

Comparative genomics and evolution of eukaryotic phospholipid biosynthesis

Athanasios Lykidis

Genome Biology Program, DOE-Joint Genome Institute, 2800 Mitchell Drive,

Walnut Creek, CA 94598, U.S.A.

Tel: 925-2965842; Fax: 925-2965850; Email: alykidis@lbl.gov

Abstract

Phospholipid biosynthetic enzymes produce diverse molecular structures and are often present in multiple forms encoded by different genes. This work utilizes comparative genomics and phylogenetics for exploring the distribution, structure and evolution of phospholipid biosynthetic genes and pathways in 26 eukaryotic genomes. Although the basic structure of the pathways was formed early in eukaryotic evolution, the emerging picture indicates that individual enzyme families followed unique evolutionary courses. For example, choline and ethanolamine kinases and cytidylyltransferases emerged in ancestral eukaryotes, whereas, multiple forms of the corresponding phosphatidyltransferases evolved mainly in a lineage specific manner. Furthermore, several unicellular eukaryotes maintain bacterial-type enzymes and reactions for the synthesis of phosphatidylglycerol and cardiolipin. Also, base-exchange phosphatidylserine synthases are widespread and ancestral enzymes. The multiplicity of phospholipid biosynthetic enzymes has been largely generated by gene expansion in a lineage specific manner. Thus, these observations suggest that phospholipid biosynthesis has been an actively evolving system.

Finally, comparative genomic analysis indicates the existence of novel phosphatidyltransferases and provides a candidate for the uncharacterized eukaryotic phosphatidylglycerol phosphate phosphatase.

1. Introduction

Phospholipids are not only the major constituents of biological membranes but they are also active participants in the control of diverse cellular functions such as regulation of enzyme and transcriptional activity, signal transduction and information processing, protein folding, translocation and trafficking etc. Phosphatidylcholine (PtdCho), phosphatidylethanolamine (PtdEtn), phosphatidylinositol (PtdIns), phosphatidylserine (PtdSer), phosphatidylglycerol (PG), cardiolipin (CL), sphingomyelin (SM) and inositolphosphorylceramide (IPC) are the major phospholipids in eukaryotes and their percentages differ among cell types and organisms.

Figure 1 outlines the phospholipid biosynthetic pathways that have been described in eukaryotic organisms [1-4]. First, phospholipid biosynthesis starts with the successive acylation of *sn*-glycerol-3-phosphate (G3P) to produce phosphatidic acid (PtdOH). These two steps are catalyzed by specific acyltransferases. Subsequently, PtdOH is distributed to either CDP-diacylglycerol (CDP-DAG) or diacylglycerol (DAG). The partition of PtdOH between CDP-DAG and DAG is directed by the action of CDP-DAG synthetases (CDSs) and PtdOH phosphatases (PAPs). Once formed, DAG is converted to both PtdCho and PtdEtn by the action of CDP-choline:diacylglycerol and CDP-ethanolamine:diacylglycerol phosphatidyltransferases. PtdEtn can be also synthesized via the decarboxylation of PtdSer. On the other hand, CDP-DAG is directly converted to PtdIns and phosphatidylglycerol phosphate (PGP). PGP is subsequently

dephosphorylated to phosphatidylglycerol (PG). Next, two reactions have been described that lead to the formation of cardiolipin (CL): one operating in bacteria that utilizes two molecules of PG and one operating in eukaryotes synthesizing CL from CDP-DAG and PG. PtdSer can be synthesized either directly from CDP-DAG or via a base-exchange mechanism from PtdCho and PtdEtn. Finally, PtdIns and PtdCho serve as donors of phosphoinositol and phosphocholine, respectively, for the formation of the sphingophospholipids inositolphosphorylceramides (IPC) and sphingomyelin (SM).

In this article we employ comparative genomics and phylogenetics to analyze the structure and evolution of the phospholipid biosynthetic genes and pathways in eukaryotic organisms. The significant number of available eukaryotic genomes from different phylogenetic clades provides the opportunity to explore the evolutionary history of phospholipid biosynthetic genes. The development of model organisms with genetic modifications in specific phospholipid metabolism genes necessitates the elucidation of their evolutionary relationships. This will allow for correct extrapolation and proper evaluations of the significance and relevance of the phenotypes that are observed in respect from one organism to the others. Also, analysis of the genomic data allows for the full reconstruction of an organism's metabolism. Metabolic reconstruction is the development of a detailed overview of an organism's metabolism obtained by analysis of the genome sequence. The application of this methodology to any metabolic subsystem necessarily raises questions pertaining to the completeness of the pathways, the existence of novel relevant genes, and their correlation to "orphan" steps in the pathways. We report here the existence of additional genes that may participate in phospholipid metabolism and provide candidates for eukaryotic "orphan" steps, e.g. the conversion of PGP to CL.

2. Data sources and methodology

In the present study we use the genomic sequences for 26 eukaryotes: 6 animals: *Homo sapiens* (Hs) [5, 6], *Mus musculus* (Mm) [7], *Xenopus tropicalis* (Xt) (unpublished, <http://genome.jgi-psf.org/Xentr4/Xentr4.home.html>), *Takifugu rubripes* (Tr) [8], *Danio rerio* (Dr) (unpublished, http://www.sanger.ac.uk/Projects/D_rerio/), and *Ciona intestinalis* (Ci) [9]; 1 insect: *Drosophila melanogaster* (Dm) [10]; 1 nematode: *Caenorhabditis elegans* (Ce) [11]; 1 plant: *Arabidopsis thaliana* (At) [12]; 1 algae: *Chlamydomonas reinhardtii* (Crh) (unpublished, <http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>); 5 fungi: *Saccharomyces cerevisiae* (Sc) [13], *Schizosaccharomyces pombe* (Sp) [14], *Neurospora crassa* (Nc) [15], *Aspergillus fumigatus* Af293 (Af) and *Cryptococcus neoformans* (Cn) [16]; 1 ciliate, *Tetrahymena thermophila* (Tt) [17]; 3 apicomplexans: *Plasmodium falciparum* (Pf) [18], *Theileria annulata* (Ta) [19] and *Cryptosporidium hominis* (Ch) [20]; 1 slime mold: *Dictyostelium discoideum* (Dd) [21], 1 amoeba: *Entamoeba histolytica* (Eh) [22]; 3 kinetoplastids: *Leishmania major* (Lm) [23], *Trypanosoma cruzi* (Tc) [24], and *Trypanosoma brucei* (Tb) [25]; 1 parabasalid, *Trichomonas vaginalis* (Tv) (unpublished, <http://www.tigr.org/parasiteProjects.shtml>); and 1 diplomonad, *Giardia lamblia* (Gl) (unpublished, <http://gmod.mbl.edu/perl/site/giardia?page=intro>). Figure 2 illustrates consensus phylogenetic relationships between the organisms used in the current study [26]. Genome data was retrieved either from NCBI or Ensemble (<http://www.ensembl.org>) or the individual genome project websites.

The queries were either BLAST searches or Hidden Markov Models (HMMs). HMMs are statistical models of the primary structure consensus of a sequence family. One can use these models to do sensitive database searches for proteins that contain domains similar to the query

model [27]. Multidirectional sequence relationships within multifunctional protein families were built using MCL clustering [28, 29]. The collected amino acid sequences were aligned using CLUSTALW and unrooted protein parsimony phylogenetic trees were constructed using the Phylip set of programs (<http://evolution.genetics.washington.edu/phylip.html>).

3. CDP-diacylglycerol synthetases (CDSs)

CDS genes have been identified from a series of eukaryotic organisms: yeast [30], fly [31], *P. falciparum* [32], mouse [33], and human [34-36]. CDS homologues are present in all eukaryotic genomes sequenced to date (Fig. 3).

There is evidence that CDS expression regulates the phosphoinositide signal transduction pathways. Experiments by Weeks *et al* [34] suggested that even modest overexpression of human CDS1 levels amplifies cellular signaling systems. *Drosophila* mutants defective in the CDS isoform expressed in the photoreceptor cannot sustain a light-activated current due to rapid depletion of PtdIns-4,5-bisphosphate [31]. Furthermore, CDS overexpression increases the amplitude of the light response, indicating that CDS regulates the availability of PtdIns-4,5-bisphosphate to phospholipase C. Nonetheless, analysis of the *D. melanogaster* genome indicates the existence of only one CDS copy (as well as in genomes of other insects, e.g. mosquito and *Drosophila pseudoobscura*). Also, analysis of the Expressed Sequence Tags (EST) database does not reveal any multiple transcripts originating from this gene. When multiple transcripts arise from an individual gene they are usually captured in the EST projects. The above observations question the mechanism by which the mutant flies make PtdIns and

other anionic phospholipids and raise the possibility for the existence of a second *CDS* gene with no apparent sequence similarity to the characterized *CDS*s.

C. elegans also has one copy of the *CDS* gene. The RNAi phenotypes associated with the nematode *CDS* are larval arrest and lethality [37] and abnormal locomotion [38]. The *A. thaliana* genome has 5 *CDS* copies, zebrafish has 3, fugu 1 whereas human and mouse have 2. Figure 4 presents the evolutionary trajectory of the *CDS* proteins in eukaryotic organisms. The emergence of four copies in *A. thaliana* is the result of a lineage-specific expansion of the eukaryotic version of *CDS*. The fifth form of *CDS* in *A. thaliana* (gil4895237) is very divergent from the others and exhibits significant similarity to *CDS*s from cyanobacteria (40% identity and 60% similarity, data not shown). This is indicative of its direct bacterial origin. According to the currently available data, the duplication event that separated mammalian *CDS1* from *CDS2* occurred during the transition from insects to fish (500-900 million years ago). Two copies of zebrafish *CDS* are related to human *CDS2* whereas *Takifugu* has only one *CDS* copy corresponding to human *CDS2*. Overall, the data indicate that the multiplicity of *CDS* isoforms has been generated by gene duplication and expansion within major eukaryotic branches.

4. Phosphatidic acid phosphatases

The conversion of PtdOH to DAG is a key branching point in phospholipid biosynthesis which regulates the distribution of the glycerol backbone between PtdCho-PtdEtn and PtdIns-PG-CL. Two general classes of PAPs have been described in the literature. The Mg^{2+} -dependent N-ethylmaleimide (NEM) sensitive form designated as PAP1 [39, 40] and the Mg^{2+} -independent NEM-insensitive form designated as PAP2. Early biochemical work [41]

implicated PAP1 in the *de novo* synthesis of phospholipids and triacylglycerols (TAG). It was shown that this activity is stimulated by fatty acids and acyl-CoA esters, which cause the enzyme to translocate from the cytosol to the endoplasmic reticulum. Recent work has shown that the *S. cerevisiae* lipin homolog (PAH1) is a PAP1 [42]. Lipin was identified through positional cloning as the gene mutated in the fatty liver dystrophy mouse [43]. Null mutation in mouse *lipin1* causes severe deficiency in adipose tissue mass, development of insulin resistance and progressive peripheral neuropathy [44, 45]. On the other hand overexpression of lipin1 in either skeletal muscle or adipose tissue promotes obesity [46]. Deletion of *PAH1* in yeast causes accumulation of PtdOH and reduced levels of TAG consistent with its role as a PAP1 [42]. RNAi knockdown of the *C. elegans* unique lipin gene is embryonic lethal [38, 47] and causes larval arrest and abnormal locomotion [48].

Figure 5 shows the phylogenetic distribution of lipin homologues in 26 eukaryotic genomes and Figure 6 shows the evolutionary relationships between them. Lipin homologues are present in almost all eukaryotic genomes, an observation consistent with the housekeeping enzyme activity proposed for lipin. Only *E. histolytica* and *G. liamblia* do not have a lipin homolog in the current assemblies. As we will discuss later, *G. liamblia* does not appear to have CDP-choline and CDP-ethanolamine pathways that require the conversion of PtdOH to DAG consistent with the absence of a PAP1 enzyme. On the contrary, these pathways appear to be present in *E. histolytica* which is inconsistent with the absence of a PAP1 enzyme. Examination of the other two ongoing *Entamoeba* genomes (*Entamoeba dispar* and *Entamoeba invadens*) using TBLASTN searches fails to identify any genomic areas that potentially code for lipin homologues (data not shown). On the other hand, *E. histolytica* contains 14 PAP2 copies in its genome (Figure 7) and it may be the case that a PAP2 enzyme operates in *de novo*

dephosphorylation of PtdOH. Most organisms have one copy of the *lipin* gene and the multiplicity of lipin homologues observed in the human genome arose after the divergence of vertebrates. The data in Figure 6 indicate that (1) the gene duplications producing multiple lipin homologues occurred during vertebrate divergence and (2) multiple paralogs have been retained in all vertebrates until today indicative of their unique contributions to cell and organism physiology.

The PAP2 enzymes have been mostly implicated in signaling pathways [49-51]. Nonetheless, members of the PAP2 family catalyze the formation of IPCs in yeast [52] and protozoa [53] as well as the synthesis of SM in higher eukaryotes [54]. The above publications refer to the existence of multiple genes coding for these enzymes in select eukaryotic genomes and contain comments on their phylogenetic relationships. For the purpose of the current discussion one should note that the multiplicity of these isoforms appears to be generated by lineage-specific expansion.

A separate phospholipid biosynthetic reaction catalyzed by a PAP2 member has been described in bacteria and it is the conversion of PGP to PG catalyzed by the *pgpB* gene [55, 56]. *E. coli* has a second enzyme that participates in PGP dephosphorylation determined as PgpA [57]. Moreover, genetic experiments strongly support the existence of a third *E. coli* gene that catalyzes the dephosphorylation of PGP [58]. Although PgpA and PgpB enzymes catalyze the same reaction they have no sequence similarity: PgpA homologues are restricted to bacteria whereas PgpB belongs to the universal PAP2 family of enzymes. In contrast to bacteria no eukaryotic PGP phosphatase has been described. Therefore, we explored the possibility of a PAP2-type protein being a candidate for the missing eukaryotic PGPP enzyme. Given the almost universal distribution of such enzyme activity we would expect that the encoding gene

will be highly conserved and will have a widespread distribution. Figure 7 shows the distribution of PAP2 enzymes in 26 eukaryotic genomes. We observe a broad distribution of PAP2 proteins and only *Theileria* does not contain any coding sequence with the PAP2 signature motif. The number of PAP2 enzymes per genome ranges from 2 in apicomplexan up to 25 in zebrafish. However, further analysis of these genes reveals that although there are uncharacterized PAP2 members they exhibit limited phylogenetic distribution in select eukaryotic branches (data not shown) and it is unlikely they code for the housekeeping PGPP activity. These observations suggest that the eukaryotic PGPP enzyme is probably not of a PAP2 origin, although certain PAP2 enzymes may exhibit non specific PGPP activity. The most promising PGPP candidate is a phosphatase domain fused to the *S. pombe* CLS gene. This will be described in a subsequent section.

The effect of PAP2 enzymes on *de novo* phospholipid biosynthesis was studied in *S. cerevisiae*. Deletion of the *lpp1* and *dpp1* genes in *S. cerevisiae* caused a decrease in the cellular levels of PtdIns and significant increases in PtdOH and PtdEtn levels [59]. These observations may be explained by the increase in PtdSer synthase activity which directs the CDP-DAG pool towards PtdSer and PtdEtn causing a decrease in PtdIns levels [59]. Nevertheless, *S. cerevisiae* phospholipid metabolism possesses some unique characteristics when compared to mammalian cells, especially regarding the metabolism of PtdOH and DAG: 1) the conversion of PtdOH to DAG is not the primary route to PtdCho and PtdEtn [3], and 2) *S. cerevisiae* does not have a DAG kinase indicating that the inter-conversion of PtdOH and DAG is distinct when compared to other eukaryotic cells. The topology of the *S. cerevisiae* LPP1 and DPP1 enzymes locates the phosphatase active site on the cytosolic side of the membranes [60]; therefore, changes in the enzyme levels may affect *de novo* phospholipid metabolism. This is in contrast to the

mammalian LPP enzymes whose active sites are facing either the lumen of intracellular organelles or the exterior of the plasma membrane and, therefore, they are unlikely to participate in *de novo* biosynthesis [61-63].

5. Choline and ethanolamine kinases (CKs and EKs)

CKs and EKs (collectively CEKs) catalyze the first step in the CDP-choline and CDP ethanolamine pathways for the synthesis of PtdCho and PtdEtn, respectively. The first CK gene was characterized by Uchida et al [64]. Subsequently, a second gene encoding CK was identified [65, 66]. Based on the dual specificity of these enzymes towards choline and ethanolamine it was considered that ethanolamine specific kinases may not exist. However, work on the *Drosophila* easily shocked gene pointed to the reality of ethanolamine specific enzymes as well as to their potentially significant physiological role [67]. Subsequent work demonstrated the existence of two mammalian EK enzymes, EK1 and EK2 [68, 69]. Mice deficient in EK2 showed reproductive defects due to extensive placental thrombosis [69]. *S. cerevisiae* also harbors two genes with overlapping CK and EK functions [70, 71]. *Saccharomyces cki1Δ*, *eki1Δ* mutants and *cki1Δeki1Δ* double mutants did not have a growth phenotype, although only the *cki1Δ* mutants suppressed the lethal phenotype of a *sec14* mutant [71]. Multiple CKs have been also identified in *C. elegans* [72].

