Multiple Horizontal Gene Transfer Events and Domain Fusions Have Created Novel Regulatory and Metabolic Networks in the Oomycete Genome

Paul Francis Morris1*, Laura Rose Schlosser1, Katherine Diane Onasch1, Tom Wittenschlaeger1, Ryan Austin2, Nicholas Provart2

1 Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio, United States of America, 2 Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada

Abstract

Complex enzymes with multiple catalytic activities are hypothesized to have evolved from more primitive precursors. Global analysis of the Phytophthora sojae genome using conservative criteria for evaluation of complex proteins identified 273 novel multifunctional proteins that were also conserved in P. ramorum. Each of these proteins contains combinations of protein motifs that are not present in bacterial, plant, animal, or fungal genomes. A subset of these proteins were also identified in the two diatom genomes, but the majority of these proteins have formed after the split between diatoms and oomycetes. Documentation of multiple cases of domain fusions that are common to both oomycetes and diatom genomes lends additional support for the hypothesis that oomycetes and diatoms are monophyletic. Bifunctional proteins that catalyze two steps in a metabolic pathway can be used to infer the interaction of orthologous proteins that exist as separate entities in other genomes. We postulated that the novel multifunctional proteins of oomycetes could function as potential Rosetta Stones to identify interacting proteins of conserved metabolic and regulatory networks in other eukaryotic genomes. However ortholog analysis of each domain within our set of 273 multifunctional proteins against 39 sequenced bacterial and eukaryotic genomes, identified only 18 candidate Rosetta Stone proteins. Thus the majority of multifunctional proteins are not Rosetta Stones, but they may nonetheless be useful in identifying novel metabolic and regulatory networks in oomycetes. Phylogenetic analysis of all the enzymes in three pathways with one or more novel multifunctional proteins was conducted to determine the probable origins of individual enzymes. These analyses revealed multiple examples of horizontal transfer from both bacterial genomes and the photosynthetic endosymbiont in the ancestral genome of Stramenopiles. The complexity of the phylogenetic origins of these metabolic pathways and the paucity of Rosetta Stones relative to the total number of multifunctional proteins suggests that the proteome of oomycetes has few features in common with other Kingdoms.

Citation: Morris PF, Schlosser LR, Onasch KD, Wittenschlaeger T, Austin R, et al. (2009) Multiple Horizontal Gene Transfer Events and Domain Fusions Have Created Novel Regulatory and Metabolic Networks in the Oomycete Genome. PLoS ONE 4(7): e6133. doi:10.1371/journal.pone.0006133

Editor: Berend Snel, Utrecht University, Netherlands

Received December 10, 2008; Accepted June 3, 2009; Published July 2, 2009

Copyright: © 2009 Morris et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a NSF grant (0731969) to PFM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pmorris@bgsu.edu

Introduction

Phytophthora species are destructive plant pathogens of a wide range of plant species. P. sojae is a host specific pathogen of soybeans, and protection from crop losses is mediated only by an extensive breeding program to introgress new resistance genes into soybean cultivars from wild germplasm [1]. P. infestans is the most serious pathogen of potatoes on a worldwide basis and control of this windborne pathogen is achieved only by a significant expenditure of pesticides [2]. While one stage of their life history includes a hyphal growth form, they are classified in the Kingdom Stramenopiles, which also include photosynthetic organisms such as brown algae and diatoms. The Stramenopiles are in turn, members of a larger group the chromalveolates, which include alveolate pathogen such as Plasmodium falciparum and Toxoplasma gondii and free-living ciliates such as Tetrahymena thermophila and Paramecium tetraurelia. Members of this superclade are thought to have been derived from a eukaryotic protist that acquired a red algal species by endosymbiosis [3]. One of the hallmarks of the genome sequence analysis of P. sojae and P. ramorum was the finding that at least 855 of the oomycete genes had a probable cyanobacterial or red algal origin [4]. Phylogenetic analyses of several chromalveolate genomes have now shown that endosymbiotic transfer of genes to the host nucleus has had a significant impact on the evolution of these genomes [4,5,6].

The completion of sequences from several protist genome projects has vastly changed our understanding of this diverse group of organisms. Domain recombination events, the lateral transfer of domains from different kingdoms, and lineage-specific expansion of particular protein families have all been identified as distinctive features of several protozoan lineages [7,8,9,10,11]. Independent phylogenetic strategies have also identified several instances of ancient horizontal transfer from bacterial genomes to stramenopiles [5,12,13].

Composite proteins are a common feature in eukaryotic genomes with the number of composite proteins increasing in proportion to genome size [14,15]. In the human and mouse genomes 29% of proteins have such a structure [15]. Our
examination of the _P. sojae_ genome identified several multifunctional genes that appeared to have arisen from novel gene fusion events. Aside from their novelty, proteins with BLAST hits to two or more proteins in other organisms have been posited as Rosetta Stones since the fusion of two or more catalytic domains as part of a metabolic or regulatory pathway has been used to infer the association of orthologous proteins containing single domains that exist as separate entities in other genomes [16,17,18]. Gene fusions of enzymes in already characterized metabolic pathways such as the pentafunctional ARO1 enzyme in the aromatic amino acid biosynthetic pathway of fungi are positive examples of these kinds of associations. Gene fusion also enforces the co-regulation of two domains and co-regulated genes in multienzyme complexes. However certain domains of eukaryotic proteins are said to be “promiscuous” because they appear in combinations with other domains in proteins associated with signal transduction pathways [19]. Nonetheless, there is broad conservation of domain combinations in multifunctional (MF) proteins from sequenced genomes [20]. We hypothesized that some of the multifunctional proteins in the _P. sojae_ genome would include proteins involved in metabolic or regulatory pathways that are conserved across Kingdoms. Analysis of these proteins might also reveal examples of interkingdom gene fusions [21] where one of the domains was acquired from either the ancestral photosynthetic endosymbiont or of interkingdom gene fusions [21] where one of the domains was acquired from either the ancestral photosynthetic endosymbiont or a bacterial genome. Here, we have used reciprocal shortest distance (RSD) algorithm [22] which is a conservative approach to identifying potential orthologs in 39 eukaryotic and bacterial genomes.

