Supplementary information

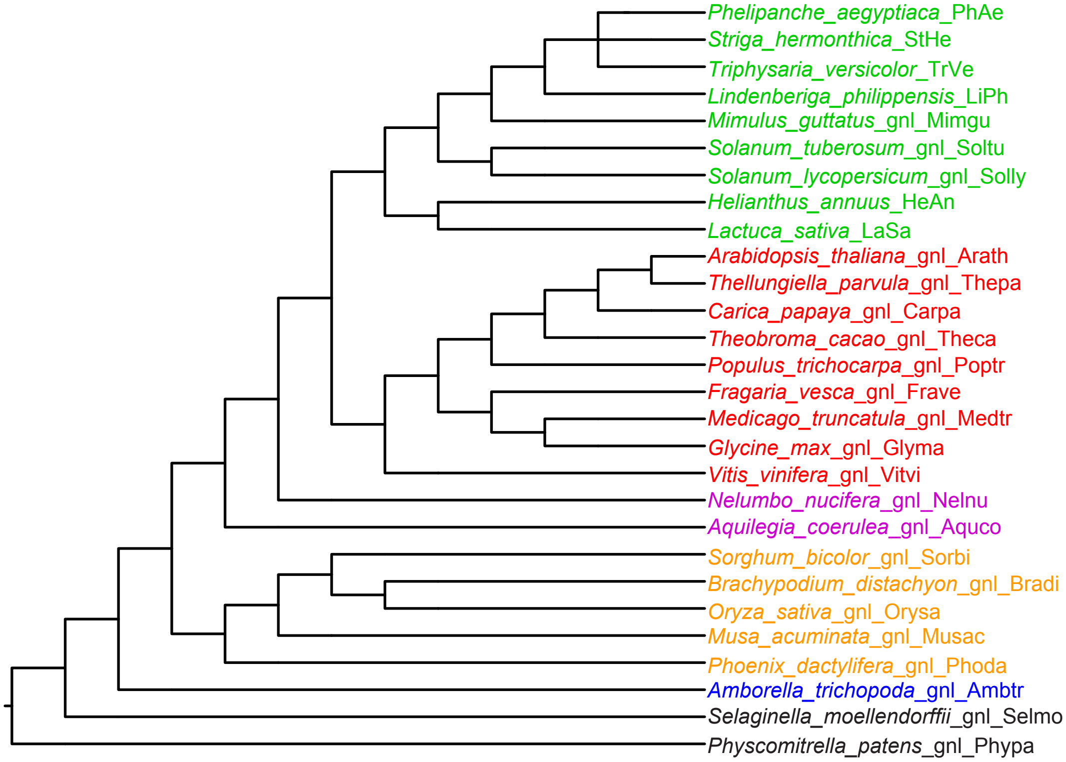


Figure S1. The species tree for HGT discovery. This tree includes 22 sequenced plant genomes containing "gnl_" followed by the first three letters of the genus name, and first two letters of the species name. It also includes four transcriptomes generated in this study – one nonparasitic outgroup *Lindenbergia*, and three parasitic plants *Triphysaria*, *Striga*, and *Phelipanche*, as well as two additional asterids with extensive EST sequence datasets – *Helianthus* and *Lactuca*. Color codes: asterid (green); rosid (red); basal eudicot (purple); monocot (gold); basal angiosperm (blue); outgroup (black).