Figure 8 shows the distribution of CEK family members in 26 eukaryotic organisms and Figure 9 depicts their phylogenetic relationships. The content of eukaryotic genomes in CEKs varies between one and nine members. Several unicellular eukaryotes, including two fungi, have only one member of this enzyme class. The two fungi with a single kinase component are *S.*

pombe and *C. neoformans*. As it will be obvious from the subsequent discussion these two organisms have the other components of both CDP-choline and CDP-ethanolamine pathways indicating that the single kinase is apparently a bifunctional enzyme. *T. vaginalis* and *G. liamblia* also have one copy of a *CEK* gene. However, *G. liamblia* when compared to *T. vaginalis* does not have the other two components of the CDP-choline and CDP-ethanolamine pathways (cytidylyltransferase and phosphatidyltransferase) suggesting that its CEK may be operating in a different pathway. *T. annulata* has nine members of this family and seven of them (TA14375, TA14380, TA14385, TA14390, TA14405, TA14410, and TA14415) are located adjacent on the chromosome. *C. elegans* and *A. thaliana* have seven and five CEKs, respectively. The phylogenetic analysis (Figure 9) indicates that most of these isoforms are related to CKs. In the case of *C. elegans* six proteins (B0285.8, B0285.9, B0285.10, F22F7.5, C28D4.2, C52B9.1) group together forming a robust phylogenetic clade indicating their emergence by amplification of a single gene. B0285.8, B0285.9, and B0285.10 are located adjacent on chromosome III. *A. thaliana* has four CK homologues that form a coherent phylogenetic clade indicating also their emergence from a single gene. In both cases these enzymes are clearly related to CKs. Both *C. elegans* and *A. thaliana* have one member of the EK family (T27A10.3 and gil15225800, respectively). RNAi experiments targeting T27A10.3 have no observed phenotype. From the six *Caenorhabditis* CK members only B0285.8, B0285.9 have RNAi associated phenotypes. RNAi targeting of B0285.8 reduces fat content [73] whereas targeting of B0285.9 causes maternal sterility [74]. The separation between CK and EK families appears to be an early event in eukaryotic evolution: kinetoplastid enzymes are most related to CKs whereas *Dictyostelium* and *Entamoeba* enzymes are included in the EK family. On the other hand, each gene follows unique evolutionary courses within major lineages with the most

notable example the significant expansion of the CK members in *Theileria*, *Caenorhabditis* and *Arabidopsis*. Overall, the data is consistent with the idea that the multiplicity of enzyme isoforms is generated by lineage-specific expansion of certain genes.

6. Choline and ethanolamine cytidylyltransferases (CCTs and ECTs).

CCTs and ECTs constitute the second step in the CDP-choline and CDP-ethanolamine pathways for the synthesis of PtdCho and PtdEtn, respectively. Genes encoding CCT enzymes have been characterized in several eukaryotic organisms. Human and mouse have two distinct genes: *Pcyt1a* coding for CCT α and *Pcyt1b* coding for CCT β 1, - β 2, and - β 3 isoforms [75-79]. *S. cerevisiae* has one *CCT* gene [80]. *CCT* genes from *C. elegans* [81], *D. melanogaster* [82], *A. thaliana* [83] and *P. falciparum* [84] have also been reported. *ECT* genes have been identified in several eukaryotic species: yeast [85], human [86], rat [87] and the algae *C. reinhardtii* [88]. Disruption of *CCT* genes has distinct phenotypes: deletion of mouse CCT α results in early embryonic lethality [89] whereas mice with impaired CCT β 2 expression are viable but have gonadal dysfunction and reduced fertility [90]. Disruption of the *D. melanogaster CCT1* gene results also in a reproductive phenotype [91] attributed to aberrant membrane trafficking of ovarian growth and differentiation factor receptors [92]. *S. cerevisiae CCT* and *ECT* mutants are not essential for viability [80, 86] indicative of the dominant role of the CDP-DAG route to phospholipid biosynthesis in this organism [93]. However, only *CCT* mutants suppressed the lethal *sec14* phenotype [93].

Figure 10 illustrates the phylogenetic distribution of CCT and ECT enzymes in the 26 eukaryotic genomes. The distinction between CCT and ECT is based mainly on the number of

cytidylyltransferase domains present in a coding sequence: if 1 domain is present the enzyme is considered as a CCT whereas if 2 domains are present it is assigned as an ECT. *T. annulata* and *P. falciparum* constitute an apparent exception since the corresponding CCTs have two cytidylyltransferase domains. The *P. falciparum* CCT enzyme characterized by Yeo *et al* [84] consisted of one cytidylyltransferase domain, however, the genome sequence of all *Plasmodia* and *Theilerias* suggests that the gene has actually two domains and is expressed as one gene. Most multicellular organisms have the common profile of 2 CCTs and one ECT (*H. sapiens*, *M. musculus*, *X. tropicalis*, *D. melanogaster*, *A. thaliana*). However, fugu fish and worms have one additional copy of both CCT and ECT (the locus tags for the three *C. elegans* CCT genes are F08C6.2, Y18H1A.11, and F28A10.10, whereas for the two ECTs are Y37E3.11 and C39D10.3). In the case of *C. elegans* searching through the EST data confirms the expression of 4 genes. The third CCT gene (F28A10.10) has no corresponding ESTs and, therefore, its expression is uncertain. The *C. elegans* genes have been targeted in whole genome RNAi experiments which mostly scored phenotypes related to maternal sterility, movement abnormality, lethality and morphology [37, 38, 48, 94]. Worms with compromised expression of F08C6.2 (which is the CCT previously characterized [81]) were arrested at late larval stage and had abnormal locomotion whereas the second CCT gene (Y18H1A.11) gave no obvious changes for the checked phenotypes. RNAi targeting of both ECT genes also gave no change in phenotype indicating a possible redundancy in their function. The above data are available at the *C. elegans* database, <http://www.wormbase.org>.

Most of the unicellular eukaryotes have a common distribution of one CCT and one ECT gene except *E. histolytica* and *T. thermophila* which appear to have two copies of ECT-like genes. Kinetoplastid protozoa (*T. cruzi*, *T. brucei*, and *L.major*) have two genes, one with two

domains (ECT) and one with a single cytidylyltransferase domain fused to a phosphatidyltransferase. The cytidylyltransferase (CT) – phosphatidyltransferase (PdT) fusion enzymes are: *T. cruzi*, gil71409321; *T. brucei*, gil71748856; *L. Major*, gil68125996. The existence of these fusion proteins in ancestral eukaryotes raises the possibility that they may constitute the ancestral form of CDP-choline and CDP-ethanolamine pathways. *T. vaginalis* has four coding sequences with a cytidylyltransferase domain: one that has two adjacent domains and it is an apparent ECT whereas the other three sequences have similarities to the nucleotidyltransferases involved in NAD and coenzyme A biosynthesis. Furthermore, phylogenetic analysis suggests that *T. vaginalis* has an ECT and does not have a CCT (Figure 11). This conclusion is in agreement with experimental evidence that failed to label choline phospholipids (PtdCho, LysoPtdCho and SM) with [³²P]orthophosphate [95] suggesting that a CDP-choline pathway is not present in this organism. *G. lamblia* has three coding sequences with a single nucleotidyltransferase sequence signature. None of these appear to be either a CCT or an ECT rather they are related to the nucleotidyltransferases involved in NAD and coenzyme A biosynthesis. As we will discuss in the subsequent section on phosphatidyltransferases, *G. lamblia* is also missing CPT and EPT homologues. These observations suggest that *G. lamblia* has no operational CDP-choline and CDP-ethanolamine pathways.

The existence of CCTs and ECTs in a range of eukaryotic organisms allows us to approach questions about their evolutionary origin: which gene or domain appeared first in evolution and what has been the evolutionary course of this family? To investigate these questions we constructed a phylogenetic tree of the cytidylyltransferase domains of all CCT and ECT family members. Since ECT proteins consist of two cytidylyltransferase domains we designated domain A the N-terminal one and domain B the C-terminal. Each domain was

handled as a separate entry in the analysis. Figure 11 shows the evolutionary relationships between the 3 cytidylyltransferase domains: the CCT domain, the ECT domain A and the ECT domain B. The topology of the tree indicates that ECT domain B is clearly related to ECT domain A and appeared early in eukaryotic evolution since homologues exist in ancestral eukaryotes such as *T. vaginalis*. An intriguing observation relates to the location of the kinetoplastid cytidylyltransferase domain originating from the CT-PdT fusion proteins. This CT domain is phylogenetically related to ECT domain B. This supports a hypothesis that the CCT and ECT enzymes emerged from ancestral eukaryotic CT-PdT fusion proteins. ECT domain B was generated from the CT domain of the fusion protein and was subsequently duplicated to give rise to ECT domain A. According to this model, CCTs originated from ECT domain A, which is also consistent with the higher similarity between them (data not shown). As it will be discussed in the subsequent section and consistent with the above model, the PdT domain of the kinetoplastid CT-PdT fusion proteins appears to be the origin of the CEPT family of eukaryotic enzymes (Figure 14).

Furthermore, the phylogenetic analysis indicates that the generation of multiple CCT isoforms in higher eukaryotes also occurred in a lineage-specific manner. For example, the *Drosophila* CCT1 and CCT2 enzymes arose from gene duplication within the *Drosophila* clade after the separation of the insect lineage. The same holds true for the CCT genes of *Caenorhabditis*. The duplication event that gave rise to mammalian CCT α and CCT β forms occurred approximately 400 million years ago after the separation of the fish lineage.

7. Phosphatidyltransferases (PdTs)

7.1 Phosphatidyltransferases determine the diversity of phospholipid structures.

PdTs catalyze the final phospholipid producing reaction and belong to two large families in terms of sequence characteristics: CDP-alcohol phosphatidyltransferases (CA-PdTs) and phospholipase D-like phosphatidyltransferases (PLD-PdTs). A CA-PdT (PGS) catalyzes the synthesis of PGP in bacteria [96]. The eukaryotic PGSs (e.g. from *S. cerevisiae* and Chinese hamster ovary cells) that have been reported previously belong to the PLD-PdT family [97, 98], although the plant enzyme is a CA-PdT and resembles the bacterial ones [99, 100]. PGP is dephosphorylated to PG [101], which in turn is the precursor to CL [101]. Two reactions have been described that synthesize CL. The bacterial reaction as exemplified by *E. coli* utilizes two molecules of PG and is catalyzed by a PLD-PdT enzyme [102]. In contrast, the eukaryotic enzymes, isolated from yeast and human, belong to the CA-PdT family [103-106]. CA-PdTs in *Bacillus subtilis*, *Helicobacter pylori* and *S. cerevisiae* catalyze the synthesis of PtdSer [107-109], while *E. coli* utilizes a PLD-PdT enzyme [110]. PtdIns is synthesized by a CA-PdT-type PIS and several eukaryotic as well as bacterial (*Mycobacterium tuberculosis*) enzymes have been characterized that belong to the CA-PdT family [35, 111-114]. The synthesis of PtdCho from CDP-DAG and choline has been recently described in *Sinorhizobium meliloti* and is catalyzed by a CA-PdT [115] although it seems that this reaction is confined in bacteria and there are no obvious eukaryotic cases for this reaction. Finally, the two phosphotransferase reactions operating in the eukaryotic CDP-choline and CDP-ethanolamine pathways (CPT and EPT, collectively C/EPT) are catalyzed by CA-PdTs [116-120].

7.2 CDP-alcohol phosphatidyltransferases

The genome sequences of 26 eukaryotes were searched with an HMM motif specific for CA-PdTs. A total of 129 genes were identified with a density range of 2 to 8 genes per genome (Figure 12). *E. histolytica* and *T. vaginalis* have the largest number (8) of CA-PdT copies in their genomes whereas *P. falciparum*, *T. annulata* and *C. hominis* had only two copies. MCL clustering of the 126 CA-PdT genes classifies them in 13 clusters in terms of sequence similarity (Figure 13). The first cluster contains 63 proteins with representatives from all genomes except *G. lamblia* and includes the verified yeast and human *CPT*, *CEPT* and *EPT* genes. In addition, it includes a set of novel CA-PdTs exemplified by the human *Sell* gene (discussion on the *Sell* gene follows in subsequent paragraphs). The absence of C/EPT homologues in *G. lamblia* is consistent with the corresponding absence of cytidylyltransferases and indicates that the CDP-choline and CDP-ethanolamine pathways are not present in this organism. The second cluster corresponds to PIS enzymes and contains 29 sequences with representatives from all 26 genomes. *A. thaliana*, *C. reinhardtii* and *T. cruzi* have two *PIS* copies. CLS sequences constitute the third cluster, which contains 14 sequences from 14 genomes. The fourth cluster contains the yeast PSS enzyme and includes sequences only from the 5 fungi genomes. The fifth cluster contains the 3 CAPdT-type PGSs found in plants (two from *A. thaliana* and one from the algae *C. reinhardtii*). In addition, it contains three *Trichomonas* and one *Giardia* CA-PdTs suggesting the existence of bacterial-type CAPdT-PGSs in these organisms. There are seven additional clusters with one to four members. The sixth cluster contains three unique CA-PdTs from kinetoplastids *T. cruzi*, *T. brucei* and *L. major* (gil71664582, gil71748950, and gil68125941, respectively). These proteins are 415 amino acids long and exhibit remote sequence similarity to other eukaryotic CA-PdTs although they clearly belong to this family. The seventh cluster contains the PdT domain of the kinetoplastid CT-PdT fusion proteins

described earlier. The eighth cluster contains two proteins from *D. rerio* and *T. rubripes* with significant similarity to bacterial PSSs and we will elaborate on them later. The final four clusters (9-12) are one-member sets with sequences from *D. melanogaster*, *G. lamblia*, *C. reinhardtii* and *T. rubripes*. The *Drosophila* CG4585 protein is 392 amino acids long and it is an insect specific PdT since it has homologues in all insect genomes sequenced until now with no apparent homologues in other eukaryotic genomes. The *G. liamblia* protein is related to bacterial PSSs. The *Takifugu* coding sequence is a unique form of CA-PdT with one related sequence from another marine organism *Strongylocentrotus purpuratus* (gil72044581). The *Chlamydomonas* protein is a unique protein with low similarity score to the CAPdT HMM domain and it may not represent a real PdT.

7.2.1 C/EPTs and a novel family of eukaryotic PdTs.

The first cluster in Figure 13 contains all the characterized eukaryotic C/EPTs. In addition, it contains a group of uncharacterized CA-PdTs as exemplified by the human Sell protein. Sell (gil50083289) was first identified as a selenoprotein in a screen to characterize the selenoprotein complement of the human genome [121]. The Sell protein is 397 amino acids long and its gene spans 9 exons on human chromosome 2. Based on Unigene information *Sell* is widely expressed in human and mouse tissues and in levels comparable to the ones of CPT and CEPT (data not shown). To investigate the evolutionary history and relationships between the CPT, EPT and Sell groups of proteins we performed phylogenetic analysis of 63 proteins by calculating protein parsimony trees. Figure 14 presents the results of the phylogenetic analysis. The tree topology indicates that Sell is the human representative of a large CA-PdT subfamily.

Sell proteins form a robust phylogenetic group and their copy number differs among genomes. Human, mouse, frog, fish and nematodes have only one *Sell* copy, whereas the fly and the sea squirt have two. It is noteworthy that two Sell homologues can be identified in a kinetoplastid genome, i.e. *T. cruzi*. Because of the absence of Sell homologues in the phylogenetic space between *Trypanosomes* and multicellular eukaryotes we suggest that Sell enzymes emerged during the evolution of multicellularity and their existence in *T. cruzi* may be attributed to horizontal transfer. RNAi phenotypes associated with the *C. elegans* homolog of Sell are slow growth [37] and larval arrest [122]. RNAi targeting of the two *C. elegans* C/EPT homologues gives distinct phenotypes: F22E10.5 does have any obvious phenotype whereas Y49A3A.1 is embryonic lethal or has abnormal fat content [37, 38, 47, 73]. We hypothesize that the most probable function for Sell is an ethanolamine-specific phosphotransferase involved in the CDP-Etn pathway. The members of the CDP-Etn pathway have appeared early in eukaryotic evolution as evidenced by the phylogenetic analysis of EK and ECT. It is conceivable that a specific EPT also emerged during eukaryotic evolution. Of course the possibility of an as yet uncharacterized function for Sell cannot be excluded.

The structure of the tree in Figure 14 provides also information on the relationship between the experimentally studied yeast and human CEPTs. It is evident that yeast CPT is most closely related to yeast EPT rather to the human CPT and CEPT proteins. Therefore, the multiplicity of the CEPT genes arose from lineage specific duplications. Human CPT1 and CEPT isoforms originated from a duplication event approximately 400 million years ago during the emergence of vertebrates. The C/EPT group of PdTs has undergone a significant expansion in *E. histolytica* where we recognize seven copies, although it is not clear whether there is any physiological significance to this observation.

The PdT domains of the kinetoplastid CT-PdT fusion enzymes form a coherent phylogenetic set clearly associated with C/EPTS (495/500) and support the hypothesis that they may constitute the evolutionary origin of the CDP-Cho and CDP-Etn pathways.

7.2.2 CDP-alcohol phosphatidyltransferase-type PSSs, PGSs and CLSs.

A eukaryotic CAPdT-PSS has been characterized in *S. cerevisiae* [109] and its homologues can be easily recognized in all sequenced fungal genomes. A CAPdT-PSS has been also reported in wheat [123] however, a homolog cannot be identified in other plants, e.g. *A. thaliana* and rice. Also, TBLASTN searches of all available plant genomic DNA using as a query the reported wheat PSS did not identify any region with significant similarity. On the contrary, wheat PSS has significant relationship (approximately 70% identity, and 80% similarity) to yeast PSSs. As it will be discussed later *A. thaliana* and rice have homologues of the mammalian serine-exchange enzymes. This appears to be the mechanism of PtdSer biosynthesis, at least in these plants.