**Materials and Methods**

**Sequence retrieval and analysis**

Predicted proteins sequences of _P. sojae_ were retrieved from DOE-JGI website, subjected to BLAST analysis against the nonredundant protein database, and InterProScan [23]. The output of these analyses was loaded into a MySQL database. To identify novel MF proteins, the database was first queried for protein models with multiple predicted protein domains where the length of the best BLAST hit was more than 100 amino acids smaller than the _P. sojae_ model. A second independent query of protein models with two or more Pfam hits was also used to identify additional examples of novel MF proteins that were not detected in the first strategy. A total of 2236 models were subjected to manual curation using the DOE-JGI browser and other web-based tools. Proteins with any of the following characteristics were excluded from further analysis: 1. Proteins containing domains with homology to transposable elements. 2. Multi-exonic protein models without EST support unless they also met one of the following criteria: A: a single domain that spanned one or more exons. B: proteins were members of a multi-genie family in _P. sojae_ in which the majority of family members contained no introns.

Draft sequences were visually inspected to remove introns not needed to maintain an open reading frame, and to identify a stop site for proteins judged to be a fusion of two adjacent proteins. Each of the domains that were identified by BLASTP as showing homology to an individual protein in other species were split into separate models for RSD analysis. The majority of these novel MF proteins had BLAST hits to only two different types of proteins in other species. A subset of the MF proteins were found to have significant BLASTP hits to only one of the domains and were excluded from RSD analysis to identify Rosetta Stone candidates.

Protein sequences for _E. huxleyi_ V1, _M. brevicollis_ V1.0, _N. hematococca_ V2, _P. ramorum_ V1, _P. sojae_ V1, _P. patens ssp patens_ V1, _P. tricornutum_ V2, _P. stipites_ V2, _T. pseudonana_ V3 were downloaded from the DOE Joint Genome Initiative website at http://www.jgi.doe.gov/ (2/23/08). Protein sequences for _Aedes aegypti, Anopheles gambiae, Ashbya gossypii, Aspergillus fumigatus, Camarohabitus braggiae, C. elegans, Candida glabrata, Danio rerio, Deburryanum hansenii, Drosophila melanogaster, D. pseudoobscura, Escherichia coli, Kluyveromyces lactis, Leishmania major, Magnaporthe grisea, Mus musculus, Neurospora crassa, Phascolarctia nodorum, Plasmadium yoelli, Rattus norvegicus, Saccharomyces cerevisiae, Streptomyces coelicolor, Schizosaccharomyces pombe, Tetradon nigroviridis, and Xenina lipolecta were downloaded from the Expasy website [24] at ftp://ftp.expasy.org/ (9/27/08). Protein sequences for _Globobacter violaceus, Nostoc commune_, and _Thermosynthesococcus elongates_ were downloaded from Cyanobase at http://bacteria. kazusa.or.jp/cyanobase/ (9/28/08). Protein sequences for _Arabidopsis thaliana_ (TAIR8) were downloaded from http://www.arabidopsis.org/_. Proteins sequences from the _Vitis vinifera_ genome were downloaded from NCBI http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy (2/23/08). The _Cyanidioschyzon merolae_ proteome was downloaded from the _C. merolae_ webpage at http://merolae.bio.i.s.tokyo.ac.jp/_. The _Oryza sativa_ proteome (Build 3) was downloaded from the rice annotation project database website at http://rapdb.dna.afic.go.jp/_. Sequences for _Plasmodium falciparum_ were obtained from PlasmoDB http://www.plasmodb.org/ plasmo/ (9/26/08).

The RSD program was downloaded and installed on a Mac desktop computer following instructions outlined in the RSD tutorial. For this analysis, the curated split models of _P. sojae_ proteins were added in replacement of the identified multifunctional proteins. Predicted orthologs that were generated by RSD analysis for each split model were loaded into a Filemaker database. Potential Rosetta Stone candidates were identified by querying the database for genomes with orthologs to two or more of the split models associated with a multifunctional protein.

**Analysis of orthologs in _P. ramorum, Phaeodactorum tricornutum and Thalassiosira pseudonana_ genomes**

To determine whether domain fusion events resulting in the formation of novel MF proteins occurred prior to the split of oomycetes and diatoms, both split and complete models of _P. sojae_ MF proteins were used in separate ortholog searches of the _P. ramorum_ and diatom genomes. In cases where RSD analysis identified hits to only one of the domains, the _P. sojae_ MF proteins were used in a TBLASTN analysis to determine if synteny of domains was retained in the assembled diatom genomes. The diatom gene model which came closest to describing the scaffold region showing homology to the _P. sojae_ gene was listed as the probable ortholog.

**Analysis of sequences from Metabolic Pathways**

The MetaCyc tool [25] (http://metacyc.org/) was used to identify _P. sojae_ gene models from several metabolic pathways with multiple examples of novel MF proteins. Homologous sequences to _P. sojae_ genes in other kingdoms were identified by BLASTP analysis of the SwissProt database and aligned using ClustalW [26]. Maximum likelihood analysis was done using PHYML v2.4.4 [27]. ML analysis was done with bootstrapping (100 replicates) using the WAG evolutionary model, assuming four substitution rate categories, and using the program defaults for estimating the gamma distribution parameter and optimization of trees. Phylogenetic trees were drawn as radial phylograms using Dendroscope [28].

**Results**

The existence of novel MF proteins in the _P. sojae_ genome was first observed in the course of examining protein models on the
JGI Genome Browser. Several instances were noted of protein models where multiple motifs were contained within a single exon, or closely spaced exons with EST support. In each case, BLAST analysis of N-terminal and C-terminal regions gave hits to dissimilar kinds of proteins. Notable examples of such proteins included nine phosphatidylinositol-4-phosphate 5 kinases with TM domains characteristic of G-coupled receptors. A second type of fused protein involved a potassium channel that in other eukaryotes functions in vivo as a tetramer [29]. In the oomycetes, a single exon contains the four motifs characteristic of each monomer.

A schematic outline of the strategy used to identify novel MF proteins in *P. sojae* is shown in Fig. 1. Manual annotation of 1570 *P. sojae* models that were at least 100 amino acids longer than the best BLASTP hit against the nonredundant database identified 191 novel multifunctional proteins. An additional 29 models were identified by manual annotation of 615 models with multiple PFAM hits. Finally, 51 MF proteins were identified by serendipitous discovery or annotation of proteins that were members of a family of novel multifunctional proteins. The 273 proteins in this data set (Table S1) were assembled to test the hypothesis that the orthologs to the domains in these proteins functioned together in metabolic or regulatory networks [17,18,16]. An additional 130 predicted genes were excluded from this set because they may simply be two adjacent proteins that were fused into a single protein by the gene prediction software, and there was no EST support for those models. The total number of novel multifunctional proteins is expected to increase as refinements to gene prediction software and additional EST data identifies both multi-exonic proteins and adjacent genes on the genome as single gene models. Novel MF proteins are conserved in the *P. ramorum* genome. Using the curated set of 273 *P. sojae* proteins, RSD analysis identified 210 orthologous proteins in *P. ramorum*, and BLAST analysis and visual inspection of the *P. ramorum* genome to check for synteny using the DOE-JGI browser confirmed the presence of *P. ramorum* orthologs for all 273 proteins (Table S1).