Fig. S2.1 - orthogroup 17

Fig. S2.2 - orthogroup 218

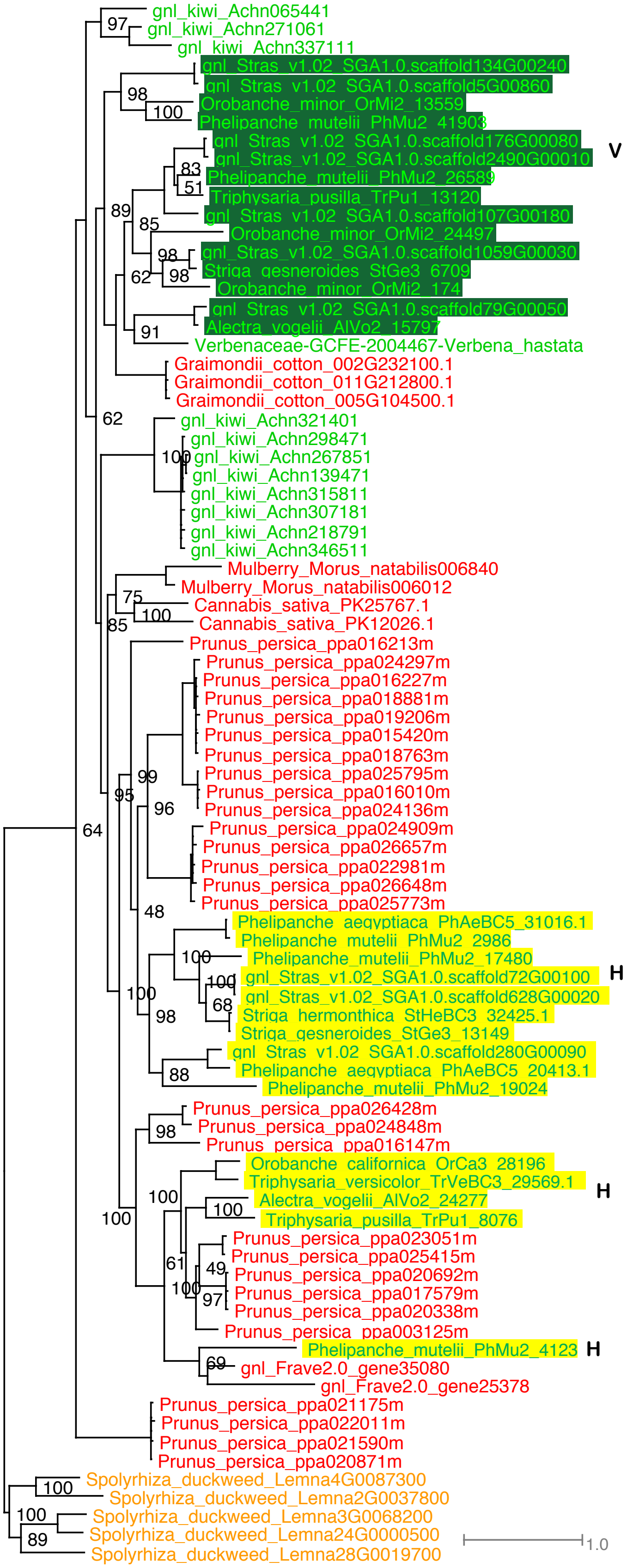


Fig. S2.3- orthogroup 226

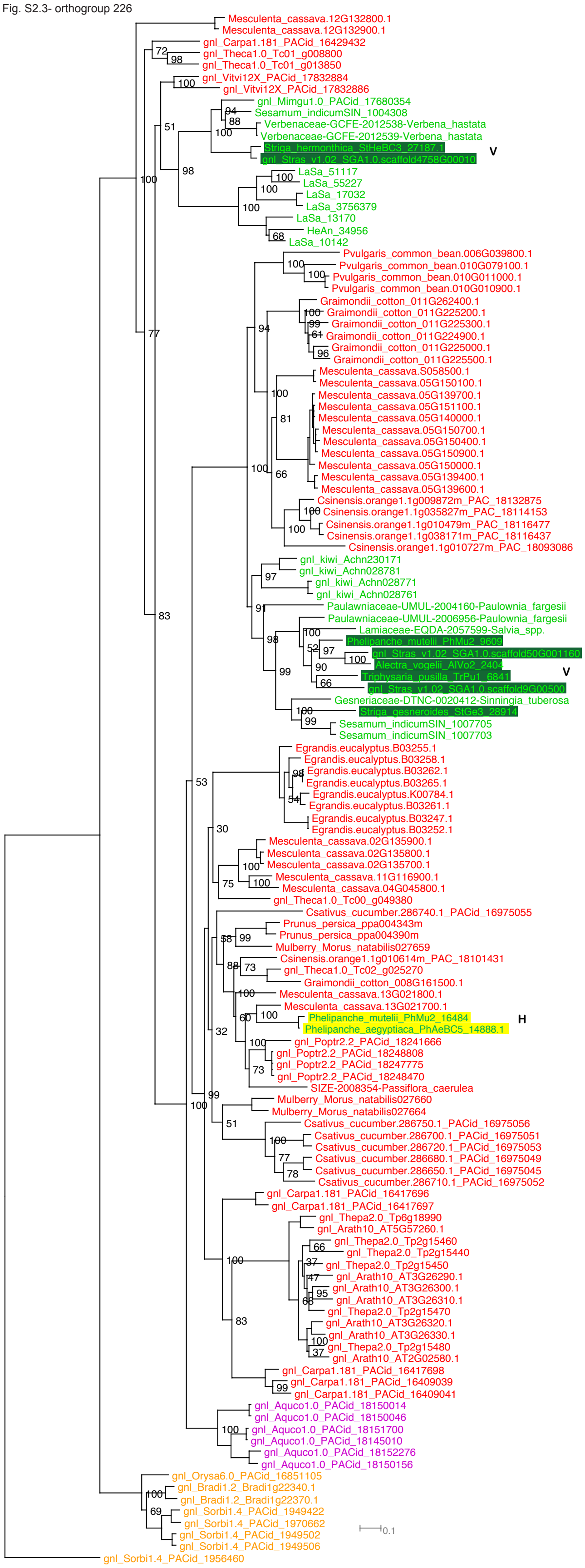


Fig. S2.4 - orthogroup 806

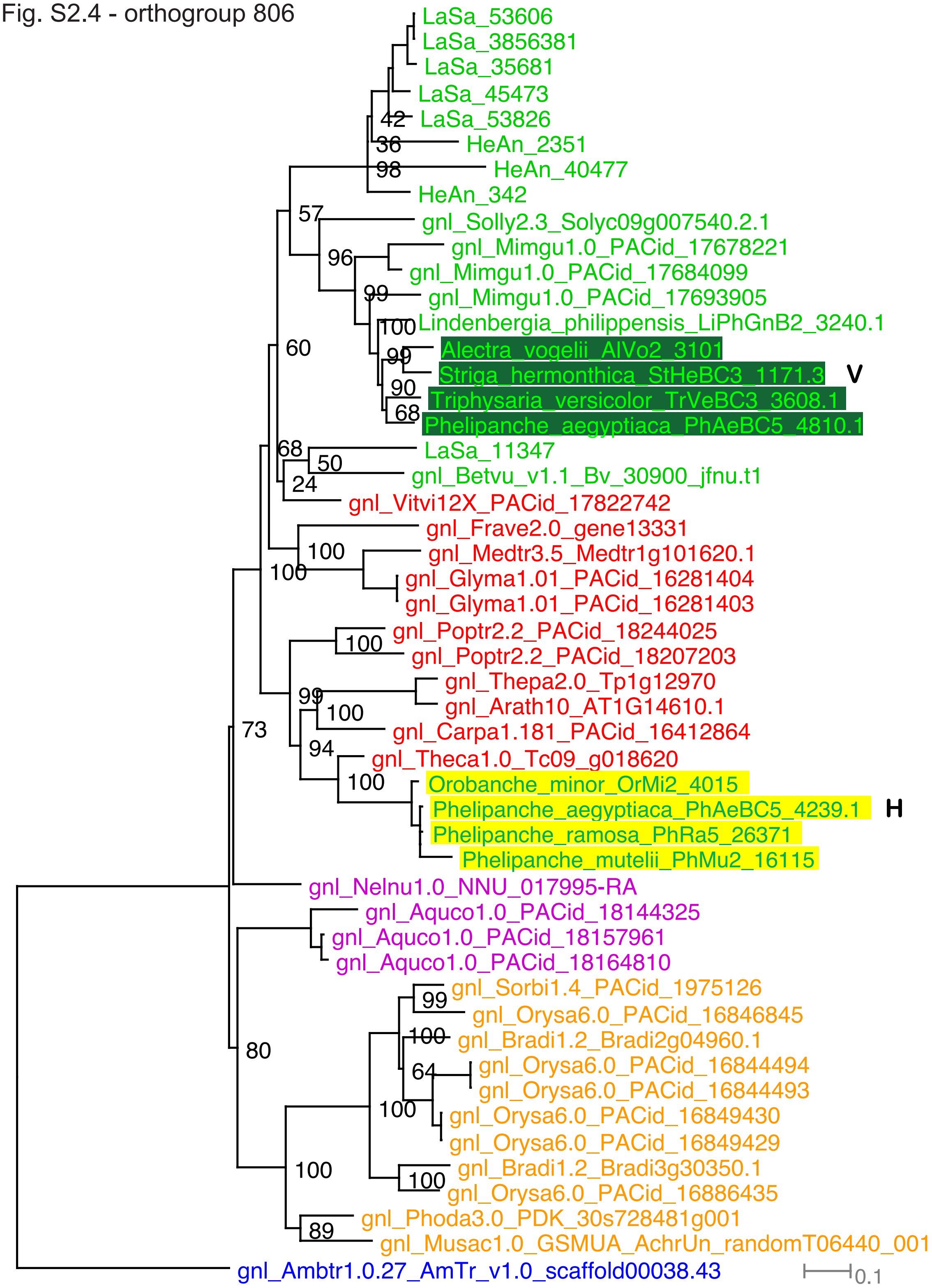


Fig. S2.5 - orthogroup 1021

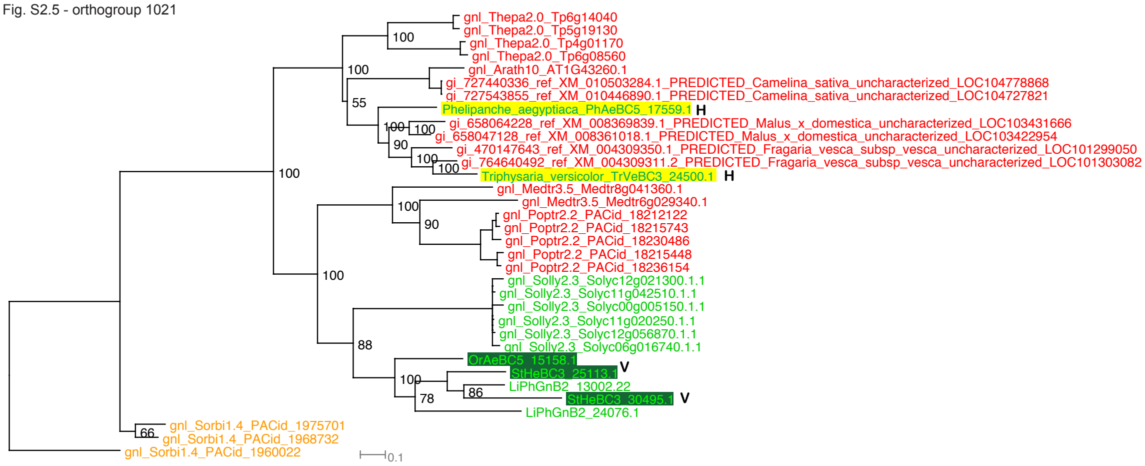


Fig. S2.6 - orthogroup 1226

