Apicomplexa ([SI Table 3](#), TP11). Similarly, gain of pathogenesis genes was also observed (q -value <0.05) at other TPs within the Apicomplexa lineage (TP4, TP6, and TP7). We also noted a significant gain of gene families with metalloproteinase activity at TP6, TP7, and TP11. For the malaria

parasites (TP6), this can be explained by the fact that they hydrolyze host proteins (hemoglobin), using acid cysteine, aspartic, and metalloproteases. Hemoglobin is thereby processed into individual amino acids, which are subsequently used for their own protein synthesis (22). However, other Apicomplexa have also gained these metalloproteases, which suggests that they also are able to hydrolyze their host proteins and can afford losing gene families involved in amino acid biosynthesis (SI Table 3, TP11).

Phytophthora, unlike the apicomplexans, seems to have acquired a huge number of novel gene families (3,458 families, TP21). We found no significant bias toward certain biological processes or functions for these newly acquired gene families, most probably because of the lack of functional annotation for many of these specific families. However, in *Phytophthora ramorum*, we observe a significant gain of gene families with potassium channel activity (SI Table 3, TP20), which can be explained by the fact that potassium ions play a key role in the locomotion and encystment of the *Phytophthora* zoospores (23). The process of encystment is directly correlated with the oomycete pathogenicity because this process allows the zoospore to perpetuate its life cycle at the expense of the infected organism.

The ciliates, which have by far the largest number of predicted genes (Fig. 1), have acquired many new gene families. Both ciliates show an enrichment of gene families with protein kinase activity and phosphotransferase activity (SI Table 3, TP12 and TP13), which is most likely correlated with the important process of exocytosis within the ciliates, because reversible protein phosphorylation is essential in controlling exocytosis (24–26). Compared with the large number of novel gene families, the number of gene families in ciliates that have been lost is quite limited (339 families, TP14).

All chromalveolates and plants are bikonts (i.e., ancestrally biciliate, or having two cilia), this in contrast to the Metazoa and Fungi, which are unikonts. However, during evolution, in all major eukaryote groups except the ciliates, different organisms have lost cilia and became secondary uniciliates (27). Our analyses indicate that among the chromalveolates, the diatoms, but also the Apicomplexa have also lost gene families responsible for cilium biogenesis (SI Table 2, TP18 resp. TP11), which is in concordance with the Apicomplexa lacking cilia and (at least the centric) diatoms (such as *Thalassiosira*) that became secondary uniciliates (27).

We also investigated the fraction of the lineage-specific gene families for the different major chromalveolate lineages—i.e., the alveolates (Fig. 1, TP15), Chromista (TP22), and Apicomplexa (TP11)—that has been maintained in all extant species (i.e., lineage-specific core gene families). These gene families probably contain the genes responsible for the shared, specific features of these lineages. However, for the alveolates, only 5 gene families of the 61 (or 8%) are still present in all extant species. These genes play a role in protein amino acid ADP-ribosylation, lipid metabolism, and electron transport. Within the Apicomplexa, 37 families (22%) of the 167 acquired families have been preserved. A significant fraction [q -value of 0.16, p -value <0.05 (data not shown)] of these gene families is involved in pathogenesis, which suggests that all Apicomplexa share a general set of pathogenicity genes. For the Chromista, 83 gene families (33%) are still present in all extant species, with a significant enrichment in prostaglandin metabolism and regulation of cell redox homeostasis [$q = 0.37$, $p < 0.05$ (data not shown)]. Finally, we observed a gain of 155 gene families in the ancestors of the chromalveolates (TP23) of which only two have survived in all extant species but, unfortunately, without informative gene descriptions available (i.e., hypothetical proteins).

The parsimony analysis also involves explicit reconstruction of the gene sets of ancestral chromalveolate genomes. Under this approach, 3,633 gene families were assigned to the last common

ancestor of the chromalveolate supergroup (TP23). The last common ancestor of the Apicomplexa, in contrast, had the smallest number of gene families (i.e., 1,629), which implies that genome reduction had already occurred early in the evolution of these parasites.