Cluster 7 in Figure 13 contains two genes from fish genomes (exemplified by zebrafish coding sequence gil68402375) that have high degree of similarity to bacterial PSSs, e.g. the *Bacillus* enzymes (for the purposes of the current discussion we designate these genes BLA-PSS, for *Bacillus-Like Animal PSS*). The possibility that these genes are the result of a bacterial contamination artifact is minimal since the gene is present in all three fish genomes sequenced (*Danio*, *Fugu*, and *Tetraodon*) at different times in different places. BLAST searches, using the aforementioned zebrafish protein as query, identify similar coding sequences in other animal genomes including human. Although a human gene model for BLA-PSS is not defined in the

current assemblies (as of October 2006), EST sequences are easily identified indicating the presence and expression of BLA-PSS in mammals. The three fish BLA-PSS proteins contain a complete CA-PdT active site, DG-X₂-AR-X₈-G-X₃-D-X₃-D (Figure 15) [124]. However, BLA-PSS proteins from higher animals have incomplete CA-PdT active sites indicating a fast evolutionary adaptation of this gene and the possibility of a divergent function. The human and mouse versions of BLA-PSS exhibit limited patterns of expression. Human BLA-PSS ESTs have been identified in brain, liver, heart and testes whereas corresponding mouse transcripts were found mainly in brain, testes and female genital tissue (data not shown). Based on sequence comparisons it is reasonable to accept that the fish BLA-PSS proteins exhibit some kind of CA-PdT activity, presumably PSS. The function of the human/mouse BLA-PSS cannot be accurately predicted because of the divergent active site. Nevertheless, the possibility that it is a PdT is really intriguing. Thus, it is clear that BLA-PSSs have evolved from a CA-PdT and it will be interesting to determine their functions and examine their relationships to CA-PdTs.

Figure 12 shows that there are three eukaryotic CA-PdTs with PGS activity from *A. thaliana* and *C. reinhardtii*. The participation of a CA-PdT enzyme in the PGS reaction is common in bacteria but the majority of eukaryotes utilize a PLD-type enzyme. CAPdT-PGS amino acid sequences exhibit relative similarity to CAPdT-CLS enzymes, in the range of 30% identity and 50% similarity. This suggests that they may be evolutionary related. To investigate the relations between the PGS, CLS, PIS and PSS CA-PdT groups we constructed a phylogenetic tree in which we also included select bacterial sequences for PGS, PSS, PIS and CLS. The data in Figure 16 indicate that the sequences of *A. thaliana*, *T. vaginalis* and *G. liamblia* PGSs branch from the bacterial PGS group. Moreover, the eukaryotic CAPdT-CLS sequences are also branching from the PGS clade suggesting that the eukaryotic CAPdT-CLS genes have evolved

from the prokaryotic type CAPdT-PGSs present in ancestral eukaryotes. Several eukaryotes do not have a CAPdT-type CLS but rather they have a PLD-type CLS (see following chapter). Eukaryotic PIS genes form a coherent and robust phylogenetic group and their source lies undetermined at the origin of the eukaryotic domain. It is important to note that eukaryotic and prokaryotic PISs appear to be evolutionary unrelated. These activities probably emerged independently in the two domains. The fungi PSSs are clearly evolutionary related to bacterial PSSs and they probably constitute a horizontal transfer event.

A very interesting observation lies on the 570 amino acids long sequence of the *Schizosaccharomyces* CLS (gil2414601). This enzyme is a fusion between two domains: a phosphatase and a PdT. This fusion event is unlikely to be an erroneous outcome of computational genome assembly since it is also present in the second *Schizosaccharomyces* genome undergoing sequencing, *Schizosaccharomyces japonicus* (data not shown) . Its amino terminal region contains a phosphatase domain of the HAD superfamily described by COG0647 as NagD phosphatase. Homologues of this domain exist in all sequenced eukaryotic genomes and in several bacteria. The existence of a protein with fused domains is very strong evidence that these domains are functionally linked. The function linked to eukaryotic CL synthesis with no characterized representative is the dephosphorylation of PGP to PG. As mentioned earlier, bacteria utilize a type-II PAP to catalyze this reaction but our analysis does not point to a specific eukaryotic PAP2 candidate for the PGPP step. Based on the above observations, we suggest that the NagD region of gil2414601 constitutes a candidate for the uncharacterized eukaryotic phosphatidylglycerol phosphate phosphatase (cPGPP). Other eukaryotic genomes encode proteins with a single NagD domain and length of approximately 350 amino acids. The human genome has one correspondent cPGPP gene which is located on chromosome 22q11 and

specifically in the region responsible for the cat eye syndrome which contains a total of 14 genes [125]. Cat eye syndrome is characterized by chromosomal abnormalities in the 22q11 region and exhibits variable clinical features, particularly congenital malformations. Further biochemical experiments are required to validate the prediction that these NagD proteins are indeed PGPPs. In addition to catalyzing the PGP dephosphorylation, these genes may also participate in the dephosphorylation of PtdOH and may represent the residual Mg^{2+} -dependent PAP1 activity observed in *S. cerevisiae PAH1* mutants [42].

7.3 PLD-type phosphatidyltransferases

A separate class of PdTs is encoded by enzymes that have sequence signatures resembling the ones found in PLDs. BLAST and HMM searches of the 26 eukaryotic genomes identify 126 sequences containing PLD sequence motifs (Figure 17) which are clustered in 11 groups based on their overall sequence similarity (data not shown). The first cluster contains 50 sequences and constitutes the group of the signaling PLDs that includes the well-characterized human PLD1 and PLD2 as well as the yeast SPO14 proteins [126]. The second and fourth clusters contain 23 and 7 uncharacterized proteins with PLD sequence motifs, respectively. There are two clusters that contain phospholipid synthesis enzymes: the third cluster contains 19 sequences from 19 genomes and includes the experimentally verified PGS from *S. cerevisiae* [97]; the fifth cluster contains six sequences from *Trypanosoma*, *Leishmania*, *Theileria*, *Plasmodium*, *Cryptosporidium* and *Dictyostelium* with significant homology to prokaryotic CLSs. Therefore, we conclude that these organisms utilize the “prokaryotic”-type CLS reaction between two PG molecules for the synthesis of CL. This is an interesting observation indicating

the evolutionary survival of the prokaryotic reaction for CL formation into the eukaryotic kingdom. *G. lamblia*, *T. vaginalis* and *E. histolytica* do not have any form of recognizable CLS. Although these genomes are not complete, the lack of CLS is consistent with the lack of mitochondria in these organisms. CL is located in mitochondria and these amitochondriate eukaryotes apparently have not evolved (or have lost) the machinery for its synthesis. *E. histolytica* does not have either a CA-PdT or a PLD like PGS and this is in agreement with experimental results suggesting that its membranes do not have PG [127]. Six PLD-PdT clusters (6-11) are one and two member sets with uncharacterized proteins from unicellular eukaryotes.

Figure 18 summarizes the occurrence of the various PdT forms in eukaryotic genomes. It is evident that besides the C/EPT and PIS reactions that are uniformly encoded by homologous genes other phospholipid biosynthetic reactions have utilized genes from different origins. Overall, there are at least two uncharacterized CAPdT families with unique evolutionary positioning: the Sell proteins found mostly in multicellular eukaryotes and *T. cruzi* and the three unique kinetoplastid PdT reported in the sixth cluster of Figure 13.

8. Base-exchange phosphatidylserine synthases (PSSE)

Higher eukaryotes synthesize PtdSer by a base-exchange mechanism in which serine substitutes for the choline and ethanolamine groups of PtdCho and PtdEtn, respectively. In mammals, two enzymes have been identified: PSSE1 which catalyzes the exchange reaction with PtdCho [128, 129] and PSSE2 which utilizes PtdEtn [130, 131]. PSSE2-deficient mice do not have severe developmental abnormalities or alterations in phospholipid content [132] whereas CHO-K1 cells lacking both PSSE forms are PtdSer auxotrophs [133].

HMM searches of 26 eukaryotic genomes reveal the existence of PSSE homologues in the majority of eukaryotic genomes (Figure 19). Further phylogenetic analysis specifies the evolutionary history of the PSSE family (Figure 20). Essentially, only fungi and *G. liamblia* utilize the CAPdT-PSS mechanism. Besides fungi, the only genomes that do not have a PSSE homolog are *T. annulata* and *T. vaginalis*. Two multicellular eukaryotes (*Fugu* and *Ciona*) appear to have three PSSE genes, whereas others (*Drosophila* and *Arabidopsis*) appear to have only one. *C. elegans* has two PSSE genes (T27E9.5 and ZC506.3) and RNAi experiments show that reduction in T27E9.5 expression causes embryonic and post-embryonic developmental abnormalities whereas reduction in ZC506.3 levels causes a small phenotype (<http://www.wormbase.org>). However, the tree topology indicates with high statistical confidence that both *C. elegans* enzymes are evolutionary related to the human PSSE1 and arise from species-specific duplication. Nonetheless, they apparently acquired specific functions related to the worm physiology. The *Drosophila* enzyme also belongs to the PSSE2 subfamily, whereas the *Arabidopsis* and apicomplexan genes are part of the PSSE1 group. An open question relates to the specificity of these enzymes towards PtdCho and/or PtdEtn and whether it corresponds to the specificity of their human counterparts. If this is true then different eukaryotes would utilize either PtdCho or PtdEtn as a PtdSer precursor depending on the presence of a PSSE1 or PSSE2 isoform. Further biochemical experiments are required to clarify this point.

The above results indicate that PSSE enzymes are abundant in the eukaryotic kingdom they appeared early in eukaryotic evolution and the separation between PSSE1 and PSSE2 is a relatively ancient evolutionary event.

9. Phosphatidylserine decarboxylase (PSD)

PSD catalyzes the decarboxylation of PtdSer to PtdEtn. This is the only route for PtdEtn formation in prokaryotes and PSD encoding genes have been characterized from *E. coli* [134] and *B. subtilis* [135]. In eukaryotes, the reaction has been studied in *S. cerevisiae* whose genome harbors two *PSD* genes, *PSD1* [136, 137] and *PSD2* [138, 139]. The two yeast isoforms are localized in different subcellular compartments: PSD1 resides in mitochondria, whereas PSD2 is found in the vacuole and Golgi compartments. Disruption of both *PSD1* and *PSD2* genes causes yeast to become ethanolamine auxotrophs, whereas deletion of the single mouse *PSD* gene is embryonic lethal [140]. *PSD* genes have been also cloned and characterized from plants [141] and *P. falciparum* [142]. RNAi phenotypes associated with the single *C. elegans PSD* gene include abnormal embryonic and post-embryonic development, slow growth and an unusual transparent appearance of the worms [37, 38, 48, 143].

Figure 21 depicts the phylogenetic distribution of the PSD enzymes in eukaryotes. PSD is found in the majority of eukaryotic genomes with the exception of *E. histolytica* and *T. vaginalis* which are missing mitochondria. These two organisms apparently utilize exclusively the CDP-ethanolamine pathway for PtdEtn biosynthesis; the existence of a second ECT enzyme in *E. histolytica* may relate to the absence of a PSD. There is significant variability in the number of PSD genes in each genome. *X. tropicalis* is the only vertebrate with two PSD copies. *A. thaliana* and *D. discoideum* have three PSD copies, whereas *S. cerevisiae* appears to be an exception in the fungi domain with only two copies. Other fungi have anywhere between three and five *PSD* copies.

Figure 22 depicts the phylogenetic relationships between the different PSD groups. The tree also includes select bacterial PSD proteins because of the relatively high overall similarity between eukaryotic and bacterial PSDs. Two well-separated clusters representing the previously studied PSD1 and PSD2 enzymes are formed. The distribution of individual genes in these clusters is not uniform. For example, *A. thaliana* has two PSD2 and one PSD1 related genes, whereas one *Dictyostelium* protein belongs to the PSD2. In addition, there is a clear formation of a third PSD class (PSD3) that includes two genes from *A. fumigatus* and *N. crassa*, and one from *C. neoformans*. Although the PSD3 enzymes branch off the PSD2 clade with high statistical significance (493/500), it is clear that they constitute a separate PSD group. For example, comparison of the *A. fumigatus* gil66844383 protein to *S. cerevisiae* PSD2 enzyme reveals limited overall identity and similarity, which is indicative of their divergent evolutionary paths. Moreover, the PSD3 forms lack the calcium-dependent membrane targeting C2 domains found in the PSD2 proteins. While processing this sequence data, we observed that the fungal PSD3 enzymes exhibit high similarity to bacterial PSD proteins. For example, the above mentioned *Aspergillus* protein has 49% identity and 65% overall similarity to a protein from *Chlorobium phaeobacteroides*. Overall there are currently 23 bacterial species that have enzymes with high sequence similarity to PSD3 and apparently constitute a case of horizontal transfer. It will be interesting to examine whether the proposed PSD3 isoforms have distinct contributions to the physiology of the corresponding organisms. In addition, Figure 22 indicates that PSD enzymes from ancient eukaryotes (*G. liamblii*, *Trypanosomes* etc) are clearly associated with PSD2 suggesting that the appearance of PSD1 is a subsequent evolutionary event.

10. Phosphatidylethanolamine methyltransferase (PEMTs)

The methylation of PtdEtn is a well-established route for the synthesis of PtdCho. It is quite abundant in bacteria [144] and it has been studied in several eukaryotes. *S. cerevisiae* has two enzymes that catalyze the methylation of PtdEtn to PtdCho: the PEMT1 protein (gil6321596) which is 869 amino acids long and the PEMT2 protein (gil129766) which contains 206 amino acids. Mammalian cells contain only a PEMT2-type enzyme [145] and disruption of the murine *PEMT2* gene results in viable mice [146]. In the absence of nutritional choline, *PEMT2*^{-/-} mice developed severe liver pathology with concomitant decrease in PtdCho levels [147].

Figure 23 presents the phylogenetic distribution of PEMT enzymes. In contrast to other phospholipid biosynthetic enzymes, PEMTs exhibit a sporadic distribution throughout eukaryotes suggestive of a selective evolutionary survival of this pathway. PEMT homologues cannot be identified in insects, nematodes, plants as well as in several unicellular protists.

11. Summary

The availability of ever increasing eukaryotic genomes provides the opportunity to examine the evolutionary history of phospholipid biosynthetic genes. The existence of multiple genes coding for phospholipid biosynthetic enzymes necessitates the development of an evolutionary context for their emergence. The most important theme emerging from the current analysis is the generation of enzyme multiplicity by lineage-specific duplications and expansion. This hypothesis implies that usually there is no one-to-one evolutionary correspondence between phospholipid biosynthetic genes from different organisms and, hence, specific phenotypes

observed in genetically modified model organisms may not be directly applicable to other species. For example, in most cases the duplicate human gene is not a direct descendant of a fly or yeast gene rather it has evolved in a vertebrate-specific manner. On the contrary, human genes are usually directly correlated with mouse genes in terms of evolutionary history.

A large number of phospholipid biosynthesis genes and enzymes from various eukaryotic organisms have been identified and characterized during the last decade. However, examination of the available genomic sequences points to the existence of novel eukaryotic enzymes (particularly phosphatidyltransferases and phosphatases) that presumably participate in phospholipid biosynthesis. Experimental interrogation of these genes is necessary and may reveal novel pathways and regulatory mechanisms operating in phospholipid biosynthesis.

Phospholipid biosynthesis is a cellular process that is controlled both temporally and spatially and responds to various intracellular and environmental stimuli. The evolution of multiple enzyme forms that will regulate phospholipid formation is depicted in the genomic sequences. Future work will establish the specific contributions of each enzyme class to the physiology of individual organisms.

Figure Legends.

Figure 1. Eukaryotic pathways for phospholipid biosynthesis. CDSs and PAPs regulate the distribution of phosphatidic acid (PtdOH) between CDP-diacylglycerol (CDP-DAG) and diacylglycerol (DAG). Phosphatidyltransferases utilize CDP-DAG and DAG to synthesize phospholipids. DAG serves also as a substrate for DAG acyltransferase (DGAT) during the synthesis of triacylglycerols. CCT, phosphocholine cytidyltransferase; CDS, CDP-DAG

synthetase; CK, choline kinase; CLS, cardiolipin synthase; CPT, CDP-choline:diacylglycerol phosphotransferase; DGAT, diacylglycerol acyltransferase; ECT, phosphoethanolamine cytidyltransferase; EK, ethanolamine kinase; EPT, CDP-ethanolamine:diacylglycerol phosphotransferase; ICS, inositol phosphorylceramide synthase; PAP, phosphatidate phosphatase; PEMT, phosphatidylethanolamine methyltransferase; PIS, phosphatidylinositol synthase; PGPP, phosphatidylglycerol phosphate phosphatase; PGS, phosphatidylglycerol phosphate synthase; PSS, phosphatidylserine synthase; PSD, phosphatidylserine decarboxylase; PSSE, phosphatidylserine synthase via base-exchange; SMS, sphingomyelin synthase.

Figure 2. Phylogeny of eukaryotes. An unrooted phylogeny of major eukaryotic lineages utilized in the current analysis was adapted from more detailed reports, [26] and references therein.

Figure 3. Phylogenetic distribution of CDSs. Genome data were queried with either a CDS HMM using HMMER (Pfam domain PF01148, CTP-transf-1) or CDS protein sequences using BLASTP and TBLASTN.

Figure 4. Evolution of eukaryotic CDSs. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 421 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers for proteins included are indicated.

Figure 5. Phylogenetic distribution of lipin. Genome data were queried with either a lipin HMM using HMMER (Pfam domain PF08235, LNS2) or lipin protein sequences using BLASTP and TBLASTN.

Figure 6. Evolution of lipins. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 563 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers for proteins included are indicated.