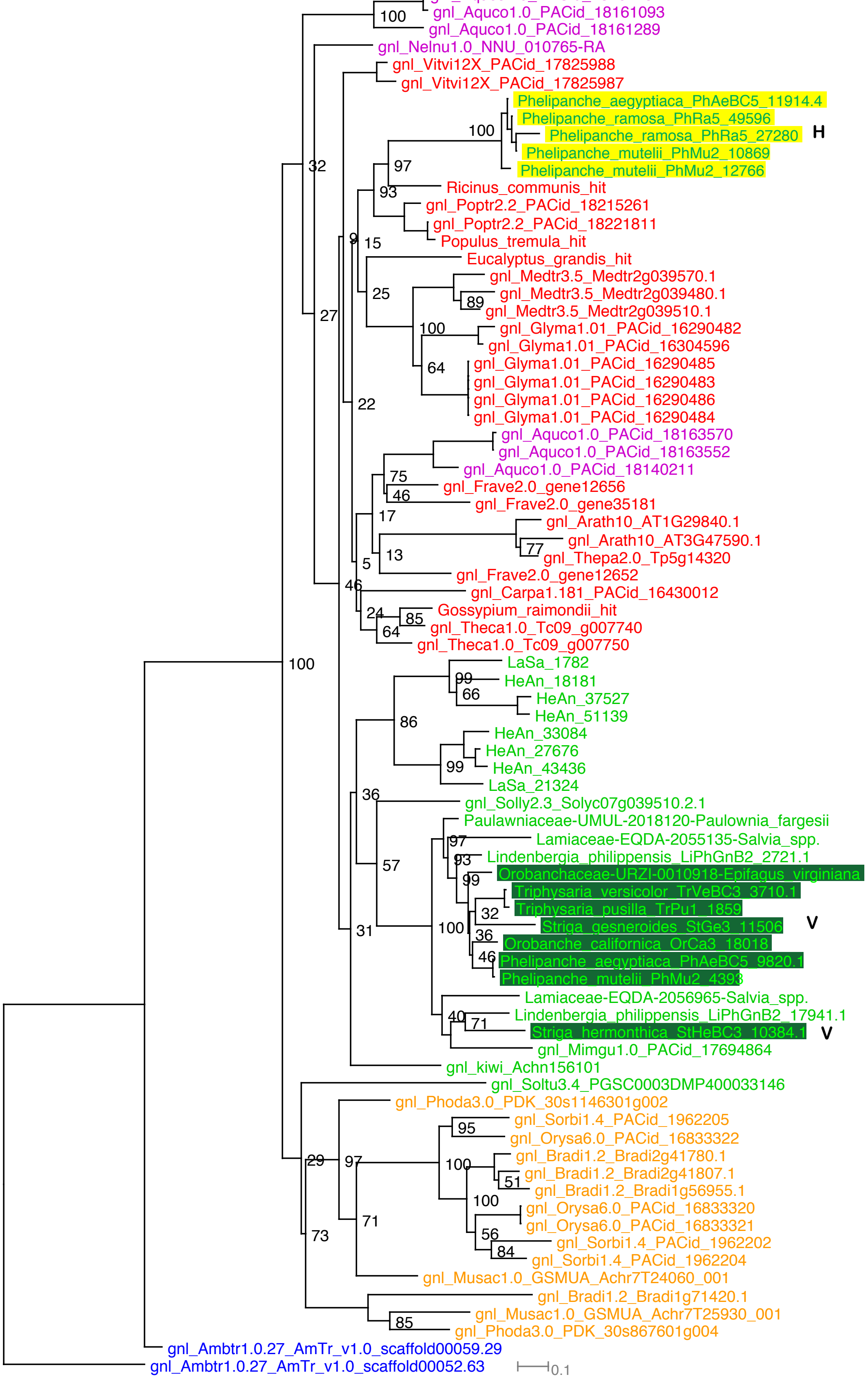


Fig. S2.7 - orthogroup 1685

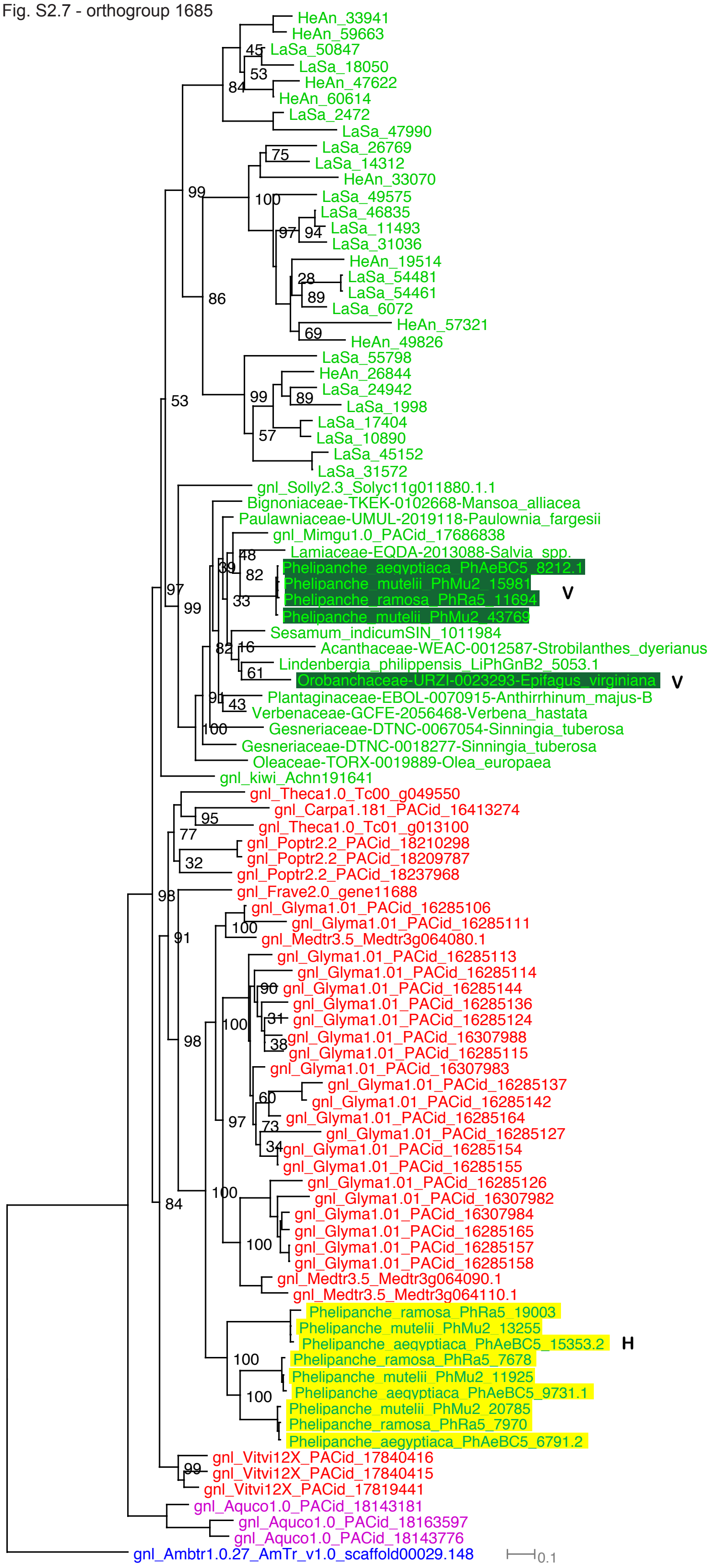


Fig. S2.8 - orthogroup 1886

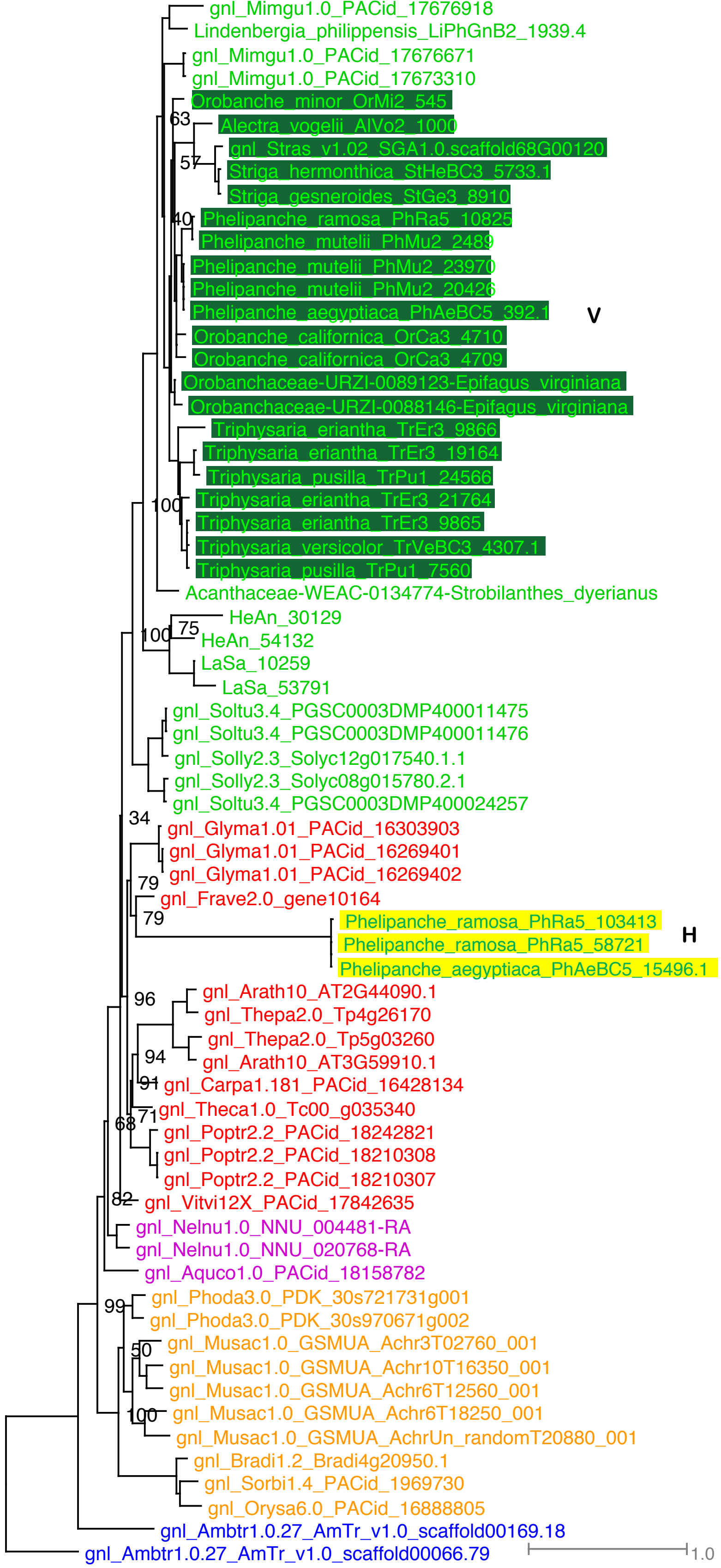


Fig. S2.9 - orthogroup 2270

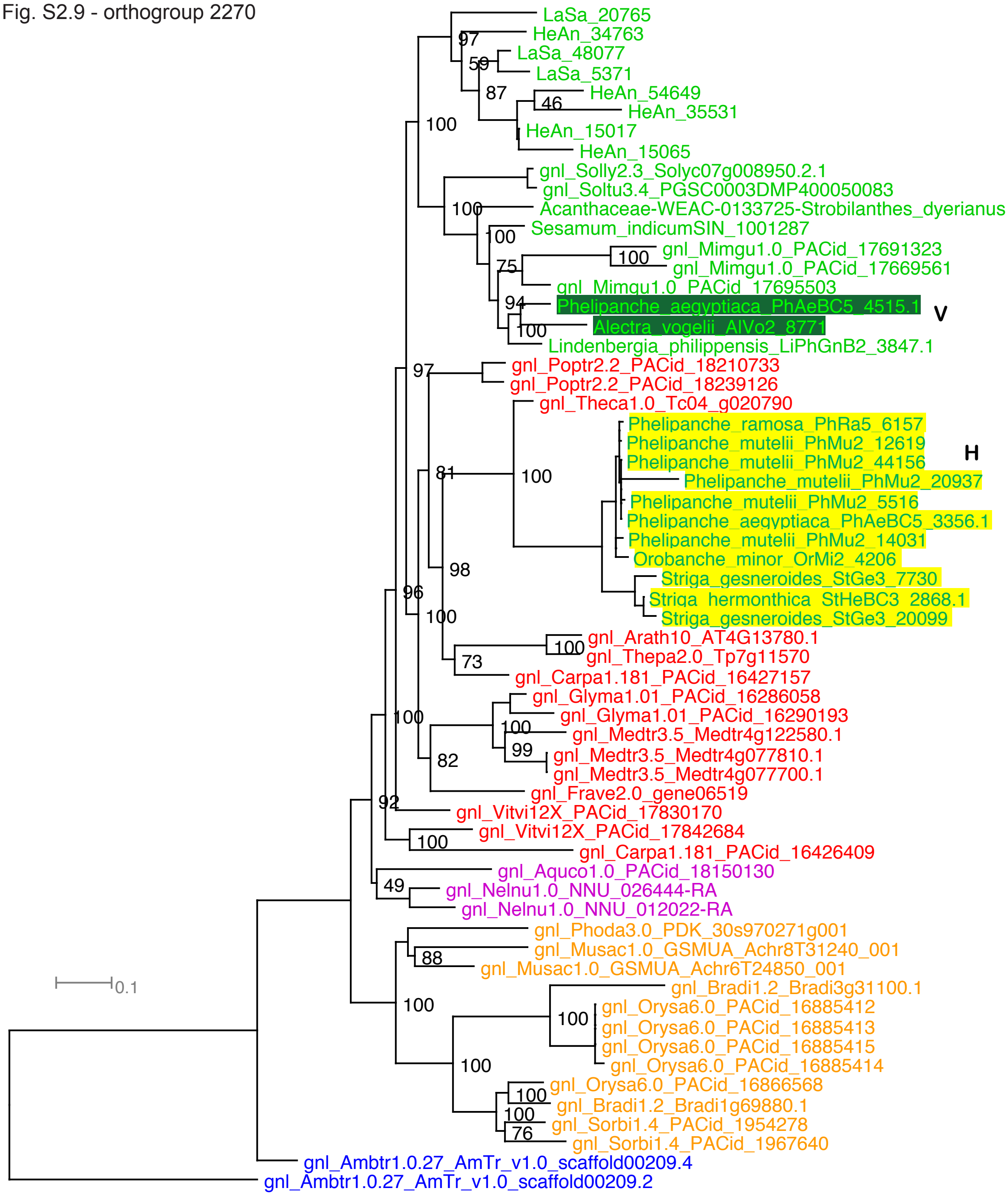


Fig. S2.10 - orthogroup 2376

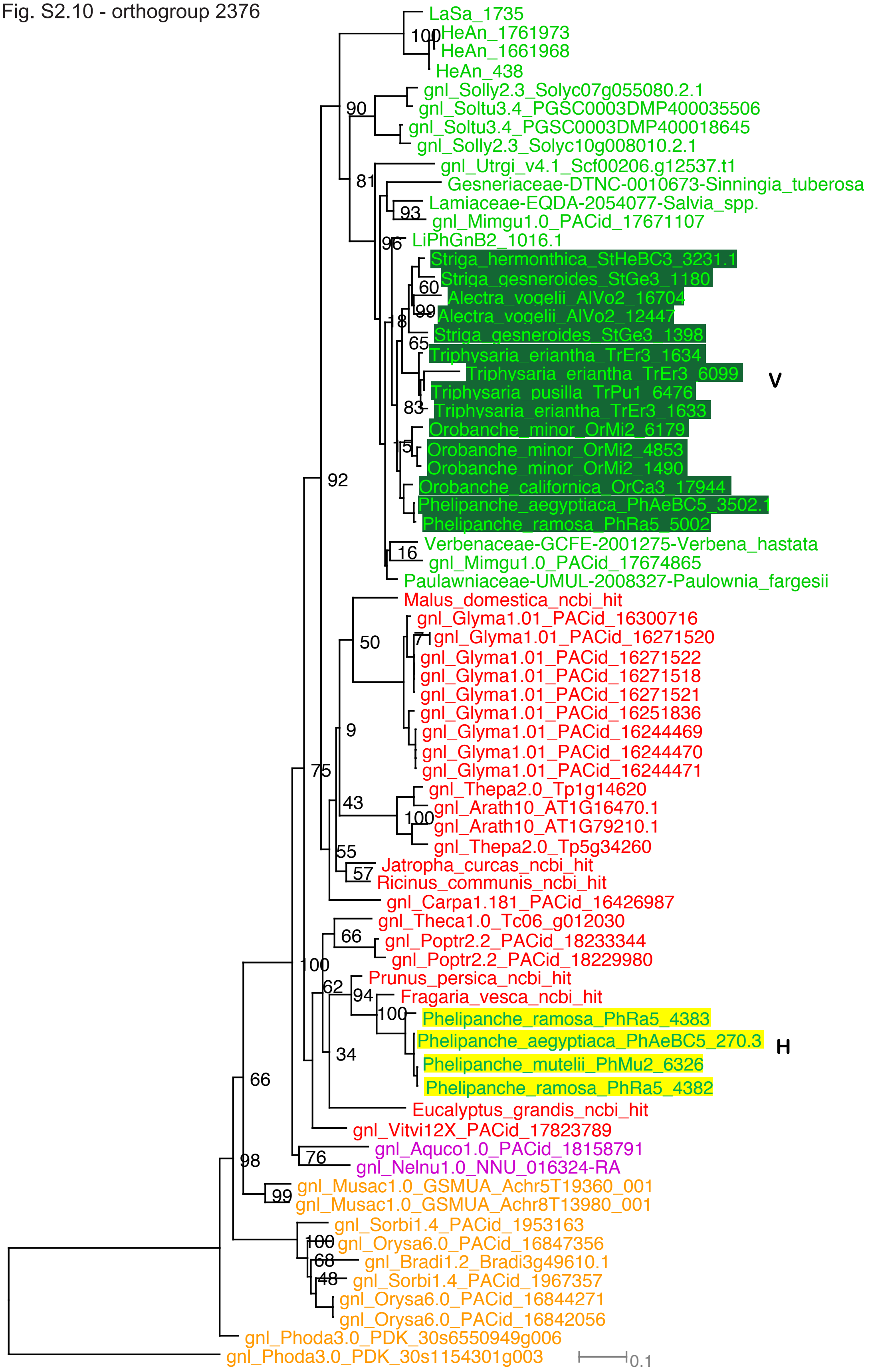