Gene Family Expansions. The results discussed so far focused on the presence or absence of gene families but not on the actual gene family sizes. Comparing the size of all gene families makes it possible to determine whether the variations in total gene number between the different species can, at least partially, be explained by gene amplification in certain gene families. Therefore, we also investigated lineage-specific gene family expansions during chromalveolate evolution. First, the 25% most variable multigene families were identified (see *Methods*). To compare these different phylogenetic profiles, the gene copy numbers were transferred into z -scores, which are expressed as standard deviations from their means (i.e., mean copy number over the species containing the gene family). This way, positive z -scores indicate gene copy numbers that are greater than the corresponding mean gene family size. To be able to observe different expansion patterns, these z -score profiles were hierarchically clustered. We removed the gene families that were only expanded in the outgroups and the gene families only expanded in one or both of the ciliates because their massive duplication events (28, 29) reduce the expansion signal of the other chromalveolates. This analysis yielded 119 gene families (see *Methods*) (Fig. 2). Again, detailed information on these gene families can be found at <http://bioinformatics.psb.ugent.be>.

As can be observed in Fig. 2, many of the z -score profiles show *Phytophthora*-specific expansions. Strikingly, many of these expanded families are involved in degrading the plant cell wall polysaccharides, which facilitates the pathogen invasion process. For example, the xylosidases (Fig. 2, families 259 and 842), which hydrolyze the xyloglucan network, help the physical penetration process by sufficient loosening of the host cell wall (30). The polygalacturonases (family 639), on the other hand, are involved in degrading the pectin backbone embedded in the plant cell wall. Similarly, the expanded endoglucanases (family 473) and exoglucanases (family 747) are involved in the degradation of cellulose. Moreover, the latter two enzymes are shown to be differentially expressed during the life cycle of the potato-pathogen *Phytophthora infestans*, as well as during the infection of potatoes (31). Additionally, the β -(1, 3)-endoglucanases are cell wall-associated in *P. ramorum*, suggesting an important role for this gene family in the interaction with the host plants (32). Finally, in *P. sojae*, these genes were found to be specifically expressed during infection (33).

Another group of gene families that are expanded in the *Phytophthora* genomes are families involved in the protection against oxidative stress, such as for example, the glutathione S -transferases (Fig. 2, families 3403 and 990), glutaredoxin (family 3430), and the catalase-peroxidase (family 220). These families were classified as putative pathogenicity genes by Torto-Alalibo and colleagues (33). Finally, the expansion of two other gene families can be explained by their role in the process of pathogenesis: first, the glucokinases (family 2414) that are expressed during the first 6 h of interaction with its host *Glycine max* (34), and second, the lipases (family 845), which are also shown to be potentially infection related (11).

In *Plasmodium* spp., three families are specifically expanded (Fig. 2, families 1526, 3, and 1997). One family, which is an Apicomplexa-specific gene family of trophozoite antigens R45, is only expanded in *P. falciparum*. This has also been observed by Schneider and Mercereau-Puijalon (35), who suggested that expansion of this family might contribute to *P. falciparum* virulence. The second gene family, a family of erythrocyte membrane-associated antigens, is expanded in both *Plasmodium* species and also plays a role in

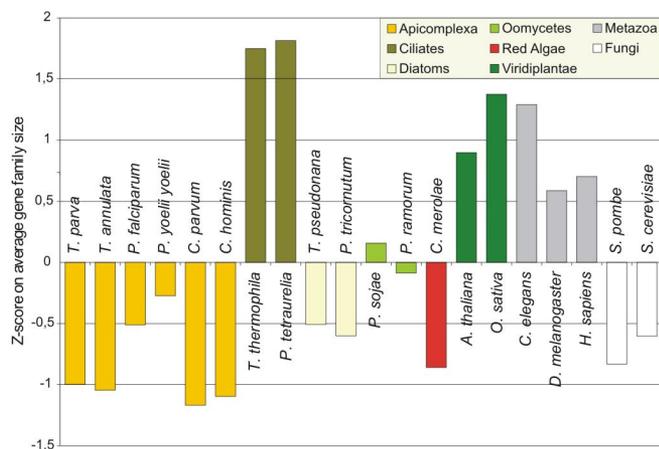


Fig. 3. Differences in average gene family size for different eukaryotes. For all organisms included in the current study, the average gene family size, converted into a z-score, is plotted on the ordinate. See text for details.

finally the other Apicomplexa. This order is in accordance with the number of genes in the corresponding genomes. To our knowledge, so far no evidence was found for whole-genome duplications within the oomycetes, but our analyses indicate the existence of a large number of duplicated blocks (especially in *P. sojae*), as well as a high fraction of tandem duplications (SI Table 4). This observation will need some further study. In addition, the pathogen *Plasmodium falciparum* has a considerable part of its genome located in duplicated blocks [i.e., 21% (SI Table 4)]. However, whereas the duplicated blocks in the other genomes are randomly distributed over the whole genome, in *P. falciparum* they seem to be preferentially located at the telomeres and occasionally at the centromeres (data not shown).

