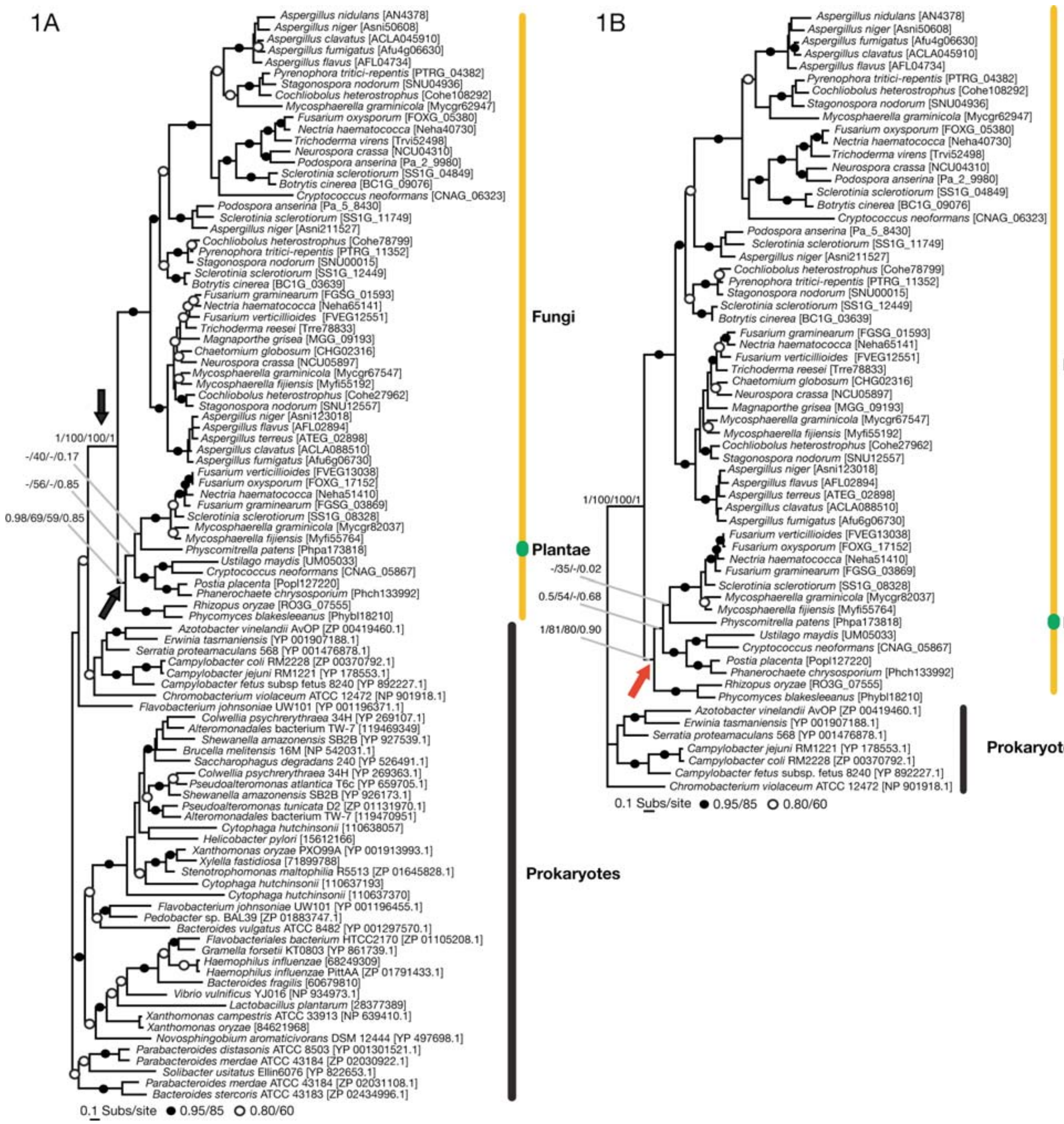
Supplemental Figure 1. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 1A.

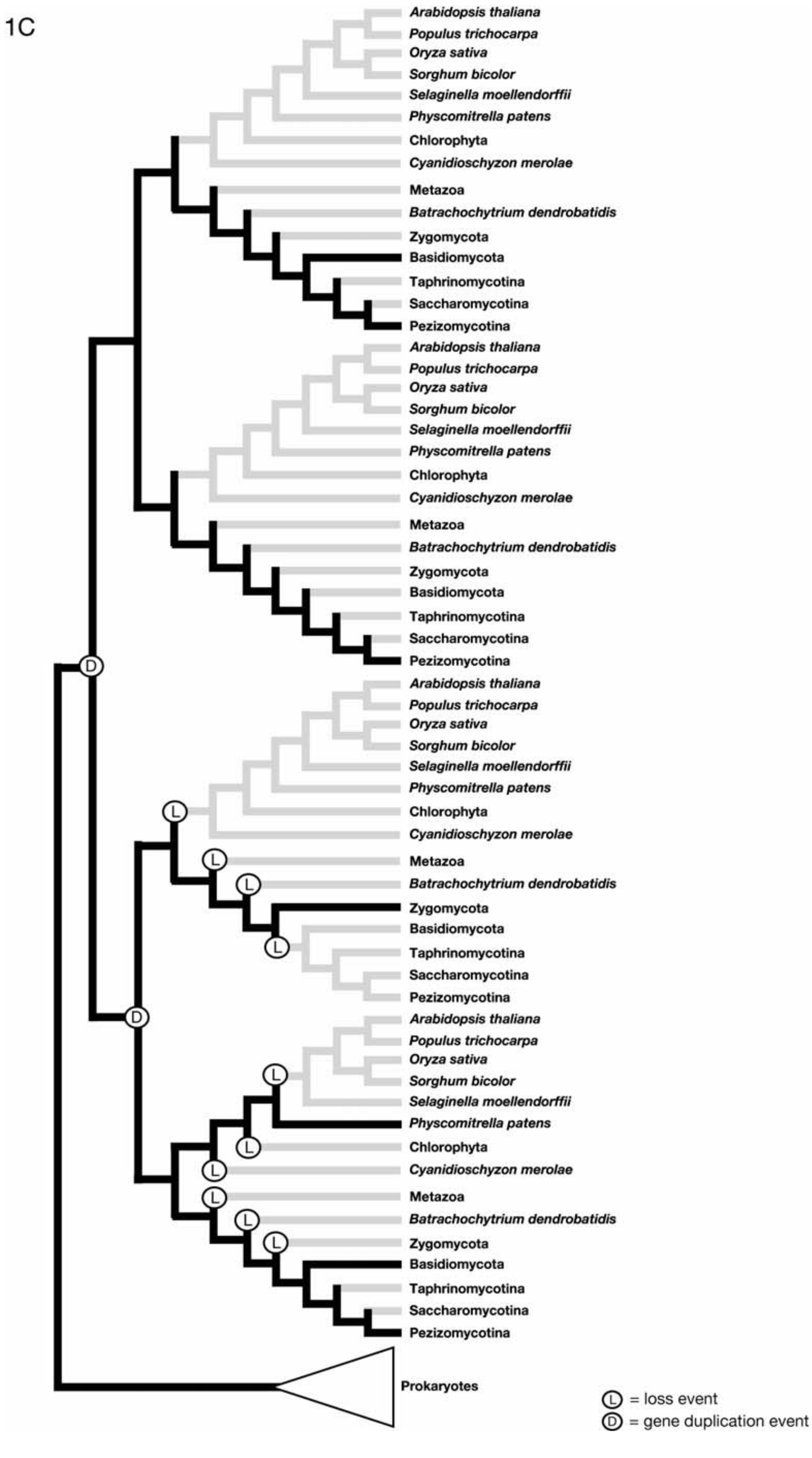
Supplemental Figure 1A-C. Phylogeny of the putative L-fucose permease encoding gene family, demonstrating a candidate fungi-to-plant gene transfer. Our comparative genomic analyses demonstrated that this protein family is, with the exception of the single plant gene, restricted to a diverse group of prokaryotes and the fungi. This taxon sampling suggests a prokaryote-to-fungi gene transfer. We also detected a putative homologue of this protein encoded by the *Physcomitrella patens* genome. This plant gene grouped with moderate to strong support with and within the fungal phylogenetic group (black arrows mark the key branching relationships – **1A**). To test the topological support for placement of the *Physcomitrella* gene we performed a second phylogenetic analyses (**1B**) removing distantly related prokaryote sequences and adjusting the alignment character sampling. Phylogenetic analyses based upon this second alignment demonstrated stronger bootstrap support for the placement of the plant gene within the fungi (81/80% bootstrap support – marked by a red arrow). To test further the placement of *Physcomitrella* gene within the Fungi we constrained a monophyletic branching order of the fungi and calculated alternative tree topologies using distance and parsimony methods (Swofford, 2002). For each alternative topology we re-calculated branch lengths using ML (Foster, 2004) and compared the resulting topologies with the MrBayes (Ronquist and Huelsenbeck, 2003) and PhyML (Guindon and Gascuel, 2003) topologies using the AU and SH test in Consel (Shimodaira and Hasegawa, 2001). We found that we could reject all three alternative topologies, with fungal monophyly enforced, at the 2.9% confidence level or lower. This strongly suggests that the plant sequence branches within the Fungi, probably as a result of a fungi-to-plant horizontal gene transfer.

The phylogeny was calculated from an alignment of 98 sequences and 341 amino acid characters (**1A**) and an alignment of 62 sequences and 349 amino acid characters (**1B**). Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix and a Γ distribution ($\alpha = 1.13$) model of site rate heterogeneity were the most appropriate parameters for the **1A** data set, while a WAG substitution matrix and a Γ distribution ($\alpha = 1.16$) were the most appropriate parameters for the **1B** data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML, with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test the topological result further, we also ran a MrBayes analysis and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short-hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML+RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The sources for all the genome sequences used in the pipeline analysis are listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated, we have provided additional supplementary material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication events and gene loss (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**1C**). These trees were based on an underlying eukaryotic species phylogeny. Because there is uncertainty about the relative branching order of many eukaryotic groups, we restricted the underlying eukaryotic species phylogeny to identified species relationships among the Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006) and the kingdom Plantae. As such, this analysis underestimates the number of gene duplication and loss events required for the alternative hypothesis of hidden paralogy. Only duplication (D) and loss (L) events required to invoke the hidden paralogy are marked. For the L-fucose permease the taxon distribution is highly restricted to fungi, prokaryotes and one plant. This taxon distribution itself suggests HGT. However, for hidden paralogy to explain the branching of the plant within the fungal clade, given the taxon sampling available for this analysis, a minimum of 10 independent gene loss events and 2 gene duplication events are required. This compares to the scenario of a single fungi-to-plant HGT event.


1A
Aspergillus nidulans [AN4378]
Aspergillus niger [Asni50608]
Aspergillus clavatus [ACLA045910]
Aspergillus fumigatus [Afu4g06630]
Aspergillus flavus [AFL04734]
Pyrenophora tritici-repentis [PTRG_04382]
Stagonospora nodorum [SNU04936]
Cochliobolus heterostrophus [Cohe108292]
Mycosphaerella graminicola [Mycgr62947]
Fusarium oxysporum [FOXG_05380]
Nectria haematococca [Neha40730]
Trichoderma virens [Trvi52498]
Neurospora crassa [NCU04310]
Podospora anserina [Pa_2_9980]
Sclerotinia sclerotiorum [SS1G_04849]
Botrytis cinerea [BC1G_09076]
Cryptococcus neoformans [CNAG_06323]
Podospora anserina [Pa_5_8430]
Sclerotinia sclerotiorum [SS1G_11749]
Aspergillus niger [Asni211527]
Cochliobolus heterostrophus [Cohe78799]
Pyrenophora tritici-repentis [PTRG_11352]
Stagonospora nodorum [SNU00015]
Sclerotinia sclerotiorum [SS1G_12449]
Botrytis cinerea [BC1G_03639]
Fusarium graminearum [FGSG_01593]
Nectria haematococca [Neha65141]
Fusarium verticillioides [FVEG12551]
Trichoderma reesei [Trre78833]
Magnaporthe grisea [MGG_09193]
Chaetomium globosum [CHG02316]
Neurospora crassa [NCU05897]
Mycosphaerella graminicola [Mycgr67547]
Mycosphaerella fijiensis [Myfi55192]
Cochliobolus heterostrophus [Cohe27962]
Stagonospora nodorum [SNU12557]
Aspergillus niger [Asni123018]
Aspergillus flavus [AFL02894]
Aspergillus terreus [ATEG_02898]
Aspergillus clavatus [ACLA088510]
Aspergillus fumigatus [Afu6g06730]
1/100/100/1
-/40/-/0.17
-/56/-/0.85
0.98/69/59/0.85
Fusarium verticillioides [FVEG13038]
Fusarium oxysporum [FOXG_17152]
Nectria haematococca [Neha51410]
Fusarium graminearum [FGSG_03869]
Sclerotinia sclerotiorum [SS1G_08328]
Mycosphaerella graminicola [Mycgr82037]
Mycosphaerella fijiensis [Myfi55764]
Physcomitrella patens [Phpa173818]
Ustilago maydis [UM05033]
Cryptococcus neoformans [CNAG_05867]
Postia placenta [Popl127220]
Phanerochaete chrysosporium [Phch133992]
Rhizopus oryzae [RO3G_07555]
Phycomyces blakesleeanus [Phybl18210]
Azotobacter vinelandii AvOP [ZP 00419460.1]
Erwinia tasmaniensis [YP 001907188.1]
Serratia proteamaculans 568 [YP 001476878.1]
Campylobacter coli RM2228 [ZP 00370792.1]
Campylobacter jejuni RM1221 [YP 178553.1]
Campylobacter fetus subsp fetus 8240 [YP 892227.1]
Chromobacterium violaceum ATCC 12472 [NP 901918.1]
Flavobacterium johnsoniae UW101 [YP 001196371.1]
Colwellia psychrerythraea 34H [YP 269107.1]
Alteromonadales bacterium TW-7 [119469349]
Shewanella amazonensis SB2B [YP 927539.1]
Brucella melitensis 16M [NP 542031.1]
Saccharophagus degradans 240 [YP 526491.1]
Colwellia psychrerythraea 34H [YP 269363.1]
Pseudoalteromonas atlantica T6c [YP 659705.1]
Shewanella amazonensis SB2B [YP 926173.1]
Pseudoalteromonas tunicata D2 [ZP 01131970.1]
Alteromonadales bacterium TW-7 [119470951]
Cytophaga hutchinsonii [110638057]
Helicobacter pylori [15612166]
Xanthomonas oryzae PXO99A [YP 001913993.1]
Xylella fastidiosa [71899788]
Stenotrophomonas maltophilia R5513 [ZP 01645828.1]
Cytophaga hutchinsonii [110637193]
Cytophaga hutchinsonii [110637370]
Flavobacterium johnsoniae UW101 [YP 001196455.1]
Pedobacter sp. BAL39 [ZP 01883747.1]
Bacteroides vulgatus ATCC 8482 [YP 001297570.1]
Flavobacteriales bacterium HTCC2170 [ZP 01105208.1]
Gramella forsetii KT0803 [YP 861739.1]
Haemophilus influenzae [68249309]
Haemophilus influenzae PittAA [ZP 01791433.1]
Bacteroides fragilis [60679810]
Vibrio vulnificus YJ016 [NP 934973.1]
Lactobacillus plantarum [28377389]
Xanthomonas campestris ATCC 33913 [NP 639410.1]
Xanthomonas oryzae [84621968]
Novosphingobium aromaticivorans DSM 12444 [YP 497698.1]
Parabacteroides distasonis ATCC 8503 [YP 001301521.1]
Parabacteroides merdae ATCC 43184 [ZP 02030922.1]
Solibacter usitatus Ellin6076 [YP 822653.1]
Parabacteroides merdae ATCC 43184 [ZP 02031108.1]
Bacteroides stercoris ATCC 43183 [ZP 02434996.1]
Fungi
Plantae
Prokaryotes
0.1 Subs/site ● 0.95/85 ○ 0.80/60
1B
Aspergillus nidulans [AN4378]
Aspergillus niger [Asni50608]
Aspergillus fumigatus [Afu4g06630]
Aspergillus clavatus [ACLA045910]
Aspergillus flavus [AFL04734]
Pyrenophora tritici-repentis [PTRG_04382]
Cochliobolus heterostrophus [Cohe108292]
Stagonospora nodorum [SNU04936]
Mycosphaerella graminicola [Mycgr62947]
Fusarium oxysporum [FOXG_05380]
Nectria haematococca [Neha40730]
Trichoderma virens [Trvi52498]
Neurospora crassa [NCU04310]
Podospora anserina [Pa_2_9980]
Sclerotinia sclerotiorum [SS1G_04849]
Botrytis cinerea [BC1G_09076]
Cryptococcus neoformans [CNAG_06323]
Podospora anserina [Pa_5_8430]
Sclerotinia sclerotiorum [SS1G_11749]
Aspergillus niger [Asni211527]
Cochliobolus heterostrophus [Cohe78799]
Pyrenophora tritici-repentis [PTRG_11352]
Stagonospora nodorum [SNU00015]
Sclerotinia sclerotiorum [SS1G_12449]
Botrytis cinerea [BC1G_03639]
Fusarium graminearum [FGSG_01593]
Nectria haematococca [Neha65141]
Fusarium verticillioides [FVEG12551]
Trichoderma reesei [Trre78833]
Chaetomium globosum [CHG02316]
Neurospora crassa [NCU05897]
Magnaporthe grisea [MGG_09193]
Mycosphaerella graminicola [Mycgr67547]
Mycosphaerella fijiensis [Myfi55192]
Cochliobolus heterostrophus [Cohe27962]
Stagonospora nodorum [SNU12557]
Aspergillus niger [Asni123018]
Aspergillus flavus [AFL02894]
Aspergillus terreus [ATEG_02898]
Aspergillus clavatus [ACLA088510]
Aspergillus fumigatus [Afu6g06730]
1/100/100/1
Fusarium verticillioides [FVEG13038]
Fusarium oxysporum [FOXG_17152]
Nectria haematococca [Neha51410]
Fusarium graminearum [FGSG_03869]
-/35/-/0.02
Sclerotinia sclerotiorum [SS1G_08328]
Mycosphaerella graminicola [Mycgr82037]
0.5/54/-/0.68
Mycosphaerella fijiensis [Myfi55764]
Physcomitrella patens [Phpa173818]
1/81/80/0.90
Ustilago maydis [UM05033]
Cryptococcus neoformans [CNAG_05867]
Postia placenta [Popl127220]
Phanerochaete chrysosporium [Phch133992]
Rhizopus oryzae [RO3G_07555]
Phycomyces blakesleeanus [Phybl18210]
Azotobacter vinelandii AvOP [ZP 00419460.1]
Erwinia tasmaniensis [YP 001907188.1]
Serratia proteamaculans 568 [YP 001476878.1]
Campylobacter jejuni RM1221 [YP 178553.1]
Campylobacter coli RM2228 [ZP 00370792.1]
Campylobacter fetus subsp. fetus 8240 [YP 892227.1]
Chromobacterium violaceum ATCC 12472 [NP 901918.1]
Fungi
Plantae
Prokaryotes
0.1 Subs/site ● 0.95/85 ○ 0.80/60