Fig. S2.11 - orthogroup 3861

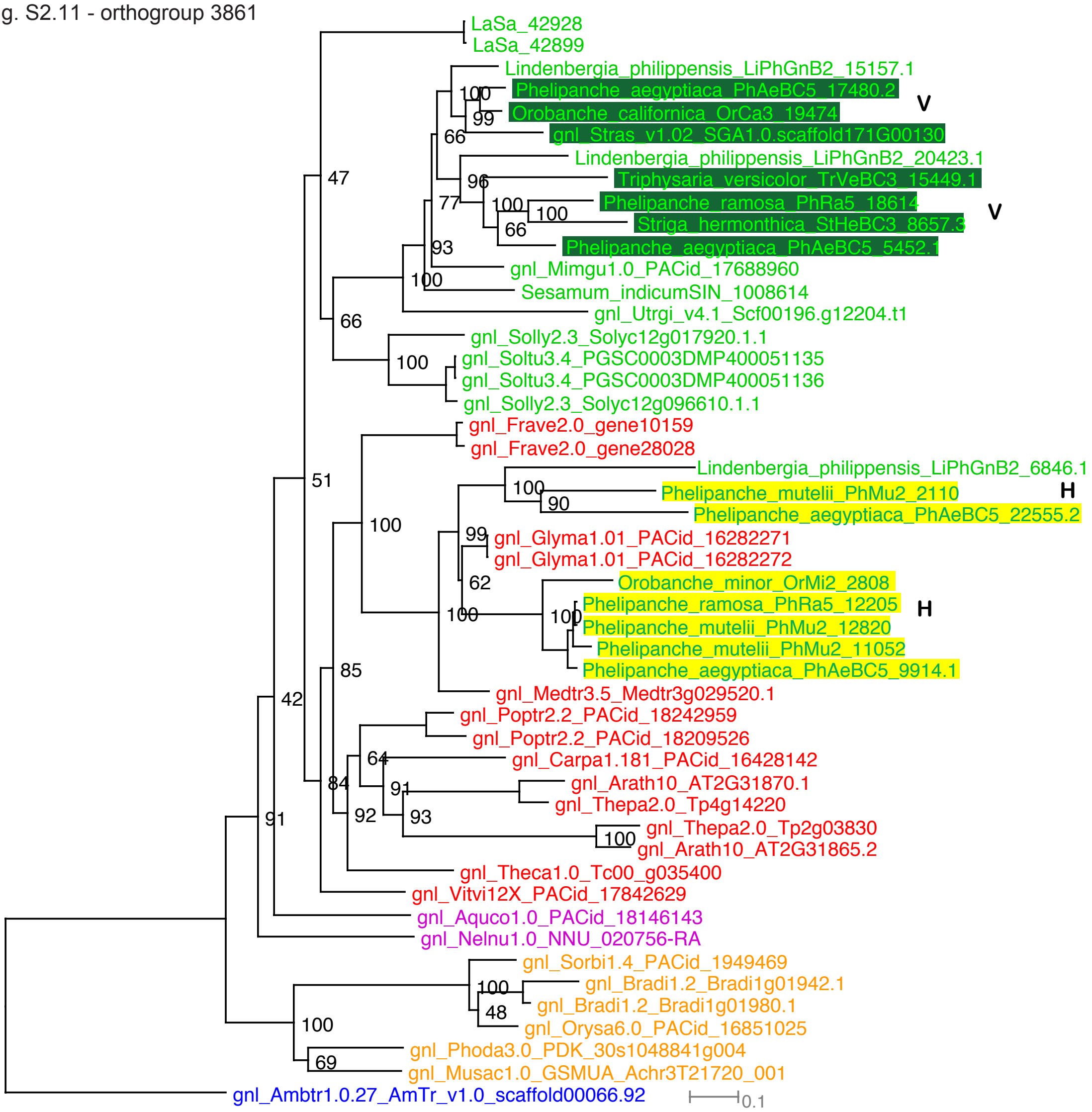


Fig. S2.12 - orthogroup 4067

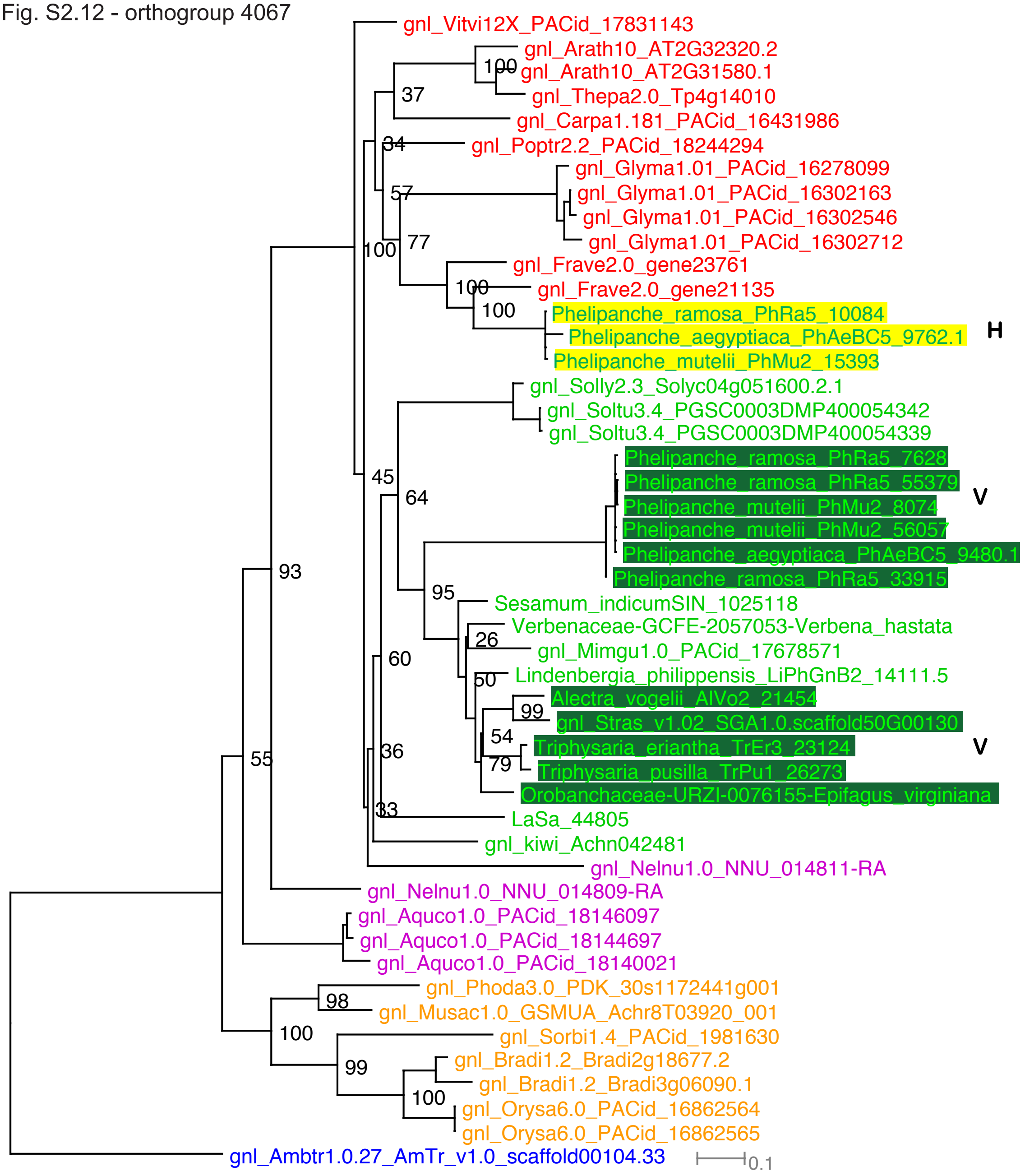


Fig. S2.13 - orthogroup 4572

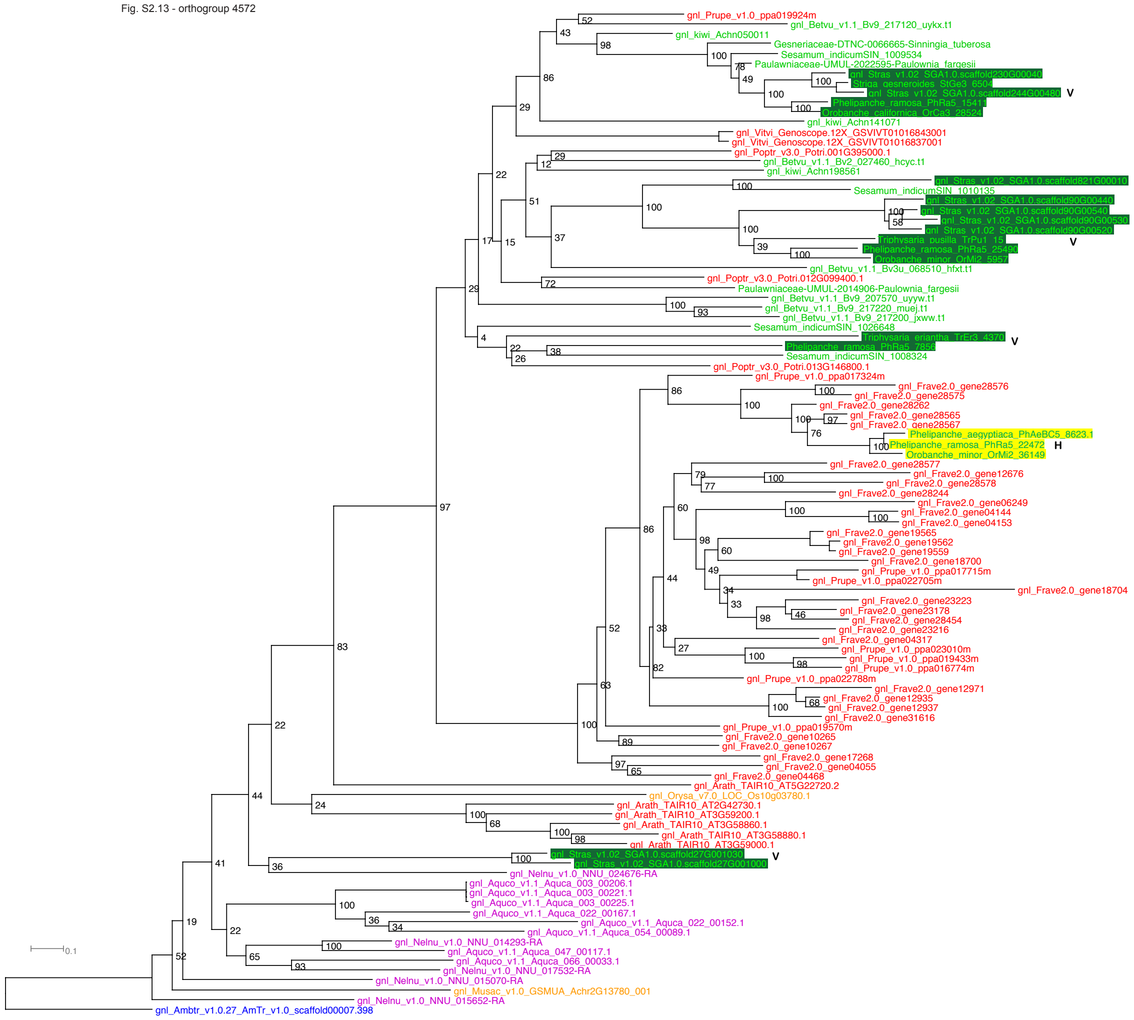


Fig. S2.14 - orthogroup 4598

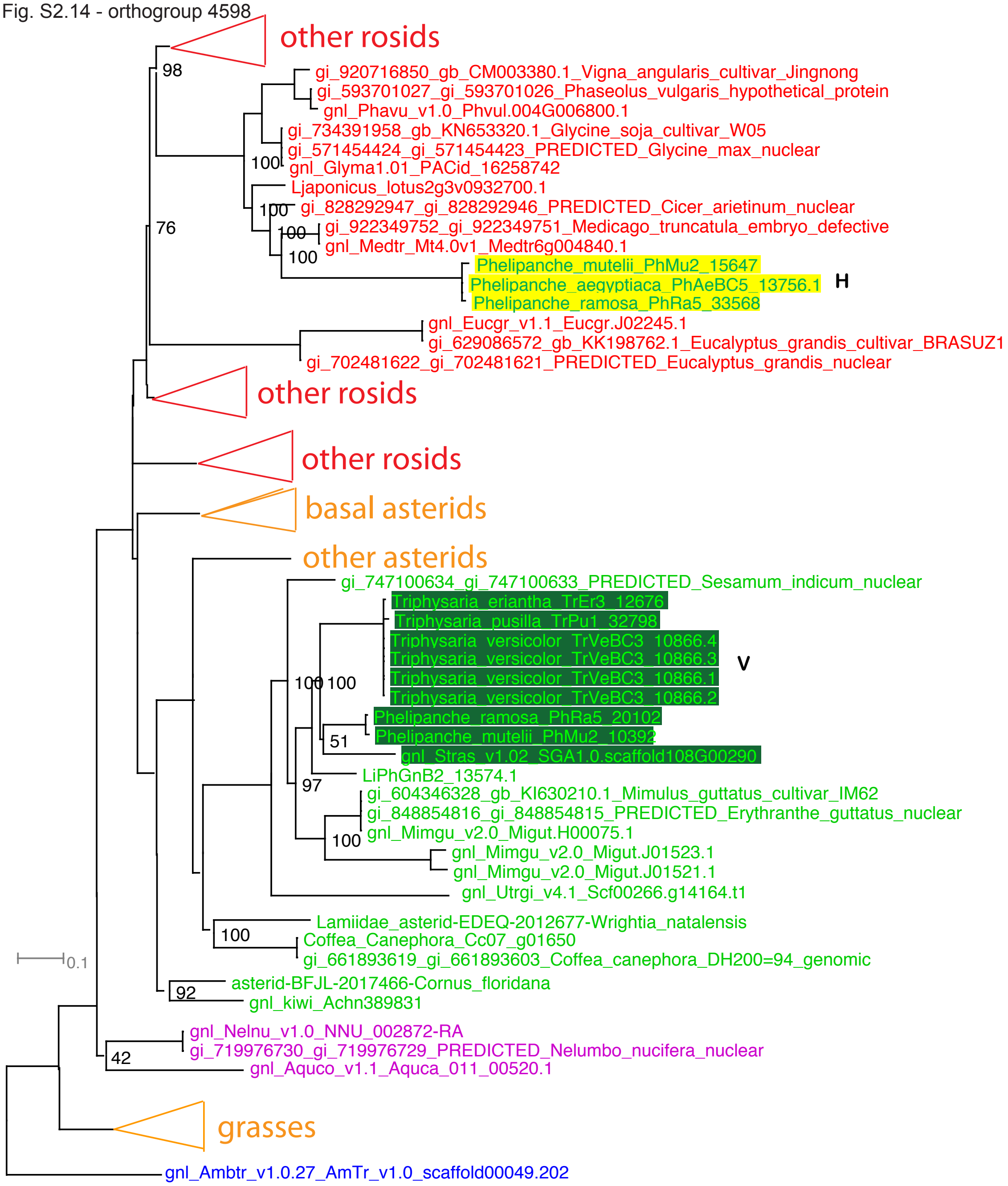


Fig. S2.15- orthogroup 5002