Origin of the Chromalveolate Orphans. Besides the set of genes that could be assigned to gene families, a substantial number of genes were uncovered that lack homologs in the eukaryotic data set. These so-called “orphans” represent *de novo* species-specific genes, falsely predicted gene models, or genes acquired through horizontal gene transfer (HGT) from other species (e.g., Bacteria, Viruses). To investigate the extent to which these genes have been obtained through HGT, all orphan genes and putative species-specific families (SSFs) were compared against a nonredundant protein database. After selection of highly similar sequences [using a relative bit score threshold (see SI Methods)], taxonomic information for all homologs was retrieved.

One first observation is that the orphans of four species (namely both *Theileria* species, *Cryptosporidium parvum*, and *Plasmodium falciparum*) do not show any homology with other genes at all. Whereas a small set of orphans (between 1 and 45) from *Plasmodium yoelii yoelii*, *Cryptosporidium hominis*, both *Phytophthora* species, and ciliates shows homology with bacterial species from a limited number of phyla, for both diatoms >100 orphan genes match a wide range of phyla (data not shown). In general, the fraction of orphans homologous to sequences from Archaea or Viruses is much smaller compared with genes showing homology with Bacterial genes (mostly Proteobacteria), suggesting that HGT between Bacteria and chromalveolates is most common. As for orphan genes, in most species only a minority of the SSFs have bacterial homologs, with *Thalassiosira* as an exception, where 42% (28/67) of all SSFs again match a wide range of phyla (data not shown).

We set out to document the genome evolution of a wide diversity of chromalveolate organisms. To this end, we in-

ferred, by using a Dollo parsimony approach, both gene loss and gain events that have occurred along the chromalveolate tree. The great differences in genome content reflect the ancient history of the chromalveolate assemblage. Indeed, starting from an estimated 3,600 gene families in the ancestor of all chromalveolates, the number of gene families in extant chromalveolates ranges from <2,500 to >7,000, while the number of predicted genes ranges from ≈3,800 to almost 40,000. Extensive gene loss and, contrarily, gene family expansions explain the large differences in lifestyle but also many specific adaptations and the high degree of specialization some organisms within this eukaryotic “superkingdom” have undergone.

Methods

Gene Family Clustering. To delineate gene families, a similarity search was performed (all-against-all BLASTP; E-value cutoff E-5) with all proteins from 12 different chromalveolate species, plus the protein sequences from eight outgroup organisms. For more information regarding the dataset, see SI Methods. Gene families were constructed with MCLBLASTLINE (inflation factor of 2.0) based on the BLASTP analyses. The performance of the MCLBLASTLINE protein clustering method was evaluated and compared with two other clustering methods (see SI Methods). We also evaluated the possible effects of errors in annotation and fast-evolving genes on gene family clustering (see SI Methods).

Functional Annotation of Genes and Gene Families. Genes and gene families were functionally annotated by using Gene Ontology (GO) (20). In a first step, all proteins were annotated by using Blast2GO to assign the proteins to GO categories (43). Proteins mapped to a particular GO category were also explicitly included into all parental categories. GO annotation per family was obtained by listing the GO labels for all of the genes of that family. A weight, equal to the percentage of genes with GO annotation within the same subcategory (molecular function, cellular component, biological process) that carried this label, was attached to all of the GO labels. Only GO labels with a weight greater than 30% were considered as representative for the family. GO labels occurring in fewer than 10 gene families were discarded. The statistical significance of functional GO enrichment was evaluated by using the hypergeometric distribution, whereas multiple hypotheses testing was done by using FDR (44).

Loss and Acquisition of Gene Families. The parsimonious evolutionary scenario, which included the loss and acquisition of gene families mapped onto the branches of the chromalveolate phylogenetic tree, was inferred by using the DOLLOP program of the PHYLIP package (45). The DOLLOP program is based on the Dollo parsimony principle, which assumes irreversibility of character loss (19) and has been used in similar studies describing the evolution of KOGs along the eukaryotic phylogenetic tree (46), and studying the intron gain and loss in paralogs (47).