1C
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
Chlorophyta
Cyanidioschyzon merolae
Metazoa
Batrachochytrium dendrobatidis
Zygomycota
Basidiomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
Chlorophyta
Cyanidioschyzon merolae
Metazoa
Batrachochytrium dendrobatidis
Zygomycota
Basidiomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
D
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
L
Chlorophyta
Cyanidioschyzon merolae
L
Metazoa
L
Batrachochytrium dendrobatidis
Zygomycota
Basidiomycota
L
Taphrinomycotina
Saccharomycotina
Pezizomycotina
D
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
L
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
L
Chlorophyta
L
Cyanidioschyzon merolae
L
Metazoa
L
Batrachochytrium dendrobatidis
L
Zygomycota
Basidiomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
Prokaryotes
L = loss event
D = gene duplication event

REFERENCES

Foster, P.G. (2004). Modeling compositional heterogeneity. Syst. Biol. **53,** 485-495.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL:~for assessing the confidence of phylogenetic tree selection. Bioinformatics **17,** 1246-1247.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. (Sunderland, Massachusetts: Sinauer Associates).

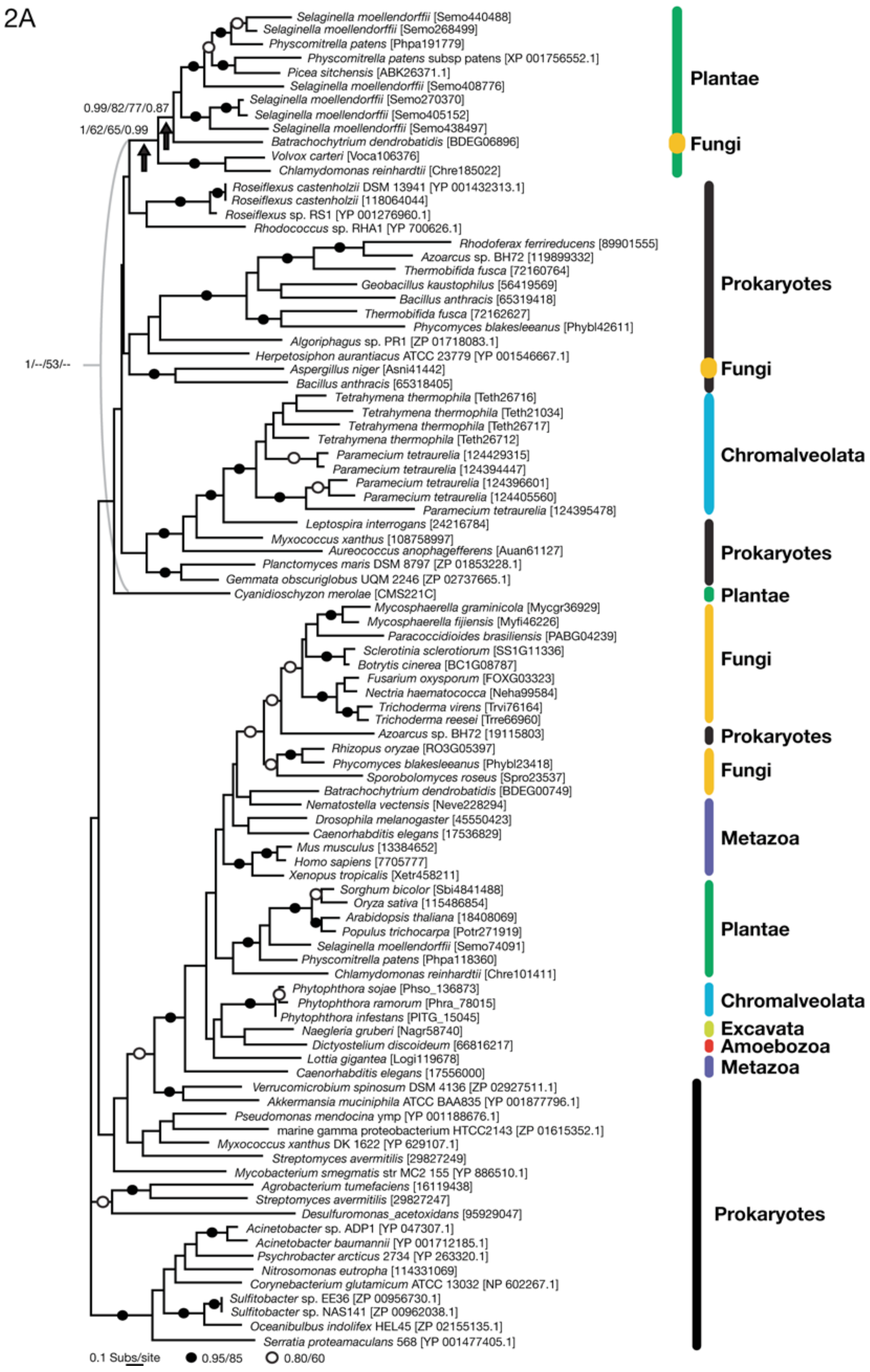
Supplemental Figure 2. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 1B.

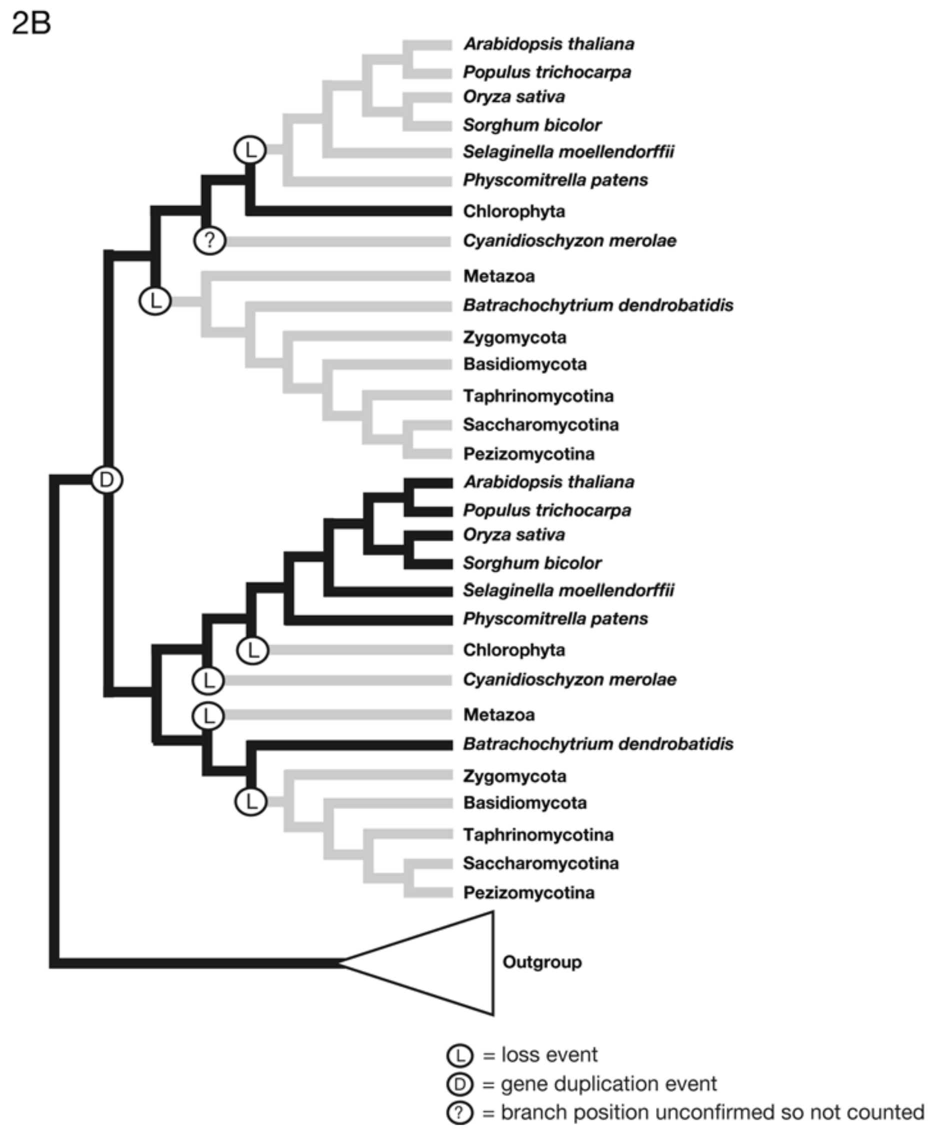
Figure 2. Phylogenetic analyses of the putative zinc-binding alcohol dehydrogenase protein encoding gene family, demonstrating a putative horizontal gene transfer event from the 'Plantae' to the genome of the 'chytrid' fungus *Batrachochytrium dendrobatidis*. The 'chytrid' gene groups with moderate to strong support within a 'Plantae' phylogenetic group sister to the land plants, but rooted by the green algae. Such a branching relationship suggests an ancient gene transfer event from an early land plant lineage to the 'chytrid' fungi (black arrows mark the key branching relationships – **2A**). To test further the placement of *Batrachochytrium* gene within the 'Plantae' clade, we constrained a monophyletic branching order of the land plant and green algae clade to the exclusion of the 'chytrid' sequence and calculated alternative tree topologies using distance and parsimony methods (Swofford, 2002). For each alternative topology we re-calculated branch lengths using ML (Foster, 2004) and compared the resulting topologies with the MrBayes (Ronquist and Huelsenbeck, 2003) and PhyML (Guindon and Gascuel, 2003) topologies using the AU and SH test in Consel (Shimodaira and Hasegawa, 2001). We found that we could reject all six alternative topologies with fungal monophyly enforced at the 2.5% confidence level or lower, strongly suggesting that the 'chytrid' sequence branches within the 'Plantae' phylogenetic cluster. Taken together, this data is consistent with a Plantae to 'chytrid' gene transfer event.

The phylogeny was calculated from an alignment of 95 sequences and 207 amino acid characters. Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix, and a Γ distribution ($\alpha = 1.57$), model of site rate heterogeneity were the most appropriate parameters for the data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML, with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test further the topological result, we also ran a MrBayes analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML + RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The sources of all the genome sequences used in the pipeline analysis are listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated, we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication events and gene loss (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**2B**). These trees were based on an underlying eukaryotic species phylogeny. Because there is uncertainty about the relative branching order of many eukaryotic groups we restricted the underlying eukaryotic species phylogeny to identified species branching relationships among the, Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006). As such this analyses underestimates the number of gene duplication and gene loss events required for the alternative hypothesis of hidden paralogy. Only gene duplication (D) and gene loss (L) events required to invoke the hidden paralogy are marked (all other loss events are not scored). For hidden paralogy to explain the branching of the fungi within the Plantae clade, given the taxon sampling available for this analysis, a minimum of 5 independent gene loss events and 1 gene duplication events are required. This compares to the scenario of a single Plantae-to-fungi HGT event.


2A
Selaginella moellendorffii [Semo440488]
Selaginella moellendorffii [Semo268499]
Physcomitrella patens [Phpa191779]
Physcomitrella patens subsp patens [XP 001756552.1]
Picea sitchensis [ABK26371.1]
Selaginella moellendorffii [Semo408776]
Selaginella moellendorffii [Semo270370]
Selaginella moellendorffii [Semo405152]
Selaginella moellendorffii [Semo438497]
Batrachochytrium dendrobatidis [BDEG06896]
Volvox carteri [Voca106376]
Chlamydomonas reinhardtii [Chre185022]
0.99/82/77/0.87
1/62/65/0.99
Roseiflexus castenholzii DSM 13941 [YP 001432313.1]
Roseiflexus castenholzii [118064044]
Roseiflexus sp. RS1 [YP 001276960.1]
Rhodococcus sp. RHA1 [YP 700626.1]
Rhodoferax ferrireducens [89901555]
Azoarcus sp. BH72 [119899332]
Thermobifida fusca [72160764]
Geobacillus kaustophilus [56419569]
Bacillus anthracis [65319418]
Thermobifida fusca [72162627]
Phycomyces blakesleeanus [Phybl42611]
Algoriphagus sp. PR1 [ZP 01718083.1]
1/--/53/--
Herpetosiphon aurantiacus ATCC 23779 [YP 001546667.1]
Aspergillus niger [Asni41442]
Bacillus anthracis [65318405]
Tetrahymena thermophila [Teth26716]
Tetrahymena thermophila [Teth21034]
Tetrahymena thermophila [Teth26717]
Tetrahymena thermophila [Teth26712]
Paramecium tetraurelia [124429315]
Paramecium tetraurelia [124394447]
Paramecium tetraurelia [124396601]
Paramecium tetraurelia [124405560]
Paramecium tetraurelia [124395478]
Leptospira interrogans [24216784]
Myxococcus xanthus [108758997]
Aureococcus anophagefferens [Auan61127]
Planctomyces maris DSM 8797 [ZP 01853228.1]
Gemmata obscuriglobus UQM 2246 [ZP 02737665.1]
Cyanidioschyzon merolae [CMS221C]
Mycosphaerella graminicola [Mycgr36929]
Mycosphaerella fijiensis [Myfi46226]
Paracoccidioides brasiliensis [PABG04239]
Sclerotinia sclerotiorum [SS1G11336]
Botrytis cinerea [BC1G08787]
Fusarium oxysporum [FOXG03323]
Nectria haematococca [Neha99584]
Trichoderma virens [Trvi76164]
Trichoderma reesei [Trre66960]
Azoarcus sp. BH72 [19115803]
Rhizopus oryzae [RO3G05397]
Phycomyces blakesleeanus [Phybl23418]
Sporobolomyces roseus [Spro23537]
Batrachochytrium dendrobatidis [BDEG00749]
Nematostella vectensis [Neve228294]
Drosophila melanogaster [45550423]
Caenorhabditis elegans [17536829]
Mus musculus [13384652]
Homo sapiens [7705777]
Xenopus tropicalis [Xetr458211]
Sorghum bicolor [Sbi4841488]
Oryza sativa [115486854]
Arabidopsis thaliana [18408069]
Populus trichocarpa [Potr271919]
Selaginella moellendorffii [Semo74091]
Physcomitrella patens [Phpa118360]
Chlamydomonas reinhardtii [Chre101411]
Phytophthora sojae [Phso_136873]
Phytophthora ramorum [Phra_78015]
Phytophthora infestans [PITG_15045]
Naegleria gruberi [Nagr58740]
Dictyostelium discoideum [66816217]
Lottia gigantea [Logi119678]
Caenorhabditis elegans [17556000]
Verrucomicrobium spinosum DSM 4136 [ZP 02927511.1]
Akkermansia muciniphila ATCC BAA835 [YP 001877796.1]
Pseudomonas mendocina ymp [YP 001188676.1]
marine gamma proteobacterium HTCC2143 [ZP 01615352.1]
Myxococcus xanthus DK 1622 [YP 629107.1]
Streptomyces avermitilis [29827249]
Mycobacterium smegmatis str MC2 155 [YP 886510.1]
Agrobacterium tumefaciens [16119438]
Streptomyces avermitilis [29827247]
Desulfuromonas_acetoxidans [95929047]
Acinetobacter sp. ADP1 [YP 047307.1]
Acinetobacter baumannii [YP 001712185.1]
Psychrobacter arcticus 2734 [YP 263320.1]
Nitrosomonas eutropha [114331069]
Corynebacterium glutamicum ATCC 13032 [NP 602267.1]
Sulfitobacter sp. EE36 [ZP 00956730.1]
Sulfitobacter sp. NAS141 [ZP 00962038.1]
Oceanibulbus indolifex HEL45 [ZP 02155135.1]
Serratia proteamaculans 568 [YP 001477405.1]
0.1 Subs/site
● 0.95/85
○ 0.80/60
Plantae
Fungi
Prokaryotes
Fungi
Chromalveolata
Prokaryotes
Plantae
Fungi
Prokaryotes
Fungi
Metazoa
Plantae
Chromalveolata
Excavata
Amoebozoa
Metazoa
Prokaryotes



REFERENCES

Foster, P.G. (2004). Modeling compositional heterogeneity. Syst. Biol. **53,** 485-495.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-

Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, **T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O.** (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL:~for assessing the confidence of phylogenetic tree selection. Bioinformatics **17,** 1246-1247.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. (Sunderland, Massachusetts: Sinauer Associates).