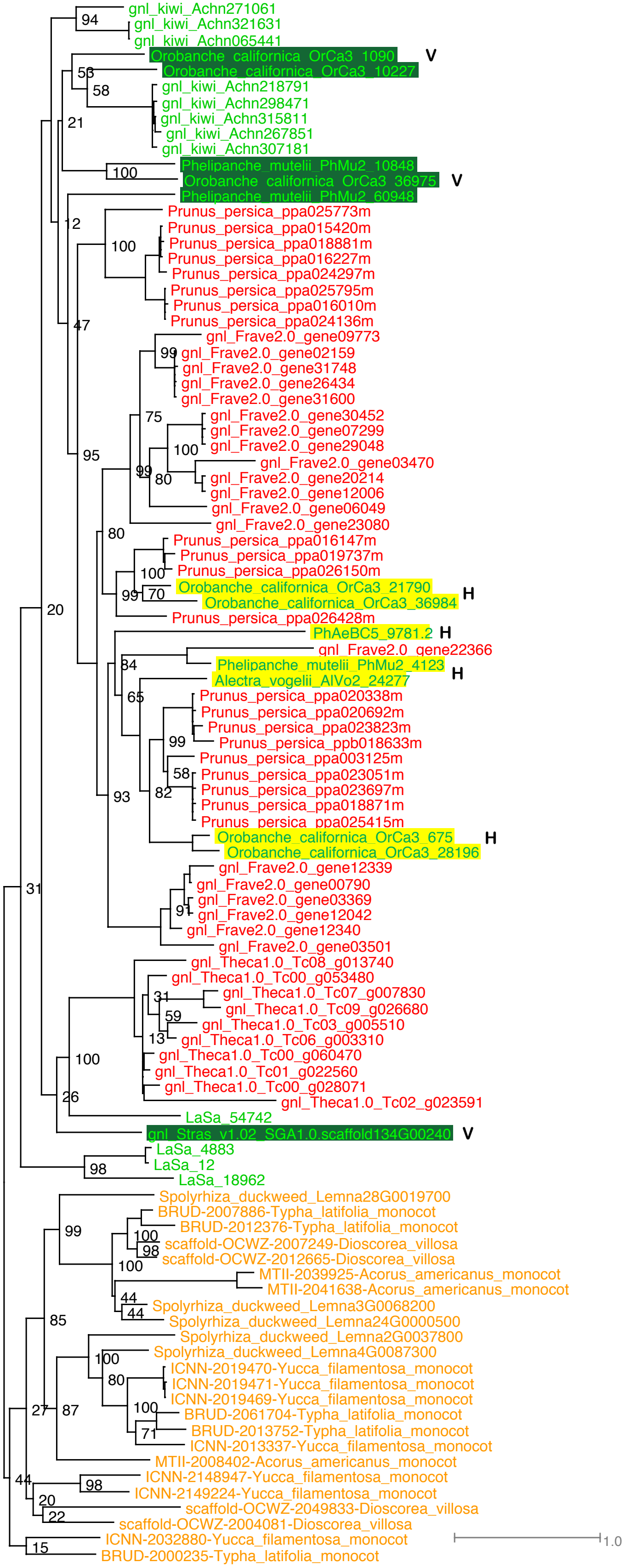


Fig. S2.16 - orthogroup 5896

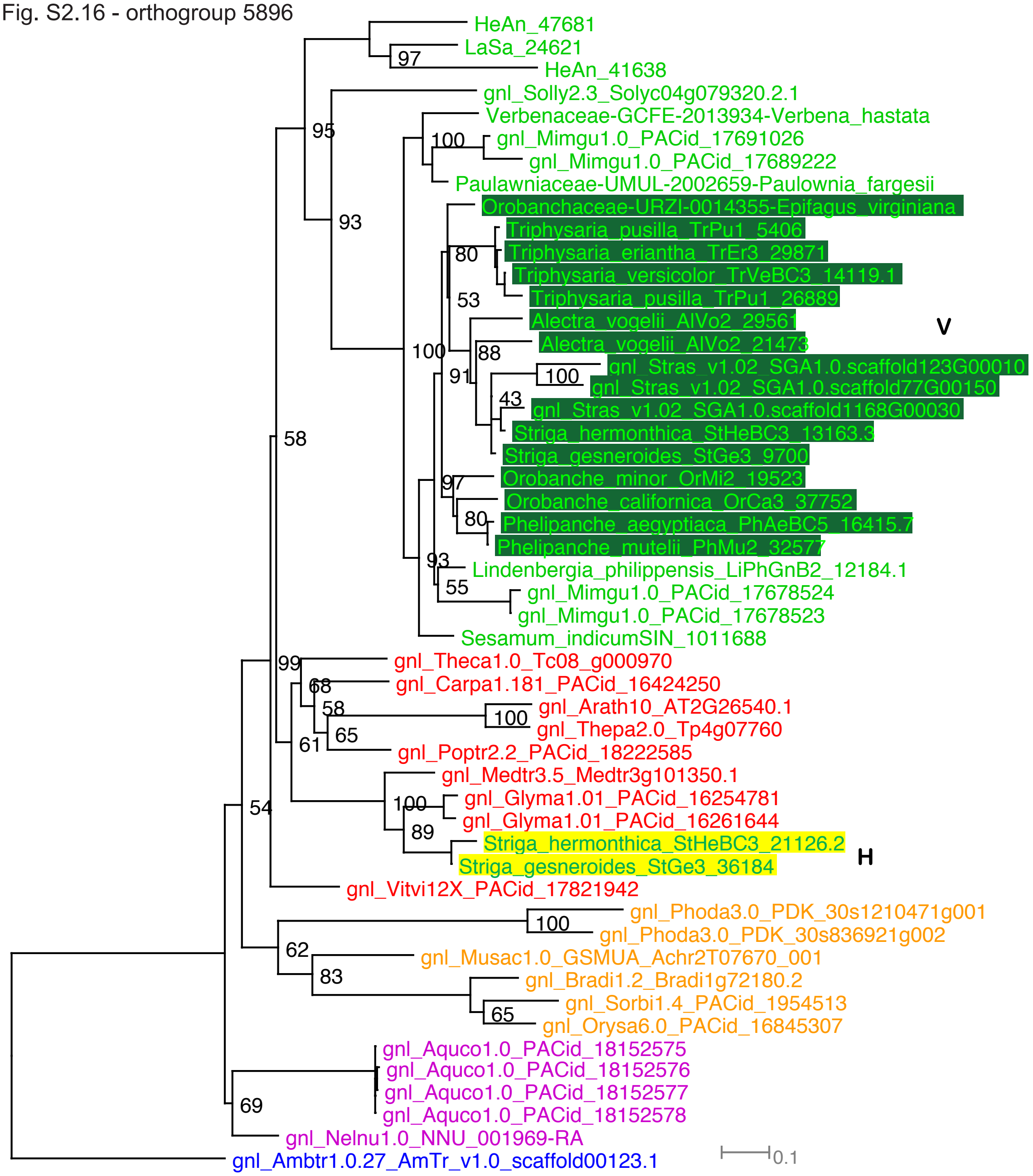


Fig. S2.17 - orthogroup 8888

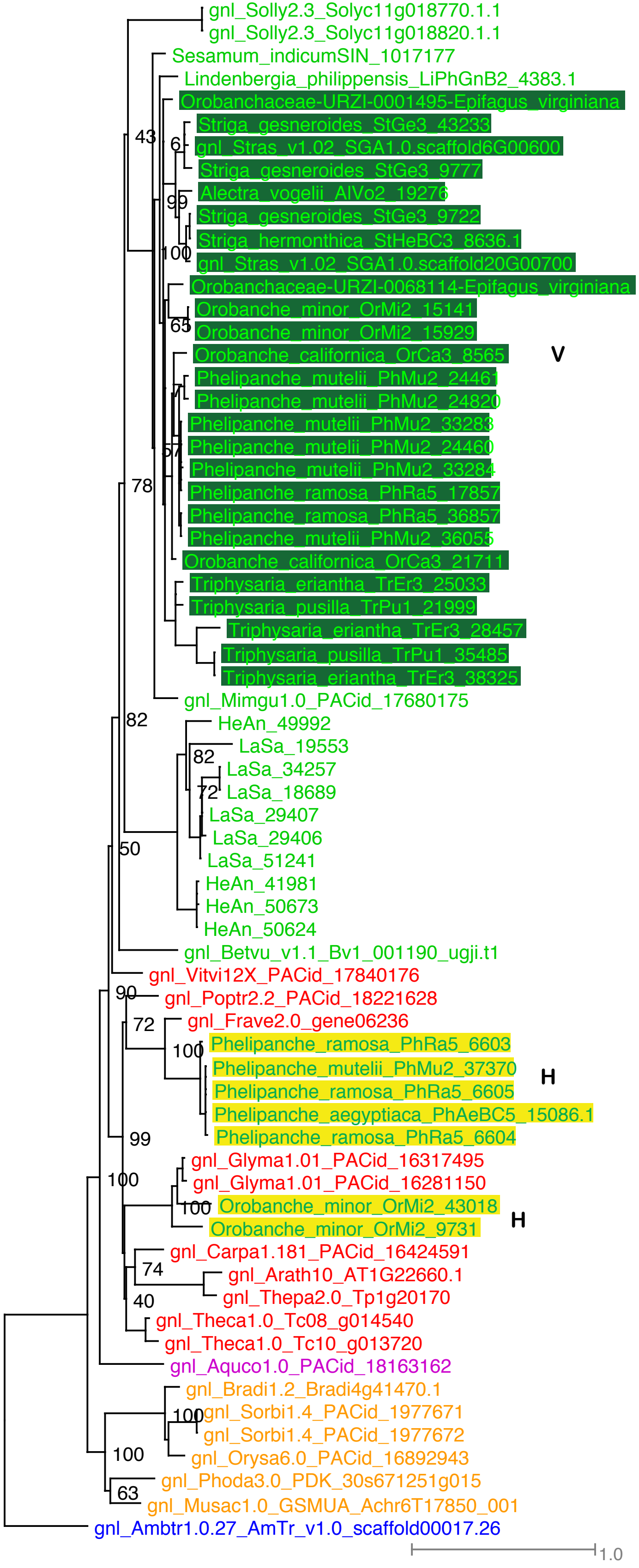


Fig. S2.18 - orthogroup 9613

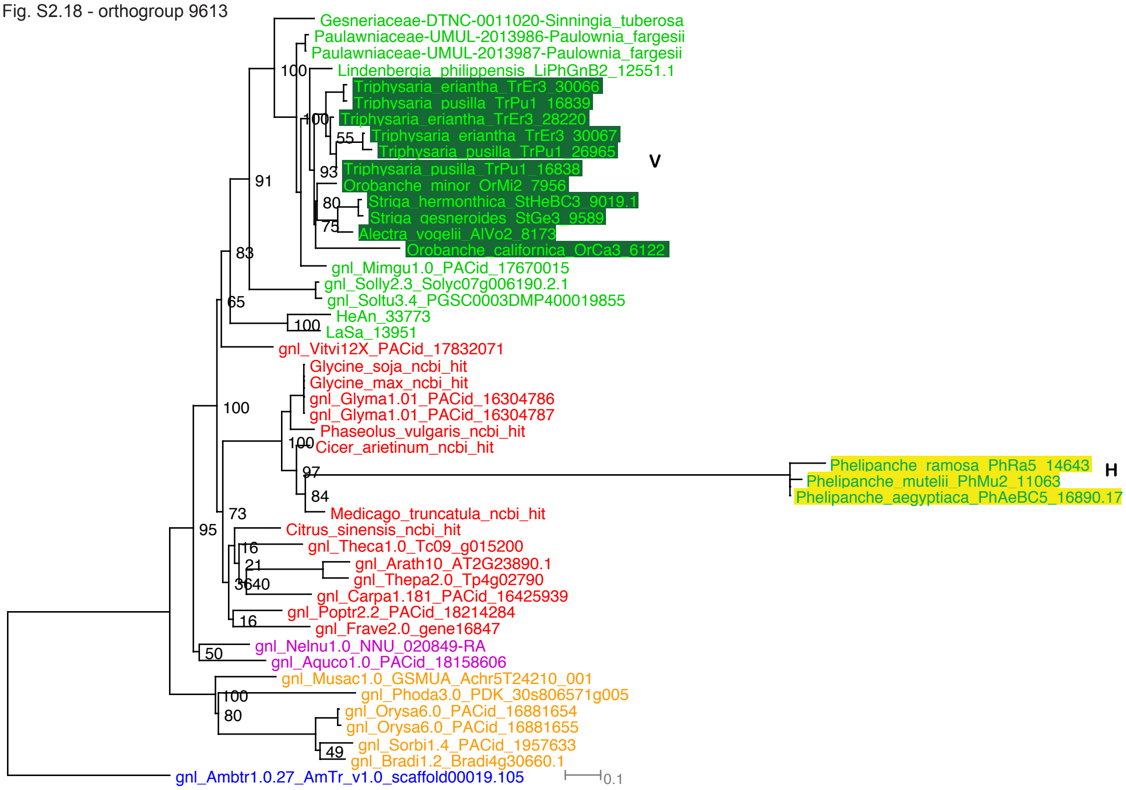


Fig. S2.19 - orthogroup 10050

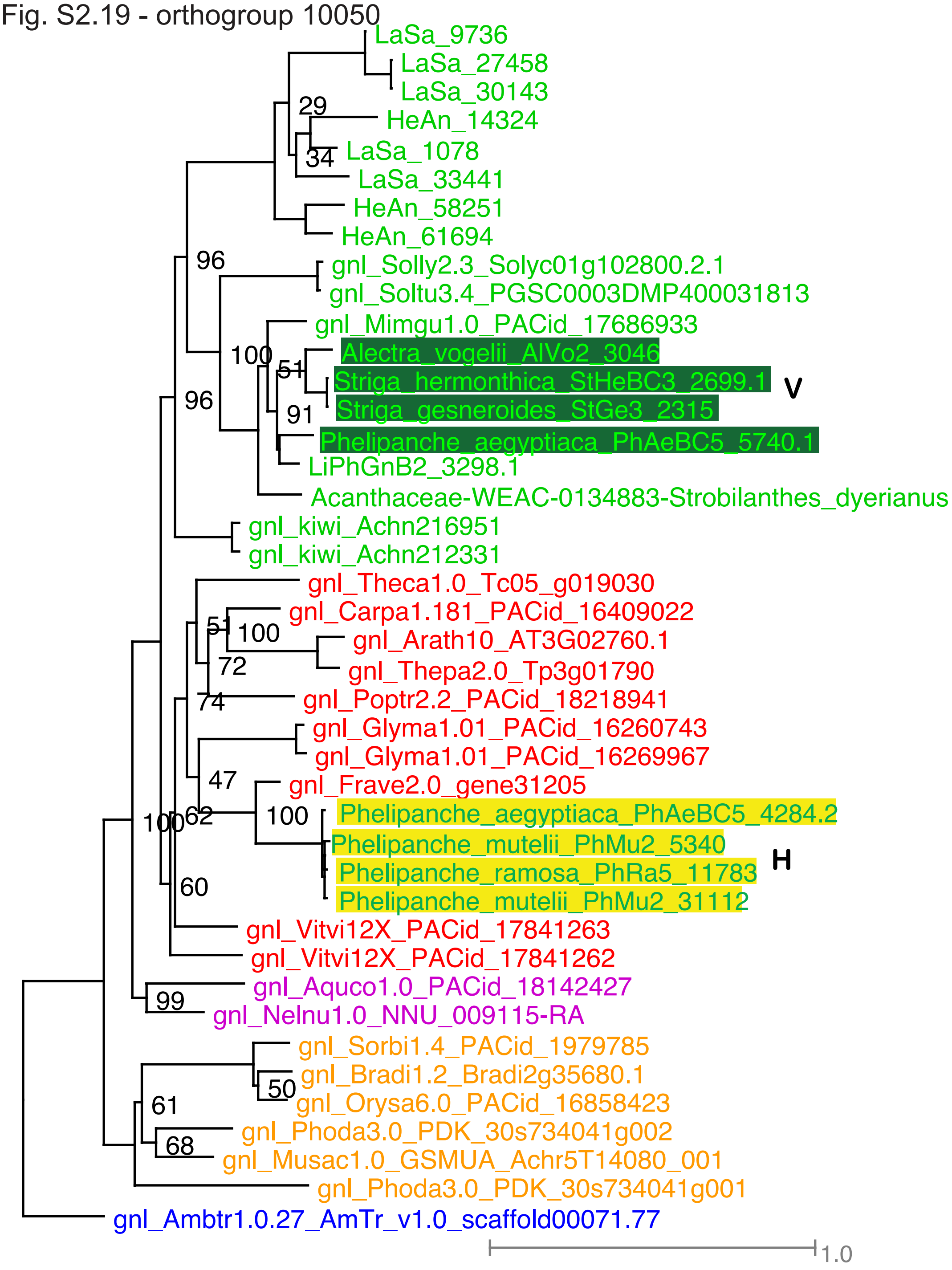