Based on visual inspection of domain descriptions, 47 of the 273 gene models are derived from gene fusions of homologous domains. The largest family of such proteins consists of nine tetrameric calcium-activated potassium channels. MF proteins were catalogued as metabolic if one or more of the domains were associated with metabolism. All remaining MF proteins had domains associated with ion transport, signaling, or protein-protein interactions, and were catalogued as involved in cellular regulation. Sixty three MF proteins have motifs suggesting that they function in metabolic pathways. Thus the majority of novel MF proteins have functional domains that are typically associated with proteins in signaling pathways.

**RSD analysis**

RSD analysis has identified 18 potential Rosetta Stones with interactors in 39 different species representing organisms from bacteria, plant, fungal, and animal kingdoms (Table S2). The complete list of orthologs to each of the MF proteins is given in Table S3. Five Rosetta Stones identified predicted interacting proteins in the human genome, but none of these are presently supported by experimental data. [http://hwiki.fzk.de/wiki/index.php/Main_Page]. Five Rosetta Stones are MF proteins that catalyze consecutive steps in well-characterized metabolic pathways. These include Ps112102, which has protein domains defining catalytic activities for the first three steps in sulfate assimilation (orthologs in 10 genomes) and Ps131776 which has domains predicted glycerol acyltransferase and glycerol acyl reductase activities (orthologs in11 genomes). The human orthologs of Ps131776 are localized in the peroxisome and the

![Figure 1. Search strategy to identify novel multifunctional proteins in the Phytophthora sojae genome.](http://www.plosone.org/doi/10.1371/journal.pone.0006133.g001)
ortholog to the N terminal domain is required for the conversion of fatty acids to fatty alcohols [30]. Plants contain homologs to only the first domain of this MF protein and orthologs to it were identified in A. thaliana, O. sativa and P. patens. Expression of this gene in A. thaliana is critical for the development of a normal pollen cell wall [31]. While the majority of novel MF proteins in P. sojae contain domains characteristic of proteins associated with scaffold or regulatory pathways, RSD analysis identified only nine candidate Rosetta Stones with domains associated with regulatory pathways. The majority of these proteins identified orthologs in only one or two species.

Since our set of novel MF proteins included 63 examples with domains associated with metabolism, we next sought to determine if this level of gene fusion in metabolic pathways was typical for other eukaryotic genomes. A recent survey of plant genomes identified only 22 bifunctional proteins [32] in metabolic pathways that are also present in oomycetes, and there are P. sojae models with overlapping metabolic activities for 19 of these plant enzymes. P. sojae also contains homologs for 20 of the 29 human fusion proteins of prokaryotic origins [33]. Table 1 lists 39 bifunctional enzymes in P. sojae that have homologs in plant, fungal, or human genomes. The presence of these bifunctional enzymes in multiple kingdoms may be indicative of either ancient gene fusions in a common eukaryotic ancestor or independent fusion events. However Ps135354 is a clear example of an independent fusion event. The fusion of orotate phosphoribosyltransferase (OPRT) and orotidine 5’ monophosphate decarboxylase (OMPDC), the last two steps in the pyrimidine biosynthetic pathway has the domain order OPRT-OMPDC in plants and metazoa, and OMPDC-OPRT in oomycetes, diatoms, kinetoplastids and some cyanobacteria [34]. Phylogenetic analysis indicated these novel gene fusions to be independent evolutionary events [34], and our analysis of the two Phytophthora sequences along with those of the diatom genomes supported those conclusions (not shown).

**Formation of some novel MF proteins preceded the split between diatoms and oomycetes**

To determine if the formation of MF proteins was initiated prior to the split between diatoms and oomycetes, we analyzed the diatom genomes for orthologous sequences to the P. sojae MF proteins. Table 2 shows 21 examples where synteny of domain order has been retained in the diatom and oomycete genomes. Since these proteins represent a small fraction of the 273 novel MF proteins in this data set, the majority of gene fusion events that resulted in the formation of these proteins in the oomycete genomes occurred after the split between diatoms and oomycetes.

**Lysine biosynthesis**

The MetaCyc pathway database presently lists six biochemical strategies for the synthesis of lysine, a clear indication that multiple, distinct biosynthetic strategies are possible [25]. In oomycetes, synthesis of lysine is carried out by five enzymes that catalyze six enzymatic reactions (Fig. 2). Mechanistically, the biosynthetic strategy of oomycetes follows that of plants, although, phylogenetic analysis shows that the oomycete sequences cluster with sequences from bacterial or archaean domains of life. Analysis of the two diatom genomes indicates that they also follow the same enzymatic strategy as oomycetes, although only some of diatom genes share the same phylogenetic origins with oomycetes.

Phylogenetic analysis of aspartate kinase, the first enzyme in this pathway shows that while the diatom sequences group with those from plant genomes, oomycete sequences are more closely related to bacterial bifunctional enzymes with diaminopimelate decarboxylase and aspartate kinase activity (Fig. S1). Phylogenetic analysis of aspartate semialdehyde dehydrogenase, the second step in this pathway show that both diatoms and oomycete genes are grouped together in the same clade with strong bootstrap support, and are distinct from a clade of cyanobacterial and plant sequences. However the Stramenopile sequences also cluster with strong bootstrap support to sequences from three archaea genomes (Fig. S2). Perhaps the simplest explanation for the grouping of archaea genes with stramenopiles would be a horizontal transfer to archaea from the ancestral bacterial genome that also transferred this gene to Stramenopiles. Phylogenetic analysis of dihydrodipicolinate synthase, the third enzyme in the oomycete lysine biosynthetic pathway, indicates that while the oomycete sequences are clustered away from other eukaryotic genomes, they do not form a strongly supported cluster with bacterial sequences (Fig. S3). The next two steps in the pathway are catalyzed by a unique fused protein with dihydrodipicolinate reduce and diaminopimelate aminotransferase activities. These domains have close homologs in only bacterial genomes (Fig. S4–Fig. S5).