Figure 7. Phylogenetic distribution of PAP2. Genome data were queried with either a PAP2 HMM using HMMER (Pfam domain PF01569) or PAP2 protein sequences using BLASTP and TBLASTN.

Figure 8. Phylogenetic distribution of CEKs. Genome data were queried with either a CEK HMM using HMMER (Pfam domain PF01633, Choline-kinase) or CEK protein sequences using BLASTP and TBLASTN.

Figure 9. Evolution of CEKs. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 414

unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated.

Figure 10. Phylogenetic distribution of CCTs and ECTs. Genome data were queried with either a CCT HMM using HMMER (Pfam domain PF01467, CTP-transf-2) or CCT and ECT protein sequences using BLASTP and TBLASTN.

Figure 11. Evolution of cytidylyltransferase domains. The analysis was performed only on the cytidylyltransferase domains. Proteins that contain two domains were considered as two separate entries: domA refers to the amino terminal domain and domB refers to the carboxyl terminal domain. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 176 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated.

Figure 12. Phylogenetic distribution of CAPdTs. Genome data were queried with either a CDP-alcohol phosphatidyltransferase HMM using HMMER (Pfam domain PF01066) or protein sequences using BLASTP and TBLASTN.

Figure 13. Clustering of CAPdTs. 132 CAPdT sequences were processed with the MCL algorithm. Activity assignment to individual clusters is based on the inclusion of sequences with specific enzyme activities.

Figure 14. Evolution of CEPTs. A phylogenetic tree of the 62 C/EPT sequences clustered in Figure 14. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 376 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated.

Figure 15. Comparison of the fish BLA-PSS predicted amino acid sequences with *B. subtilis* PSS and similar mammalian genes. The CAPdT active site region is underlined. It is present in the fish proteins but has been substantially modified in higher eukaryotes. NCBI Accession Numbers are: *D. rerio*, gil68402375, *Tetraodon nigroviridis*, gil47229076; *M. musculus*, gil58037399; *Bos taurus*, gil113911913; *Macaca mulatta*, gil109003372; and *Pan troglodytes*, gil114556073; *T. rubripes* does not have an NCBI accession number; *B. subtilis* PSS, gil 730414.

Figure 16. Evolution of PIS, PGS, CLS, PSS. 51 sequences from the PIS, CLS, PGS and PSS clusters were combined with selected PGS, PSS, PIS and CLS sequences of bacterial origin. A total of 63 sequences were used for the construction of the phylogenetic tree. Amino acid

sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 228 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated. Ec, *Escherichia coli*; Bs, *Bacillus subtilis*; Hp, *Helicobacter pylori*; Pa, *Pseudomonas aeruginosa*; Mt, *Mycobacterium tuberculosis*; Pm, *Prochlorococcus marinus*.

Figure 17. Phylogenetic distribution of PLDs. Genome data were queried with either a PLD HMM using HMMER (Pfam domain PF00614) or protein sequences using BLASTP and TBLASTN.

Figure 18. Distribution of PdTs types and activities in eukaryotic genomes. * *E. histolytica* has a total of seven genes classified in the C/EPT cluster.

Figure 19. Phylogenetic distribution of PSSEs. Genome data were queried with either a PSSE HMM using HMMER (Pfam domain PF03034) or protein sequences using BLASTP and TBLASTN.

Figure 20. Evolution of PSSEs. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 296 unambiguously aligned amino acid positions. Numbers at nodes indicate the frequencies of

branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated.

Figure 21. Phylogenetic distribution of PSDs. Genome data were queried with either a PSD HMM using HMMER (Pfam domain PF02666) or protein sequences using BLASTP and TBLASTN.

Figure 22. Evolution of PSDs. Amino acid sequence alignments were performed using ClustalW. Alignments were manually edited to remove ambiguous positions. The phylogenetic tree was created using the protein parsimony method in the Phylip package using 528 aligned amino acid positions. Numbers at nodes indicate the frequencies of branch associations on the basis of 500 bootstrap repetitions. Originating genome, accession numbers and/or locus tags for proteins included are indicated. Coding sequences from the following bacteria have been included: *Chlorobium phaeobacteroides* BS1 (CphaeBS1), *Lactobacillus acidophilus* NCFM (LacidNCFM), *Lactobacillus johnsonii* NCC 533 (LjohnsNCC533), *Vibrio angustum* S14 (Vang), *Flavobacterium johnsoniae* UW101 (FjohnsUW101), *Burkholderia cenocepacia* HI2424 (BcencepHI2424), *Burkholderia cenocepacia* AU1054 (BcencepAU1054), *Burkholderia pseudomallei* 1655 (Bpsedom1655), *Burkholderia mallei* ATCC 23344 (BmalATCC23344), *Burkholderia xenovorans* LB400 (BxenLB400), *Anaeromyxobacter dehalogenans* 2CP-C (Adeh2CP-C), *Campylobacter jejuni* subsp. *jejuni* HB93-13 (CjejHB9313), *Campylobacter jejuni* subsp. *Doylei* 269.97 (Cjej269.97), *Campylobacter jejuni* RM1221 (CjejRM1221), *Polaromonas* sp. JS666 (PolarJS666), *Coxiella burnetii* Dugway 7E9-12 (Coxbur), *Dechloromonas aromatica* RCB (DaromDCB), *Myxococcus xanthus* DK 1622 (MxanDK1622),

Helicobacter pylori HPAG1 (HpylHPAG1), *Acidovorax avenae* subsp. *citrulli* AAC00-1 (AcitrAAC00-1), *Alcanivorax burkumensis* SK2 (AborkSK2), *Vibrio fischeri* ES114 (VfisES114), *Bacillus subtilis* strain 168 (Bs), *Escherichia coli* K12 (K12), *Haemophilus influenzae* R2846 (Hinfl).

Figure 23. Phylogenetic distribution of PEMTs. Genome data were queried with either a PEMT HMM using HMMER (Pfam domain PF04191) or protein sequences using BLASTP and TBLASTN.

References

- [1] C. Kent, Annu Rev Biochem 64 (1995) 315-43.
- [2] W. Dowhan, Annu Rev Biochem 66 (1997) 199-232.
- [3] G.M. Carman and S.A. Henry, Prog Lipid Res 38 (1999) 361-99.
- [4] A. Lykidis and S. Jackowski, Prog Nucleic Acid Res Mol Biol 65 (2001) 361-93.
- [5] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Nature 409 (2001) 860-921.
- [6] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P.

- Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., *Science* 291 (2001) 1304-51.
- [7] R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyra, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, E.K. Karlsson, et al., *Nature* 420 (2002) 520-62.
- [8] S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J.M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M.D. Gelpke, J. Roach, T. Oh, I.Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S.F. Smith, M.S. Clark, Y.J. Edwards, N. Doggett, A. Zharkikh, S.V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y.H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar and S. Brenner, *Science* 297 (2002) 1301-10.
- [9] P. Dehal, Y. Satou, R.K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D.M. Goodstein, N. Harafuji, K.E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I.A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H.G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D.N. Keys, S. Haga, H. Hayashi, K. Hino, K.S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B.I. Lee, K.W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M.M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P.D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh and D.S. Rokhsar, *Science* 298 (2002) 2157-67.
- [10] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L.

- Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, et al., *Science* 287 (2000) 2185-95.
- [11] *Science* 282 (1998) 2012-8.
- [12] *Nature* 408 (2000) 796-815.
- [13] *Nature* 387 (1997) 5.
- [14] V. Wood, R. Gwilliam, M.A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E.J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M.A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R.G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schafer, S. Muller-Auer, C. Gabel, M. Fuchs, A. Dusterhoft, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T.M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dreano, S. Gloux, et al., *Nature* 415 (2002) 871-80.
- [15] J.E. Galagan, S.E. Calvo, K.A. Borkovich, E.U. Selker, N.D. Read, D. Jaffe, W. FitzHugh, L.J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C.B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M.A. Nelson, M. Werner-Washburne, C.P. Selitrennikoff, J.A. Kinsey, E.L. Braun, A. Zelter, U. Schulte, G.O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R.L. Metzenberg, D.D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R.J. Pratt, S.A. Osmani, C.P. DeSouza, L. Glass, M.J. Orbach, J.A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D.O. Natvig, L.A. Alex, G. Mannhaupt, D.J. Ebbrole, M. Freitag, I. Paulsen, M.S. Sachs, E.S. Lander, C. Nusbaum and B. Birren, *Nature* 422 (2003) 859-68.
- [16] B.J. Loftus, E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I.J. Anderson, J.A. Fraser, J.E. Allen, I.E. Bosdet, M.R. Brent, R. Chiu, T.L. Doering, M.J. Donlin, C.A. D'Souza, D.S. Fox, V. Grinberg, J. Fu, M. Fukushima, B.J. Haas, J.C. Huang, G. Janbon, S.J. Jones, H.L. Koo, M.I. Krzywinski, J.K. Kwon-Chung, K.B. Lengeler, R. Maiti, M.A. Marra, R.E. Marra, C.A. Mathewson, T.G. Mitchell, M. Pertea, F.R. Riggs, S.L. Salzberg, J.E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway,

- C.A. Specht, B.B. Suh, A. Tenney, T.R. Utterback, B.L. Wickes, J.R. Wortman, N.H. Wye, J.W. Kronstad, J.K. Lodge, J. Heitman, R.W. Davis, C.M. Fraser and R.W. Hyman, *Science* 307 (2005) 1321-4.
- [17] J.A. Eisen, R.S. Coyne, M. Wu, D. Wu, M. Thiagarajan, J.R. Wortman, J.H. Badger, Q. Ren, P. Amedeo, K.M. Jones, L.J. Tallon, A.L. Delcher, S.L. Salzberg, J.C. Silva, B.J. Haas, W.H. Majoros, M. Farzad, J.M. Carlton, R.K. Smith, Jr., J. Garg, R.E. Pearlman, K.M. Karrer, L. Sun, G. Manning, N.C. Elde, A.P. Turkewitz, D.J. Asai, D.E. Wilkes, Y. Wang, H. Cai, K. Collins, B.A. Stewart, S.R. Lee, K. Wilamowska, Z. Weinberg, W.L. Ruzzo, D. Wloga, J. Gaertig, J. Frankel, C.C. Tsao, M.A. Gorovsky, P.J. Keeling, R.F. Waller, N.J. Patron, J.M. Cherry, N.A. Stover, C.J. Krieger, C. del Toro, H.F. Ryder, S.C. Williamson, R.A. Barbeau, E.P. Hamilton and E. Orias, *PLoS Biol* 4 (2006) e286.
- [18] M.J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E. Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M.S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Perte, J. Allen, J. Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S. Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser and B. Barrell, *Nature* 419 (2002) 498-511.
- [19] A. Pain, H. Renauld, M. Berriman, L. Murphy, C.A. Yeats, W. Weir, A. Kerhornou, M. Aslett, R. Bishop, C. Bouchier, M. Cochet, R.M. Coulson, A. Cronin, E.P. de Villiers, A. Fraser, N. Fosker, M. Gardner, A. Goble, S. Griffiths-Jones, D.E. Harris, F. Katzer, N. Larke, A. Lord, P. Maser, S. McKellar, P. Mooney, F. Morton, V. Nene, S. O'Neil, C. Price, M.A. Quail, E. Rabinowitsch, N.D. Rawlings, S. Rutter, D. Saunders, K. Seeger, T. Shah, R. Squares, S. Squares, A. Tivey, A.R. Walker, J. Woodward, D.A. Dobbelaere, G. Langsley, M.A. Rajandream, D. McKeever, B. Shiels, A. Tait, B. Barrell and N. Hall, *Science* 309 (2005) 131-3.
- [20] P. Xu, G. Widmer, Y. Wang, L.S. Ozaki, J.M. Alves, M.G. Serrano, D. Puiu, P. Manque, D. Akiyoshi, A.J. Mackey, W.R. Pearson, P.H. Dear, A.T. Bankier, D.L. Peterson, M.S. Abrahamsen, V. Kapur, S. Tzipori and G.A. Buck, *Nature* 431 (2004) 1107-12.
- [21] L. Eichinger, J.A. Pachebat, G. Glockner, M.A. Rajandream, R. Sugang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B.A. Konfortov, F. Rivero, A.T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M.A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E.C. Cox, R.L. Chisholm, R. Gibbs, W.F. Loomis, M. Platzter, R.R. Kay, J. Williams, P.H. Dear, A.A. Noegel, B. Barrell and A. Kuspa, *Nature* 435 (2005) 43-57.
- [22] B. Loftus, I. Anderson, R. Davies, U.C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R.P. Hirt, B.J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene,

- J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M.A. Quail, E. Rabbino-witsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S.E. Stroup, S. Bhattacharya, A. Lohia, P.G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N.M. El-Sayed, W.A. Petri, Jr., C.G. Clark, T.M. Embley, B. Barrell, C.M. Fraser and N. Hall, *Nature* 433 (2005) 865-8.
- [23] A.C. Ivens, C.S. Peacock, E.A. Worthey, L. Murphy, G. Aggarwal, M. Berriman, E. Sisk, M.A. Rajandream, E. Adlem, R. Aert, A. Anupama, Z. Apostolou, P. Attipoe, N. Bason, C. Bauser, A. Beck, S.M. Beverley, G. Bianchetti, K. Borzym, G. Bothe, C.V. Bruschi, M. Collins, E. Cadag, L. Ciarloni, C. Clayton, R.M. Coulson, A. Cronin, A.K. Cruz, R.M. Davies, J. De Gaudenzi, D.E. Dobson, A. Duesterhoeft, G. Fazelina, N. Fosker, A.C. Frasch, A. Fraser, M. Fuchs, C. Gabel, A. Goble, A. Goffeau, D. Harris, C. Hertz-Fowler, H. Hilbert, D. Horn, Y. Huang, S. Klages, A. Knights, M. Kube, N. Larke, L. Litvin, A. Lord, T. Louie, M. Marra, D. Masuy, K. Matthews, S. Michaeli, J.C. Mottram, S. Muller-Auer, H. Munden, S. Nelson, H. Norbertczak, K. Oliver, S. O'Neil, M. Pentony, T.M. Pohl, C. Price, B. Purnelle, M.A. Quail, E. Rabbino-witsch, R. Reinhardt, M. Rieger, J. Rinta, J. Robben, L. Robertson, J.C. Ruiz, S. Rutter, D. Saunders, M. Schafer, J. Schein, D.C. Schwartz, K. Seeger, A. Seyler, S. Sharp, H. Shin, D. Sivam, R. Squares, S. Squares, V. Tosato, C. Vogt, G. Volckaert, R. Wambutt, T. Warren, H. Wedler, J. Woodward, S. Zhou, W. Zimmermann, D.F. Smith, J.M. Blackwell, K.D. Stuart, B. Barrell, et al., *Science* 309 (2005) 436-42.
- [24] N.M. El-Sayed, P.J. Myler, D.C. Bartholomeu, D. Nilsson, G. Aggarwal, A.N. Tran, E. Ghedin, E.A. Worthey, A.L. Delcher, G. Blandin, S.J. Westenberger, E. Caler, G.C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D.A. Campbell, M. Carrington, J. Crabtree, H. Darban, J.F. da Silveira, P. de Jong, K. Edwards, P.T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A.C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, S. Kluge, H. Koo, D. Lacerda, M.J. Levin, H. Lorenzi, T. Louie, C.R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J.C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J.L. Ramirez, J. Rinta, L. Robertson, S.L. Salzberg, D.O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A.J. Simpson, E. Sisk, M.T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P.N. Ward, B. Wickstead, J. Wortman, O. White, C.M. Fraser, K.D. Stuart and B. Andersson, *Science* 309 (2005) 409-15.
- [25] M. Berriman, E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D.C. Bartholomeu, N.J. Lennard, E. Caler, N.E. Hamlin, B. Haas, U. Bohme, L. Hannick, M.A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U.C. Alsmark, C. Arrowsmith, R.J. Atkin, A.J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.J. Chillingworth, C. Churcher, L.N. Clark, C.H. Corton, A. Cronin, R.M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M.C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B.R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A.X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P.J. Mooney, S. Moule, D.M. Martin, G.W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M.A. Quail, E. Rabbino-witsch, M.A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S.