Supplemental Figure 3. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 1C.

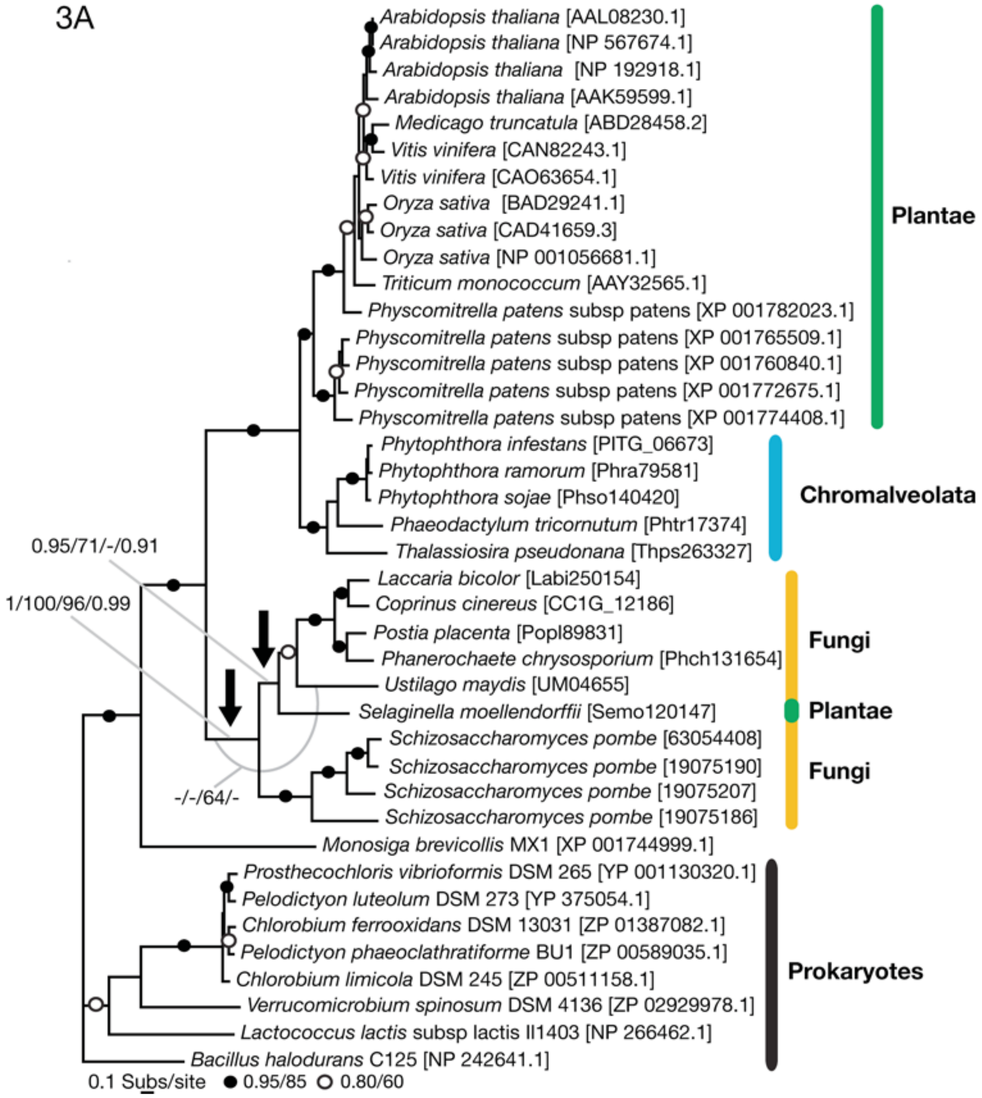
Figure 3. Phylogenetic analyses of a putative membrane transporter protein encoding gene family, demonstrating a putative horizontal gene transfer event from the fungi to the lycophyte *Selaginella* genome. The phyML and RAxML demonstrated slightly different topologies, shown by grey lines on Figure **3A**, however both analyses placed the *Selaginella* clade with 71 and 64% bootstrap support respectively (marked by a black arrow) within the fungal clade. This combined with the strong bootstrap support (100/96% bootstrap support – marked by a black arrow - **3A**) for the grouping of *Selaginella* with the fungi, separately from all other plants suggests support for a putative fungi-to-plant horizontal gene transfer. To specifically test the phylogenetic support for the branching of the *Selaginella* gene within the fungal clade, we constrained a monophyletic branching order for the fungal clade and calculated alternative tree topologies using distance and parsimony methods (Swofford, 2002). For each alternative topology we re-calculated branch lengths using ML (Foster, 2004) and compared the resulting topologies with the MrBayes (Ronquist and Huelsenbeck, 2003) and PhyML (Guindon and Gascuel, 2003) topologies using the AU and SH test in Consel (Shimodaira and Hasegawa, 2001). We found that we could reject two of the alternative topologies at the 2.4% significance level using AU and SH test. However, one alternative topology that placed the *Selaginella* sequence with the prokaryotes could not be rejected at the 5% confidence level using the AU test (AU score = 0.168). However, this alternative topology was strongly rejected by the SH test at the 0.1% confidence level. This combined with the strong bootstrap support for placement of the *Selaginella* sequence with the fungi (100/96%) strongly suggests that the *Selaginella* sequence branches with the ascomycete clade most likely as a product of a fungi-to-plant horizontal gene transfer.

The phylogeny was calculated from an alignment of 40 sequences and 354 amino acid characters. Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix and a Γ distribution ($\alpha = 2.22$) model of site rate heterogeneity were the most appropriate parameters for the data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML, with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test further the topological result, we also ran a MrBayes analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML + RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

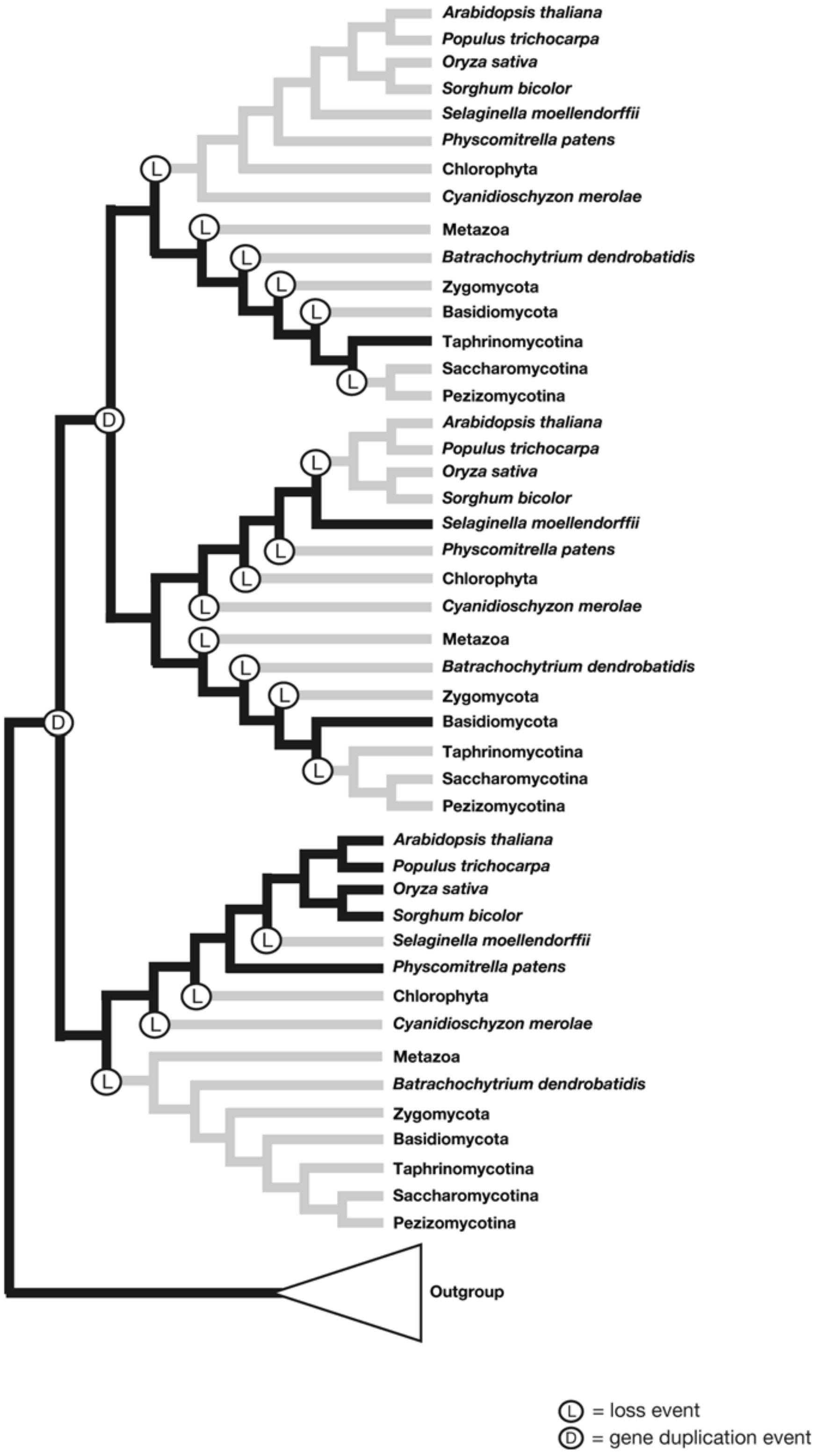
The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication and gene loss (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**3B**). These trees were based on an underlying eukaryotic species phylogeny. Because there is uncertainty about the relative branching order of many eukaryotic groups we restricted the underlying eukaryotic species phylogeny to strongly supported branching relationships among the Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006). As such this analyses underestimates the number of gene duplication and gene loss events required for the alternative hypothesis of hidden paralogy. Only duplication (D) and loss (L) events required to invoke the hidden paralogy are marked (all other loss events are not scored). For hidden paralogy to explain the branching of the plant within the fungal clade, given the taxon sampling available for this analysis, a minimum of 17 independent gene loss events and 2 gene duplication events are required. This compares to the scenario of a single fungi-to-plant HGT event.

3A


Arabidopsis thaliana [AAL08230.1]
Arabidopsis thaliana [NP 567674.1]
Arabidopsis thaliana [NP 192918.1]
Arabidopsis thaliana [AAK59599.1]
Medicago truncatula [ABD28458.2]
Vitis vinifera [CAN82243.1]
Vitis vinifera [CAO63654.1]
Oryza sativa [BAD29241.1]
Oryza sativa [CAD41659.3]
Oryza sativa [NP 001056681.1]
Triticum monococcum [AAY32565.1]
Physcomitrella patens subsp patens [XP 001782023.1]
Physcomitrella patens subsp patens [XP 001765509.1]
Physcomitrella patens subsp patens [XP 001760840.1]
Physcomitrella patens subsp patens [XP 001772675.1]
Physcomitrella patens subsp patens [XP 001774408.1]
Plantae
Phytophthora infestans [PITG_06673]
Phytophthora ramorum [Phra79581]
Phytophthora sojae [Phso140420]
Phaeodactylum tricornutum [Phtr17374]
Thalassiosira pseudonana [Thps263327]
Chromalveolata
0.95/71/-/0.91
1/100/96/0.99
Laccaria bicolor [Labi250154]
Coprinus cinereus [CC1G_12186]
Postia placenta [Popl89831]
Phanerochaete chrysosporium [Phch131654]
Ustilago maydis [UM04655]
Selaginella moellendorffii [Semo120147]
Schizosaccharomyces pombe [63054408]
Schizosaccharomyces pombe [19075190]
Schizosaccharomyces pombe [19075207]
Schizosaccharomyces pombe [19075186]
-/-/64/-
Fungi
Plantae
Fungi
Monosiga brevicollis MX1 [XP 001744999.1]
Prosthecochloris vibrioformis DSM 265 [YP 001130320.1]
Pelodictyon luteolum DSM 273 [YP 375054.1]
Chlorobium ferrooxidans DSM 13031 [ZP 01387082.1]
Pelodictyon phaeoclathratiforme BU1 [ZP 00589035.1]
Chlorobium limicola DSM 245 [ZP 00511158.1]
Verrucomicrobium spinosum DSM 4136 [ZP 02929978.1]
Lactococcus lactis subsp lactis Il1403 [NP 266462.1]
Bacillus halodurans C125 [NP 242641.1]
Prokaryotes
0.1 Subs/site
● 0.95/85
○ 0.80/60

3B



REFERENCES

Foster, P.G. (2004). Modeling compositional heterogeneity. Syst. Biol. **53,** 485-495.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL:~for assessing the confidence of phylogenetic tree selection. Bioinformatics **17,** 1246-1247.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. (Sunderland, Massachusetts: Sinauer Associates).

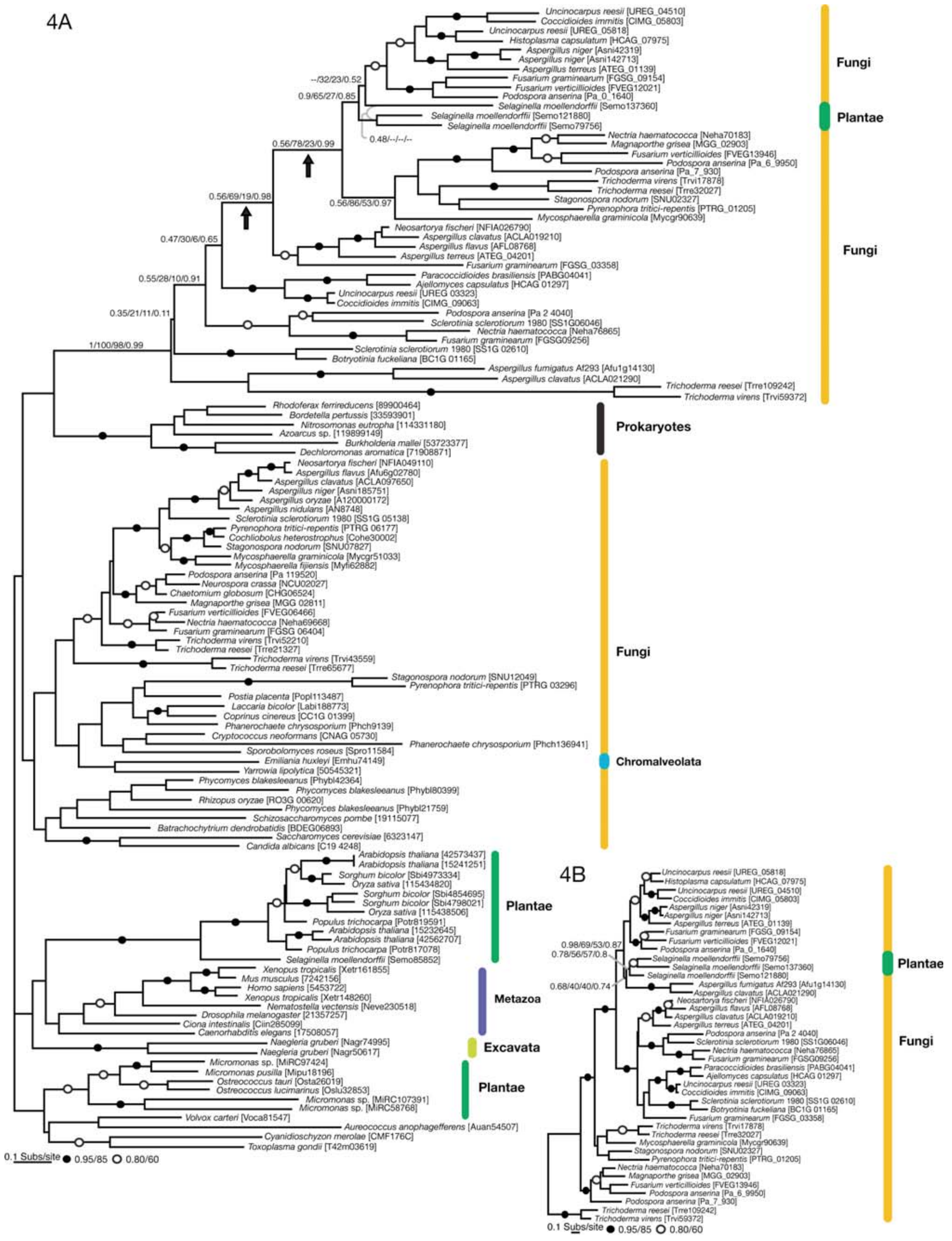
Supplemental Figure 4. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 2.