Species/Lineage-Specific Gene Family Expansions. For all gene families (orphans and SSFs excluded), the mean gene family size and standard deviation were calculated. The 25% most variable profiles, based on the standard deviations, were extracted. The matrix of these profiles was transformed into a matrix of z-scores to center and normalize the data. Subsequently, these profiles were hierarchically clustered (complete linkage clustering) by using Pearson correlation as a distance measure. The clustering and visualization was done by using Genesis (48). A description was added to all families based on the most frequently occurring gene descriptions in the corresponding gene family. Gene families only expanded in the outgroup were removed, in addition to those expanded only in one or both of the ciliates because of their massive gene duplication. Transposon-derived gene families were removed because the distribution of such families is likely to be a consequence of whether the gene models were derived from a repeat-masked genome sequence and therefore may be artifactual.

Detection of Block Duplications and Tandem Repeats. The detection of tandem and block duplications within the 12 chromalveolate genomes was done with i-ADHoRe (49). After removal of transposable elements, the MCL pairs were used to define the homologous relations between the genes of every genome, which served as input for the i-ADHoRe algorithm. The following parameters were used in the i-ADHoRe analysis: gap size of 20 genes, *q*-value of 0.8,

probability cutoff at 0.01, minimum of three homologs to define a block, and higher-level multiplicon detection disabled (level 2 only).

Origin of the Chromalveolate Orphans. For the proteins that had no homologs in our dataset (i.e., the orphans or one gene representative per SSF), a similarity search (BLASTP, E-value cutoff E-5) was performed against the nonredundant protein database at NCBI (release 160.0). Valid hits were extracted by using a relative bit score (BS) threshold and taxonomic filtering of putative homologs (see *SI Methods*). Similar results were found

by applying different relative BS thresholds [range 0.6–0.9 (data not shown)].

ACKNOWLEDGMENTS. We thank Tine Blomme and Steven Maere for helpful discussions and Thomas Van Parys for technical assistance. C.M. is indebted to the Institute for the Promotion of Innovation by Science and Technology (Flanders, Belgium) for a predoctoral fellowship. K.V. is a postdoctoral fellow of the Research Foundation-Flanders. The sequence data of *Phaeodactylum tricornutum* were produced by the U.S. Department of Energy Joint Genome Institute www.jgi.doe.gov.