In bacteria, the conversion of L,L diaminopimelate to mesodiaminopimelate by epimerase is a necessary step because mesodiaminopimelate is an essential component of the bacterial peptidoglycan cell wall. Plants and diatoms also have an epimerase and, while they do not make peptidoglycans, nine homologous genes to bacterial homologs of the peptidoglycan pathway play an important role in chloroplast cell division in mosses and potentially other plants [35]. We searched P. ramorum, P. sojae and P. infestans genomes using plant and bacterial epimerases in a TBlastN search and found no evidence for epimerase activity. Thus we postulate that oomycetes convert L-diaminopimelate directly to lysine without an intermediate step. The oomycete genomes each contain three candidate genes with PFAM domains for ornithine, diaminopimelate, or arginine (ORN/DAP/ARG) decarboxylase activity, suggesting that one or more of these proteins function as diaminopimelate decarboxylase, the last step in the pathway. Two of these proteins in each of the oomycete genomes cluster with sequences from bacterial genomes (Fig. 3). The third pair of genes (Ps108220 and Pr71432) is in a separate clade of diverse bacterial sequences, but the oomycete sequences form a cluster with sequences from Archaea genomes (Fig. 3). The clustering of archaical sequences in this clade of bacterial and oomycete genes is suggestive of a horizontal transfer to archaea from the ancestral bacterial genome that also transferred this domain to oomycetes. Notably, this clade also includes the bacterial bifunctional proteins (aspartate kinase-diaminopimelate decarboxylase) from X. axyridis, L. pneumophila and S. ruher. The N terminal domain of these bacterial genes also clustered with oomycete genes with predicted aspartate kinase activity (Fig. S1). The N-terminal domain of the P. sojae gene Ps128004 and its ortholog in P. ramorum Pr72228 is a conserved domain of unknown function that clusters with sequences from bacterial genomes (Fig. 3).

**Serine biosynthesis**

One of the potential serine biosynthetic strategies of oomycetes involves the reduction of phosphoglycerate to make 3-phosphoglyceropyruvate, a phosphoserine aminotransferase to make phosphoserine, and a phosphoserine phosphatase to produce serine. In oomycetes, the first two steps in this pathway are carried out by a bifunctional enzyme (Ps142688 and Pr80380). Phylogenetic analysis of the N terminal domain shows that the oomycete domains cluster with genes from metazoans (Fig. S7). Phylogenetic analysis of both the C-terminal domain of the bifunctional enzyme, and the oomycete phosphoserine phosphatase gene show
strong support for a bacterial origin of these genes (Figs. S8–S9). An alternate biosynthetic pathway leading to serine could involve the amination of glyoxylate, but in this case, too, there is one aminotransferase enzyme in each genome (Ps109249 and Pr77437) and phylogenetic analysis suggests that these proteins were acquired by an ancient horizontal transfer event from a bacterial genome (Fig. S10).

Sulfate assimilation

Phylogenetic analysis of the sulfate assimilation pathway in eukaryotes has indicated a very complex pattern of inheritance across the Kingdoms with multiple examples of gene fusions and horizontal transfer events (Patron et al. 2008). In fungi and plants, the reduction of sulfate and the subsequent transfer of the sulfhydryl group to O-acetyl-serine resulting in the formation of

Table 1. Multifunctional metabolic enzymes in oomycete and other eukaryotic genomes.

| Gene_ID   | Metabolic Process              | Catalytic activities                                      | A. thaliana | H. sapiens | S. cerevisiae |
|-----------|--------------------------------|-----------------------------------------------------------|-------------|------------|--------------|
| Ps108209  | amino acid catabolism          | Methyl malonyl Coenzyme A mutase                          | none        | P22033     | none         |
| Ps108248  | aspartate metabolism           | Aspartate kinase-homoserine dehydrogenase                 | Q9SA18      | none       | none         |
| Ps108383  | proline synthesis              | Delta 1-pyrroline-5-carboxylate synthetase (P5CS)         | P54887      | none       | P54885       |
| Ps108492  | purine synthesis               | AIR synthase related protein                              | A9Y5J1      | A8KB8N7    | Q45SU25      |
| Ps108690  | glutathione                    | 1-cys-peroxiredoxin                                       | O04005      | P30041     | P34227       |
| Ps108696  | tryptophan                     | Anthranilate synthase component II                        | Q9ULB2      | None       | P00937       |
| Ps108681  | fucose                         | GDP-L-fucose synthase                                     | O49213      | Q13630     | none         |
| Ps108780  | glycolysis                     | Pyruvate carboxylase                                      | none        | P11498     | P32327       |
| Ps108801  | lysine catabolism              | Lysine-ketoglutarate reductase-saccharopine dehydrogenase | Q9SMZ4      | Q9UDR5     | P38999       |
| Ps108913  | Lipid biosynthesis             | ATP citrate lyase                                         | none        | P35936     | none         |
| Ps108918  | Acetyl CoA carboxylase         | none                                                      | Q38970      | Q13085     | P32874       |
| Ps108989  | mRNA synthesis                 | RNA polymerase II                                         | P18616      | P24928     | P04050       |
| Ps109060  | histidine synthesis            | Glutamine amidotransferase glutamine cyclase              | Q5SZ30      | none       | P33734       |
| Ps109110  | riboflavin metabolism          | GTP cyclohydrolase II                                    | Q9SJY9      | none       | Q98258       |
| Ps109141  | folic acid                     | 7,8-Dihydro-6-hydroxymethylperin-pyrrophosokinase,        | Q9SZV3      | none       | P53848       |
| Ps109146  | Mo cofactor biosynthesis       | Molybdenum cofactor biosynthesis protein 1                | Q39055      | Q9N2BB-5  | none         |
| Ps109150  | ketone metabolism              | Succinyl-CoA-3-ketoacid-coenzyme A transferase            | none        | P55809     | none         |
| Ps109206  | urea cycle                     | Carbamoyl-phosphate synthetase III;                       | Q9SZV3      | none       | Q2VOX0       |
| Ps109271  | mRNA synthesis                 | RNA polymerase                                            | P38420      | B40H3      | P08518       |
| Ps109291  | fatty acid degradation         | Enoyl-CoA hydratase; Long chain 3-hydroxyacyl-CoA dehydrogenase | Q9ZP16   | P40939     | P28817       |
| Ps109321  | dihydrofololate                | Dihydrofolate reductase-thymidylate synthase              | Q55762      | P04818     | None         |
| Ps109425  | urea cycle                     | Carbamoyl phosphate synthetase                            | Q42601      | P31327     | P07259       |
| Ps109486  | purine biosynthesis            | Phosphoribosylaminomimidazole carboxamidine formyltransferase activity | Q8RW75 | P31939     | P54113       |
| Ps109918  | fatty acid metabolism          | Acetyl-coenzyme A carboxylase                             | Q84MM8      | none       | P36022       |
| Ps123952  | PABA                           | P-aminobenzoate synthase-glutamine-amidotransferase       | Q9ZV26      | P49915     | P49915       |
| Ps128107  | Mo cofactor biosynthesis       | Molybdopterin biosynthesis                                | Q39054      | Q9NOX3     | none         |
| Ps129332  | adenine biosynthesis           | ADE1                                                      | P52420      | Q59HH3     | P07244       |
| Ps131873  | glucose metabolism             | Phosphofructo-2 kinase, fructose 2,6 bispophatase         | Q9MB58      | Q16875     | P32604       |
| Ps133964  | multi-functional TRNA          | Glutamyl prolyl TRNA synthetase                           | O82462      | Q8NAJ6     | P46555       |
| Ps134345  | biotin synthesis               | Bifunctional diaminopelargonate synthase-dethiobiotin synthetase | B0F481 | none       | none         |
| Ps134889  | sucrose metabolism             | Phosphoribosylfructo-2 kinase, fructose 2,6 bispophatase   | Q9MB58      | B286L2     | P40433       |
| Ps135279  | histidine biosynthesis         | Phosphoribosyl-ATP pyrophosphorylhydrodase–Phosphoribosyl-AMP cyclodrolase | novel     | novel      | novel         |
| Ps135354  | pyrimidine metabolism          | Orotidine 5’-phosphate decarboxylase-Orotate              | none        | none       | none         |
| Ps136410  | Aspartate biosynthesis         | Aspartate kinase-homoserine dehydrogenase                 | Q9SA18      | None       | P10869       |
| Ps137041  | RNA polymerase                 | DNA-directed RNA polymerase III                           | Q9LVH0      | Q14802     | P040501      |
| Ps137320  | Topoisomerase/gyrase            | Topoisomerase/gyrase                                      | P30182      | P11388     | Q8TG58       |
| Ps141174  | glucose metabolism             | Phosphoribofructo-2 kinase, fructose 2,6 bispophatase     | Q9MB58      | Q16877     | P32604       |
| Ps143339  | amino acid biosynthesis        | Pentafunctional aromatic amino acid synthesis             | None        | None       | P08566       |
| Ps155429  | Aspartate biosynthesis         | Aspartate kinase-homoserine dehydrogenase                 | Q9SA18      | None       | P10869       |