- Sharp, M. Simmonds, A.J. Simpson, L. Tallon, C.M. Turner, A. Tait, A.R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M.D. Adams, T.M. Embley, K. Gull, E. Ullu, J.D. Barry, A.H. Fairlamb, F. Opperdoes, B.G. Barrell, J.E. Donelson, N. Hall, C.M. Fraser, et al., *Science* 309 (2005) 416-22.
- [26] T.M. Embley and W. Martin, *Nature* 440 (2006) 623-30.
- [27] A. Bateman, L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats and S.R. Eddy, *Nucleic Acids Res* 32 (2004) D138-41.
- [28] A.J. Enright, S. Van Dongen and C.A. Ouzounis, *Nucleic Acids Res* 30 (2002) 1575-84.
- [29] A.J. Enright, V. Kunin and C.A. Ouzounis, *Nucleic Acids Res* 31 (2003) 4632-8.
- [30] H. Shen, P.N. Heacock, C.J. Clancey and W. Dowhan, *J Biol Chem* 271 (1996) 789-95.
- [31] L. Wu, B. Niemeyer, N. Colley, M. Socolich and C.S. Zuker, *Nature* 373 (1995) 216-22.
- [32] D. Martin, L. Gannoun-Zaki, S. Bonnefoy, P. Eldin, K. Wengelnik and H. Vial, *Mol Biochem Parasitol* 110 (2000) 93-105.
- [33] S.L. Inglis-Broadgate, L. Ocaka, R. Banerjee, M. Gaasenbeek, J.P. Chapple, M.E. Cheetham, B.J. Clark, D.M. Hunt and S. Halford, *Gene* 356 (2005) 19-31.
- [34] R. Weeks, W. Dowhan, H. Shen, N. Balantac, B. Meengs, E. Nudelman and D.W. Leung, *DNA Cell Biol* 16 (1997) 281-9.
- [35] A. Lykidis, P.D. Jackson, C.O. Rock and S. Jackowski, *J Biol Chem* 272 (1997) 33402-9.
- [36] M. Volta, A. Bulfone, C. Gattuso, E. Rossi, M. Mariani, G.G. Consalez, O. Zuffardi, A. Ballabio, S. Banfi and B. Franco, *Genomics* 55 (1999) 68-77.
- [37] F. Simmer, C. Moorman, A.M. van der Linden, E. Kuijk, P.V. van den Berghe, R.S. Kamath, A.G. Fraser, J. Ahringer and R.H. Plasterk, *PLoS Biol* 1 (2003) E12.
- [38] R.S. Kamath, A.G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D.P. Welchman, P. Zipperlen and J. Ahringer, *Nature* 421 (2003) 231-7.
- [39] A. Martin, P. Hales and D.N. Brindley, *Biochem J* 245 (1987) 347-55.
- [40] Z. Jamal, A. Martin, A. Gomez-Munoz and D.N. Brindley, *J Biol Chem* 266 (1991) 2988-96.
- [41] A. Gomez-Munoz, G.M. Hatch, A. Martin, Z. Jamal, D.E. Vance and D.N. Brindley, *FEBS Lett* 301 (1992) 103-6.
- [42] G.S. Han, W.I. Wu and G.M. Carman, *J Biol Chem* (2006).
- [43] M. Peterfy, J. Phan, P. Xu and K. Reue, *Nat Genet* 27 (2001) 121-4.
- [44] C.A. Langner, E.H. Birkenmeier, O. Ben-Zeev, M.C. Schotz, H.O. Sweet, M.T. Davisson and J.I. Gordon, *J Biol Chem* 264 (1989) 7994-8003.
- [45] K. Reue, P. Xu, X.P. Wang and B.G. Slavin, *J Lipid Res* 41 (2000) 1067-76.
- [46] J. Phan and K. Reue, *Cell Metab* 1 (2005) 73-83.
- [47] B. Sonnichsen, L.B. Koski, A. Walsh, P. Marschall, B. Neumann, M. Brehm, A.M. Alleaume, J. Artelt, P. Bettencourt, E. Cassin, M. Hewitson, C. Holz, M. Khan, S. Lazik, C. Martin, B. Nitzsche, M. Ruer, J. Stamford, M. Winzi, R. Heinkel, M. Roder, J. Finell, H. Hantsch, S.J. Jones, M. Jones, F. Piano, K.C. Gunsalus, K. Oegema, P. Gonczy, A. Coulson, A.A. Hyman and C.J. Echeverri, *Nature* 434 (2005) 462-9.

- [48] J.F. Rual, J. Ceron, J. Koreth, T. Hao, A.S. Nicot, T. Hirozane-Kishikawa, J. Vandenhaute, S.H. Orkin, D.E. Hill, S. van den Heuvel and M. Vidal, *Genome Res* 14 (2004) 2162-8.
- [49] D.N. Brindley and D.W. Waggoner, *J Biol Chem* 273 (1998) 24281-4.
- [50] D.N. Brindley, D. English, C. Pilquil, K. Buri and Z.C. Ling, *Biochim Biophys Acta* 1582 (2002) 33-44.
- [51] Y.J. Sigal, M.I. McDermott and A.J. Morris, *Biochem J* 387 (2005) 281-93.
- [52] M.M. Nagiec, E.E. Nagiec, J.A. Baltisberger, G.B. Wells, R.L. Lester and R.C. Dickson, *J Biol Chem* 272 (1997) 9809-17.
- [53] P.W. Denny, H. Shams-Eldin, H.P. Price, D.F. Smith and R.T. Schwarz, *J Biol Chem* 281 (2006) 28200-9.
- [54] K. Huitema, J. van den Dikkenberg, J.F. Brouwers and J.C. Holthuis, *Embo J* 23 (2004) 33-44.
- [55] T. Ichio and C.R. Raetz, *J Bacteriol* 153 (1983) 722-30.
- [56] T. Ichio, *J Bacteriol* 170 (1988) 5117-24.
- [57] T. Ichio, *J Bacteriol* 170 (1988) 5110-6.
- [58] C.R. Funk, L. Zimniak and W. Dowhan, *J Bacteriol* 174 (1992) 205-13.
- [59] D.A. Toke, W.L. Bennett, J. Oshiro, W.I. Wu, D.R. Voelker and G.M. Carman, *J Biol Chem* 273 (1998) 14331-8.
- [60] G.S. Han, C.N. Johnston and G.M. Carman, *J Biol Chem* 279 (2004) 5338-45.
- [61] R. Jasinska, Q.X. Zhang, C. Pilquil, I. Singh, J. Xu, J. Dewald, D.A. Dillon, L.G. Berthiaume, G.M. Carman, D.W. Waggoner and D.N. Brindley, *Biochem J* 340 (Pt 3) (1999) 677-86.
- [62] D.W. Waggoner, J. Xu, I. Singh, R. Jasinska, Q.X. Zhang and D.N. Brindley, *Biochim Biophys Acta* 1439 (1999) 299-316.
- [63] Q.X. Zhang, C.S. Pilquil, J. Dewald, L.G. Berthiaume and D.N. Brindley, *Biochem J* 345 Pt 2 (2000) 181-4.
- [64] T. Uchida and S. Yamashita, *J Biol Chem* 267 (1992) 10156-62.
- [65] C. Aoyama, K. Nakashima and K. Ishidate, *Biochim Biophys Acta* 1393 (1998) 179-85.
- [66] C. Aoyama, K. Nakashima, M. Matsui and K. Ishidate, *Biochim Biophys Acta* 1390 (1998) 1-7.
- [67] P. Pavlidis, M. Ramaswami and M.A. Tanouye, *Cell* 79 (1994) 23-33.
- [68] A. Lykidis, J. Wang, M.A. Karim and S. Jackowski, *J Biol Chem* 276 (2001) 2174-9.
- [69] Y. Tian, P. Jackson, C. Gunter, J. Wang, C.O. Rock and S. Jackowski, *J Biol Chem* 281 (2006) 28438-49.
- [70] K. Hosaka, T. Kodaki and S. Yamashita, *J Biol Chem* 264 (1989) 2053-9.
- [71] K. Kim, K.H. Kim, M.K. Storey, D.R. Voelker and G.M. Carman, *J Biol Chem* 274 (1999) 14857-66.
- [72] P. Gee and C. Kent, *Biochim Biophys Acta* 1648 (2003) 33-42.
- [73] K. Ashrafi, F.Y. Chang, J.L. Watts, A.G. Fraser, R.S. Kamath, J. Ahringer and G. Ruvkun, *Nature* 421 (2003) 268-72.
- [74] F. Piano, A.J. Schetter, M. Mangone, L. Stein and K.J. Kemphues, *Curr Biol* 10 (2000) 1619-22.
- [75] W. Tang, G.A. Keesler and I. Tabas, *J Biol Chem* 272 (1997) 13146-51.
- [76] G.B. Kalmar, R.J. Kay, A.C. LaChance and R.B. Cornell, *Biochim Biophys Acta* 1219 (1994) 328-34.

- [77] G.B. Kalmar, R.J. Kay, A. Lachance, R. Aebersold and R.B. Cornell, *Proc Natl Acad Sci U S A* 87 (1990) 6029-33.
- [78] A. Lykidis, K.G. Murti and S. Jackowski, *J Biol Chem* 273 (1998) 14022-9.
- [79] A. Lykidis, I. Baburina and S. Jackowski, *J Biol Chem* 274 (1999) 26992-7001.
- [80] Y. Tsukagoshi, J. Nikawa and S. Yamashita, *Eur J Biochem* 169 (1987) 477-86.
- [81] J.A. Friesen, M.F. Liu and C. Kent, *Biochim Biophys Acta* 1533 (2001) 86-98.
- [82] B.A. Helmink and J.A. Friesen, *Biochim Biophys Acta* 1683 (2004) 78-88.
- [83] S.B. Choi, K.W. Lee and S.H. Cho, *Mol Cells* 7 (1997) 58-63.
- [84] H.J. Yeo, J. Sri Widada, O. Mercereau-Puijalon and H.J. Vial, *Eur J Biochem* 233 (1995) 62-72.
- [85] R. Min-Seok, Y. Kawamata, H. Nakamura, A. Ohta and M. Takagi, *J Biochem (Tokyo)* 120 (1996) 1040-7.
- [86] A. Nakashima, K. Hosaka and J. Nikawa, *J Biol Chem* 272 (1997) 9567-72.
- [87] B.A. Bladergroen, M. Houweling, M.J. Geelen and L.M. van Golde, *Biochem J* 343 Pt 1 (1999) 107-14.
- [88] W. Yang, C.B. Mason, S.V. Pollock, T. Lavezzi, J.V. Moroney and T.S. Moore, *Biochem J* 382 (2004) 51-7.
- [89] L. Wang, S. Magdaleno, I. Tabas and S. Jackowski, *Mol Cell Biol* 25 (2005) 3357-63.
- [90] S. Jackowski, J.E. Rehg, Y.M. Zhang, J. Wang, K. Miller, P. Jackson and M.A. Karim, *Mol Cell Biol* 24 (2004) 4720-33.
- [91] T. Gupta and T. Schupbach, *Development* 130 (2003) 6075-87.
- [92] U. Weber, C. Eroglu and M. Mlodzik, *Dev Cell* 5 (2003) 559-70.
- [93] A.E. Cleves, T.P. McGee, E.A. Whitters, K.M. Champion, J.R. Aitken, W. Dowhan, M. Goebel and V.A. Bankaitis, *Cell* 64 (1991) 789-800.
- [94] I. Maeda, Y. Kohara, M. Yamamoto and A. Sugimoto, *Curr Biol* 11 (2001) 171-6.
- [95] D.H. Beach, G.G. Holz, Jr., B.N. Singh and D.G. Lindmark, *Mol Biochem Parasitol* 44 (1991) 97-108.
- [96] S.C. Dryden and W. Dowhan, *J Bacteriol* 178 (1996) 1030-8.
- [97] S.C. Chang, P.N. Heacock, C.J. Clancey and W. Dowhan, *J Biol Chem* 273 (1998) 9829-36.
- [98] K. Kawasaki, O. Kuge, S.C. Chang, P.N. Heacock, M. Rho, K. Suzuki, M. Nishijima and W. Dowhan, *J Biol Chem* 274 (1999) 1828-34.
- [99] C. Xu, H. Hartel, H. Wada, M. Hagio, B. Yu, C. Eakin and C. Benning, *Plant Physiol* 129 (2002) 594-604.
- [100] E. Babiychuk, F. Muller, H. Eubel, H.P. Braun, M. Frentzen and S. Kushnir, *Plant J* 33 (2003) 899-909.
- [101] M.L. Greenberg and J.M. Lopes, *Microbiol Rev* 60 (1996) 1-20.
- [102] B.E. Tropp, L. Ragolia, W. Xia, W. Dowhan, R. Milkman, K.E. Rudd, R. Ivanisevic and D.J. Savic, *J Bacteriol* 177 (1995) 5155-7.
- [103] S.C. Chang, P.N. Heacock, E. Mileykovskaya, D.R. Voelker and W. Dowhan, *J Biol Chem* 273 (1998) 14933-41.
- [104] D. Chen, X.Y. Zhang and Y. Shi, *Biochem J* 398 (2006) 169-76.
- [105] R.H. Houtkooper, H. Akbari, H. van Lenthe, W. Kulik, R.J. Wanders, M. Frentzen and F.M. Vaz, *FEBS Lett* 580 (2006) 3059-64.
- [106] B. Lu, F.Y. Xu, Y.J. Jiang, P.C. Choy, G.M. Hatch, C. Grunfeld and K.R. Feingold, *J Lipid Res* 47 (2006) 1140-5.

- [107] Z. Ge and D.E. Taylor, *J Bacteriol* 179 (1997) 4970-6.
- [108] M. Okada, H. Matsuzaki, I. Shibuya and K. Matsumoto, *J Bacteriol* 176 (1994) 7456-61.
- [109] S. Yamashita and J. Nikawa, *Biochim Biophys Acta* 1348 (1997) 228-35.
- [110] A. DeChavigny, P.N. Heacock and W. Dowhan, *J Biol Chem* 266 (1991) 5323-32.
- [111] J. Nikawa, T. Kodaki and S. Yamashita, *J Biol Chem* 262 (1987) 4876-81.
- [112] S. Collin, A.M. Justin, C. Cantrel, V. Arondel and J.C. Kader, *Eur J Biochem* 262 (1999) 652-8.
- [113] H.W. Xue, K. Hosaka, G. Plesch and B. Mueller-Roeber, *Plant Mol Biol* 42 (2000) 757-64.
- [114] M. Jackson, D.C. Crick and P.J. Brennan, *J Biol Chem* 275 (2000) 30092-9.
- [115] C. Sohlenkamp, K.E. de Rudder, V. Rohrs, I.M. Lopez-Lara and O. Geiger, *J Biol Chem* 275 (2000) 18919-25.
- [116] R.H. Hjelmstad and R.M. Bell, *J Biol Chem* 263 (1988) 19748-57.
- [117] R.H. Hjelmstad and R.M. Bell, *J Biol Chem* 262 (1987) 3909-17.
- [118] A.L. Henneberry, G. Wistow and C.R. McMaster, *J Biol Chem* 275 (2000) 29808-15.
- [119] A.L. Henneberry and C.R. McMaster, *Biochem J* 339 (Pt 2) (1999) 291-8.
- [120] W. Yang, J.V. Moroney and T.S. Moore, *Arch Biochem Biophys* 430 (2004) 198-209.
- [121] G.V. Kryukov, S. Castellano, S.V. Novoselov, A.V. Lobanov, O. Zehtab, R. Guigo and V.N. Gladyshev, *Science* 300 (2003) 1439-43.
- [122] A.G. Fraser, R.S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann and J. Ahringer, *Nature* 408 (2000) 325-30.
- [123] E. Delhaize, D.M. Hebb, K.D. Richards, J.M. Lin, P.R. Ryan and R.C. Gardner, *J Biol Chem* 274 (1999) 7082-8.
- [124] J.G. Williams and C.R. McMaster, *J Biol Chem* 273 (1998) 13482-7.
- [125] H.E. McDermid and B.E. Morrow, *Am J Hum Genet* 70 (2002) 1077-88.
- [126] M. Liscovitch, M. Czarny, G. Fiucci and X. Tang, *Biochem J* 345 Pt 3 (2000) 401-15.
- [127] S.B. Aley, W.A. Scott and Z.A. Cohn, *J Exp Med* 152 (1980) 391-404.
- [128] O. Kuge, M. Nishijima and Y. Akamatsu, *J Biol Chem* 266 (1991) 24184-9.
- [129] S.J. Stone, Z. Cui and J.E. Vance, *J Biol Chem* 273 (1998) 7293-302.
- [130] O. Kuge, K. Saito and M. Nishijima, *J Biol Chem* 272 (1997) 19133-9.
- [131] S.J. Stone and J.E. Vance, *Biochem J* 342 (Pt 1) (1999) 57-64.
- [132] M.O. Bergo, B.J. Gavino, R. Steenbergen, B. Sturbois, A.F. Parlow, D.A. Sanan, W.C. Skarnes, J.E. Vance and S.G. Young, *J Biol Chem* 277 (2002) 47701-8.
- [133] K. Saito, M. Nishijima and O. Kuge, *J Biol Chem* 273 (1998) 17199-205.
- [134] Q.X. Li and W. Dowhan, *J Biol Chem* 263 (1988) 11516-22.
- [135] K. Matsumoto, M. Okada, Y. Horikoshi, H. Matsuzaki, T. Kishi, M. Itaya and I. Shibuya, *J Bacteriol* 180 (1998) 100-6.
- [136] P.J. Trotter, J. Pedretti and D.R. Voelker, *J Biol Chem* 268 (1993) 21416-24.
- [137] C.J. Clancey, S.C. Chang and W. Dowhan, *J Biol Chem* 268 (1993) 24580-90.
- [138] P.J. Trotter and D.R. Voelker, *J Biol Chem* 270 (1995) 6062-70.
- [139] P.J. Trotter, J. Pedretti, R. Yates and D.R. Voelker, *J Biol Chem* 270 (1995) 6071-80.
- [140] R. Steenbergen, T.S. Nanowski, A. Beigneux, A. Kulinski, S.G. Young and J.E. Vance, *J Biol Chem* 280 (2005) 40032-40.
- [141] D. Rontein, W.I. Wu, D.R. Voelker and A.D. Hanson, *Plant Physiol* 132 (2003) 1678-87.
- [142] F. Baunaure, P. Eldin, A.M. Cathiard and H. Vial, *Mol Microbiol* 51 (2004) 33-46.