Figure 4. Phylogenetic analyses of the putative phospholipase / carboxylesterase family protein encoding gene family, demonstrating a putative horizontal gene transfer event from the ascomycete fungi to the lycophyte *Selaginella* genome (**4A**). Three *Selaginella* genes grouped within the ascomycetes by numerous weakly support nodes. However, two nodes were moderately supported in the phyML based bootstrap analyses (69 and 78% bootstrap support – marked by black arrows). Together, these phylogenetic relationships provide support for a putative fungi-to-plant horizontal gene transfer. The main analyses also suggested that the three *Selaginella* genes were not monophyletic. To test this possibility, we performed a second phylogenetic analyses (**4B**) removing distantly related sequences so that the resulting analyses focused on the ascomycete/*Selaginella* clade and adjusting the alignment character sampling. Phylogenetic analyses based upon this second alignment demonstrated weak support (56/57% bootstrap support) for a monophyletic *Selaginella* clade, suggesting that the putative gene transfer was a single fungal-to-plant transfer event. To specifically test the phylogenetic support for the three *Selaginella* sequences grouping within the ascomycete fungal clade- the key phylogenetic relationship for inferring a fungi-to-plant HGT event- we constrained a monophyletic branching order for the relevant ascomycete fungal clade and calculated alternative tree topologies using distance and parsimony methods (Swofford, 2002). For each alternative topology we re-calculated branch lengths using ML (Foster, 2004) and compared the resulting topologies with the MrBayes (Ronquist and Huelsenbeck, 2003) and PhyML (Guindon and Gascuel, 2003) topologies using the AU and SH test in Consel (Shimodaira and Hasegawa, 2001). We found that we could reject all five alternative topologies with fungal monophyly enforced at less than 0.1% confidence level. This strongly suggests that the *Selaginella* sequence branches within the ascomycete clade. Taken together this data suggests a fungi-to-plant horizontal gene transfer.

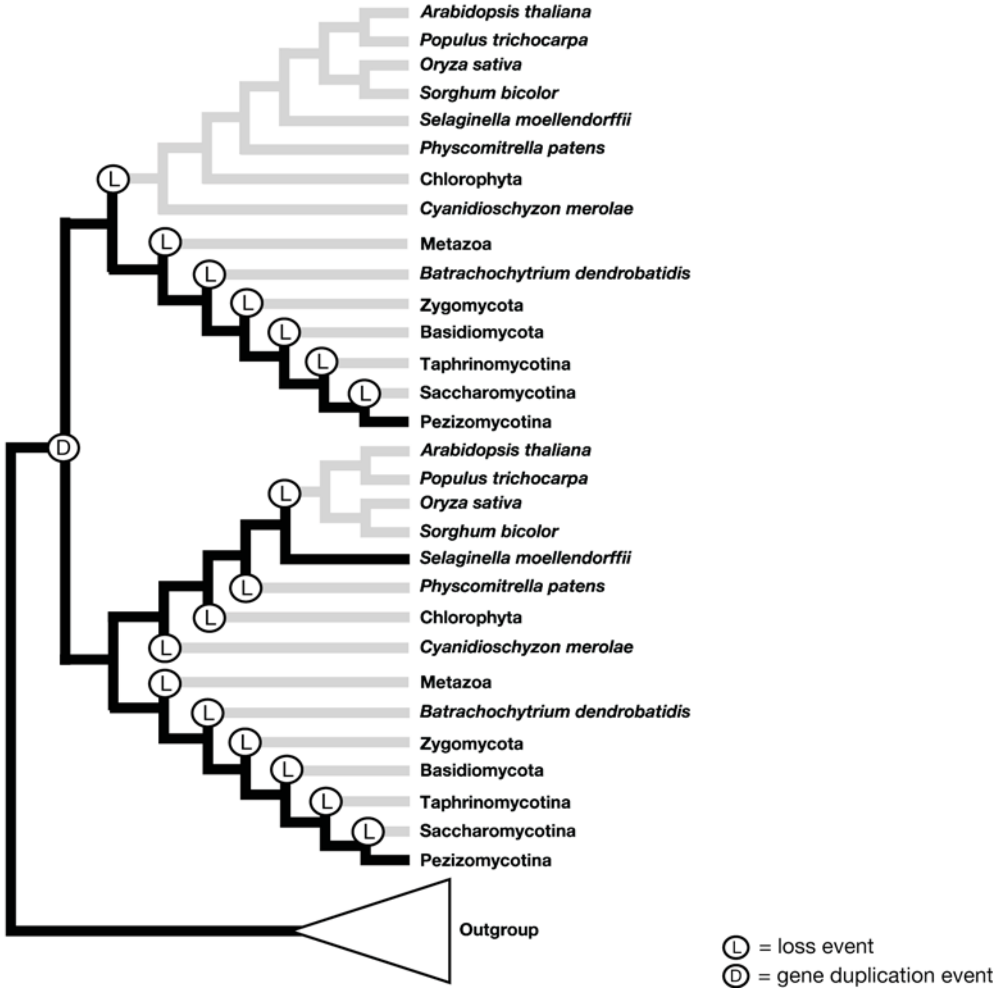
The phylogeny was calculated from an alignment of 122 sequences and 158 amino acid characters (**4A**) and an alignment of 62 sequences and 349 amino acid characters (**4B**). Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix, Γ distribution ($\alpha = 1.52$), and a proportion of invariant sites ($I = 0.03$), model of site rate heterogeneity were the most appropriate parameters for the **4A** data set. While a WAG substitution matrix, Γ distribution ($\alpha = 2.302$), and a proportion of invariant sites ($I = 0.025$) model of site rate heterogeneity were the most appropriate parameters for the **4B** data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML, with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test further the topological result, we also ran a MrBayes analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML+ RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication events and gene losses (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**4C**). These trees were based on an underlying eukaryotic species phylogeny. Because there is uncertainty about the relative branching order of many eukaryotic groups we restricted the underlying eukaryotic species phylogeny to strongly supported branching relationships among the Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006). As such this analyses underestimates the number of gene duplication and gene loss events required for the alternative hypothesis of hidden paralogy. Only duplication (D) and loss (L) events required to invoke the hidden paralogy are marked (all other loss events are not scored). For hidden paralogy to explain the branching of the plant within the fungal clade, given the taxon sampling available for this analysis, a minimum of 17 independent gene loss events and 1 gene duplication events are required, in contrast to a single HGT.


4A
Uncinocarpus reesii [UREG_04510]
Coccidioides immitis [CIMG_05803]
Uncinocarpus reesii [UREG_05818]
Histoplasma capsulatum [HCAG_07975]
Aspergillus niger [Asni42319]
Aspergillus niger [Asni142713]
Aspergillus terreus [ATEG_01139]
--/32/23/0.52
Fusarium graminearum [FGSG_09154]
Fusarium verticillioides [FVEG12021]
0.9/65/27/0.85
Podospora anserina [Pa_0_1640]
Selaginella moellendorffii [Semo137360]
Selaginella moellendorffii [Semo121880]
Selaginella moellendorffii [Semo79756]
0.48/--/--/--
Nectria haematococca [Neha70183]
0.56/78/23/0.99
Magnaporthe grisea [MGG_02903]
Fusarium verticillioides [FVEG13946]
Podospora anserina [Pa_6_9950]
Podospora anserina [Pa_7_930]
Trichoderma virens [Trvi17878]
Trichoderma reesei [Trre32027]
0.56/69/19/0.98
Stagonospora nodorum [SNU02327]
0.56/86/53/0.97
Pyrenophora tritici-repentis [PTRG_01205]
Mycosphaerella graminicola [Mycgr90639]
Neosartorya fischeri [NFIA026790]
0.47/30/6/0.65
Aspergillus clavatus [ACLA019210]
Aspergillus flavus [AFL08768]
Aspergillus terreus [ATEG_04201]
Fusarium graminearum [FGSG_03358]
Paracoccidioides brasiliensis [PABG04041]
0.55/28/10/0.91
Ajellomyces capsulatus [HCAG 01297]
Uncinocarpus reesii [UREG 03323]
Coccidioides immitis [CIMG_09063]
Podospora anserina [Pa 2 4040]
0.35/21/11/0.11
Sclerotinia sclerotiorum 1980 [SS1G06046]
Nectria haematococca [Neha76865]
Fusarium graminearum [FGSG09256]
1/100/98/0.99
Sclerotinia sclerotiorum 1980 [SS1G 02610]
Botryotinia fuckeliana [BC1G 01165]
Aspergillus fumigatus Af293 [Afu1g14130]
Aspergillus clavatus [ACLA021290]
Trichoderma reesei [Trre109242]
Trichoderma virens [Trvi59372]
Fungi
Plantae
Fungi
Rhodoferax ferrireducens [89900464]
Bordetella pertussis [33593901]
Nitrosomonas eutropha [114331180]
Azoarcus sp. [119899149]
Burkholderia mallei [53723377]
Dechloromonas aromatica [71908871]
Prokaryotes
Neosartorya fischeri [NFIA049110]
Aspergillus flavus [Afu6g02780]
Aspergillus clavatus [ACLA097650]
Aspergillus niger [Asni185751]
Aspergillus oryzae [A120000172]
Aspergillus nidulans [AN8748]
Sclerotinia sclerotiorum 1980 [SS1G 05138]
Pyrenophora tritici-repentis [PTRG 06177]
Cochliobolus heterostrophus [Cohe30002]
Stagonospora nodorum [SNU07827]
Mycosphaerella graminicola [Mycgr51033]
Mycosphaerella fijiensis [Myfi62882]
Podospora anserina [Pa 119520]
Neurospora crassa [NCU02027]
Chaetomium globosum [CHG06524]
Magnaporthe grisea [MGG 02811]
Fusarium verticillioides [FVEG06466]
Nectria haematococca [Neha69668]
Fusarium graminearum [FGSG 06404]
Trichoderma virens [Trvi52210]
Trichoderma reesei [Trre21327]
Trichoderma virens [Trvi43559]
Trichoderma reesei [Trre65677]
Fungi
Stagonospora nodorum [SNU12049]
Pyrenophora tritici-repentis [PTRG 03296]
Postia placenta [Popl113487]
Laccaria bicolor [Labi188773]
Coprinus cinereus [CC1G 01399]
Phanerochaete chrysosporium [Phch9139]
Cryptococcus neoformans [CNAG 05730]
Phanerochaete chrysosporium [Phch136941]
Sporobolomyces roseus [Spro11584]
Emiliania huxleyi [Emhu74149]
Chromalveolata
Yarrowia lipolytica [50545321]
Phycomyces blakesleeanus [Phybl42364]
Phycomyces blakesleeanus [Phybl80399]
Rhizopus oryzae [RO3G 00620]
Phycomyces blakesleeanus [Phybl21759]
Schizosaccharomyces pombe [19115077]
Batrachochytrium dendrobatidis [BDEG06893]
Saccharomyces cerevisiae [6323147]
Candida albicans [C19 4248]
Arabidopsis thaliana [42573437]
Arabidopsis thaliana [15241251]
Sorghum bicolor [Sbi4973334]
Oryza sativa [115434820]
Sorghum bicolor [Sbi4854695]
Sorghum bicolor [Sbi4798021]
Oryza sativa [115438506]
Plantae
Populus trichocarpa [Potr819591]
Arabidopsis thaliana [15232645]
Arabidopsis thaliana [42562707]
Populus trichocarpa [Potr817078]
Selaginella moellendorffii [Semo85852]
Xenopus tropicalis [Xetr161855]
Mus musculus [7242156]
Homo sapiens [5453722]
Xenopus tropicalis [Xetr148260]
Metazoa
Nematostella vectensis [Neve230518]
Drosophila melanogaster [21357257]
Ciona intestinalis [Ciin285099]
Caenorhabditis elegans [17508057]
Naegleria gruberi [Nagr74995]
Excavata
Naegleria gruberi [Nagr50617]
Micromonas sp. [MiRC97424]
Micromonas pusilla [Mipu18196]
Ostreococcus tauri [Osta26019]
Plantae
Ostreococcus lucimarinus [Oslu32853]
Micromonas sp. [MiRC107391]
Micromonas sp. [MiRC58768]
Volvox carteri [Voca81547]
Aureococcus anophagefferens [Auan54507]
Cyanidioschyzon merolae [CMF176C]
Toxoplasma gondii [T42m03619]
0.1 Subs/site
● 0.95/85
○ 0.80/60
4B
Uncinocarpus reesii [UREG_05818]
Histoplasma capsulatum [HCAG_07975]
Uncinocarpus reesii [UREG_04510]
Coccidioides immitis [CIMG_05803]
Aspergillus niger [Asni42319]
Aspergillus niger [Asni142713]
Aspergillus terreus [ATEG_01139]
Fusarium graminearum [FGSG_09154]
Fusarium verticillioides [FVEG12021]
0.98/69/53/0.87
Podospora anserina [Pa_0_1640]
0.78/56/57/0.8
Selaginella moellendorffii [Semo79756]
Selaginella moellendorffii [Semo137360]
Plantae
Selaginella moellendorffii [Semo121880]
0.68/40/40/0.74
Aspergillus fumigatus Af293 [Afu1g14130]
Aspergillus clavatus [ACLA021290]
Neosartorya fischeri [NFIA026790]
Aspergillus flavus [AFL08768]
Aspergillus clavatus [ACLA019210]
Aspergillus terreus [ATEG_04201]
Podospora anserina [Pa 2 4040]
Sclerotinia sclerotiorum 1980 [SS1G06046]
Fungi
Nectria haematococca [Neha76865]
Fusarium graminearum [FGSG09256]
Paracoccidioides brasiliensis [PABG04041]
Ajellomyces capsulatus [HCAG 01297]
Uncinocarpus reesii [UREG 03323]
Coccidioides immitis [CIMG_09063]
Sclerotinia sclerotiorum 1980 [SS1G 02610]
Botryotinia fuckeliana [BC1G 01165]
Fusarium graminearum [FGSG_03358]
Trichoderma virens [Trvi17878]
Trichoderma reesei [Trre32027]
Mycosphaerella graminicola [Mycgr90639]
Stagonospora nodorum [SNU02327]
Pyrenophora tritici-repentis [PTRG_01205]
Nectria haematococca [Neha70183]
Magnaporthe grisea [MGG_02903]
Fusarium verticillioides [FVEG13946]
Podospora anserina [Pa_6_9950]
Podospora anserina [Pa_7_930]
Trichoderma reesei [Trre109242]
Trichoderma virens [Trvi59372]
0.1 Subs/site
● 0.95/85
○ 0.80/60

4C



REFERENCES

Foster, P.G. (2004). Modeling compositional heterogeneity. Syst. Biol. **53,** 485-495.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL:~for assessing the confidence of phylogenetic tree selection. Bioinformatics **17,** 1246-1247.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. (Sunderland, Massachusetts: Sinauer Associates).

Supplemental Figure 5. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 3A.

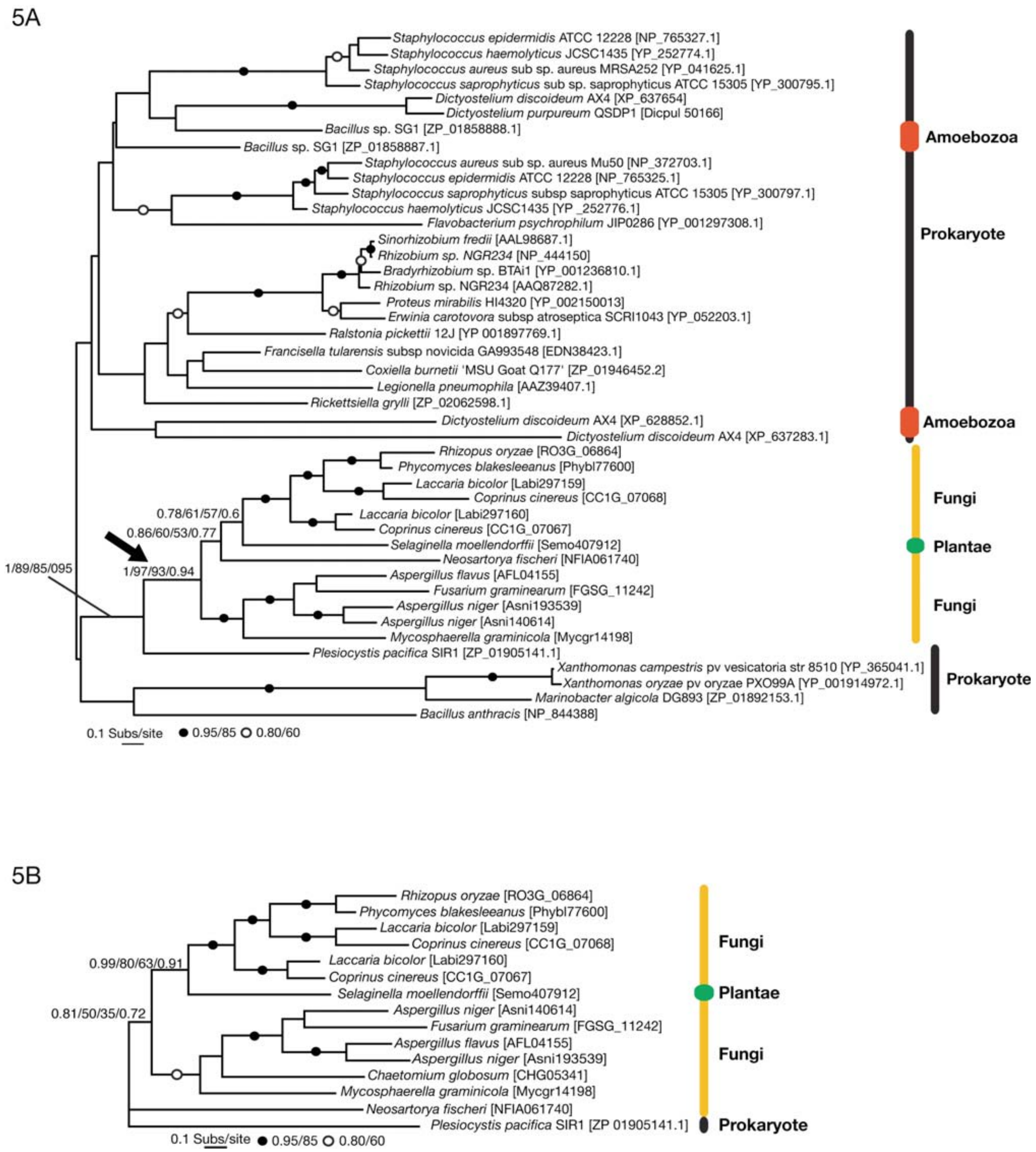
Figure 5. Phylogenetic analyses of the putative iucA / iucC protein encoding gene family, involved in siderophore biosynthesis, demonstrating a putative horizontal gene transfer event from the fungi to the lycophyte *Selaginella* genome (**5A**). This protein family, with the exception of the one plant genome and two *Dictyostelium* genomes, is restricted to a wide diversity of prokaryotes and the fungi, suggesting a prokaryote-to-fungi gene transfer. We also detected a putative homologue of this protein encoded by the plant *Selaginella moellendorffii* genome. This plant gene clustered strongly with the fungi (key node is marked with a black arrow) but with weak support for the plant gene grouping within the fungal phylogenetic group. Furthermore, the fungal tree topology showed a non-standard fungal branching order (Fitzpatrick et al., 2006; James et al., 2006) suggesting the support for the plant gene branching within the fungal phylogenetic cluster is weak. To attempt to resolve the branching position of the *Selaginella* gene within the fungi we performed a second phylogenetic analyses (**5B**) removing distantly related prokaryote sequences and adjusting the alignment character sampling. Phylogenetic analyses based upon this second alignment showed improved topology support values but still demonstrated a non-standard fungal branching order (Fitzpatrick et al., 2006; James et al., 2006). However, when the phylogenetic data is considered with the taxon distribution data, these analysis suggest the gene transfer of a prokaryote gene to a plant genome via a fungal genome.