doi:10.1371/journal.pone.0006133.t001
Cysteine is mediated by six enzymes. In animals, the first two enzymes in this pathway, ATP surfurylase and adenylsulfate kinase, form a single bifunctional enzyme. These two functional domains are present in the oomycete models but the oomycete proteins also contain an additional pyrophosphatase domain to break down the pyrophosphate that is produced as a byproduct of.

Table 2. Novel multifunctional proteins present in the P. sojae, P. tricornutum and T. pseudonana genomes.

| Genomes | Enzymes | Domains |
|---------|---------|---------|
| Ps108305 | Ps109221 | ATP surfurylase and adenylsulfate kinase |
| Ps110986 | Ps112102 | Pyrophosphatase domain |
| Ps116824 | Ps127510 | N terminal protein is upstream 5801 |
| Ps128154 | Ps129073 | Beta-ketoacyl synthase and phosphohydrolase |

Table 3. Phylogenetic associations of enzyme domains in serine biosynthesis and sulfate assimilation pathways of P. sojae and P. ramorum.

| Process | Gene Function | GENE_ID | Clade Association |
|---------|--------------|---------|------------------|
| Serine Biosynthesis | phosphoglycerate dehydrogenase | Ps142688a | metazoan |
| | phosphoserine aminotransferase | Ps142688b | bacterial |
| | phosphoserine phosphatase | Ps132144 | bacterial |
| | serine-pyruvate aminotransferase | Ps109249 | bacterial |
| Sulfate Assimilation/Metabolism | Adenyl sulfate kinase | Ps112102a | indeterminant |
| | ATP sulfurylase | Ps112102b | indeterminant |
| | pyrophosphatase | Ps112102c | indeterminant |
| | phosphoadenosine phosphosulfate reductase | Ps156997a | plant/cyano |
| | glutaredoxin | Ps156997b | bacterial |
| | sulfite reductase ECM17 | Ps139493 | fungal |
| | sulfite reductase MET10 | Ps139488 | fungal |
| | cysteine synthase | Ps109172, Ps109175 | fungal |
| | cystathionine gamma lyase | Ps109222 | metazoan |
| | cystathionine beta lyase | Ps120833 | indeterminant |

*Letter following Gene ID indicates a multifunctional protein where each of the domains was analyzed separately. Clade association was assigned when bootstrap support was >70% of 100 replicates using PHYML parameters as described in methods.

doi:10.1371/journal.pone.0006133.t002

doi:10.1371/journal.pone.0006133.t003
ATP surfurylase. Phylogenetic analysis suggests that the fusion of the three domains is an ancient event since it is also present in the two diatom genomes as well as the coccolithophore *Emiliania huxleyi*, and separate analysis of each of the domains show that they cluster in the same clades (Figs. S11, S12, S13).

The second oomycete gene in this pathway (Ps156997) clusters in a clade with cyanobacterial sequences, while higher plant sequences cluster in a separate clade and the diatom gene form a clade with moss and Selaginella genes (Fig. S14). The oomycete genes also contain an additional glutaredoxin motif that phylogenetic analysis indicates is of probable bacterial origin (Fig. S15). The *P. sojae* models Ps139493 and Ps139488 are homologous to the *S. cerevisiae* enzymes ECM17 and MET10 that carry out the reduction of sulfite to hydrogen sulfide (Fig. S16–17). *P. sojae* contains two cysteine synthases (Ps109172 and Ps109755) that are phylogenetically related to the plant cytosolic enzymes (Fig. S16–17). In oomycetes, the conversion of cystathionine to cysteine by cystathionine gamma lyase is done by an enzyme that clusters with proteins from metazoa (Fig. S19). Finally the conversion of cystathionine to homocysteine by cystathionine B lyase is achieved by an enzyme that clustered with both plant and fungal proteins (Fig. S20).

In summary the contribution of horizontal transfer from both the algal endosymbiont and bacterial genomes have enabled oomycete and diatoms to develop unique metabolic networks.