Fig. S2.20 - orthogroup 10143

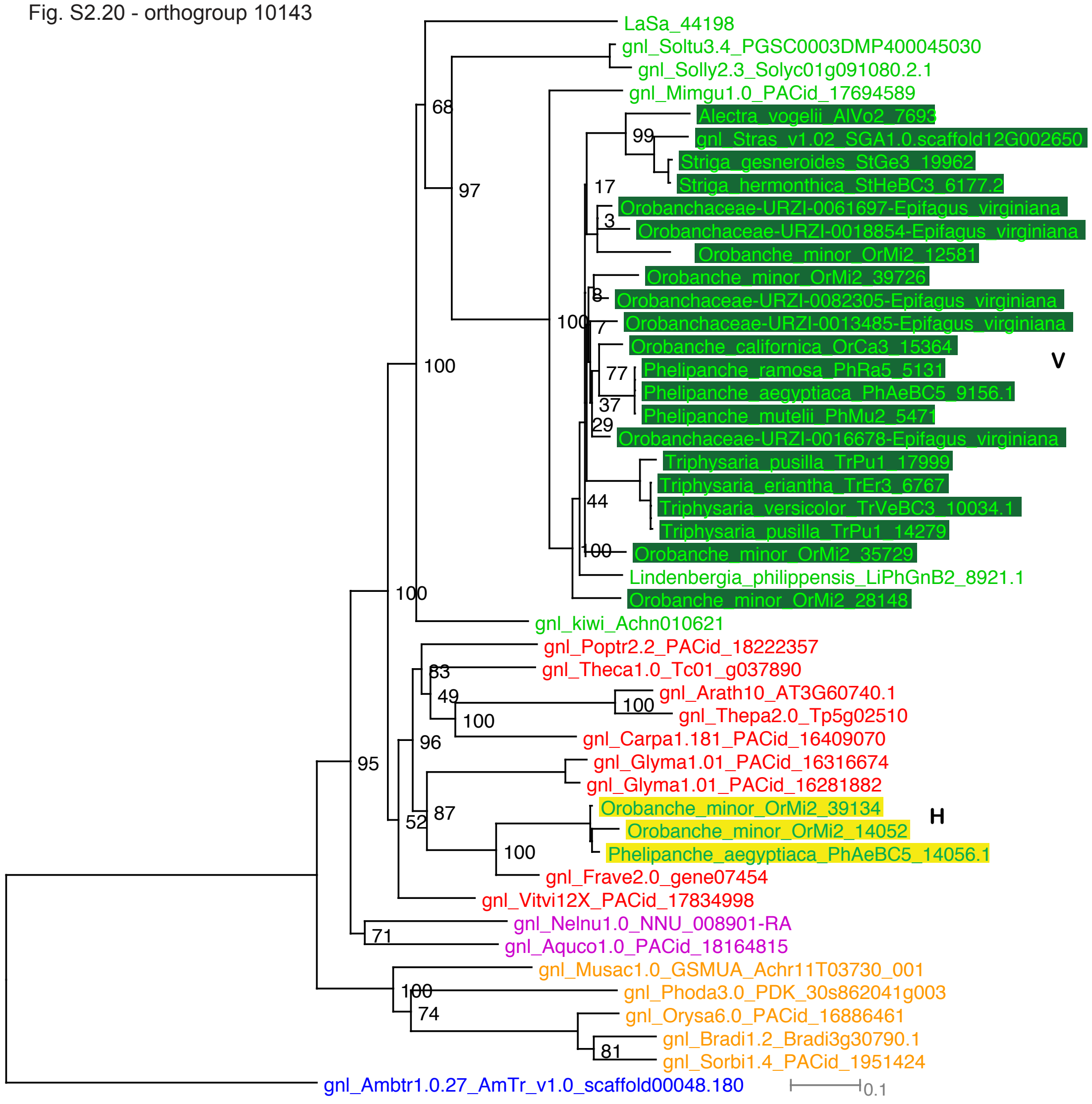


Fig. S2.21 - orthogroup 11437

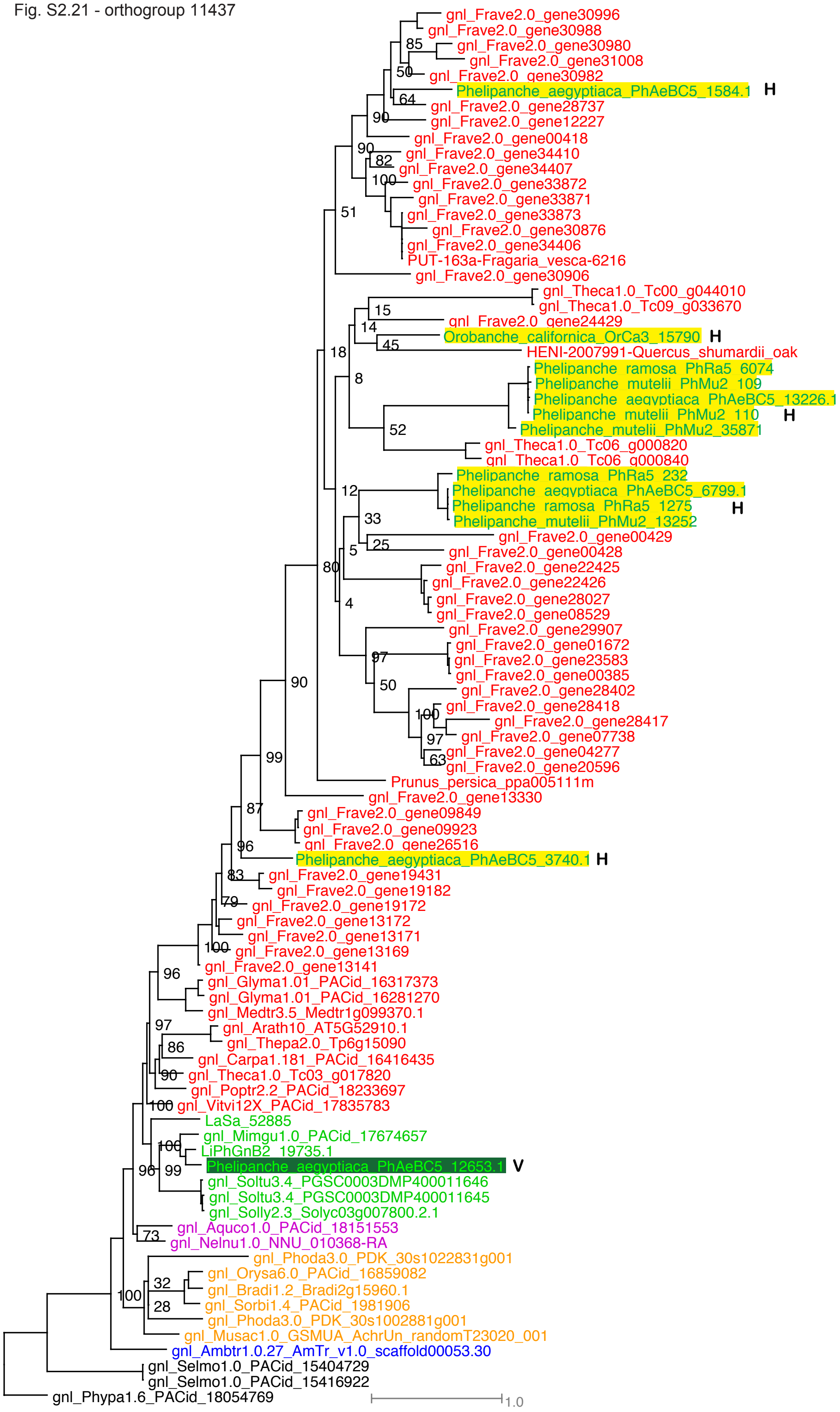


Fig. S2.22 - orthogroup 11841

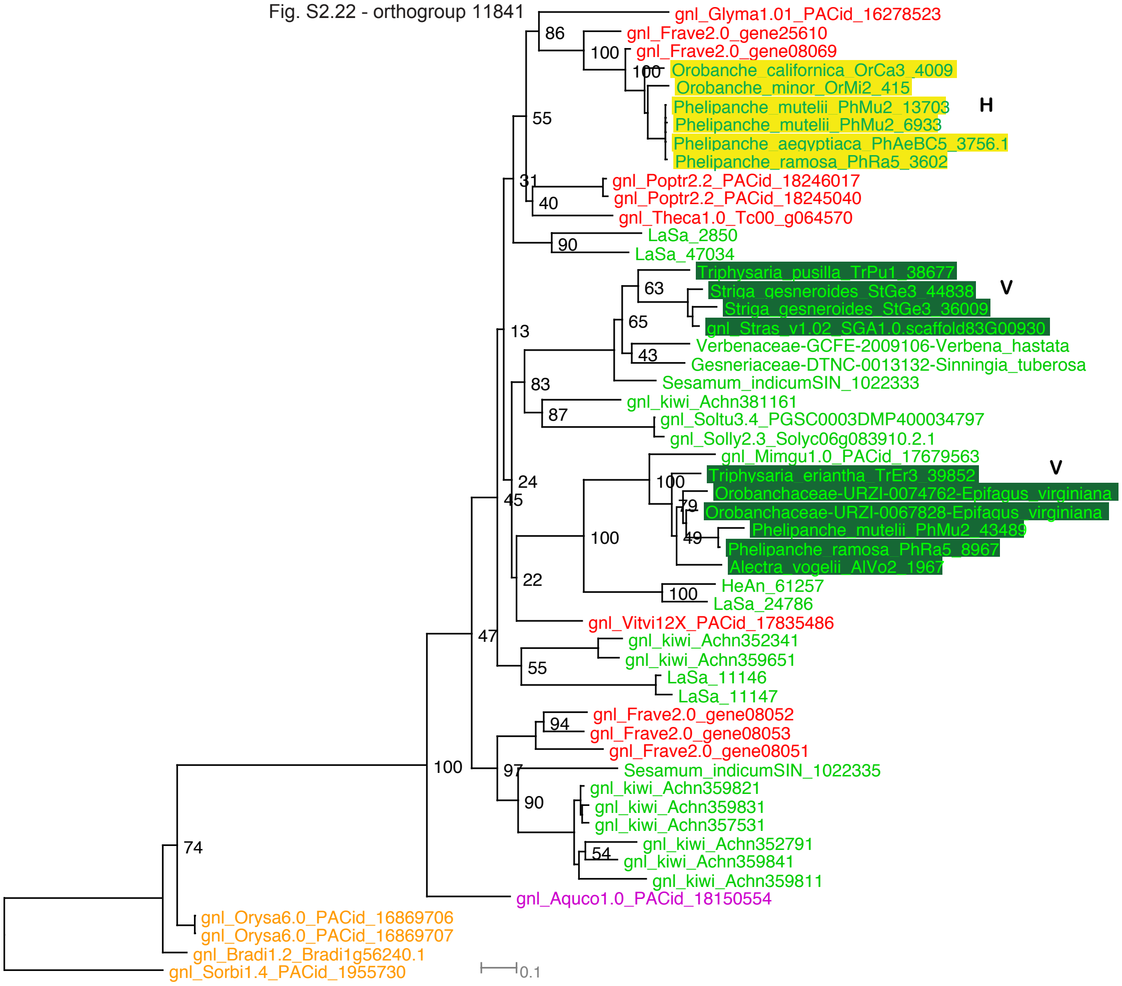


Fig. S2.23 - orthogroup 12835

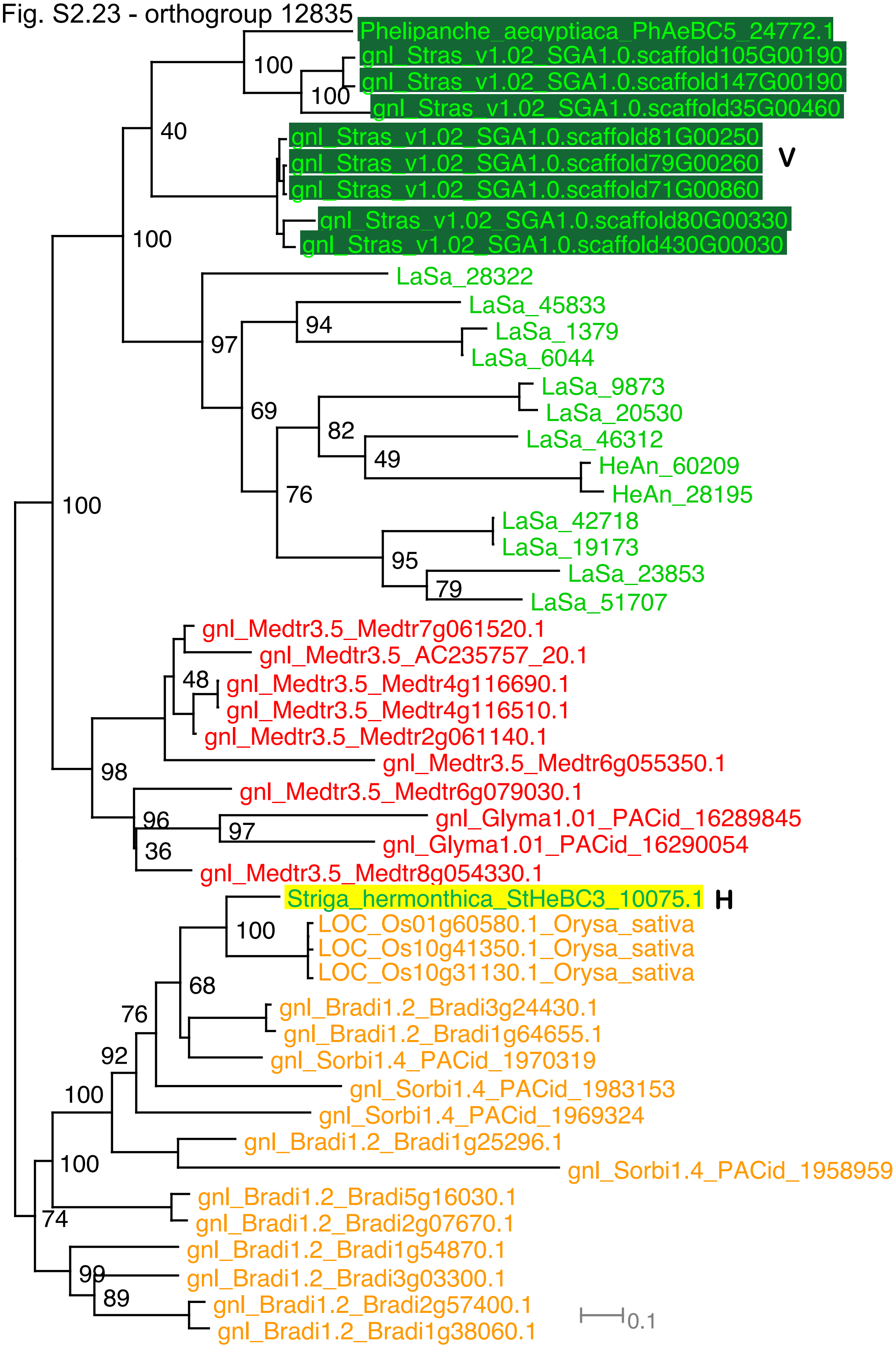


Fig. S2.24 - orthogroup 13512