- [143] P. Gonczy, C. Echeverri, K. Oegema, A. Coulson, S.J. Jones, R.R. Copley, J. Duperon, J. Oegema, M. Brehm, E. Cassin, E. Hannak, M. Kirkham, S. Pichler, K. Flohrs, A. Goessen, S. Leidel, A.M. Alleaume, C. Martin, N. Ozlu, P. Bork and A.A. Hyman, *Nature* 408 (2000) 331-6.
- [144] C. Sohlenkamp, I.M. Lopez-Lara and O. Geiger, *Prog Lipid Res* 42 (2003) 115-62.
- [145] Z. Cui, J.E. Vance, M.H. Chen, D.R. Voelker and D.E. Vance, *J Biol Chem* 268 (1993) 16655-63.
- [146] C.J. Walkey, L.R. Donohue, R. Bronson, L.B. Agellon and D.E. Vance, *Proc Natl Acad Sci U S A* 94 (1997) 12880-5.
- [147] C.J. Walkey, L. Yu, L.B. Agellon and D.E. Vance, *J Biol Chem* 273 (1998) 27043-6.

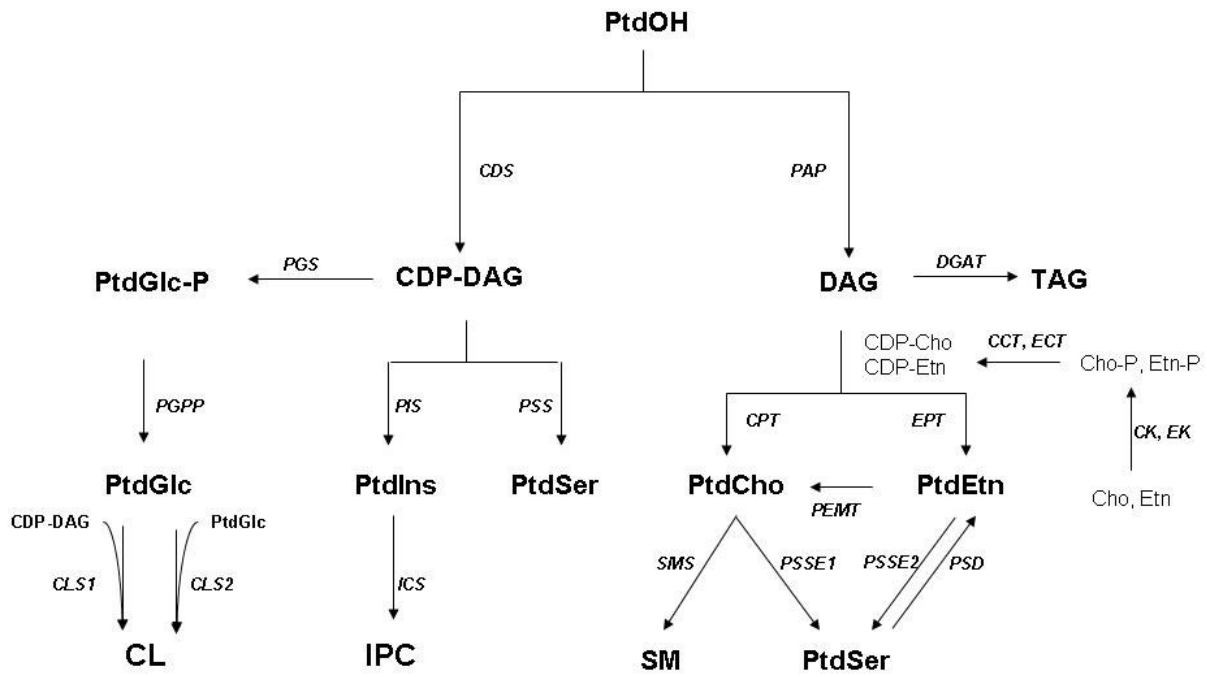


Figure 1

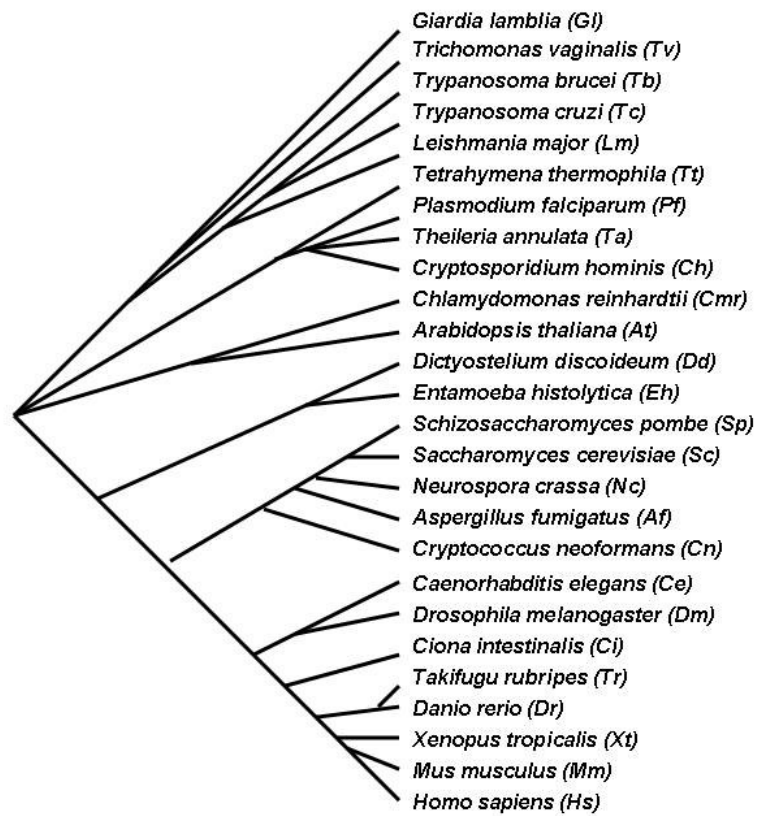


Figure 2

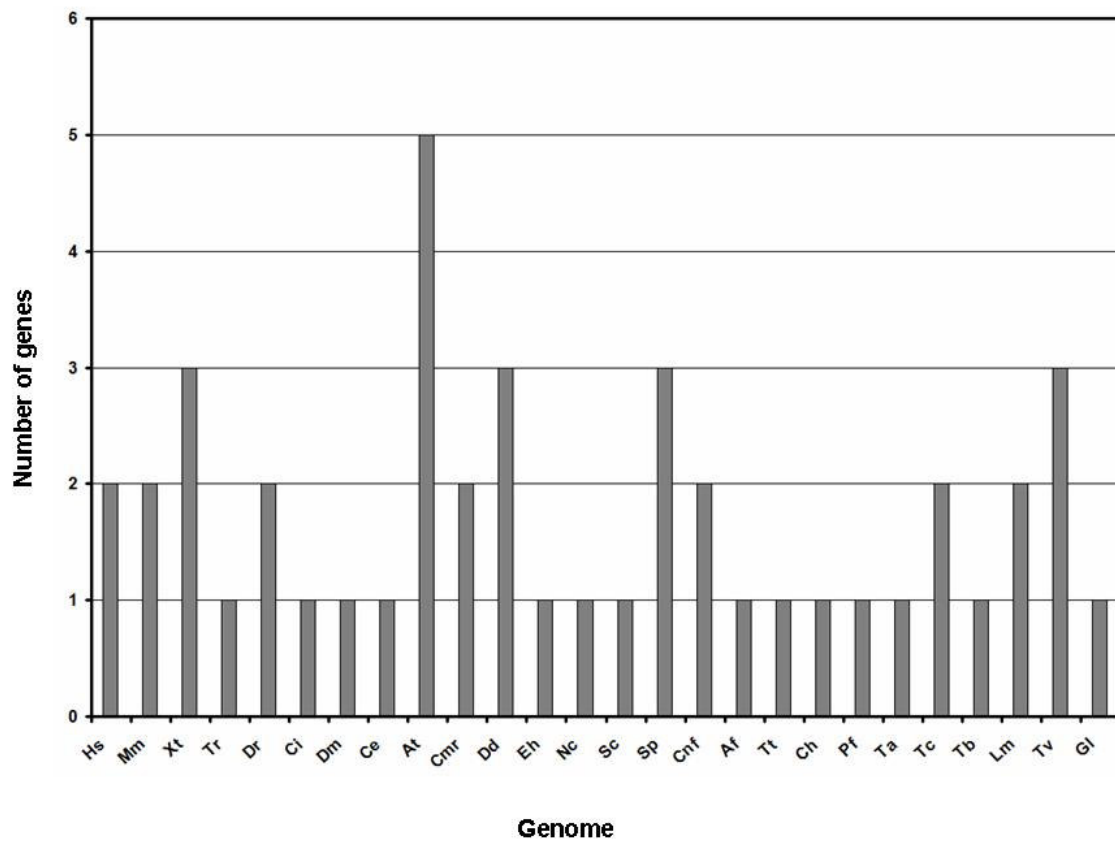


Figure 3

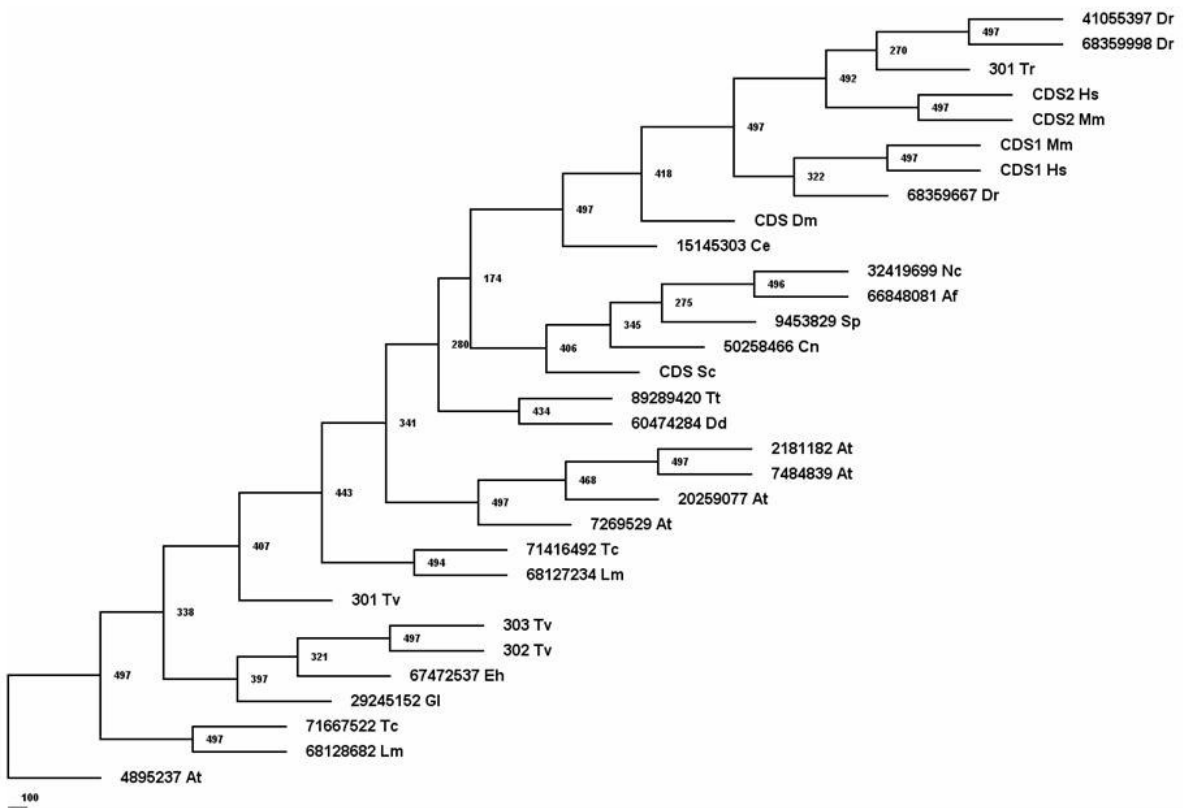


Figure 4

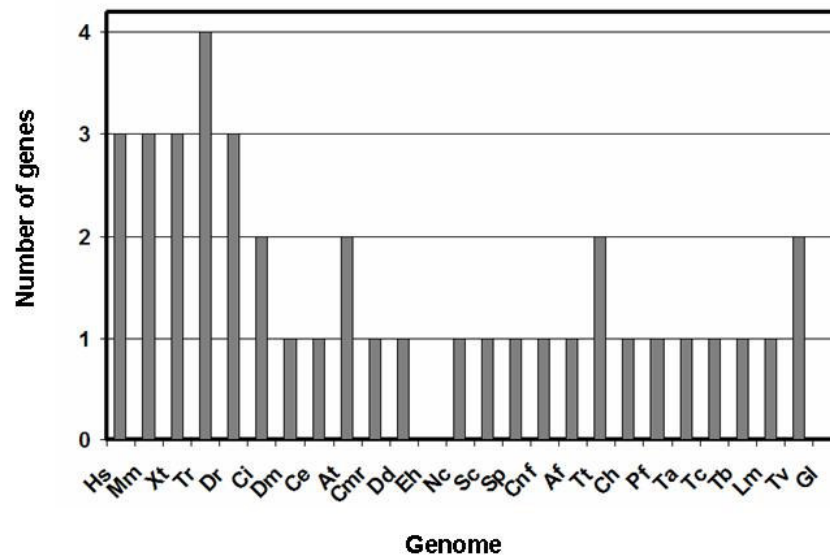


Figure 5

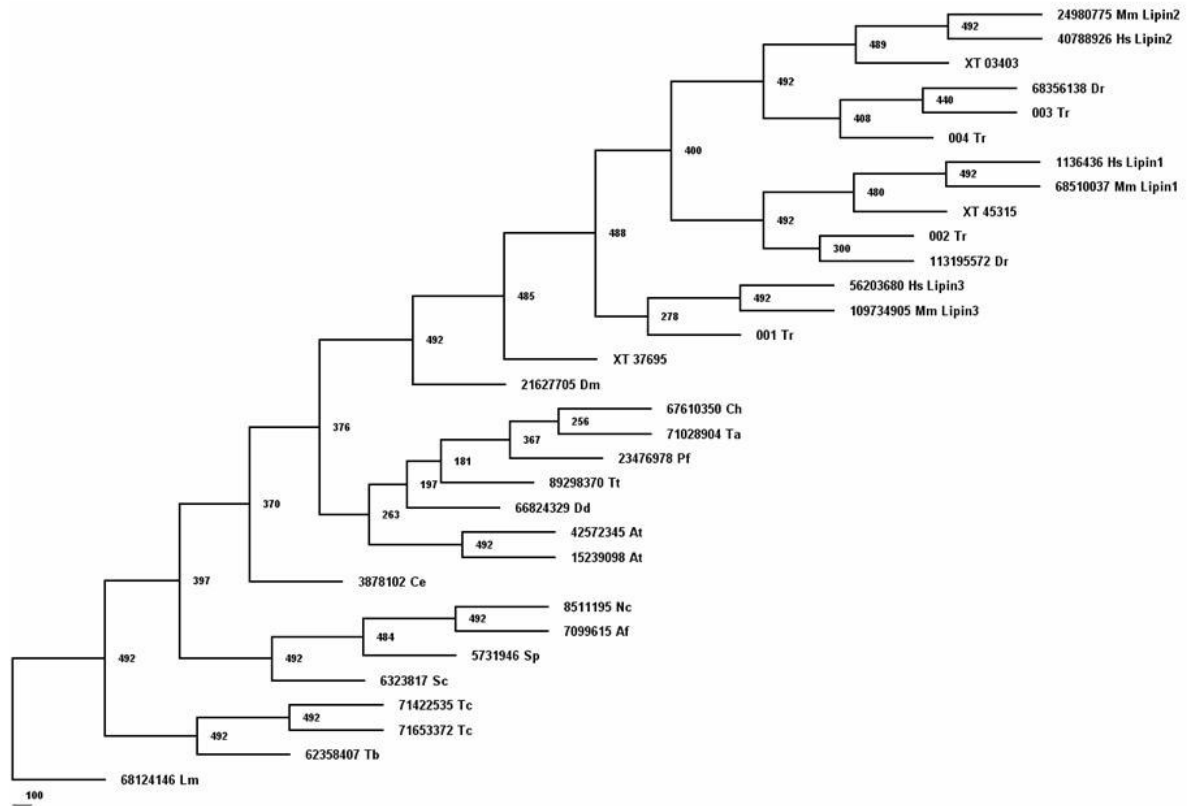


Figure 6

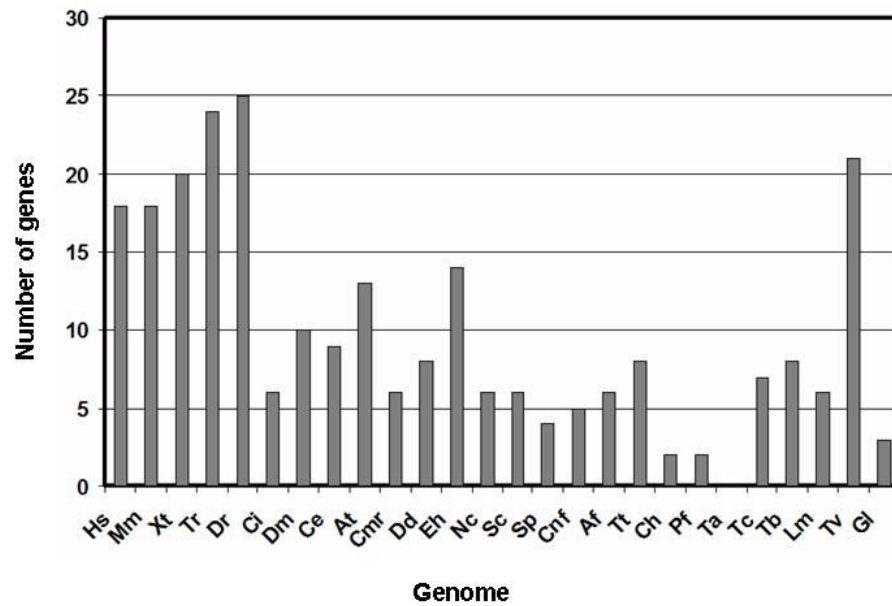


Figure 7

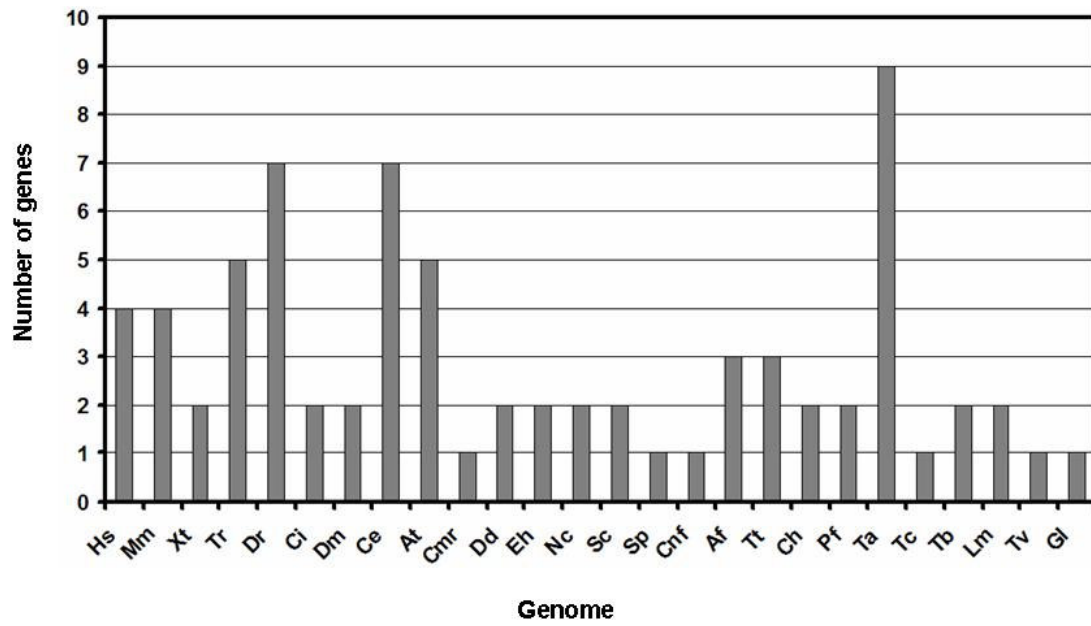


Figure 8

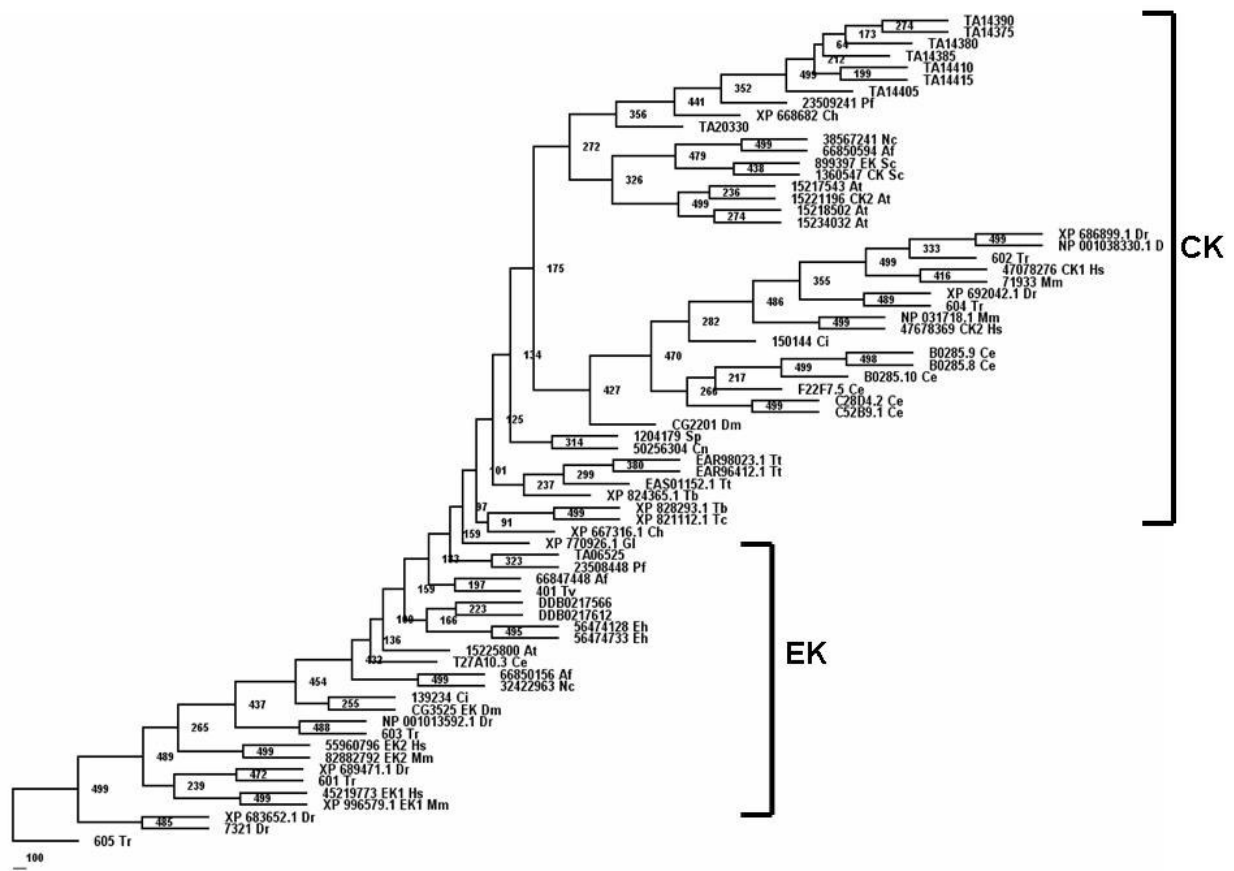


Figure 9

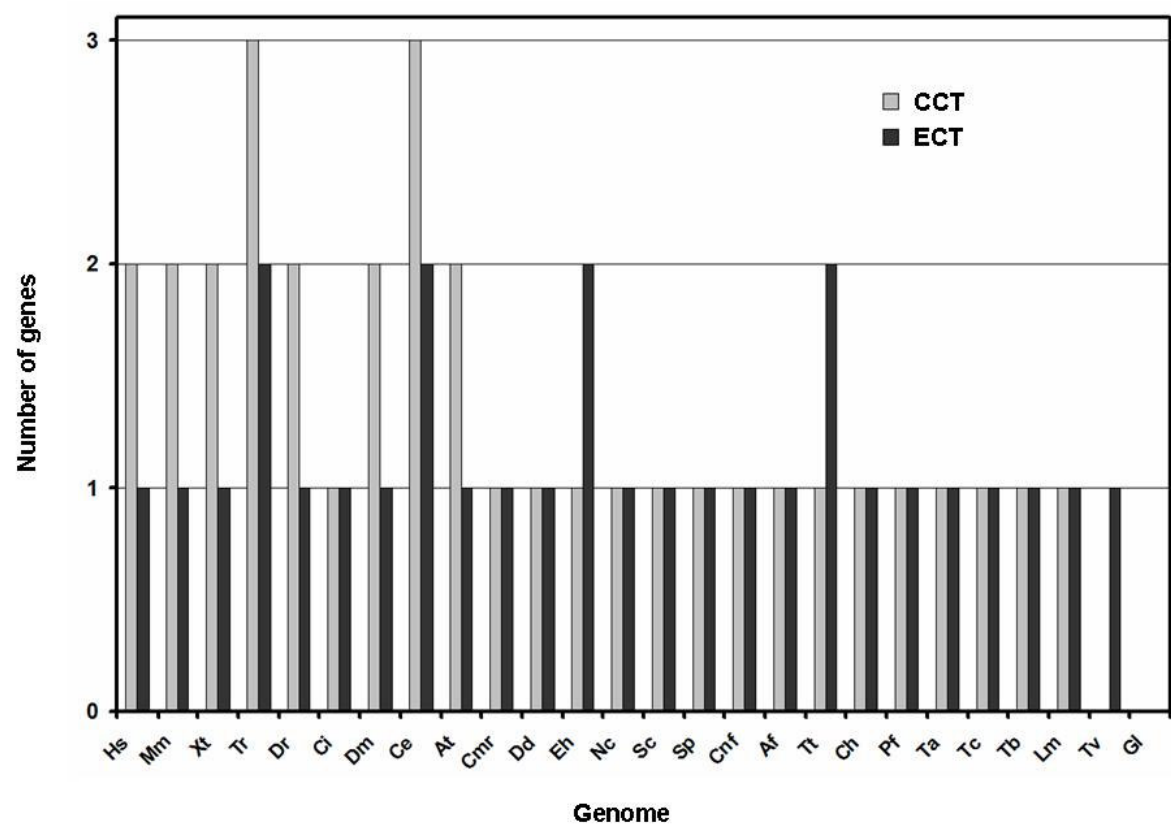


Figure 10

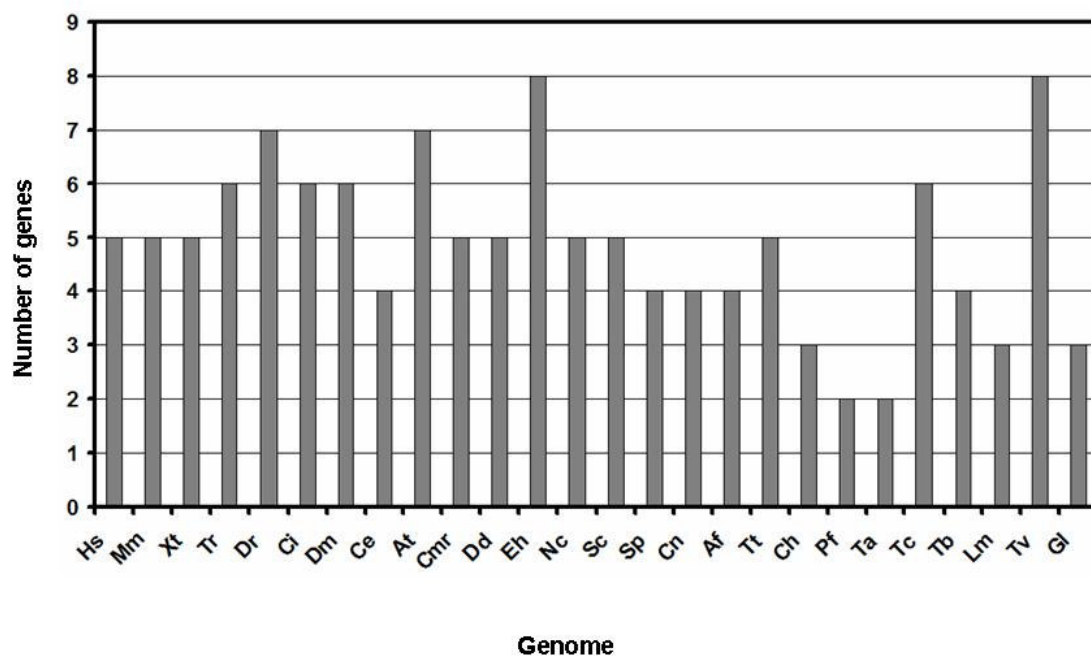


Figure 12

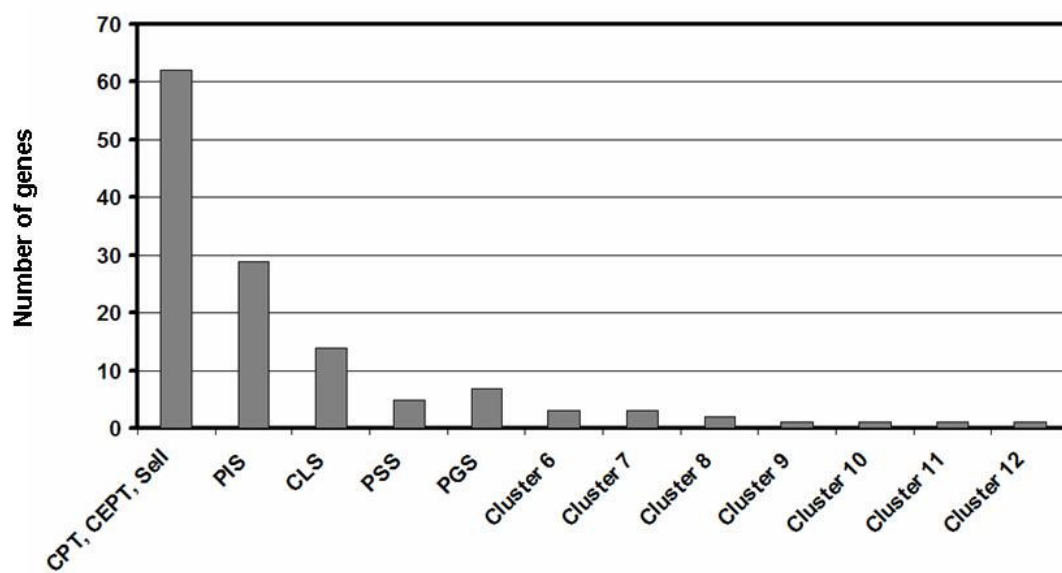


Figure 13

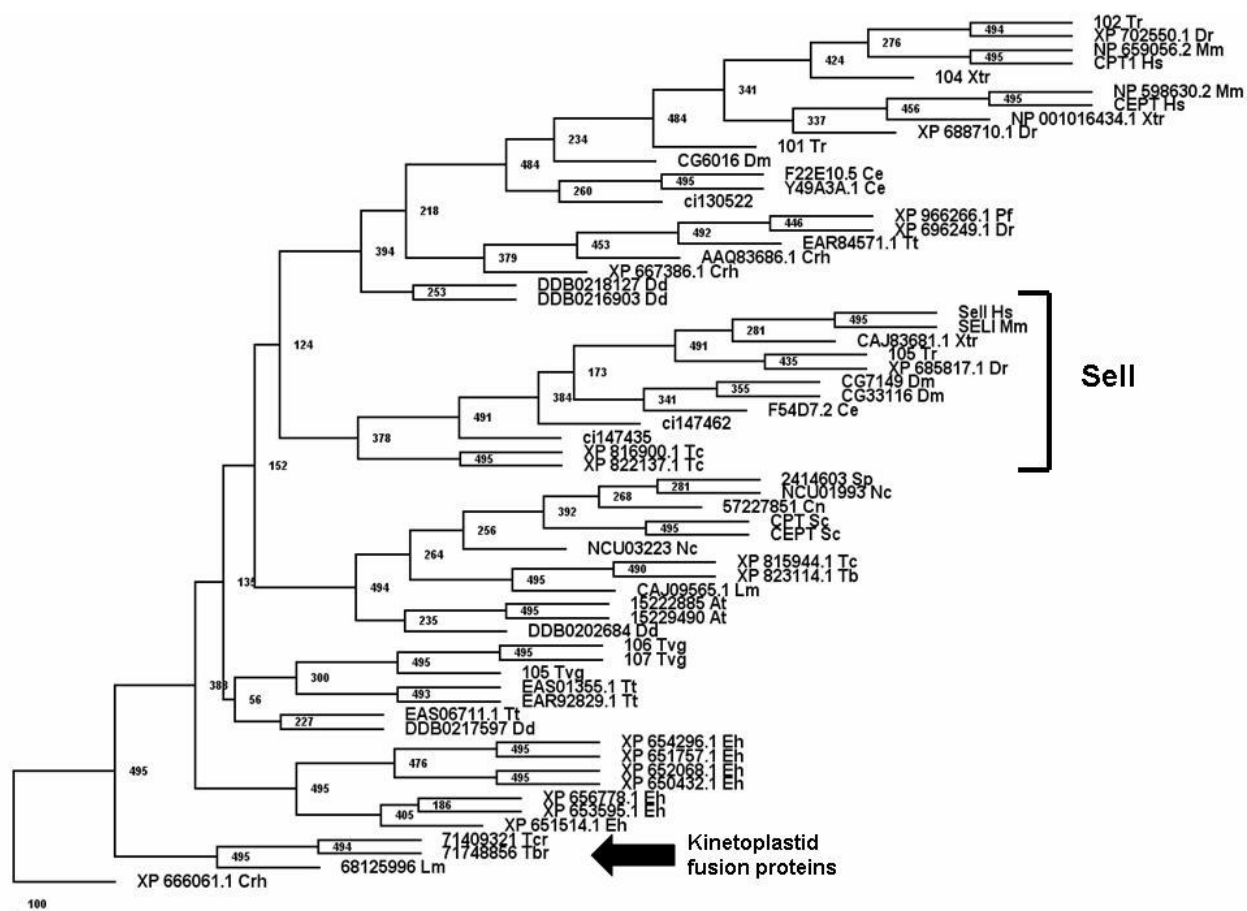


Figure 14

```

      *      20      *      40      *      60      *      80      *      100