**Discussion**

The large number of both regulatory and metabolic proteins in the oomycete genomes that are derived from novel gene fusion events emphasizes the distinctive nature of these genomes. An important feature of the oomycete genomes that aided in the identification of verifiable novel MF proteins is that the majority of protein models are defined by a single open reading frame. We noted 100% conservation of the 273 MF proteins between *P. sojae* and *P. ramorum*. Extending the analysis of novel MF proteins to other oomycete genomes will enable us to address several questions. What is the level of conservation of these proteins? Does domain fusion and rearrangement continue to be an important factor in contributing to species diversity? What is the role of intron gain or loss in the evolution of these proteins? A subset of the MF proteins described in this study are also present in the two diatom genomes, a likely indication that some of these fusion events preceded the split between diatoms and oomycetes (Table 2). Gene islands are also a feature of the *P. tricornutum and T. pseudonana* genomes, and both species have small introns. The diatom genomes have proven to be surprisingly diverse but the contribution of novel gene fusions to that diversity has not been specifically addressed [5]. Thus, in addition to the set of MF proteins that are conserved between the diatom and oomycete genomes, the diatom genome may also include several novel MF proteins.
proteins that are presently recognized as separate proteins, due to limitations in gene prediction software.

The majority of *P. sojae* MF proteins analyzed here contain domains such as calcium binding, protein binding, Zn finger or kinase networks, consistent with their role in regulatory networks. For most proteins, RSD analysis identified orthologs to only one of the domains in each MF protein (Table S2). Since the complete proteins have no direct orthologs in other eukaryotic genomes, each of the domains may interact with other proteins to form novel oomycete regulatory networks. Since these proteins are conserved in *P. ramorum* and presumably other oomycete genomes, they may also serve to identify some of the novel regulatory networks in this important group of plant pathogens.

A particularly striking feature of the oomycete genome is the unusual number of gene fusion events in primary metabolic pathways. The oomycetes have acquired the majority of bifunctional metabolic enzymes in metabolic pathways that they share with plant, fungal and animal kingdoms (Table 1), along with 63 additional models that were identified in this survey. What conditions in the evolutionary history of oomycetes drove the selection of so many gene fusion events in common metabolic pathways? Biochemical studies suggest that the enzymes involved in many metabolic processes form large macromolecular complexes [36,37] and fusion of catalytic domains represents a rare, more constrained version of such complexes. Gene fusions may further increase the efficiency of the biosynthetic process by enabling metabolic channeling between catalytic domains. However if the driving force for such events in eukaryotic genomes was simply a matter of metabolic efficiency, the existence of bifunctional metabolic proteins would be much more widespread. Selection pressure that favors fusion events of metabolic genes must balance metabolic throughput with the need to regulate these processes. While metabolic channeling can serve to improve the catalytic efficiency of the complex, transient protein-protein interactions can function to regulate the catalytic activity of other enzymes in the complex [37]. In this regard, biochemical analysis of a bifunctional enzyme in the lysine degradation pathway of *A. thaliana* serves to illustrate the importance of protein-protein binding in metabolic pathways. The bifunctional LKR/SDH locus also codes for monofunctional enzymes with lysine-glutarate reductase and saccharopine reductase activities, and the monofunctional enzymes have strikingly different catalytic activities than the fused model [38].

An alternative hypothesis is that the processes of gene duplication and sequence divergence in plant, fungal, and metazoan genomes did not produce the kinds of combinations needed to enable the domains to fuse and still retain optimal catalytic activity. Cross genomic comparisons of the enzymes in
the pyrimidine biosynthetic and sulfate assimilation pathways provided multiple examples of independent gene fusion events [34,39]. However, the absence of fusion events in most phylogenetic groups may hint that the domains of genes that catalyzed individual steps in these pathways were not sufficiently compatible to enable gene fusion to be selected for. Given this scenario, we would expect a similar high rate of domain fusion events in diatom genomes since they share a similar evolutionary history including the acquisition of bacterial genes [5]. The analysis of point mutations in the MF proteins across multiple oomycete genomes may prove to be a useful means of evaluating the relative importance of catalytic efficiency and regulatory control in selected metabolic pathways.

Lateral transfer appears to have played a major role in the evolution of proteins in Bacteria and Archaea [40,41,42]. Phylogenetic analyses now suggest that horizontal gene transfer has played a significant role in evolutionary biology of eukaryotes, and particularly protistan lineages [5,10,12,13,39,43]. That some of the MF proteins contained one or more domains with the lowest RSD scores coming from plant genomes was not surprising, since global analysis of the P. sojae and P. ramorum genomes identified 855 genes of probable origin from the photosynthetic endosymbiont to the ancestral genome of oomycetes. The presence of bacterial domains in the MF proteins was unexpected, since we did not systematically search for such proteins. Here we have described a significant contribution of bacterial domains, both as single entities, and as components of novel MF proteins in two metabolic pathways. Our observations extend those of other recent reports documenting the lateral transfer of bacterial genes into oomycete genomes [5,13]. The model shown in Fig. 4 is intended to illustrate that the processes of horizontal transfer of genes from the photosynthetic endosymbiont, along with the independent acquisition of genes from bacterial genomes continued after the split of these phylogenetic lineages and may help to account for the surprising level of diversity between diatoms and oomycetes. One measure of this diversity can be seen in the phylogenetic analysis of serine and lysine biosynthesis. Oomycetes have adopted two biosynthetic strategies for serine biosynthesis from bacteria. While both diatoms and oomycetes share a similar biosynthetic strategy for lysine, in several cases the diatoms’ genes were clustered in different clades (Fig. S1, S3 S5).

Lateral transfer of domains associated with signaling proteins from both bacterial and host genomes of various protist genomes has been cited as evidence of the diversification of regulatory networks in several protist genomes [11]. In P. sojae, the curated set of MF proteins included 171 models with two or more domains that are typically associated with signaling networks. These proteins represent nodal proteins in novel oomycete regulatory pathways. One of the predictions made to support the investment in whole genome sequencing project of oomycete pathogens was that this data would enable the identification of new targets of pesticides. An integrated approach incorporating expression analyses, proteomics, and comparative genomics is now needed to define the novel proteome of oomycetes, so that we can identify both the unique growth and pathogenic strategies of this group of pathogens.

Supporting Information

Table S1 Novel Multifunctional Proteins of P. sojae and P. ramorum. This file contains a list of the multifunctional proteins from P. sojae, and their ortholog from P. ramorum. The file also
contains a description of the domains of these proteins and the amino acid sequence of the protein domains that were subjected to ortholog analysis by reciprocal smallest distance splits. Numbers at nodes indicate bootstrap support values (100 replications). Note that the diatom genomes contain a single protein with domains which catalyze the first two steps in a predicted lysine biosynthetic pathway. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Table S2 Candidate Rosetta Stone proteins in the P. sojae genome as identified by RSD analysis. This table identifies all the multifunctional gene models in P. sojae that had orthologs to at least two of their domains in one or more species.

Table S3 List of orthologs to the multifunctional protein domains of P. sojae. This table lists all the orthologs from 39 species to the individual domains of the multifunctional proteins.