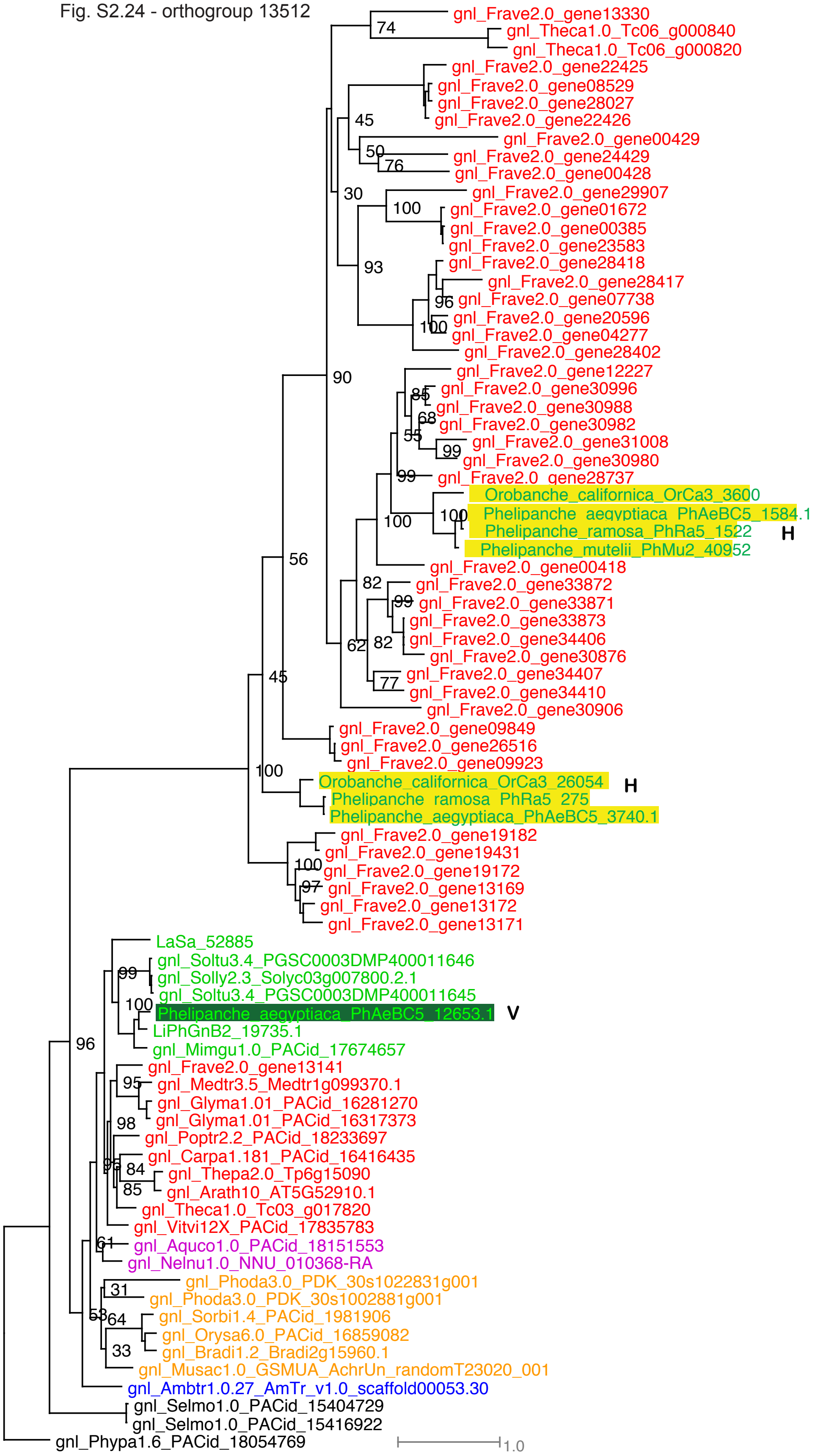


Fig. S2.25 - orthogroup 13656

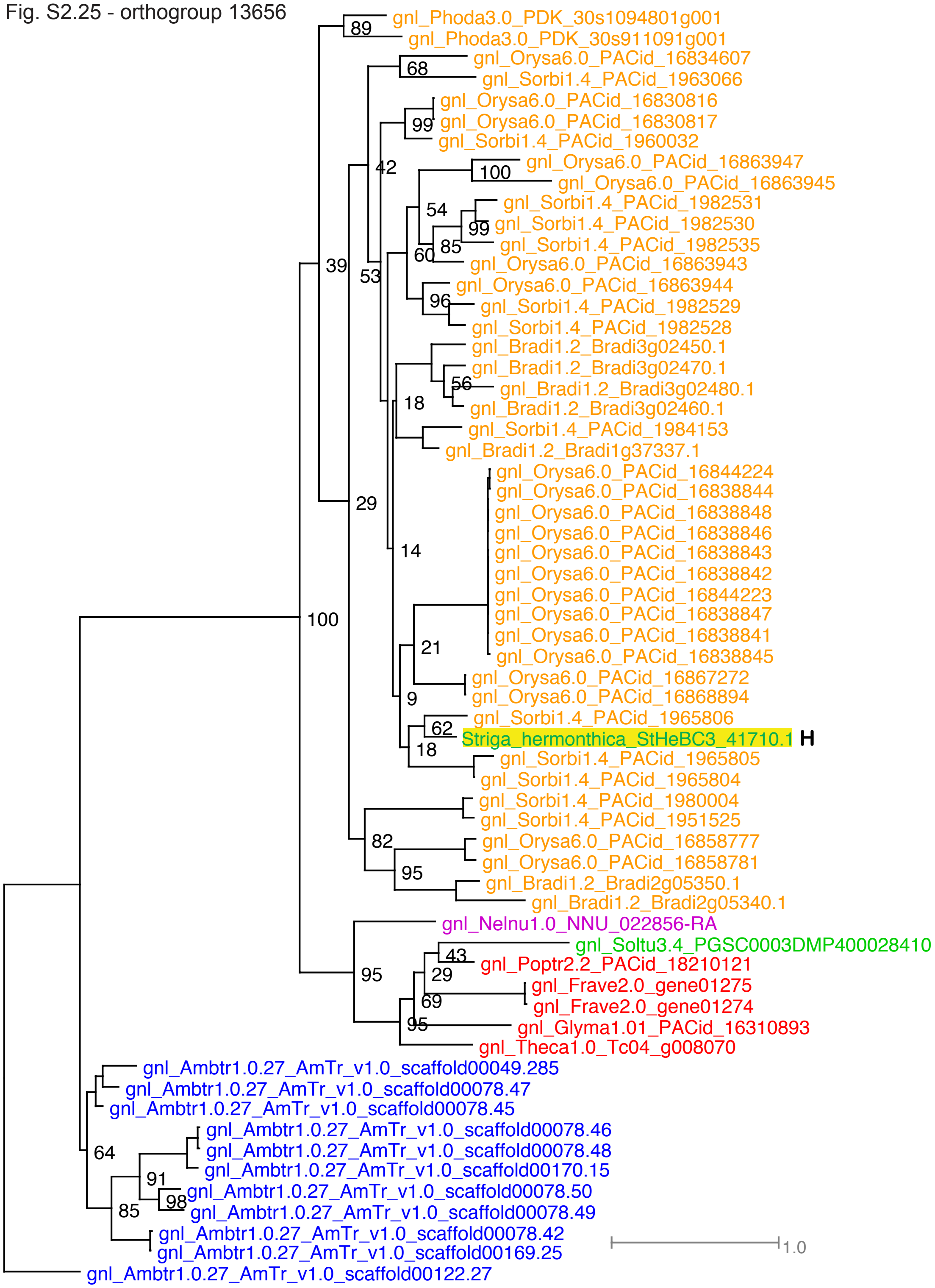


Fig. S2.26 - orthogroup 13892

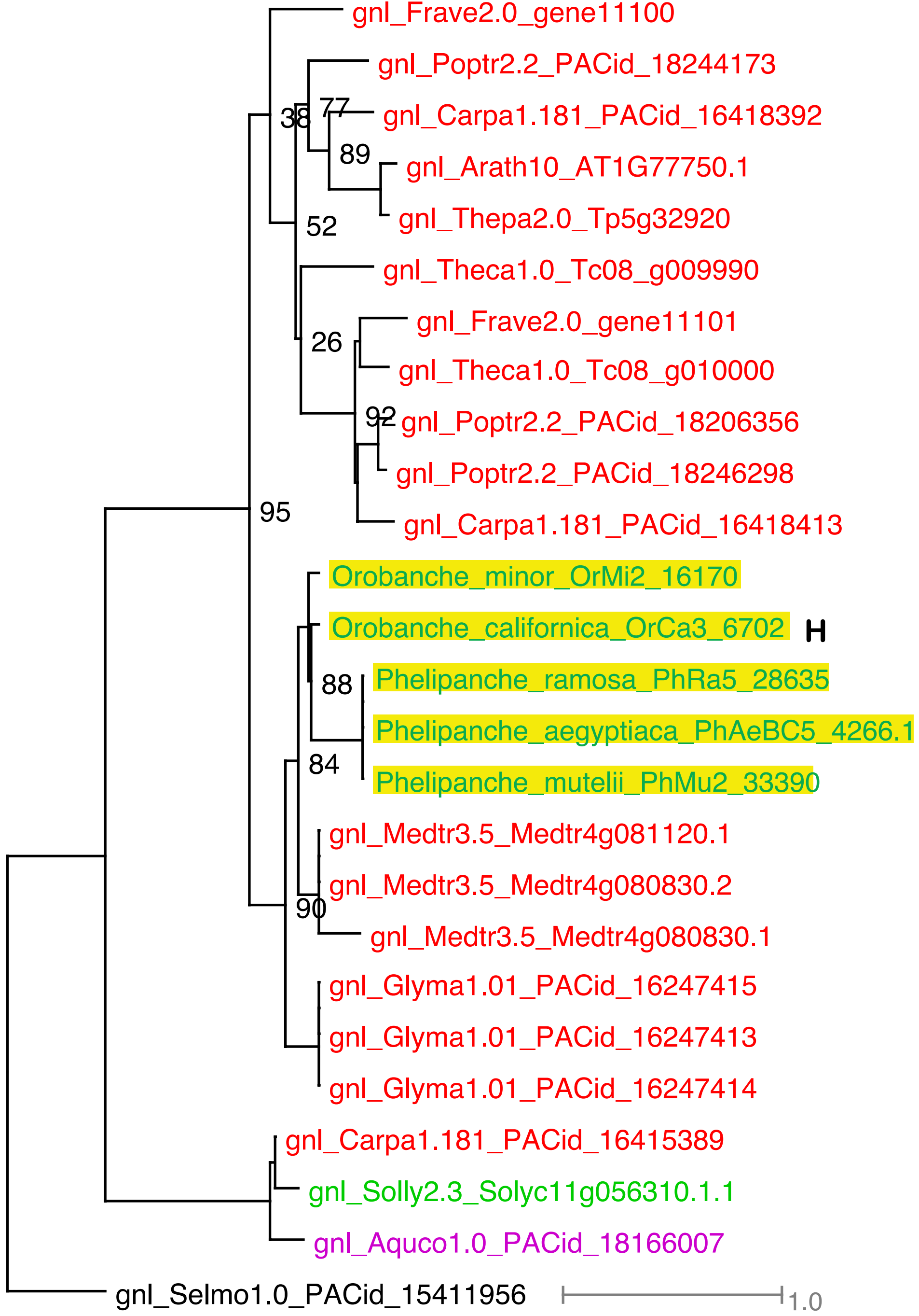


Fig. S2.27 - orthogroup 14230

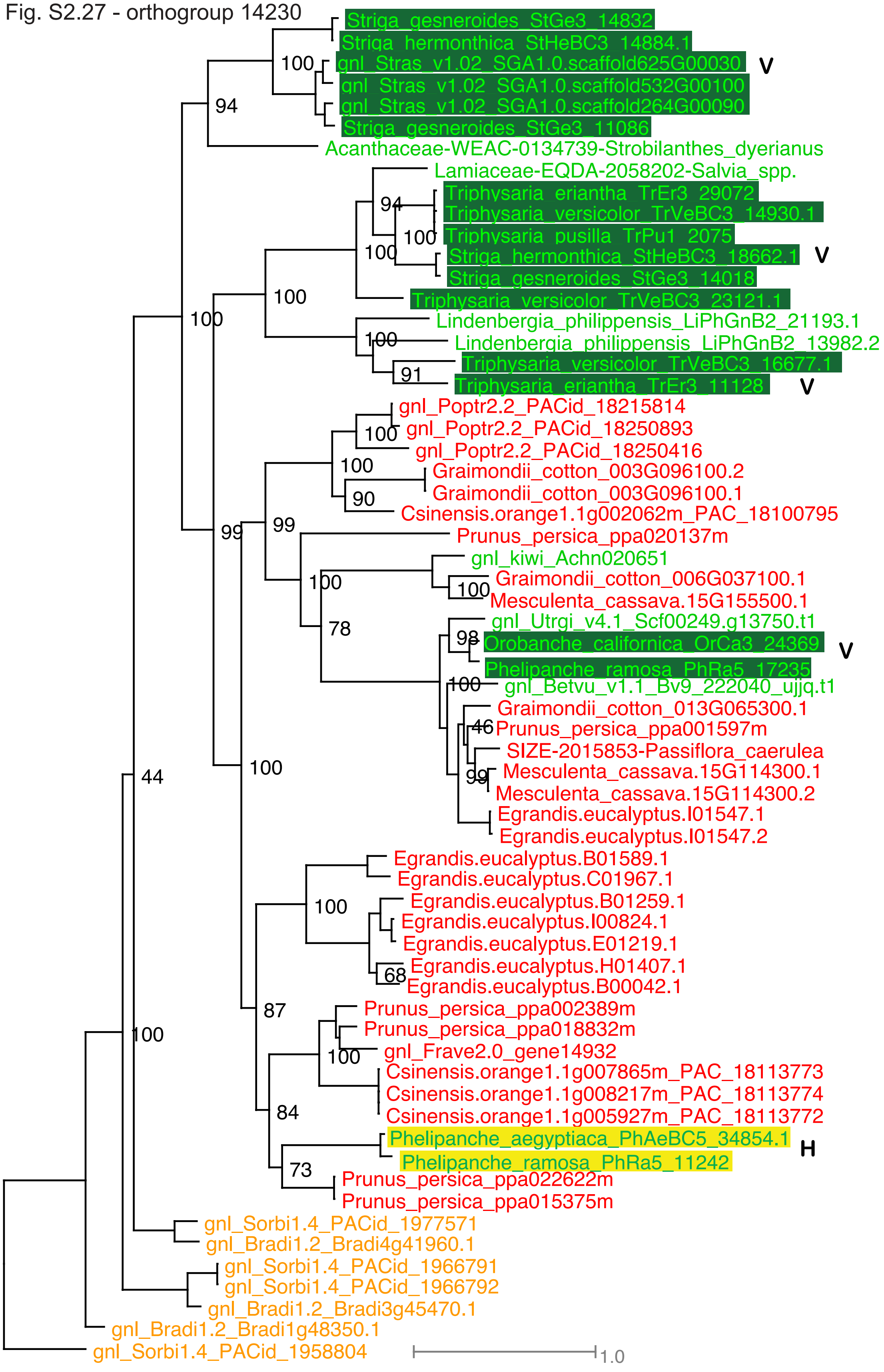


Fig. S2.28 - orthogroup 14233

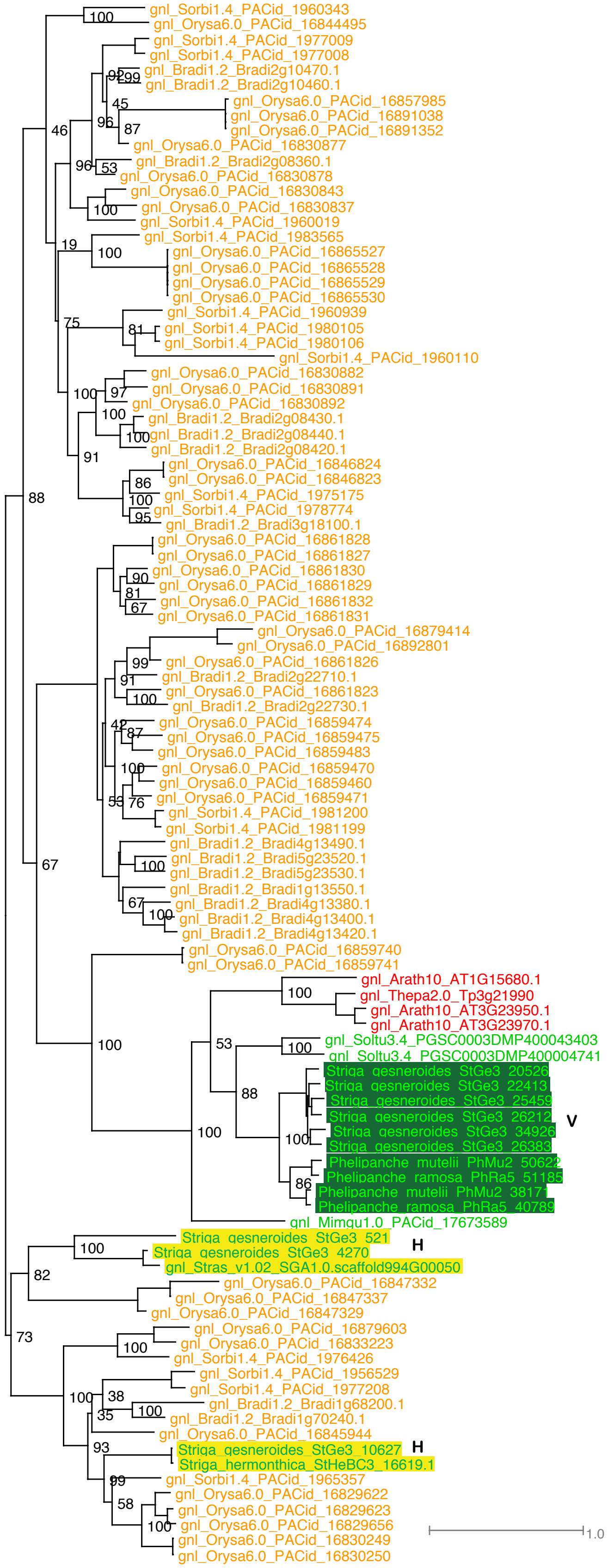