The *Dictyostelium* genes branched separately from the plant/fungal clade with strong bootstrap support, branching among the prokaryotes, consistent with a separate origin of the *Dictyostelium* genes, possibly as a separate prokaryote-to-*Dictyostelium* HGT event as suggested by (Eichinger et al., 2005). Consequently, the presence of distantly related forms of this gene in *Dictyostelium* is not an important factor for the plant-fungi HGT. However, the alignment demonstrated that the *Dictyostelium* genes were highly divergent genes and formed long branches on the subsequent phylogenetic tree. The topology support values, although placed the *Dictyostelium* genes separate from the other eukaryotic genes with strong support, they did not demonstrate strong support for the monophyly of the *Dictyostelium* genes, indeed only the RAxML analysis recovered monophyly of the four *Dictyostelium* genes in the top scoring bootstrap consensus tree with 28% bootstrap support. Futhermore, the weak bootstrap support among this part of the tree topology makes it difficult to pinpoint a possible prokaryote donor lineage for the *Dictyostelium* genes.

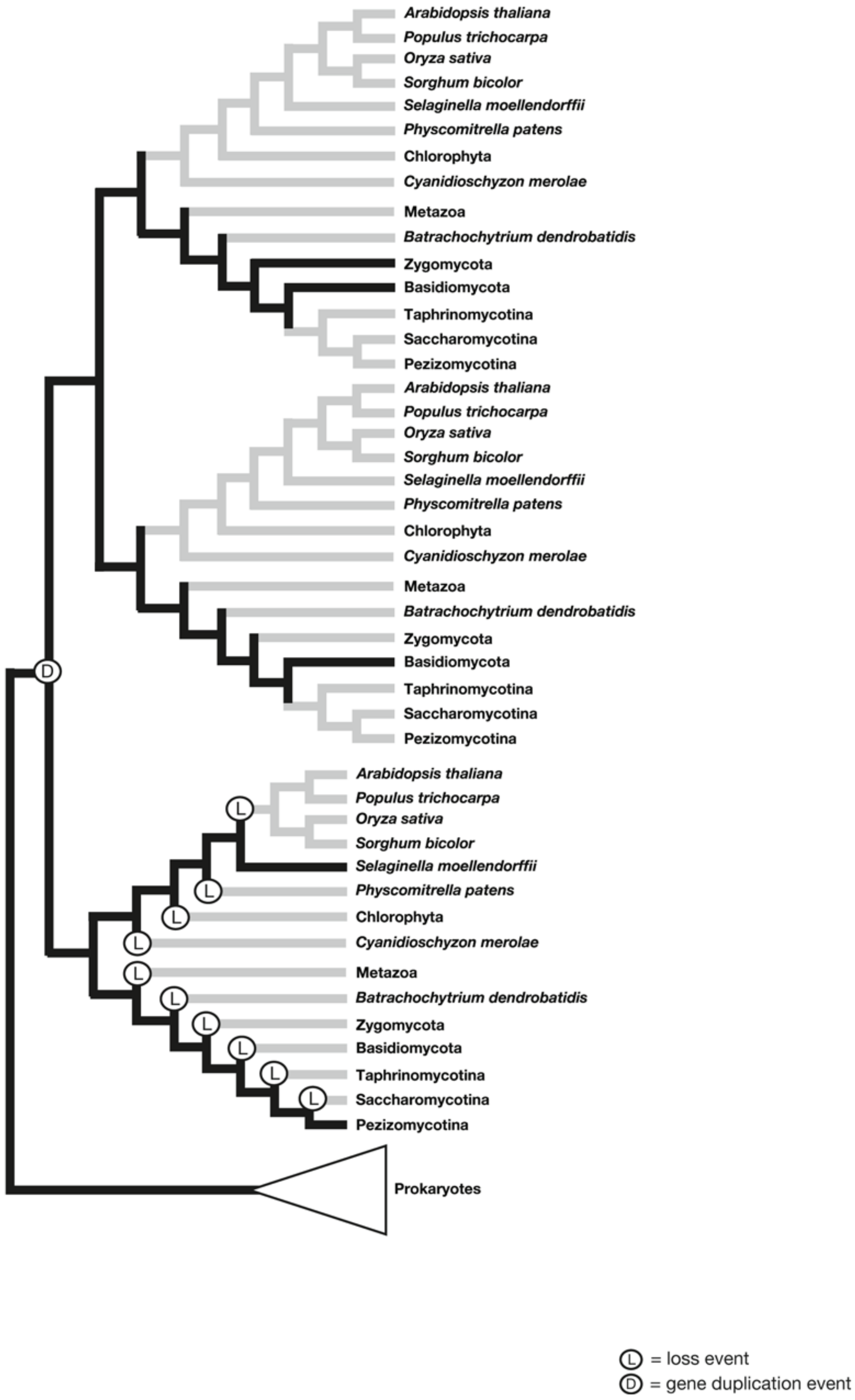
The phylogeny was calculated from an alignment of 44 sequences and 218 amino acid characters (**5A**) and an alignment of 15 sequences and 262 amino acid characters (**5B**). Modelgenerator (Keane et al., 2004) analysis demonstrated that a RtREV substitution matrix, Γ distribution (α = 45), and a proportion of invariant sites (I = 0.08) were the most appropriate parameters for the **5A** data set. While a RtREV substitution matrix, and a Γ distribution (α = 2.24), and a proportion of invariant sites (I = 0.1) were the most appropriate parameters for the **5B** data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML (Guindon and Gascuel, 2003), with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test further the topological result we also ran a MrBayes (Ronquist and Huelsenbeck, 2003) analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML+ RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication events and gene loss (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**5C**). These trees were based on an underlying eukaryotic species phylogeny. However, because the gene phylogeny showed a non-standard fungal branching order we assumed the gene family included an ancestry of two gene duplications generating three paralogue families, but only counted the duplication and loss relevant to the HGT hypothesis. Because there is uncertainty about the relative branching order of many eukaryotic groups we restricted the underlying eukaryotic species phylogeny to strongly supported branching relationships among the Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006). As such this analyses underestimates the number of gene duplication and gene loss events required for the alternative hypothesis of hidden paralogy. Only gene duplication (D) and gene loss (L) events required to invoke the hidden paralogy are marked (all other loss events are not scored). For hidden paralogy to explain the branching of the plant within the fungal clade, given the taxon sampling available for this analysis, a minimum of 10 independent gene loss events and 1 gene duplication events are required. This compares to the scenario of a single fungi-to-plant HGT event.


5A
Staphylococcus epidermidis ATCC 12228 [NP_765327.1]
Staphylococcus haemolyticus JCSC1435 [YP_252774.1]
Staphylococcus aureus sub sp. aureus MRSA252 [YP_041625.1]
Staphylococcus saprophyticus sub sp. saprophyticus ATCC 15305 [YP_300795.1]
Dictyostelium discoideum AX4 [XP_637654]
Dictyostelium purpureum QSDP1 [Dicpul 50166]
Bacillus sp. SG1 [ZP_01858888.1]
Bacillus sp. SG1 [ZP_01858887.1]
Staphylococcus aureus sub sp. aureus Mu50 [NP_372703.1]
Staphylococcus epidermidis ATCC 12228 [NP_765325.1]
Staphylococcus saprophyticus subsp saprophyticus ATCC 15305 [YP_300797.1]
Staphylococcus haemolyticus JCSC1435 [YP _252776.1]
Flavobacterium psychrophilum JIP0286 [YP_001297308.1]
Sinorhizobium fredii [AAL98687.1]
Rhizobium sp. NGR234 [NP_444150]
Bradyrhizobium sp. BTAi1 [YP_001236810.1]
Rhizobium sp. NGR234 [AAQ87282.1]
Proteus mirabilis HI4320 [YP_002150013]
Erwinia carotovora subsp atroseptica SCRI1043 [YP_052203.1]
Ralstonia pickettii 12J [YP 001897769.1]
Francisella tularensis subsp novicida GA993548 [EDN38423.1]
Coxiella burnetii 'MSU Goat Q177' [ZP_01946452.2]
Legionella pneumophila [AAZ39407.1]
Rickettsiella grylli [ZP_02062598.1]
Dictyostelium discoideum AX4 [XP_628852.1]
Dictyostelium discoideum AX4 [XP_637283.1]
Rhizopus oryzae [RO3G_06864]
Phycomyces blakesleeanus [Phybl77600]
Laccaria bicolor [Labi297159]
Coprinus cinereus [CC1G_07068]
0.78/61/57/0.6
Laccaria bicolor [Labi297160]
0.86/60/53/0.77
Coprinus cinereus [CC1G_07067]
Selaginella moellendorffii [Semo407912]
Neosartorya fischeri [NFIA061740]
1/89/85/095
1/97/93/0.94
Aspergillus flavus [AFL04155]
Fusarium graminearum [FGSG_11242]
Aspergillus niger [Asni193539]
Aspergillus niger [Asni140614]
Mycosphaerella graminicola [Mycgr14198]
Plesiocystis pacifica SIR1 [ZP_01905141.1]
Xanthomonas campestris pv vesicatoria str 8510 [YP_365041.1]
Xanthomonas oryzae pv oryzae PXO99A [YP_001914972.1]
Marinobacter algicola DG893 [ZP_01892153.1]
Bacillus anthracis [NP_844388]
0.1 Subs/site
● 0.95/85 ○ 0.80/60
Amoebozoa
Prokaryote
Amoebozoa
Fungi
Plantae
Fungi
Prokaryote
5B
Rhizopus oryzae [RO3G_06864]
Phycomyces blakesleeanus [Phybl77600]
Laccaria bicolor [Labi297159]
Coprinus cinereus [CC1G_07068]
0.99/80/63/0.91
Laccaria bicolor [Labi297160]
Coprinus cinereus [CC1G_07067]
Selaginella moellendorffii [Semo407912]
0.81/50/35/0.72
Aspergillus niger [Asni140614]
Fusarium graminearum [FGSG_11242]
Aspergillus flavus [AFL04155]
Aspergillus niger [Asni193539]
Chaetomium globosum [CHG05341]
Mycosphaerella graminicola [Mycgr14198]
Neosartorya fischeri [NFIA061740]
Plesiocystis pacifica SIR1 [ZP 01905141.1]
0.1 Subs/site
● 0.95/85 ○ 0.80/60
Fungi
Plantae
Fungi
Prokaryote

5C


Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
Chlorophyta
Cyanidioschyzon merolae
Metazoa
Batrachochytrium dendrobatidis
Zygomycota
Basidiomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
Physcomitrella patens
Chlorophyta
Cyanidioschyzon merolae
Metazoa
Batrachochytrium dendrobatidis
Zygomycota
Basidiomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
D
L
Arabidopsis thaliana
Populus trichocarpa
Oryza sativa
Sorghum bicolor
Selaginella moellendorffii
L
Physcomitrella patens
L
Chlorophyta
L
Cyanidioschyzon merolae
L
Metazoa
L
Batrachochytrium dendrobatidis
L
Zygomycota
L
Basidiomycota
L
Taphrinomycotina
L
Saccharomycotina
Pezizomycotina
Prokaryotes
Ⓛ = loss event
Ⓓ = gene duplication event

REFERENCES

Eichinger, L., Pachebat, J.A., Glockner, G., Rajandream, M.A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B.A., Rivero, F., Bankier, A.T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Madan Babu, M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M.A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E.C., Chisholm, R.L., Gibbs, R., Loomis, W.F., Platzer, M., Kay, R.R., Williams, J., Dear, P.H., Noegel, A.A., Barrell, B., and Kuspa, A. (2005). The genome of the social amoeba *Dictyostelium discoideum*. Nature **435,** 43-57.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., and Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol. Biol. **6**.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Supplemental Figure 6. Description of phylogenetic analysis, including the models used, sequence database accession numbers, alternative topology tests and assessment of hidden paralogy for the phylogeny shown on Figure 3B.

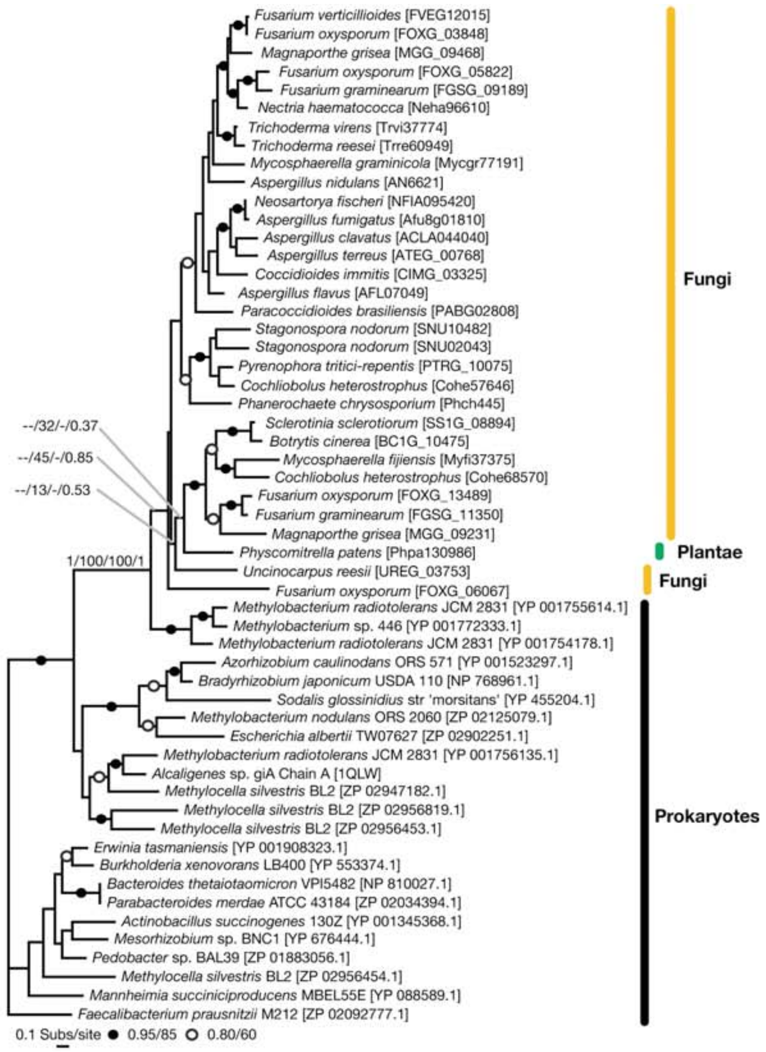
Figure 6. Phylogeny of an unknown / conserved hypothetical protein family demonstrating a candidate fungal-to-plant gene transfer (**6A**). This protein family, with the exception of the one plant genome, is restricted to a wide diversity of prokaryotes and the ascomycete fungi suggesting a prokaryote-to-fungi gene transfer. We also detected a putative homologue of this protein encoded by the *Physcomitrella patens* genome. This plant gene clustered with the fungi with weak support for the plant gene grouping within the fungi phylogenetic group. To attempt to resolve the branching position of the *Physcomitrella* gene within the fungi we performed a second phylogenetic analyses removing distantly related prokaryote sequences and adjusting the alignment character sampling (**6B**). Phylogenetic analyses based upon this second alignment strongly supported the grouping of the plant sequence with the fungi but did not strongly support the placement of the *Physcomitrella* gene within the fungal clade. However, the taxonomic distribution of this gene family demonstrates a wide paralogous distribution in the fungi and prokaryotes and a very narrow distribution in all other eukaryotes sampled with a single plant sequence detected. Taken together this suggests a fungus-to-plant gene transfer event.