Figure S1 PHYML analysis of genes with homology to the aspartate kinase domain of Ps144305. The P. sojae enzyme is the first committed step in a predicted lysine biosynthetic pathway of oomycetes. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S2 PHYML analysis of genes with homology to the aspartate semi-aldehyde dehydrogenase domain of Ps108835. The P. sojae enzyme is the second step in a predicted lysine biosynthetic pathway of oomycetes. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S3 PHYML analysis of genes with homology to the dihydricolinate synthase domain of Ps108407. The P. sojae enzyme is the third step in a predicted lysine biosynthetic pathway of oomycetes. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S4 PHYML analysis of genes with homology to the dihydricolinate reductase domain of Ps133120. The P. sojae enzyme carries out step four and step five of a predicted lysine biosynthetic pathway of oomycetes. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S5 PHYML analysis of genes with homology to the dianiminopipenolate aminotransferase domain of Ps133120. The P. sojae enzyme carries out step four and step five of a predicted lysine biosynthetic pathway of oomycetes. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S6 PHYML analysis of genes with homology to the N-terminal domain of Ps129804. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S7 PHYML analysis of genes with homology to the phosphoglycerate dehydrogenase domain of Ps142688. The P. sojae enzyme contains domains which catalyze the first two steps in a serine biosynthetic pathway. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S8 PHYML analysis of genes with homology to the phosphoserine aminotransferase domain of Ps142688. The P. sojae enzyme carries out step four and step five of a predicted lysine biosynthetic pathway. Numbers at nodes indicate bootstrap support values (100 replications). Note that the diatom genomes contain agene model in the same clade as oomycete genomes. Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S9 PHYML analysis of genes with homology to the phosphoserine phosphatase domain of Ps132144 Ps134157, and Ps157034. The P. sojae enzymes catalyze the last step in a serine biosynthetic pathway. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S10 PHYML analysis of genes with homology to the serine-pyruvate aminotransferase domain of P. sojae. The P. sojae enzyme catalyzes the synthesis of serine from pyruvate, a biosynthetic strategy independent of the P. sojae enzymes analyzed in Figs. S7, S8, S9. Numbers at nodes indicate bootstrap support values (100 replications). Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S11 PHYML analysis of genes with homology to the adenylosulfate kinase domain of Ps112102. Numbers at nodes indicate bootstrap support values (100 replications). The oomycete, diatoms and E. huxleyi genomes contain a single protein with adenylosulfate kinase, ATP sulfurylase and pyrophosphatase domains. The domains of these proteins have been subjected to separate phylogenetic analyses to determine whether they share a common phylogenetic origin. Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S12 PHYML analysis of genes with homology to the adenylosulfate kinase domain of Ps112102. Numbers at nodes indicate bootstrap support values (100 replications). The oomycete, diatoms and E. huxleyi genomes contain a single protein with adenylosulfate kinase, ATP sulfurylase and pyrophosphatase domains. The domains of these proteins have been subjected to separate phylogenetic analyses to determine whether they share a common phylogenetic origin. Gene names are color coded to indicate the major phylogenetic groups: archaea, purple; bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.

Figure S13 PHYML analysis of genes with homology to the pyrophosphatase domain of Ps112102. Numbers at nodes
indicate bootstrap support values (100 replications). The oomy-
cete, diatoms and E. huxleyi genomes contain a single protein with
adenylsulfate kinase, ATP sulfurylase and pyrophosphatase
domains. The domains of these proteins have been subjected to
separate phylogenetic analyses to determine whether they share a
common phylogenetic origin. Gene names are color coded to
indicate the major phylogenetic groups: archaea, purple; bacteria,
blue; opistokonts, brown; plants, green; stramenopiles, red.
Found at: doi:10.1371/journal.pone.0006133.s016 (3.00 MB TIF)

Figure S14  PHYML analysis of genes with homology to the
phosphoadenosine phosphosulfate reductase domain of Ps156997.
The P sojae enzyme catalyzes step four in the sulfate assimilation
pathway. Numbers at nodes indicate bootstrap support values (100
replications). Gene names are color coded to indicate the major
phylogenetic groups: archaea, purple; bacteria, blue; opistokonts,
brown; plants, green; stramenopiles, red.
Found at: doi:10.1371/journal.pone.0006133.s017 (3.00 MB TIF)

Figure S15  PHYML analysis of genes with homology to a novel
glutaredoxin domain on Ps156997. Numbers at nodes indicate
bootstrap support values (100 replications). Gene names are color
coded to indicate the major phylogenetic groups: archaea, purple;
bacteria, blue; opistokonts, brown; plants, green; stramenopiles, red.
Found at: doi:10.1371/journal.pone.0006133.s018 (3.00 MB TIF)

Figure S16  PHYML analysis of genes with homology to
Ps139493. The P. sojae enzyme is predicted to form a complex
with Ps139488 to catalyze the reduction of sulfite. Numbers at
nodes indicate bootstrap support values (100 replications). Gene
names are color coded to indicate the major phylogenetic groups:
archaea, purple; bacteria, blue; opistokonts, brown; plants, green;
stramenopiles, red.
Found at: doi:10.1371/journal.pone.0006133.s019 (3.00 MB TIF)

Figure S17  PHYML analysis of genes with homology to Ps139488. The P. sojae enzyme is predicted to form a complex
with Ps139493 to catalyze the reduction of sulfite. Numbers at
nodes indicate bootstrap support values (100 replications). Gene
names are color coded to indicate the major phylogenetic groups:
archaea, purple; bacteria, blue; opistokonts, brown; plants, green;
stramenopiles, red.
Found at: doi:10.1371/journal.pone.0006133.s020 (3.00 MB TIF)

Acknowledgments

We thank Mia Hall and Alexandra Schmacker for their assistance in
annotation of genes and assembly of data in supplementary tables.

Author Contributions

Conceived and designed the experiments: PFM TW RA NJP. Performed
the experiments: PFM LRS KDO TW. Analyzed the data: PFM LRS
KDO. Contributed reagents/materials/analysis tools: TW RA NJP. Wrote
the experiments: PFM LRS KDO TW. Analyzed the data: PFM LRS
KDO. Contributed reagents/materials/analysis tools: TW RA NJP. Wrote
the paper: PFM NJP.