1. Cavalier-Smith T (1999) Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 46:347–366.
2. Archibald JM, Keeling PJ (2002) Recycled plastids: A 'green movement' in eukaryotic evolution. *Trends Genet* 18:577–584.
3. Cavalier-Smith T (2004) In *Organelles, Genomes and Eukaryotic Evolution*, eds Hirt RP, Horner D (Taylor & Francis, London), pp 71–103.
4. Templeton TJ, et al. (2004) Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res* 14:1686–1695.
5. Billioug M, et al. (2002) *Theileria parva* epidemics: A case study in eastern Zambia. *Vet Parasitol* 107:51–63.
6. Nahlen BL, Korenromp EL, Miller JM, Shibuya K (2005) Malaria risk: Estimating clinical episodes of malaria. *Nature* 437:E3; discussion E4–E5.
7. Kaplan JE, Jones JL, Dykewicz CA (2000) Protists as opportunistic pathogens: Public health impact in the 1990s and beyond. *J Eukaryot Microbiol* 47:15–20.
8. Jahn CL, Klobutcher LA (2002) Genome remodeling in ciliated protozoa. *Annu Rev Microbiol* 56:489–520.
9. Bhattacharya D, et al. (1992) Algae containing chlorophylls a+c are polyphyletic: Molecular evolutionary analysis of the Chromophyta. *Evolution (Lawrence, Kans)* 46:1801–1817, and erratum (1993) 47:986.
10. Erwin DC, Ribeiro OK (1996) *Phytophthora Diseases Worldwide* (Am Phytopathol Soc, St Paul).
11. Tyler BM, et al. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261–1266.
12. Rizzo DM, Garbelotto M, Hansen EM (2005) *Phytophthora ramorum*: Integrative research and management of an emerging pathogen in California and Oregon forests. *Annu Rev Phytopathol* 43:309–335.
13. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972–977.
14. Ben Ali A, De Baere R, Van der Auwera G, De Wachter R, Van de Peer Y (2001) Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. *Int J Syst Evol Microbiol* 51:737–749.
15. Dacks JB, Marinets A, Doolittle WF, Cavalier-Smith T, Logsdon JM, Jr (2002) Analyses of RNA polymerase II genes from free-living protists: Phylogeny, long branch attraction, and the eukaryotic big bang. *Mol Biol Evol* 19:830–840.
16. Stechmann A, Cavalier-Smith T (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J Mol Evol* 57:408–419.
17. Van de Peer Y, De Wachter R (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J Mol Evol* 45:619–630.
18. Hackett JD, et al. (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol Biol Evol* 24:1702–1713.
19. Farris J (1977) Phylogenetic analysis under Dollo's law. *Syst Zool* 26:77–88.
20. Harris MA, et al. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258–D261.
21. Zhang Z, Green BR, Cavalier-Smith T (2000) Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: A possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* 51:26–40.
22. Rosenthal PJ (2002) Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr Opin Hematol* 9:140–145.
23. Appiah AA, van West P, Osborne MC, Gow NA (2005) Potassium homeostasis influences the locomotion and encystment of zoospores of plant pathogenic oomycetes. *Fungal Genet Biol* 42:213–223.
24. Burgoyne RD, Morgan A (1993) Regulated exocytosis. *Biochem J* 293:305–316.
25. Plattner H (1989) Regulation of membrane fusion during exocytosis. *Int Rev Cytol* 119:197–286.
26. Greengard P, Valtorta F, Czernik AJ, Benfenati F (1993) Synaptic vesicle phosphoproteins and regulation of synaptic function. *Science* 259:780–785.
27. Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* 52:297–354.
28. Aury JM, et al. (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178.
29. Eisen JA, et al. (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4:e286.
30. Costanzo S, Ospina-Giraldo MD, Deahl KL, Baker CJ, Jones RW (2006) Gene duplication event in family 12 glycosyl hydrolase from *Phytophthora* spp. *Fungal Genet Biol* 43:707–714.
31. McLeod A, Smart CD, Fry WE (2003) Characterization of 1,3- β -glucanase and 1,3;1,4- β -glucanase genes from *Phytophthora infestans*. *Fungal Genet Biol* 38:250–263.
32. Meijer HJ, et al. (2006) Identification of cell wall-associated proteins from *Phytophthora ramorum*. *Mol Plant-Microbe Interact* 19:1348–1358.
33. Torto-Alalibo TA, et al. (2007) Expressed sequence tags from *Phytophthora sojae* reveal genes specific to development and infection. *Mol Plant-Microbe Interact* 20:781–793.
34. Chen X, Shen G, Wang Y, Zheng X, Wang Y (2007) Identification of *Phytophthora sojae* genes upregulated during the early stage of soybean infection. *FEMS Microbiol Lett* 269:280–288.
35. Schneider AG, Mercereau-Puijalon O (2005) A new *Apicomplexa*-specific protein kinase family: Multiple members in *Plasmodium falciparum*, all with an export signature. *BMC Genomics* 6:30.
36. Hall N, et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 307:82–86.
37. Sijwali PS, Shenai BR, Rosenthal PJ (2002) Folding of the *Plasmodium falciparum* cysteine protease falcipain-2 is mediated by a chaperone-like peptide and not the prodomain. *J Biol Chem* 277:14910–14915.
38. Olliaro PL, Yuthavong Y (1999) An overview of chemotherapeutic targets for antimalarial drug discovery. *Pharmacol Ther* 81:91–110.
39. Dobbelaere D, Heussler V (1999) Transformation of leukocytes by *Theileria parva* and *T. annulata*. *Annu Rev Microbiol* 53:1–42.
40. Pain A, et al. (2005) Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* 309:131–133.
41. Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 99:13627–13632.
42. Vandepoele K, Van de Peer Y (2005) Exploring the plant transcriptome through phylogenetic profiling. *Plant Physiol* 137:31–42.
43. Conesa A, et al. (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
44. Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* 100:9440–9445.
45. Felsenstein J (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 266:418–427.
46. Koonin EV, et al. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* 5:R7.
47. Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV (2004) Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Res* 32:3724–3733.
48. Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: Cluster analysis of microarray data. *Bioinformatics* 18:207–208.
49. Simillion C, Vandepoele K, Saey Y, Van de Peer Y (2004) Building genomic profiles for uncovering segmental homology in the twilight zone. *Genome Res* 14:1095–1106.