The phylogeny was calculated from an alignment of 55 sequences and 174 amino acid characters (**6A**) and an alignment of 34 sequences and 247 amino acid characters (**6B**). Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix, Γ distribution ($\alpha = 2.03$), and a proportion of invariant sites ($I = 0.03$) model of site rate heterogeneity were the most appropriate parameters for the **6A** data set. While a WAG substitution matrix, Γ distribution ($\alpha = 65$), and a proportion of invariant sites ($I = 0.068$) model of site rate heterogeneity were the most appropriate parameters for the **6B** data set. The phylogenetic trees shown were calculated using the fast maximum likelihood program phyML (Guindon and Gascuel, 2003), with 1000 bootstrap replicates and SH analyses of each node (as described in the main text of the paper). To test further the topological result we also ran a MrBayes (Ronquist and Huelsenbeck, 2003) analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (phyML + RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key). For key nodes the actual support values are shown in the order Bayesian Posterior Probability / 1000 phyML bootstraps / 100 RAxML bootstraps / phyML node-by-node SH test.

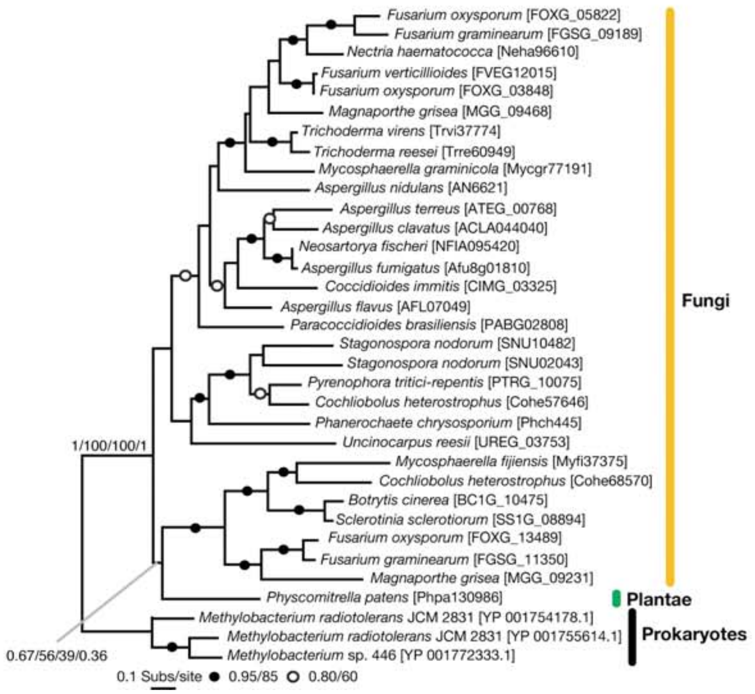
The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

To compare the HGT scenarios with an alternative hypothesis of gene duplication events and gene loss (hidden paralogy) we drew a cladogram demonstrating gene duplication and gene loss events that would be necessary to generate the phylogenetic results shown without a HGT event (**6C**). These trees were based on an underlying eukaryotic species phylogeny. Because there is uncertainty about the relative branching order of many eukaryotic groups we restricted the underlying eukaryotic species phylogeny to strongly supported branching relationships among the Plantae, the Fungi, and their sister group the metazoa (Rodriguez-Ezpeleta et al., 2005; James et al., 2006). As such this analyses underestimates the number of gene duplication and gene loss events required for the alternative hypothesis of hidden paralogy. Only duplication (D) and loss (L) events required to invoke the hidden paralogy are marked (all other loss events are not scored). For hidden paralogy to explain the branching of the plant within the fungal clade, given the taxon sampling available for this analysis, a minimum of 16 independent gene loss events are required. This compares to the scenario of a single fungi-to-plant HGT event.

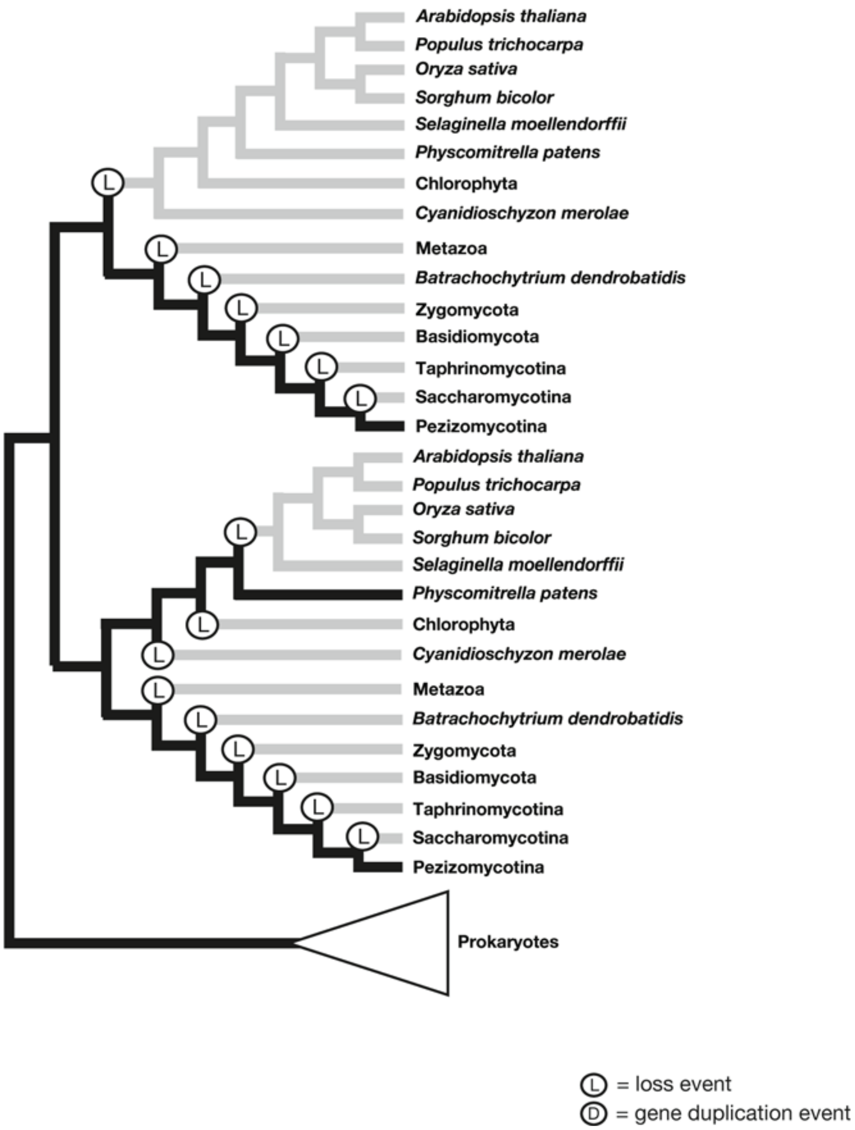
6A


Fusarium verticillioides [FVEG12015]
Fusarium oxysporum [FOXG_03848]
Magnaporthe grisea [MGG_09468]
Fusarium oxysporum [FOXG_05822]
Fusarium graminearum [FGSG_09189]
Nectria haematococca [Neha96610]
Trichoderma virens [Trvi37774]
Trichoderma reesei [Trre60949]
Mycosphaerella graminicola [Mycgr77191]
Aspergillus nidulans [AN6621]
Neosartorya fischeri [NFIA095420]
Aspergillus fumigatus [Afu8g01810]
Aspergillus clavatus [ACLA044040]
Aspergillus terreus [ATEG_00768]
Coccidioides immitis [CIMG_03325]
Aspergillus flavus [AFL07049]
Paracoccidioides brasiliensis [PABG02808]
Stagonospora nodorum [SNU10482]
Stagonospora nodorum [SNU02043]
Pyrenophora tritici-repentis [PTRG_10075]
Cochliobolus heterostrophus [Cohe57646]
Phanerochaete chrysosporium [Phch445]
Sclerotinia sclerotiorum [SS1G_08894]
Botrytis cinerea [BC1G_10475]
Mycosphaerella fijiensis [Myfi37375]
Cochliobolus heterostrophus [Cohe68570]
Fusarium oxysporum [FOXG_13489]
Fusarium graminearum [FGSG_11350]
Magnaporthe grisea [MGG_09231]
Physcomitrella patens [Phpa130986]
Uncinocarpus reesii [UREG_03753]
Fusarium oxysporum [FOXG_06067]
Methylobacterium radiotolerans JCM 2831 [YP 001755614.1]
Methylobacterium sp. 446 [YP 001772333.1]
Methylobacterium radiotolerans JCM 2831 [YP 001754178.1]
Azorhizobium caulinodans ORS 571 [YP 001523297.1]
Bradyrhizobium japonicum USDA 110 [NP 768961.1]
Sodalis glossinidius str 'morsitans' [YP 455204.1]
Methylobacterium nodulans ORS 2060 [ZP 02125079.1]
Escherichia albertii TW07627 [ZP 02902251.1]
Methylobacterium radiotolerans JCM 2831 [YP 001756135.1]
Alcaligenes sp. giA Chain A [1QLW]
Methylocella silvestris BL2 [ZP 02947182.1]
Methylocella silvestris BL2 [ZP 02956819.1]
Methylocella silvestris BL2 [ZP 02956453.1]
Erwinia tasmaniensis [YP 001908323.1]
Burkholderia xenovorans LB400 [YP 553374.1]
Bacteroides thetaiotaomicron VPI5482 [NP 810027.1]
Parabacteroides merdae ATCC 43184 [ZP 02034394.1]
Actinobacillus succinogenes 130Z [YP 001345368.1]
Mesorhizobium sp. BNC1 [YP 676444.1]
Pedobacter sp. BAL39 [ZP 01883056.1]
Methylocella silvestris BL2 [ZP 02956454.1]
Mannheimia succiniciproducens MBEL55E [YP 088589.1]
Faecalibacterium prausnitzii M212 [ZP 02092777.1]
--/32/-/0.37
--/45/-/0.85
--/13/-/0.53
1/100/100/1
Fungi
Plantae
Fungi
Prokaryotes
0.1 Subs/site ● 0.95/85 ○ 0.80/60

6B


Fusarium oxysporum [FOXG_05822]
Fusarium graminearum [FGSG_09189]
Nectria haematococca [Neha96610]
Fusarium verticillioides [FVEG12015]
Fusarium oxysporum [FOXG_03848]
Magnaporthe grisea [MGG_09468]
Trichoderma virens [Trvi37774]
Trichoderma reesei [Trre60949]
Mycosphaerella graminicola [Mycgr77191]
Aspergillus nidulans [AN6621]
Aspergillus terreus [ATEG_00768]
Aspergillus clavatus [ACLA044040]
Neosartorya fischeri [NFIA095420]
Aspergillus fumigatus [Afu8g01810]
Coccidioides immitis [CIMG_03325]
Aspergillus flavus [AFL07049]
Paracoccidioides brasiliensis [PABG02808]
Stagonospora nodorum [SNU10482]
Stagonospora nodorum [SNU02043]
Pyrenophora tritici-repentis [PTRG_10075]
Cochliobolus heterostrophus [Cohe57646]
Phanerochaete chrysosporium [Phch445]
Uncinocarpus reesii [UREG_03753]
Mycosphaerella fijiensis [Myfi37375]
Cochliobolus heterostrophus [Cohe68570]
Botrytis cinerea [BC1G_10475]
Sclerotinia sclerotiorum [SS1G_08894]
Fusarium oxysporum [FOXG_13489]
Fusarium graminearum [FGSG_11350]
Magnaporthe grisea [MGG_09231]
Physcomitrella patens [Phpa130986]
Methylobacterium radiotolerans JCM 2831 [YP 001754178.1]
Methylobacterium radiotolerans JCM 2831 [YP 001755614.1]
Methylobacterium sp. 446 [YP 001772333.1]
1/100/100/1
0.67/56/39/0.36
Fungi
Plantae
Prokaryotes
0.1 Subs/site ● 0.95/85 ○ 0.80/60

6C



REFERENCES

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-

Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443,** 818-822.

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15,** 1325-1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

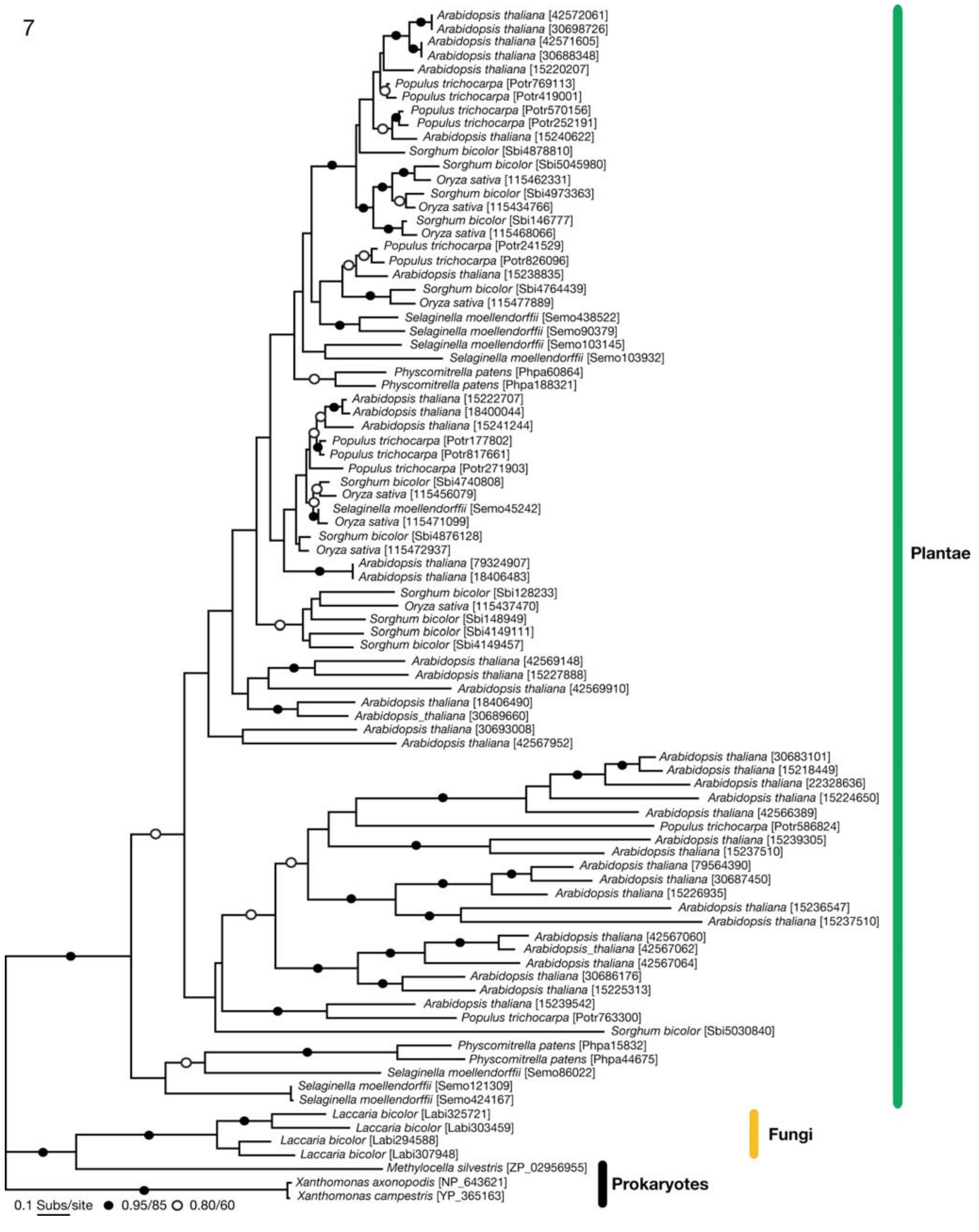
Supplemental Figure 7. Description of phylogenetic analysis, including the models used and database sequence accession numbers for the phylogenetic tree shown on 4A.

Figure 7. Phylogeny of the putative carboxy-terminal peptidase-like protein encoding gene family, demonstrating a candidate plant-to-fungal gene transfer. This protein family is mainly restricted to the land plants with each land plant genome surveyed encoding several paralogous copies of this gene family. However, we detected multiple putative homologues of this gene in the basidiomycete *Laccaria bicolor* and two divergent prokaryote lineages (*Xanthomonas* and *Methylocella*). The taxonomic distribution of this gene family demonstrates a wide paralogous distribution in the plants and very narrow distribution in the fungi and prokaryotes sampled. Taken together, this suggests a plant-to-fungus gene transfer event either involving a prokaryote intermediate, or additional transfer to prokaryotes.

The phylogeny was calculated from an alignment of 87 sequences and 210 amino acid characters. Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix, and a Γ distribution ($\alpha = 1.606$) model of site rate heterogeneity were the most appropriate parameters for this dataset. The phylogeny shown was calculated using the fast maximum likelihood program PhyML (Guindon and Gascuel, 2003), with 1000 bootstrap replicates and SH analyses (Anisimova and Gascuel, 2006) of each node (as described in the main text of the paper). To further test the topological result we also ran a MrBayes (Ronquist and Huelsenbeck, 2003) analysis and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (PhyML + RAxML).