P_troglody : -----MGLQ-EESKGTSEFFSHDQSHS
M_mulatta : -----MVGIWTFGMSALQKGGKGTSEFFSHDQSHS
B_taurus : MSGPQVSIRSTCLLHVLSAVGAGSRGGGSRERAVLTWRRPAPAWSLLCFHLGIFAVAAIPQGARNSEGGTRRSTLMSSAWTD--QRQEEAGECSEFFTKEQSHF
M_musculus : -----MLGTGDPNLHSSCILSRK-KGRQGHHMKHEQIGS
T_nigrovir : -----ASGLFRCTDRLLMDQAARQ
T_rubripes : -----MVLRAP
D_rerio : -----MTQLCALSPKRSVIPQINILSLFGAGSVLIRFCCPVVSALFQCTNKLFTDRTRQ
PSS_B_Subt : -----MILLTP-----

      *      120      *      140      *      160      *      180      *      200      *
P_troglody : INEVSWRDVTSLALSLANMVLGFSIIFPSRRKRHYASNMLVSELLDMAVRAMTSHINICSRKLCABLNDFAMVSTTFGLABALLGVGDGLSG----TIAIHYSAAS
M_mulatta : INEVSWRDVTNALSLANMVLGFSIIFPSRRKRHYASNMLVSELLDMAVRAMTSHINICSRKLCABLNDFAMVSTTFGLABALLGVGDGLSG----TIAIHYSAAS
B_taurus : INBETWRDVNALSLANMVLGFSIIFCSCKKCCSCASNMVLVSELLDMAIESVTRHLNISCSTELNDFAMVSTTFGLABALLGVGDGPLNG----DIAMMYVLAVS
M_musculus : MTBESWRDIIINALSLANMVLGFSIIFCSFRRKSYCASNMLVSELLDIAIETVTRHLNIPHKLCLELNDFAMVSTTFGLABALLGVGDGPLNG----TIAIHYVLTS
T_nigrovir : LKBEARNAANALSVANMLMGASITLSSLNGRHHAACNIVLIGLLDLADCAVARRINASSALCAKLDDEADSTTFGLASLLIRTPTDFDN----TICICYVLSVF
T_rubripes : LKBEARNAANALSVANMLMGASITLSSLNGRHHAACNIVLIGLLDLADCAVARRINASSALCAKLDDEADSTTFGLASLLIRTPTDFDN----TICICYVLSVF
D_rerio : LKBEARNAANALSVANMLMGASITLSSLNGRHHAACNIVLIGLLDLADCAVARRINASSALCAKLDDEADSTTFGLASLLIRTPTDFDN----TICICYVLSVF
PSS_B_Subt : -----MNYIFCMTLIGFPCCLLAHSLLYHNHSAVLFTFTGMPLEFFDMAARKINAVSDNCRBLLSEADLVFVGVPASMLAYSVAIYTLFPIGICALTYSICGM