References

1. Dorrance AE, McClure SA, DeSilva A (2003) Pathogenic Diversity of
Phytophthora sojae in Ohio Soybean Fields. Plant Dis 87: 139–146.
2. Juddson H (2007) Genomics of the plant pathogenic oomycete
Phytophthora: insights into biology and evolution. Adv Genet 57: 57–143.
3. Simpson AGB, Roger AJ (2004) The real kingdoms of eukaryotes. Curr Biol 14.
4. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, et al. (2006)
Phytophthora Genome Sequences Uncover Evolutionary Origins and Mecha-
nisms of Pathogenesis. Science 313: 1261–1266.
5. Bowler C, Allen AE, Badger JH, Grünwood J, Jabbari K, et al. (2008) The
Phaeodactylum genome reveals the evolutionary history of diatom genomes.
Nature 456: 239–244.
6. Reyes-Prieto A, Montaño A, Bhattacharya D (2008) Multiple genes of apparent
algal origin suggest ciliates may once have been photosynthetic. Curr Biol 18.
7. Marcotte CJV, Marcotte EM (2002) Predicting functional linkages from gene
fusion events. Proc Natl Acad Sci USA 105: 3427–3432.
8. Peixoto L, Roos D (2007) Genomic scale analysis of lateral gene transfer in
Protozoa: insights into early eukaryotic evolution, host-pathogen
interaction and drug target development. BMC Bioinformatics 8(Suppl 8): S5.
9. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ (2006)
Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic
Kingdoms. Curr Biol 16: 1857–1864.
10. Anderson JO (2006) Convergent Evolution: Gene Sharing by Eukaryotic Plant
Pathogens. Curr Biol 16: R804–R806.
11. Anantharaman V, Iyer ML, Aravind L. (2007) Comparative genomics of
protists/new insights into the evolution of eukaryotic signal transduction and
gene regulation. Annu Rev Microbiol 61.
12. Nosseko T, Bhattacharya D (2007) Horizontal gene transfer in chromalveolates.
BMC Evol Biol 7.
13. Martens C, Vandepoele K, Van de Peer Y (2000) Whole-genome analysis
reveals molecular innovations and evolutionary transitions in chromalveolates
species. Proc Natl Acad Sci USA 107: 1761–1766.
14. Durren P, Nikolić M, Sherman D (2008) Fusion and fission of genes define a
metric between fungal genomes. PLoS Comput Biol.
15. Kamburov A, Goldovsky L, Freilich S, Kapazoglou A, Kanin V, et al. (2007)
DNAinseq inferred functional association networks obtained by gene fusion
analysis. BMC Genomics 8: 460.
16. Enright A, Bieopoulou I, Kyprides N, Ouzounis C (1999) Protein interaction
maps for complete genomes based on gene fusion events. Nature 402: 86–90.
17. Marcotte CJV, Marcotte EM (2002) Predicting functional linkages from gene
fusions with confidence. Applied Bioinformatics 2: 1–8.
18. Tsoka S, Ouzounis CA (2006) Prediction of protein interactions: metabolic
enzymes are frequently involved in gene fusion. Nature Genetics 26: 138–142.
19. Baumann M, Nopoulos GB, VKE (2008) Evolution of protein domain
promiscuity in eukaryotes. Genome Res 18: 449–461.
20. Bowern-Bauer E, Beausart F, Kummerfeld SR, Tschumi SA, Weiner JJ (2005)
The evolution of domain arrangements in proteins and interaction
networks. Cell Mol Life Sci 62: 443–445.
21. Wolf YI, Kondrashov AS, Koonin EV (2000) Interkingdom gene fusions. Gene
51: 313–317.
22. Wall DP, Fraser HB, Hirsh AE. (2003) Detecting putative orthologs. Bioinformatics
19: 1710–1711.
23. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the
signature-recognition methods in InterPro. Bioinformatics 17: 417–418.
24. Roeckmann B, Baireich A, Apweiler R, Blatter MG, Eutricher A, et al. (2003)
The SWISS-PROT protein knowledgebase and its supplement TrEMBL in
2003. Nucleic Acids Res 31: 365–370.
25. Caspi R, Foerster H, Fulcher C, Kaipa P, Krummenacker M, et al. (2008) The
MetaCyc Database of metabolic pathways and enzymes and the BioCyc

collection of Pathway/Genome Databases. Nucleic Acids Res 36 Database issue: D633–D631.

26. Chenna R, Sugawara H, Koake T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment using the Clustal series of programs. Nucleic Acids Res 31: 3497–3500.

27. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.

28. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8: 460.

29. Miller C (2000) An overview of the potassium channel family. Genome Biol 1: reviews0004.0001–reviews0004.0005.

30. Cheng JB, Russell DW (2004) Mammalian Wax Biosynthesis I. 1. Identification of two fatty acyl-coenzyme a reductases with different substrate specificities and tissue distributions. J Biol Chem 279: 37757–37767.

31. Aarts MG, Hodge R, Kalantidis K, D.F, Wilson ZA, et al. (1997) The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J 12: 615–623.

32. Moore BD (2004) Bifunctional and moonlighting enzymes:lighting the way to regulatory control. Trends in Plant Sci 9: 221–228.

33. Yining Y, Chaturvedi I, Meow LK, Kangucane P, Sakharkar MK (2004) Can ends justify the means? digging deep for human fusion genes of prokaryotic origin. Frontiers in Bioscience 9: 2964–2971.

34. Makiuchi T, Nara T, Aonoura T, Hashimoto T, Aoki T (2007) Occurrence of multiple, independent gene fusion events for the fifth and sixth enzymes of pyrimidine biosynthesis in different eukaryotic groups. Gene 394: 78–86.

35. Machida M, Takechi K, Sato H, Chung SJ, Kuroiwa H, et al. (2006) Genes for the peptidylglycan synthesis pathway are essential for chloroplast division in moss. PNAS 103: 6753–6758.

36. Sree P (2000) macromolecular interactions:tracing the roots. Trends Biochem Sci 25: 150–153.

37. Winkel B, SJ (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55: 85–107.

38. Tang G, Xiaoshong Z, Gakiere B, Levanony B, Kahana A, et al. (2002) The Bifunctional LKR/SDH Locus of Plants Also Encodes a Highly Active Monofunctional Lysine-Ketoglutarate Reductase Using a Polyadenylation Signal Located within an Intron. Plant Physiol 130: 147–152.

39. Patron NJ, Durnford DG, Kopriva S (2008) Sulfate assimilation in eukaryotes: fusions, relocations and lateral transfer. BMC Evol Biol 8: 39.

40. Koivuni A, Makareva K, L. A (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55: 709–742.

41. Yanai I, Wolf YI, EV. K (2002) Evolution of gene fusions: horizontal transfer versus independent events. Genome Biology 3: research0024.

42. Wolf YI, Kondrashov AS, EV. K (2000) Interkingdom gene fusions. Genome Biology 1: RESEARCH0013.

43. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9: 605–618.