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

7


Arabidopsis thaliana [42572061]
Arabidopsis thaliana [30698726]
Arabidopsis thaliana [42571605]
Arabidopsis thaliana [30688348]
Arabidopsis thaliana [15220207]
Populus trichocarpa [Potr769113]
Populus trichocarpa [Potr419001]
Populus trichocarpa [Potr570156]
Populus trichocarpa [Potr252191]
Arabidopsis thaliana [15240622]
Sorghum bicolor [Sbi4878810]
Sorghum bicolor [Sbi5045980]
Oryza sativa [115462331]
Sorghum bicolor [Sbi4973363]
Oryza sativa [115434766]
Sorghum bicolor [Sbi146777]
Oryza sativa [115468066]
Populus trichocarpa [Potr241529]
Populus trichocarpa [Potr826096]
Arabidopsis thaliana [15238835]
Sorghum bicolor [Sbi4764439]
Oryza sativa [115477889]
Selaginella moellendorffii [Semo438522]
Selaginella moellendorffii [Semo90379]
Selaginella moellendorffii [Semo103145]
Selaginella moellendorffii [Semo103932]
Physcomitrella patens [Phpa60864]
Physcomitrella patens [Phpa188321]
Arabidopsis thaliana [15222707]
Arabidopsis thaliana [18400044]
Arabidopsis thaliana [15241244]
Populus trichocarpa [Potr177802]
Populus trichocarpa [Potr817661]
Populus trichocarpa [Potr271903]
Sorghum bicolor [Sbi4740808]
Oryza sativa [115456079]
Selaginella moellendorffii [Semo45242]
Oryza sativa [115471099]
Sorghum bicolor [Sbi4876128]
Oryza sativa [115472937]
Arabidopsis thaliana [79324907]
Arabidopsis thaliana [18406483]
Sorghum bicolor [Sbi128233]
Oryza sativa [115437470]
Sorghum bicolor [Sbi148949]
Sorghum bicolor [Sbi4149111]
Sorghum bicolor [Sbi4149457]
Arabidopsis thaliana [42569148]
Arabidopsis thaliana [15227888]
Arabidopsis thaliana [42569910]
Arabidopsis thaliana [18406490]
Arabidopsis_thaliana [30689660]
Arabidopsis thaliana [30693008]
Arabidopsis thaliana [42567952]
Arabidopsis thaliana [30683101]
Arabidopsis thaliana [15218449]
Arabidopsis thaliana [22328636]
Arabidopsis thaliana [15224650]
Arabidopsis thaliana [42566389]
Populus trichocarpa [Potr586824]
Arabidopsis thaliana [15239305]
Arabidopsis thaliana [15237510]
Arabidopsis thaliana [79564390]
Arabidopsis thaliana [30687450]
Arabidopsis thaliana [15226935]
Arabidopsis thaliana [15236547]
Arabidopsis thaliana [15237510]
Arabidopsis thaliana [42567060]
Arabidopsis_thaliana [42567062]
Arabidopsis thaliana [42567064]
Arabidopsis thaliana [30686176]
Arabidopsis thaliana [15225313]
Arabidopsis thaliana [15239542]
Populus trichocarpa [Potr763300]
Sorghum bicolor [Sbi5030840]
Physcomitrella patens [Phpa15832]
Physcomitrella patens [Phpa44675]
Selaginella moellendorffii [Semo86022]
Selaginella moellendorffii [Semo121309]
Selaginella moellendorffii [Semo424167]
Laccaria bicolor [Labi325721]
Laccaria bicolor [Labi303459]
Laccaria bicolor [Labi294588]
Laccaria bicolor [Labi307948]
Methylocella silvestris [ZP_02956955]
Xanthomonas axonopodis [NP_643621]
Xanthomonas campestris [YP_365163]
Plantae
Fungi
Prokaryotes
0.1 Subs/site
● 0.95/85
○ 0.80/60

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. **55,** 539-552.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

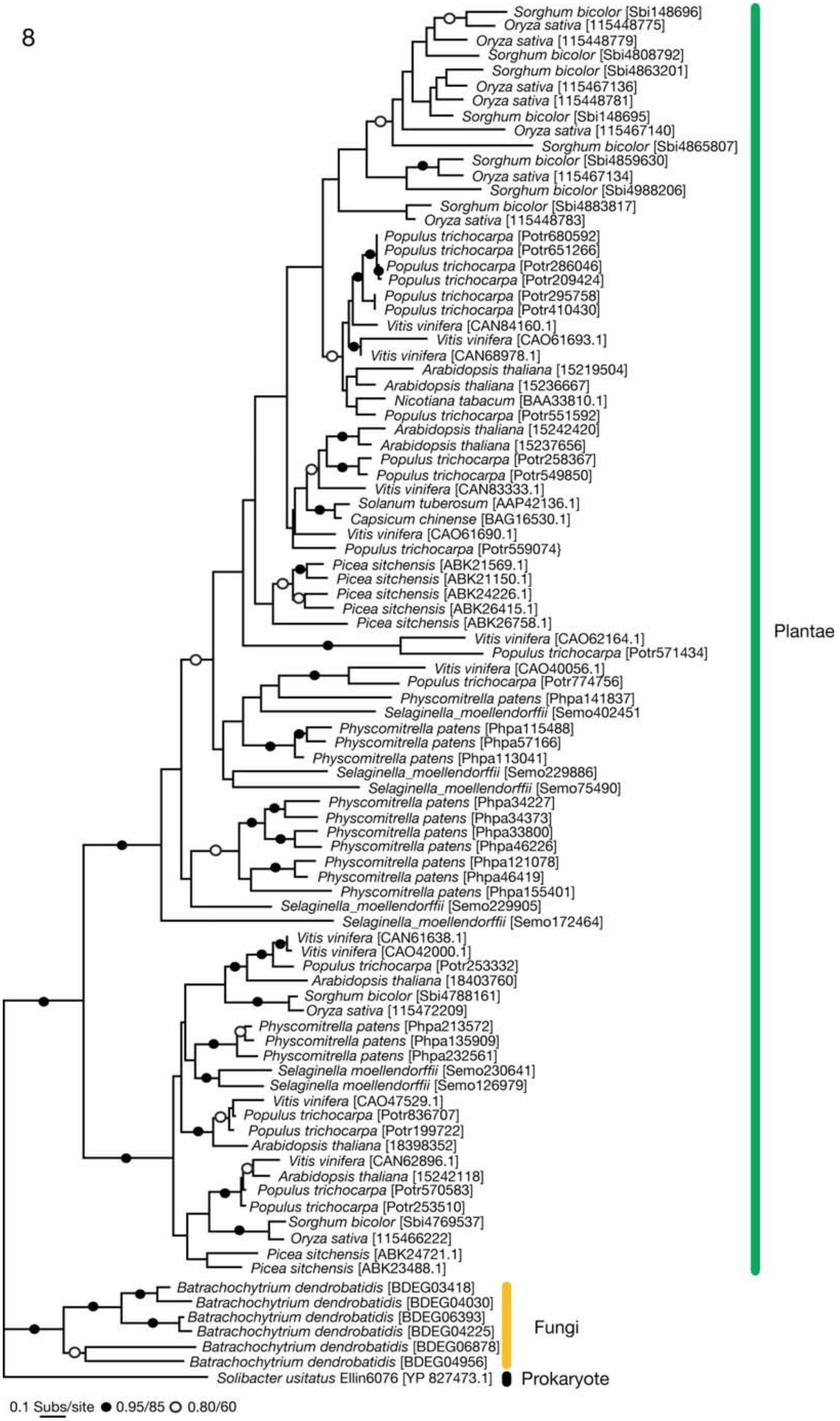
Supplemental Figure 8. Description of phylogenetic analysis, including the models used and database sequence accession numbers for the phylogenetic tree shown on 4B.

Figure 8. Phylogeny of the putative phosphate-responsive 1 protein encoding gene family, demonstrating a candidate plant-to-fungal gene transfer. This protein family is mainly restricted to the land plant with each plant genome surveyed encoding several paralogous copies of this gene family. However, we detected multiple putative homolgues of this gene in the chytrid *Batrachochytrium dendrobatidis* and one prokaryote sequence (*Solibacter usitatus*). The taxonomic distribution of this gene family demonstrates a wide paralogous distribution in the plants and very narrow distribution in the fungi and prokaryotes sampled. Taken together this suggests a plant-to-fungus gene transfer event either involving prokaryote intermediate or additional transfers to a prokaryote genome.

The phylogeny was calculated from an alignment of 93 sequences and 198 amino acid characters. Modelgenerator (Keane et al., 2004) analysis demonstrated that a WAG substitution matrix, and a Γ distribution (α = 1.381) model of site rate heterogeneity were the most appropriate parameters for this dataset. The phylogeny shown was calculated using the fast maximum likelihood program PhyML (Guindon and Gascuel, 2003), with 1000 bootstrap replicates and SH analyses (Anisimova and Gascuel, 2006) of each node (as described in the main text of the paper). To further test the topological result we also ran a MrBayes (Ronquist and Huelsenbeck, 2003) analysis and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (PhyML + RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key).

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview (Galtier et al., 1996) alignment files.

8


Sorghum bicolor [Sbi148696]
Oryza sativa [115448775]
Oryza sativa [115448779]
Sorghum bicolor [Sbi4808792]
Sorghum bicolor [Sbi4863201]
Oryza sativa [115467136]
Oryza sativa [115448781]
Sorghum bicolor [Sbi148695]
Oryza sativa [115467140]
Sorghum bicolor [Sbi4865807]
Sorghum bicolor [Sbi4859630]
Oryza sativa [115467134]
Sorghum bicolor [Sbi4988206]
Sorghum bicolor [Sbi4883817]
Oryza sativa [115448783]
Populus trichocarpa [Potr680592]
Populus trichocarpa [Potr651266]
Populus trichocarpa [Potr286046]
Populus trichocarpa [Potr209424]
Populus trichocarpa [Potr295758]
Populus trichocarpa [Potr410430]
Vitis vinifera [CAN84160.1]
Vitis vinifera [CAO61693.1]
Vitis vinifera [CAN68978.1]
Arabidopsis thaliana [15219504]
Arabidopsis thaliana [15236667]
Nicotiana tabacum [BAA33810.1]
Populus trichocarpa [Potr551592]
Arabidopsis thaliana [15242420]
Arabidopsis thaliana [15237656]
Populus trichocarpa [Potr258367]
Populus trichocarpa [Potr549850]
Vitis vinifera [CAN83333.1]
Solanum tuberosum [AAP42136.1]
Capsicum chinense [BAG16530.1]
Vitis vinifera [CAO61690.1]
Populus trichocarpa [Potr559074}
Picea sitchensis [ABK21569.1]
Picea sitchensis [ABK21150.1]
Picea sitchensis [ABK24226.1]
Picea sitchensis [ABK26415.1]
Picea sitchensis [ABK26758.1]
Vitis vinifera [CAO62164.1]
Populus trichocarpa [Potr571434]
Vitis vinifera [CAO40056.1]
Populus trichocarpa [Potr774756]
Physcomitrella patens [Phpa141837]
Selaginella_moellendorffii [Semo402451
Physcomitrella patens [Phpa115488]
Physcomitrella patens [Phpa57166]
Physcomitrella patens [Phpa113041]
Selaginella_moellendorffii [Semo229886]
Selaginella_moellendorffii [Semo75490]
Physcomitrella patens [Phpa34227]
Physcomitrella patens [Phpa34373]
Physcomitrella patens [Phpa33800]
Physcomitrella patens [Phpa46226]
Physcomitrella patens [Phpa121078]
Physcomitrella patens [Phpa46419]
Physcomitrella patens [Phpa155401]
Selaginella_moellendorffii [Semo229905]
Selaginella_moellendorffii [Semo172464]
Vitis vinifera [CAN61638.1]
Vitis vinifera [CAO42000.1]
Populus trichocarpa [Potr253332]
Arabidopsis thaliana [18403760]
Sorghum bicolor [Sbi4788161]
Oryza sativa [115472209]
Physcomitrella patens [Phpa213572]
Physcomitrella patens [Phpa135909]
Physcomitrella patens [Phpa232561]
Selaginella moellendorffii [Semo230641]
Selaginella moellendorffii [Semo126979]
Vitis vinifera [CAO47529.1]
Populus trichocarpa [Potr836707]
Populus trichocarpa [Potr199722]
Arabidopsis thaliana [18398352]
Vitis vinifera [CAN62896.1]
Arabidopsis thaliana [15242118]
Populus trichocarpa [Potr570583]
Populus trichocarpa [Potr253510]
Sorghum bicolor [Sbi4769537]
Oryza sativa [115466222]
Picea sitchensis [ABK24721.1]
Picea sitchensis [ABK23488.1]
Batrachochytrium dendrobatidis [BDEG03418]
Batrachochytrium dendrobatidis [BDEG04030]
Batrachochytrium dendrobatidis [BDEG06393]
Batrachochytrium dendrobatidis [BDEG04225]
Batrachochytrium dendrobatidis [BDEG06878]
Batrachochytrium dendrobatidis [BDEG04956]
Solibacter usitatus Ellin6076 [YP 827473.1]
Plantae
Fungi
Prokaryote
0.1 Subs/site ● 0.95/85 ○ 0.80/60

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. **55,** 539-552.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

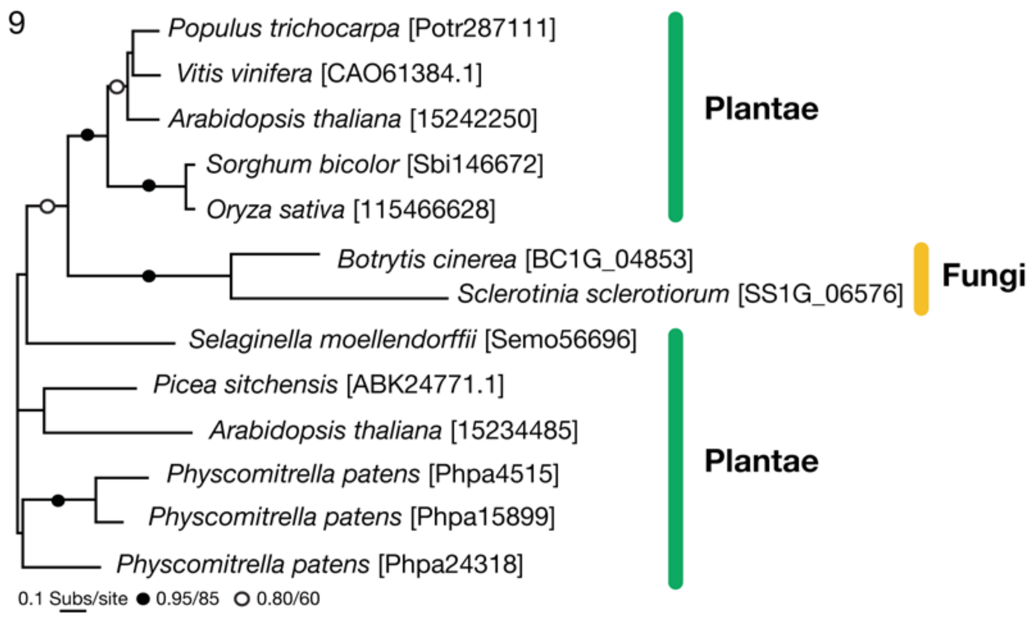
Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Supplemental Figure 9. Description of phylogenetic analysis, including the models used and database sequence accession numbers for the phylogenetic tree shown on 4C.

Figure 9. Phylogeny of unkown / conserved hypothetical protein family with similarity to zinc finger (C2H2 type) proteins, demonstrating a candidate plant-to-fungus gene transfer. This gene family was restricted to the land plants. However, we found that the closely related ascomycete fungi (*Sclerotinia sclerotiorum* and *Botrytis cinerea*) also posses putative homologues of this protein family. This taxon distribution suggests a plant-to-fungi gene transfer

The phylogeny was calculated from an alignment of 13 sequences and 222 amino acid characters. Modelgenerator (Keane et al., 2004) analysis demonstrated that a JTT substitution matrix, Γ distribution ($\alpha = 1.95$), and a proportion of invariant sites ($I = 0.19$) model of site rate heterogeneity were the most appropriate parameters for this dataset. The phylogeny shown was calculated using the fast maximum likelihood program PhyML (Guindon and Gascuel, 2003), with 1000 bootstrap replicates and SH analyses (Anisimova and Gascuel, 2006) of each node (as described in the main text of the paper). To further test the topological result we also ran a MrBayes (Ronquist and Huelsenbeck, 2003) analyses and 100 RAxML (Stamatakis, 2006) bootstrap replicates (as described in the main text of the paper). The key for each tree shows the short hand description of topology support values in the order Bayesian posterior probability / % bootstrap support (PhyML + RAxML). Shaded discs represent nodes with 'robust' topology support values, while rings demonstrate nodes with 'moderate' topology support values (actual cut off values are given on the key).

The species are labelled with an identifier code in square brackets, relating to the source of sequence data. These include GenBank protein accession codes and GI numbers, Broad Institute gene identifiers (in some cases curtailed for program compatibility reasons), and DOE JGI gene identifiers with a 4 letter species codes that we have added. The source of all the genome sequences used in the pipeline analysis is listed in Supplemental Table 1. All additional non-genome project sequences are from GenBank. As genome sequence identifiers are continually updated we have provided additional supporting material with all the sequences used as Seaview alignment files (Galtier et al., 1996).


9
Populus trichocarpa [Potr287111]
Vitis vinifera [CAO61384.1]
Arabidopsis thaliana [15242250]
Sorghum bicolor [Sbi146672]
Oryza sativa [115466628]
Plantae
Botrytis cinerea [BC1G_04853]
Sclerotinia sclerotiorum [SS1G_06576]
Fungi
Selaginella moellendorffii [Semo56696]
Picea sitchensis [ABK24771.1]
Arabidopsis thaliana [15234485]
Physcomitrella patens [Phpa4515]
Physcomitrella patens [Phpa15899]
Physcomitrella patens [Phpa24318]
Plantae
0.1 Subs/site ● 0.95/85 ○ 0.80/60

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. **55,** 539-552.

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12,** 543-548.