      220      *      240      *      260      *      280      *      300      *      320
P_troglody : FHLCFYSPC--VFSTYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
M_mulatta : FHLCFYSPC--VFSTYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
B_taurus : FHLCFYSPC--VFSTYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
M_musculus : FEMCFYSPC--ATSGYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
T_nigrovir : VLCHFPSSC--VFPMYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
T_rubripes : VLCHFPSSC--VFPMYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
D_rerio : VLCHFPSSC--VFPMYKGLPCPYASCILASTSLLTKGNRFLCCMASIMILEMMDQSYYPYKILESENKKKLVVIGGVIMLFFSPPLSLAPICMMDSTSYIFFPDAL
PSS_B_Subt : LRLSKNIEQSKLPTFTMBIFPAGMCLVILSTYTN--PTLLAIGTCCSYLMISKIKP-----HFKKHAANENLESGR--N

      *      340      *      360      *      380
P_troglody : WGKAACLSPOH-----
M_mulatta : WGKAACVSHSAEELSSSTSSCRPLSDKADNKTNNKKHLHSLPLAVDESLLALHNHR
B_taurus : WGKAACLRPFY-----
M_musculus : WGRGVRIKP-----
T_nigrovir : WSSKV-----
T_rubripes : -----
D_rerio : -----
PSS_B_Subt : -----

```

Figure 15

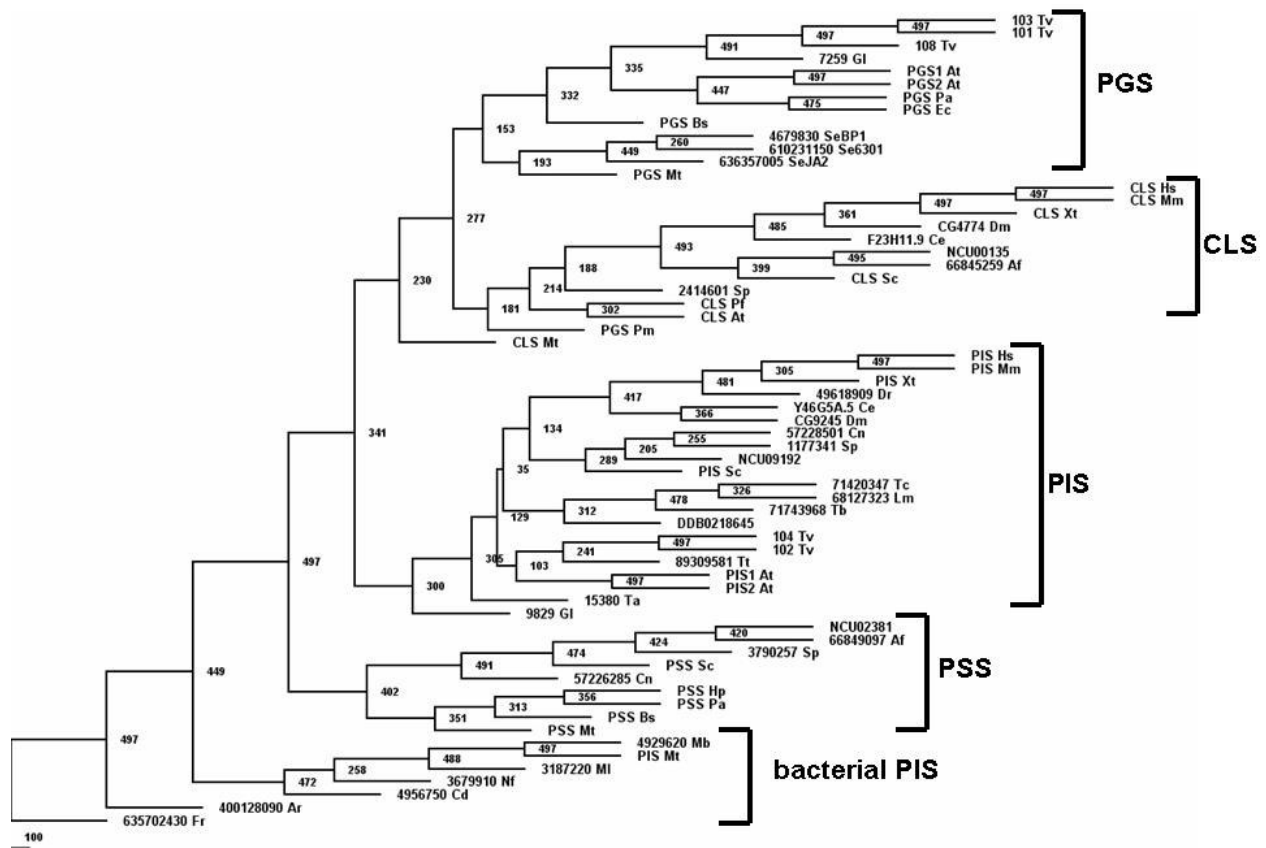


Figure 16

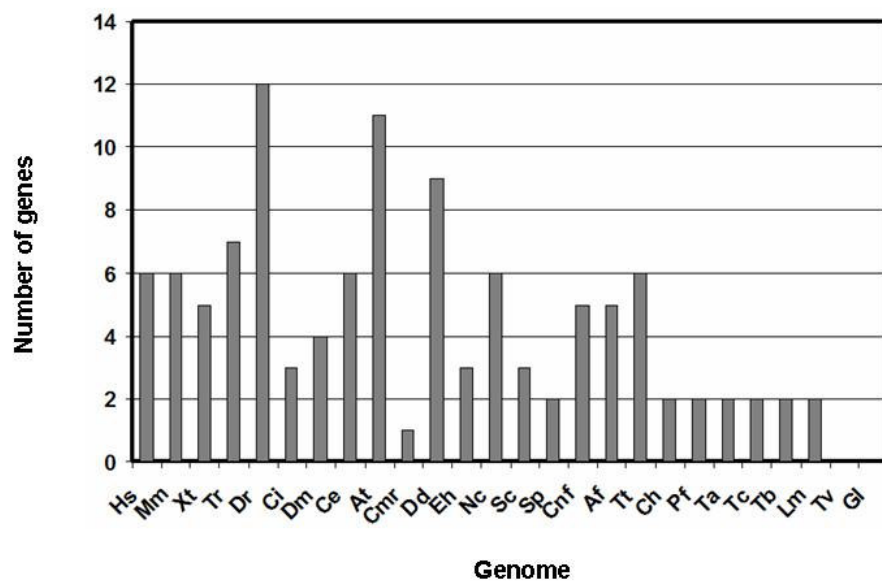


Figure 17

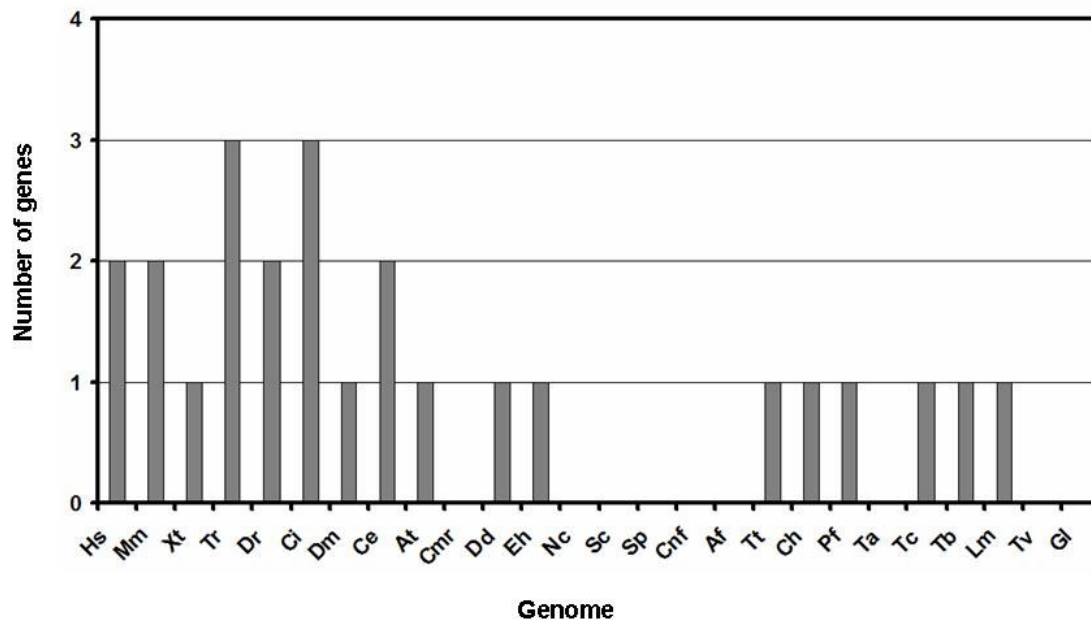


Figure 19

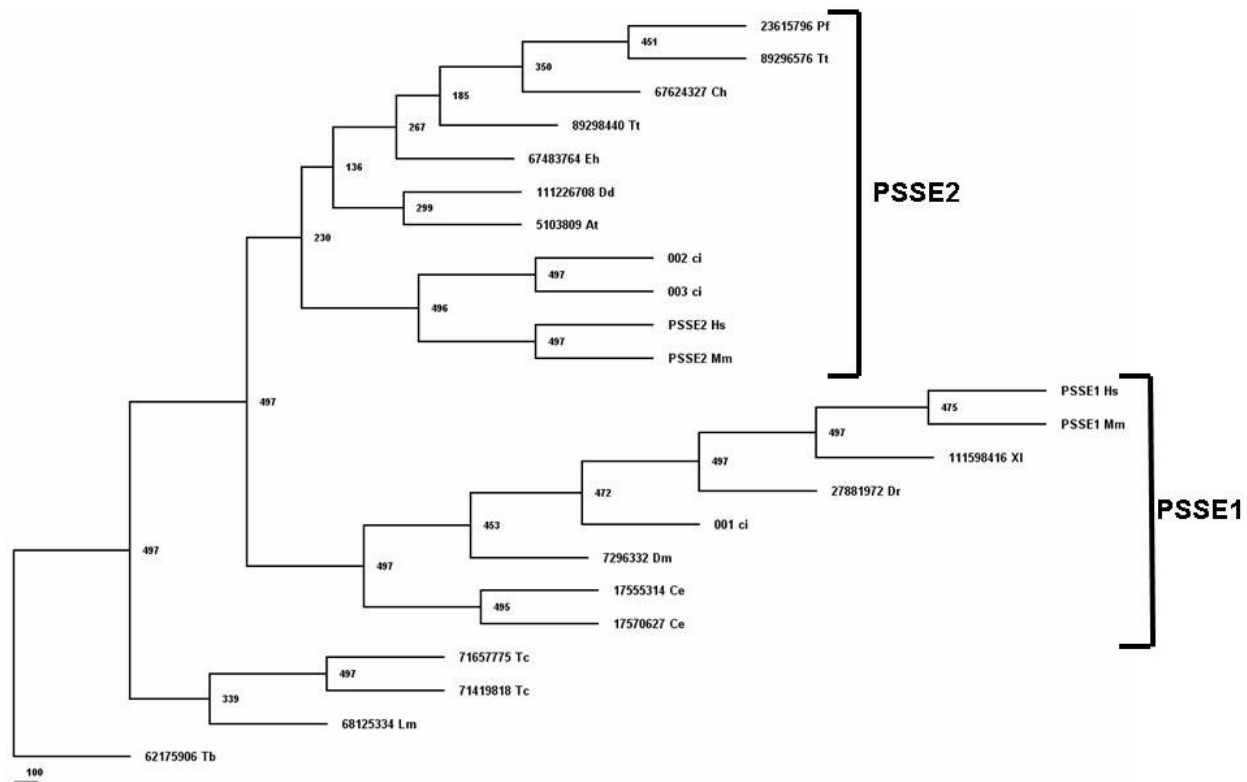


Figure 20

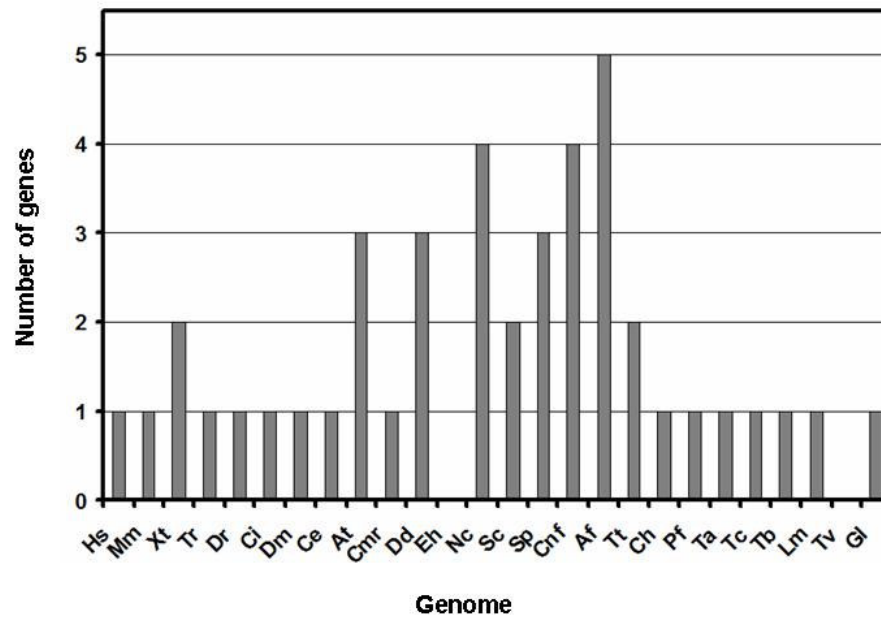


Figure 21

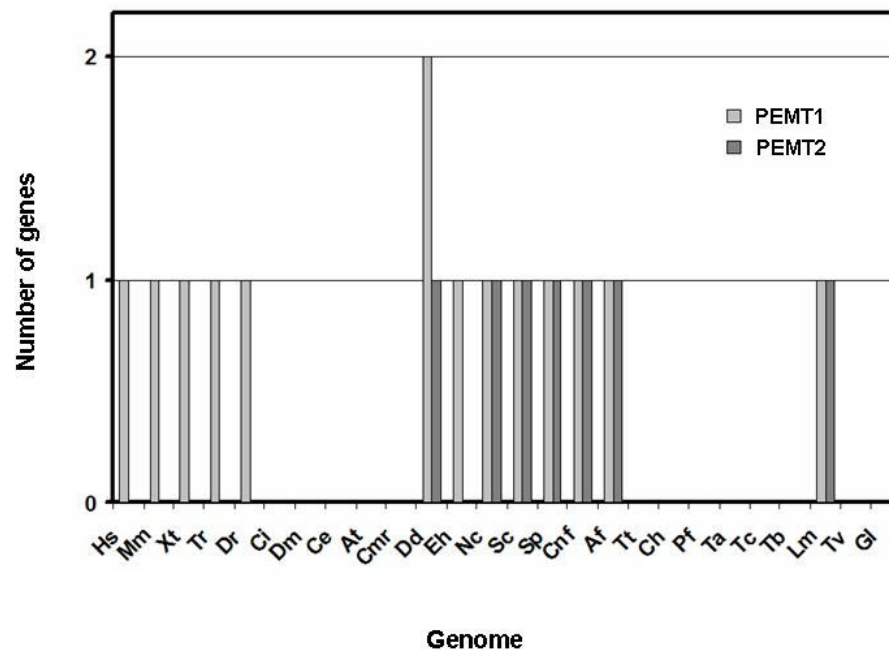


Figure 23