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

Keane, T.M., Creevey, C.J., Naughton, T.J., Pentony, M.M., Naughton, T.J., and Mcinerney, J.O. (2004). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6,** 29.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19,** 1572-1574.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688-2690.

Supplemental Table 1. Genomes used for HGT identification pipeline

Search seed genomes (Source)
Arabidopsis thaliana (http://www.ncbi.nlm.nih.gov/)
Oryza sativa (http://www.ncbi.nlm.nih.gov/)
Populus trichocarpa (http://genome.jgi-psf.org/)
Selaginella moellendorffii (http://genome.jgi-psf.org/)
Sorghum bicolour (http://genome.jgi-psf.org/)
*Physcomitrella paten*s (http://genome.jgi-psf.org/)

For every plant search seed gene that had a higher BLAST search similarity to a fungal gene a phylogenetic analyses was conducted. Genomes compared:

Fungi (Source)	Other Eukaryotes (Source)	Prokaryotes (Source = http://www.ncbi.nlm.nih.gov/)
Aspergillus clavatus (http://www.broad.mit.edu/)	*Aureococcus anophagefferens* (http://genome.jgi-psf.org/)	*Aeromonas salmonicida*
Aspergillus flavus (http://www.broad.mit.edu/)	*Caenorhabditis elegans* (http://www.ncbi.nlm.nih.gov/)	*Aeropyrum pernix*
Aspergillus fumigatus (http://www.broad.mit.edu/)	*Chlamydomonas reinhardtii* (http://genome.jgi-psf.org/)	*Agrobacterium tumefaciens*
Aspergillus nidulans (http://www.broad.mit.edu/)	*Ciona intestinalis* (http://genome.jgi-psf.org/)	*Alteromonadales bacterium TW-7*
Aspergillus niger (http://www.broad.mit.edu/)	*Cryptosporidium parvum* (http://www.ncbi.nlm.nih.gov/)	*Aquifex aeolicus*
Aspergillus oryzae (http://www.broad.mit.edu/)	*Cyanidioschyzon merolae* (http://merolae.biol.s.u-tokyo.ac.jp/)	*Archaeoglobus fulgidus*
Aspergillus terreus (http://www.broad.mit.edu/)	*Dictyostelium discoideum* (http://www.ncbi.nlm.nih.gov/)	*Azoarcus sp. BH72*
Batrachochytrium dendrobatidis (http://www.broad.mit.edu/)	*Drosophila melanogaster* (http://www.ncbi.nlm.nih.gov/)	*Aster yellows witches-broom phytoplasma AYWB*
Botrytis cinerea (http://www.broad.mit.edu/)	*Emiliania huxleyi* (http://genome.jgi-psf.org/)	*Bacillus anthracis*
Candida albicans SC5314 (http://www.broad.mit.edu/)	*Entamoeba histolytica* (http://www.tigr.org/db.shtml)	*Bacillus subtilis*
Chaetomium globosum (http://www.broad.mit.edu/)	*Homo sapiens*(http://www.ncbi.nlm.nih.gov/)	*Bacteroides fragilis*
Coccidioides immitis (http://www.broad.mit.edu/)	*Hyaloperonospora parasitica* (http://genome.wustl.edu)	*Bartonella henselae*
Cochliobolus heterostrophus (http://genome.jgi-psf.org/)	*Giardia lamblia* (http://www.tigr.org/db.shtml)	*Bifidobacterium longum*
Coprinus cinereus (http://www.broad.mit.edu/)	*Leishmania major* (http://www.ncbi.nlm.nih.gov/)	*Blastopirellula marina*
Cryptococcus neoformans (http://www.broad.mit.edu/)	*Lottia gigantea* (http://genome.jgi-psf.org/)	*Bordetella pertussis*
Encephalitozoon cuniculi (http://www.genoscope.cns.fr/spip/Projects.html)	*Micromonas pusilla* (http://genome.jgi-psf.org/)	*Borrelia burgdorferi*
Fusarium graminearum (http://www.broad.mit.edu/)	*Micromonas strain RCC299* (http://genome.jgi-psf.org/)	*Bradyrhizobium japonicum*
Fusarium oxysporum (http://www.broad.mit.edu/)	*Monosiga brevicollis* (http://genome.jgi-psf.org/)	*Buchnera aphidicola*
Fusarium verticillioides (http://www.broad.mit.edu/)		*Burkholderia mallei ATCC 23344*
		Campylobacter jejuni

Histoplasma capsulatum (http://www.broad.mit.edu/)	*Mus musculus* (http://www.ncbi.nlm.nih.gov/)	*Candidatus Protochlamydia amoebophila UWE25*
Laccaria bicolour (http://genome.jgi-psf.org/)	*Naegleria gruberi* (http://genome.jgi-psf.org/)	*Chlamydia trachomatis*
Magnaporthe grisea (http://www.broad.mit.edu/)	*Nematostella vectensis* (http://genome.jgi-psf.org/)	*Chlorobium tepidum*
Mycosphaerella fijiensis (http://genome.jgi-psf.org/)	*Ostreococcus lucimarinus* (http://genome.jgi-psf.org/)	*Clostridium perfringens*
Mycosphaerella graminicola (http://genome.jgi-psf.org/)	*Ostreococcus tauri* (http://genome.jgi-psf.org/)	*Corynebacterium glutamicum*
Nectria haematococca (http://www.broad.mit.edu/)	*Paramecium tetraurelia* (http://www.genoscope.cns.fr/spip/Projects.html)	*Cytophaga hutchinsonii*
Neosartorya fischeri (http://www.broad.mit.edu/)	*Phaeodactylum tricornutum* (http://genome.jgi-psf.org/)	*Dechloromonas aromatica*
Neurospora crassa (http://www.broad.mit.edu/)	*Phytophthora infestans* (http://www.broad.mit.edu/)	*Desulfuromonas acetoxidans*
Paracoccidioides brasiliensis (http://www.broad.mit.edu/)	*Phytophthora ramorum* (http://genome.jgi-psf.org/)	*Ehrlichia ruminantium*
Phanerochaete chrysosporium (http://genome.jgi-psf.org/)	*Phytophthora sojae* (http://genome.jgi-psf.org/)	*Escherichia coli*
Phycomyces blakesleeanus(http://genome.jgi-psf.org/)	*Plasmodium yoelii* (http://www.tigr.org/db.shtml)	*Flavobacteria bacterium BAL38*
Pichia stipitis(http://genome.jgi-psf.org/)	*Tetrahymena thermophila* (http://www.tigr.org/db.shtml)	*Geobacillus kaustophilus*
Podospora anserine (http://podospora.igmors.u-psud.fr/)	*Thalassiosira pseudonana* (http://genome.jgi-psf.org/)	*Geobacter uraniumreducens*
Postia placenta (http://genome.jgi-psf.org/)	*Toxoplasma gondii* (http://www.tigr.org/db.shtml)	*Gloeobacter violaceus*
Puccinia graminis (http://www.broad.mit.edu/)	*Trichomonas vaginalis* (http://www.tigr.org/db.shtml)	*Haemophilus influenzae*
Pyrenophora tritici-repentis (http://www.broad.mit.edu/)	*Trichoplax adhaerens* (http://genome.jgi-psf.org/)	*Halobacterium sp. NRC-1*
Rhizopus oryzae (http://www.broad.mit.edu/)	*Trypanosoma brucei* (http://www.tigr.org/db.shtml)	*Helicobacter pylori*
Saccharomyces cerevisiae (http://www.ncbi.nlm.nih.gov/)	*Trypanosoma cruzi* (http://www.tigr.org/db.shtml)	*Kineococcus radiotolerans*
Schizosaccharomyces pombe (http://www.ncbi.nlm.nih.gov/)	*Xenopus tropicalis* (http://genome.jgi-psf.org/)	*Lactobacillus plantarum*
Sclerotinia sclerotiorum (http://www.broad.mit.edu/)	*Volvox carteri* (http://genome.jgi-psf.org/)	*Leptospira interrogans*
Sporobolomyces roseus (http://genome.jgi-psf.org/)		*Magnetococcus sp. MC-1*
Stagonospora nodorum (http://www.broad.mit.edu/)		*Methanosarcina mazei*
Trichoderma reesei (http://genome.jgi-psf.org/)		*Methylobacillus flagellatus*
Trichoderma virens (http://genome.jgi-psf.org/)		*Mycobacterium tuberculosis*
Uncinocarpus reesii (http://www.broad.mit.edu/)		*Myxococcus xanthus*
Ustilago maydis (http://www.broad.mit.edu/)		*Nanoarchaeum equitans*
Yarrowia lipolytica (http://www.ncbi.nlm.nih.gov/)		*Nitrosomonas eutropha*
		Nostoc punctiforme
		Picrophilus torridus
		Porphyromonas gingivalis
		Prochlorococcus marinus
		Pyrobaculum aerophilum
		Pyrococcus abyssi
		Rhodoferax ferrireducens
		Rickettsia typhi
		Roseiflexus castenholzii
		Staphylococcus aureus
		Streptomyces avermitilis
		Sulfolobus tokodaii
		Synechococcus sp. WH 8102

Syntrophus aciditrophicus
Thermobifida fusca
Thermosynechococcus elongatus
Thermotoga maritima
Treponema denticola
Ureaplasma parvum
Vibrio cholerae
Xanthomonas oryzae
Xylella fastidiosa

Supplemental Table 2. Results of phylogenetic analysis of genes linked to the 9 plant-fungi HGTs on the genome contigs of the HGT recipient taxa. The table summarises the results of phylogenetic analysis of three open reading frames immediately 5' and 3' of each putatively transferred gene from each of the recipient genome sequences. This data is further summarised in Figure 5. One of the HGTs was present in two closely related recipient genomes (Figure 4c), therefore they were judged to be two independent samples and unlikely to be due to two identical gene contamination events. In the remaining 8 HGTs the phylogenetic and comparative genome data demonstrated that the HGT candidate gene was located on a contiguous section of chromosome sequence flanked by gene sequences where we were able to demonstrate conventional vertical inheritance, such that the gene approximated species phylogeny so that alternative hypothesise of gene ancestry were non-parsimonious. Patterns of vertical inheritance were pinpointed based upon either a tree topology showing the gene branching with species known to be close evolutionary relatives of the genome species, or that the taxon distribution of the gene was limited to close evolutionary relatives. In all cases we adjusted the phylogenetic analyses pipeline sampling threshold to include a wide as possible taxon sampling in order to address the question of gene ancestry. The table summarises the sampling thresholds used for each dataset.

HGT (Fig.)	Gene	Upstream	BLAST sampling threshold - Result	Downstream	BLAST sampling threshold - Result
1a	**Phpa(173818)**	Phpa(63772)	Unique gene (no hits at 1e-5)	Phpa(63774)	1e-5, potential TE, 1913 hits vs Ppatens genome at 1e-20
		Phpa(158455)	No tree, only one hit (Semo(409669) at 1e-5	Phpa(63775)	1e-5, potential TE, 3253 hits vs Ppatens genome at 1e-20
		Phpa(111080)	1e-20 - clusters with plants	Phpa(63776)	1e-5, potential TE, > 3429 hits vs Ppatens genome at 1e-20
1b	**BDEG_06896**	BDEG_06895	1e-5 - fungal only	BDEG_06897	1e-20 - clusters with fungi
		BDEG_06894	1e-20 - clusters with fungi	BDEG_06898	1e-15 - clusters with fungi
		BDEG_06893	1e-30 - clusters with fungi	BDEG_06899	1e-20 - clusters with fungi
1c	**Semo(120147)**	Semo(20613)	1e-20 - plants only	Semo(423569)	1e-5 - Selaginella specific gene family
		Semo(423566)	1e-20 - plants only	Semo(445894)	1e-10 - clusters with plants
		Semo(72148)	1e-20 - plants only	Semo(19624)	1e-5 - plants only
2-	**Semo(137360)**	Semo(431654)	1e-10 - only present in Selaginella and Physcomitrella	Semo(431673)	1e-10 - plants only
		no gene present		Semo(431674)	1e-5 - Selaginella only (6 genes)
		no gene present		Semo(431681)	1e-5 - Selaginella only (3 genes)
	Semo(121880)	Semo(424337)	1e-5 - Selaginella only (4 genes)	Semo(19275)	1e-5 - Selaginella specific gene family
		Semo(424334)	1e-5 - plants only	Semo(446215)	1e-20 - clusters with plants
		Semo(424333)	Unique gene (no hits at 1e-5)	Semo(121832)	1e-20 - clusters with plants
	Semo(79756)	Semo(80654)	1e-30 - clusters with plants	Semo(80323)	1e-20 - clusters with plants
		Semo(404818)	1e-5 - Selaginella only (3 genes)	Semo(73023)	1e-10 - Selaginella only (4 genes), same tree as Semo(73021
		Semo(73021)	1e-10 - Selaginella only (4 genes), same tree as Semo(7302	Semo(27379)	1e-30 - clusters with plants
3a	**Semo(407912)**	Semo(230913)	1e-30 - looks good	Semo(407913)	Unique gene (no hits at 1e-5)
		Semo(407910)	No tree, only one hit (Semo(417849) at 1e-5	Semo(407914)	1e-5 - Selaginella specific gene family
		Semo(85570)	1e-20 - plants only	Semo(407915)	1e-5 - plants only
3b	**Phpa(130986)**	Phpa(80292)	1e-5, potential TE,	Phpa(130938)	1e-20 - clusters with plants
		Phpa(80291)	Unique gene (no hits at 1e-5)	Phpa(106587)	1e-40 - clusters with plants
		Phpa(80290)	Unique gene (no hits at 1e-5)	Phpa(185688)	1e-10 - clusters with plants
4a	**Labi_325721**	Labi_318106	Unique gene (no hits at 1e-5)	Labi_325722	Unique gene (no hits at 1e-5)
		Labi_325720	1e-30 - fungal only	Labi_151444	1e-30 - fungal only
		Labi_318103	1e-30 - fungal only	Labi_318108	1e-30 - fungal only
	Labi_303459	Labi_294871	1e-30 - fungal only	Labi_303460	1e-30 - fungal only
		Labi_303457	1e-20 - basidomycete only	Labi_303461	1e-5 - Laccaria only (7 genes)
		Labi_303456	Unique gene (no hits at 1e-5)	Labi_329878	1e-5 - basidomycete only
	Labi_294588	Labi_299574	1e-30 - fungal only	Labi_299576	1e-5 - Laccaria only (5 genes)
		Labi_299573	1e-5, one hit - Coprinus cinereus	Labi_299578	Unique gene (no hits at 1e-5)
		Labi_236267	1e-30 - fungal only	Labi_328458	1e-20 - basidomycete only
	Labi_307948	Labi_307947	Unique gene (no hits at 1e-5)	Labi_295356	1e-5 - Laccaria, Coprinus, 16 Laccaria hits at 1e-5
		Labi_295355	Unique gene (no hits at 1e-5)	Labi_174921	1e-20 - clusters with fungi
		Labi_332008	1e-20 - fungal only	Labi_332014	Unique gene (no hits at 1e-5)
4b	**BDEG_03418**	BDEG_03417	Unique gene (no hits at 1e-5)	BDEG_03419	1e-5, with bacteria
		BDEG_03416	1e-20 - with Dictyostelium	BDEG_03420	1e-5, with bacteria

		BDEG_03415	1e-20 - clusters with fungi	BDEG_03421	Unique gene (no hits at 1e-5)
	BDEG_04030	BDEG_04029	1e-20 - with Cryptosporidium	BDEG_04031	Unique gene (no hits at 1e-5)
		BDEG_04028	Unique gene (no hits at 1e-5)	BDEG_04032	Unique gene (no hits at 1e-5)
		BDEG_04027	1e-20 - clusters with fungi	BDEG_04033	1e-20 - clusters with Caenorhabditis
	BDEG_06393	BDEG_06392	1e-30 - clusters with fungi	BDEG_06394	1e-30 - clusters with fungi
		BDEG_06391	1e-20 - with chlorophyte, stramenopiles, paramecium, tetrah	BDEG_06395	Unique gene (no hits at 1e-5)
		BDEG_06390	1e-30 - with Naegleria, fungi	BDEG_06396	1e-100 - clusters with metazoa
	BDEG_04225	BDEG_04224	1e-20 - clusters with Phycomyces, plants	BDEG_04226	1e-5 - with Thalassiosira
		BDEG_04223	1e-30 - clusters with Trichoplax	BDEG_04227	Unique gene (no hits at 1e-5)
		BDEG_04222A	1e-20 - clusters with Phycomyces	BDEG_04228	1e-40 - clusters with fungi
	BDEG_06878	BDEG_06877	1e-30 - clusters with Ostreococcus	BDEG_06879	Unique gene (no hits at 1e-5)
		BDEG_06876	Unique gene (no hits at 1e-5)	BDEG_06880	1e-5 - clusters with fungi
		BDEG_06875	1e-20 - clusters with fungi	BDEG_06881	1e-5 - clusters with zygomycetes
	BDEG_04956	BDEG_04955	1e-5, Batrachochytrium only (6 copies)	BDEG_04957	1e-20 - clusters with Trichomonas
		BDEG_04954	1e-5, Batrachochytrium only (6 copies)	BDEG_04958	1e-5 - clusters with Dictyostelium
		BDEG_04953	Unique gene (no hits at 1e-5)	BDEG_04959	1e-5 - clusters with Paramecium
4c	Unlikely to be contamination HGT present in two recipient genomes.				

Key
Taxon distribution of gene family suggests vertical inheritance
Phylogeny suggests vertical inheritance
Transposable elements
Gene appears unique to genome
Could not confirm vertical inheritance