Fig. S2.29- orthogroup 14624

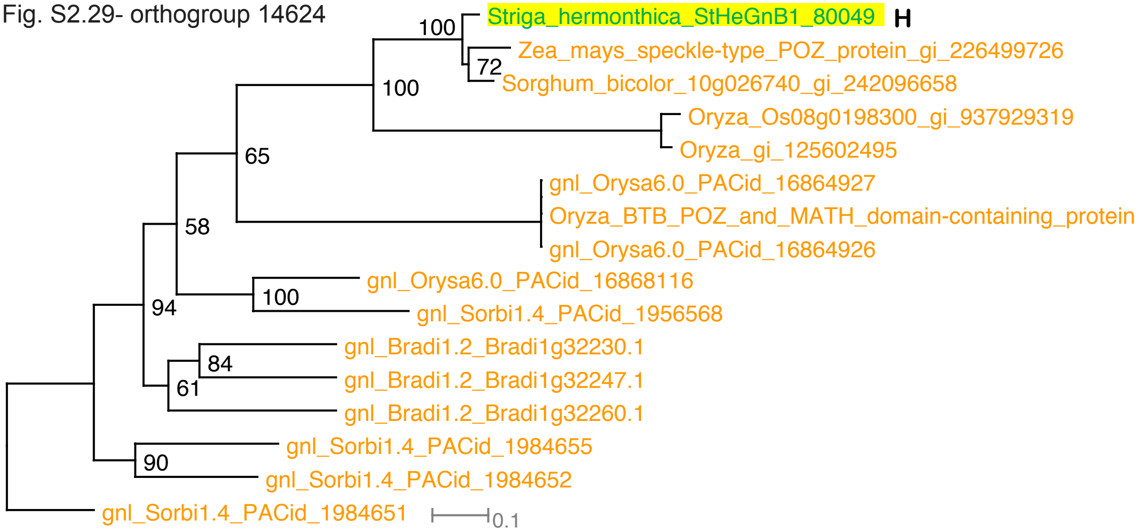


Fig. S2.30 - orthogroup 14675

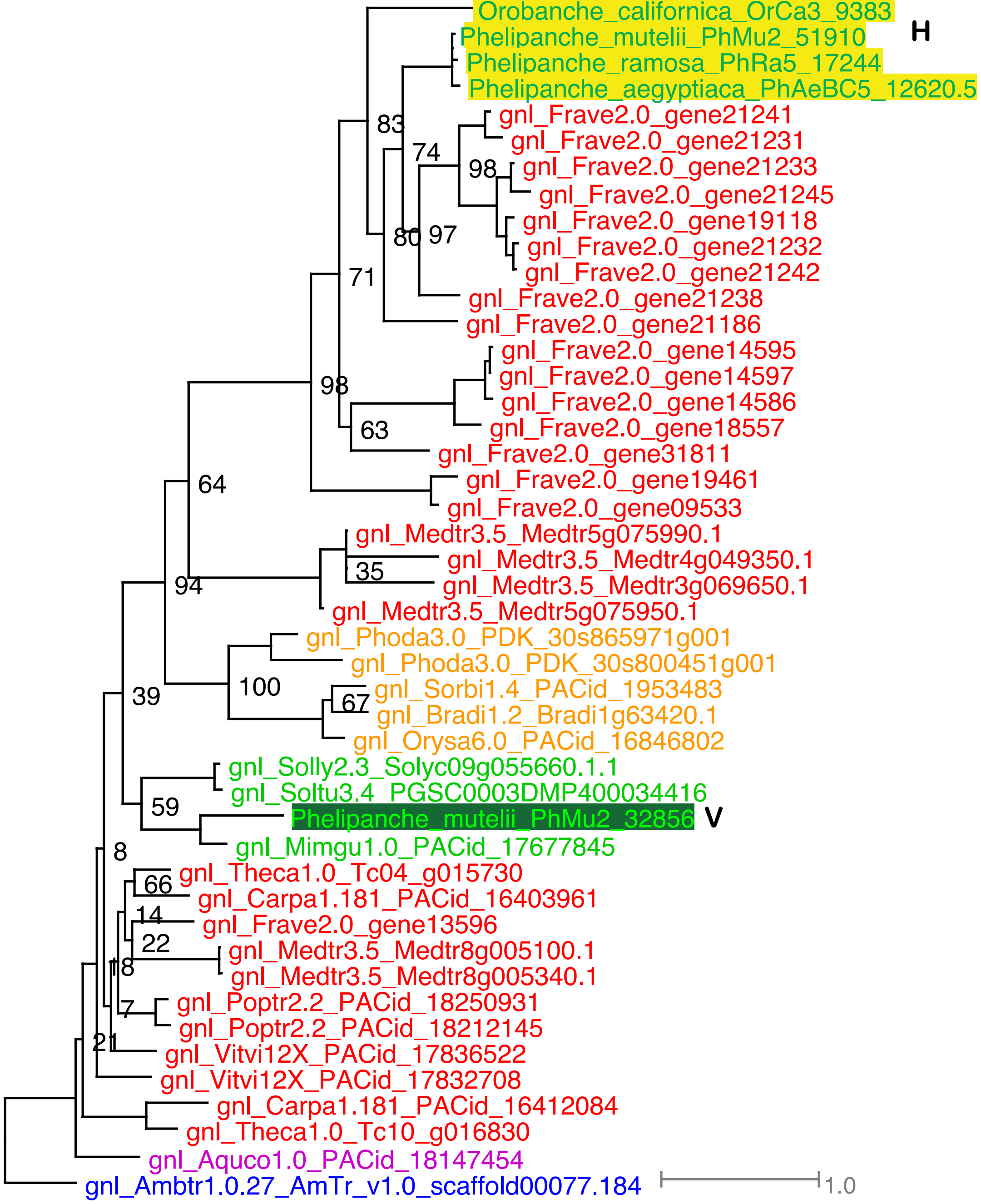


Fig. S2.31 - orthogroup 15149

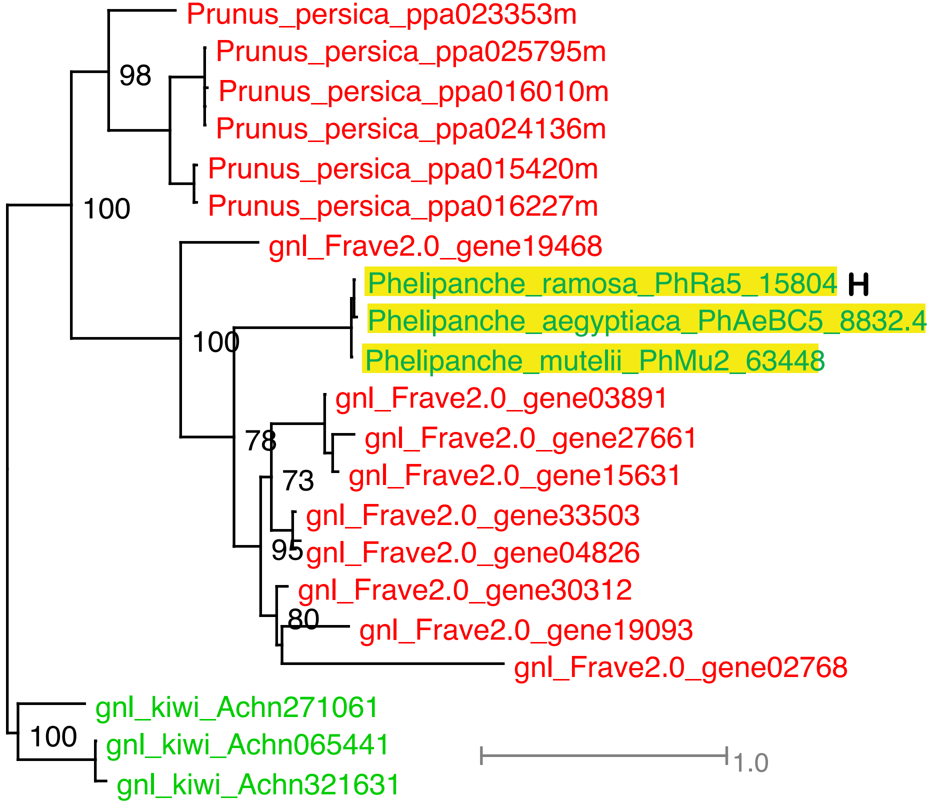


Fig. S2.32 - orthogroup 15246

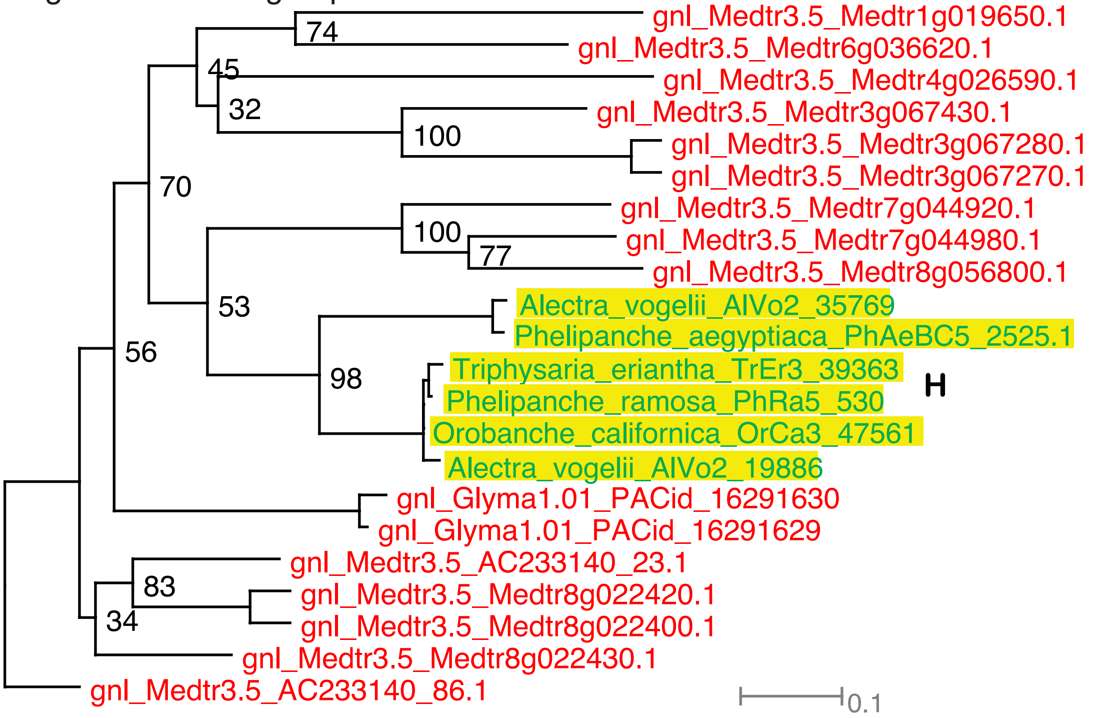


Fig. S2.33 - orthogroup 16703

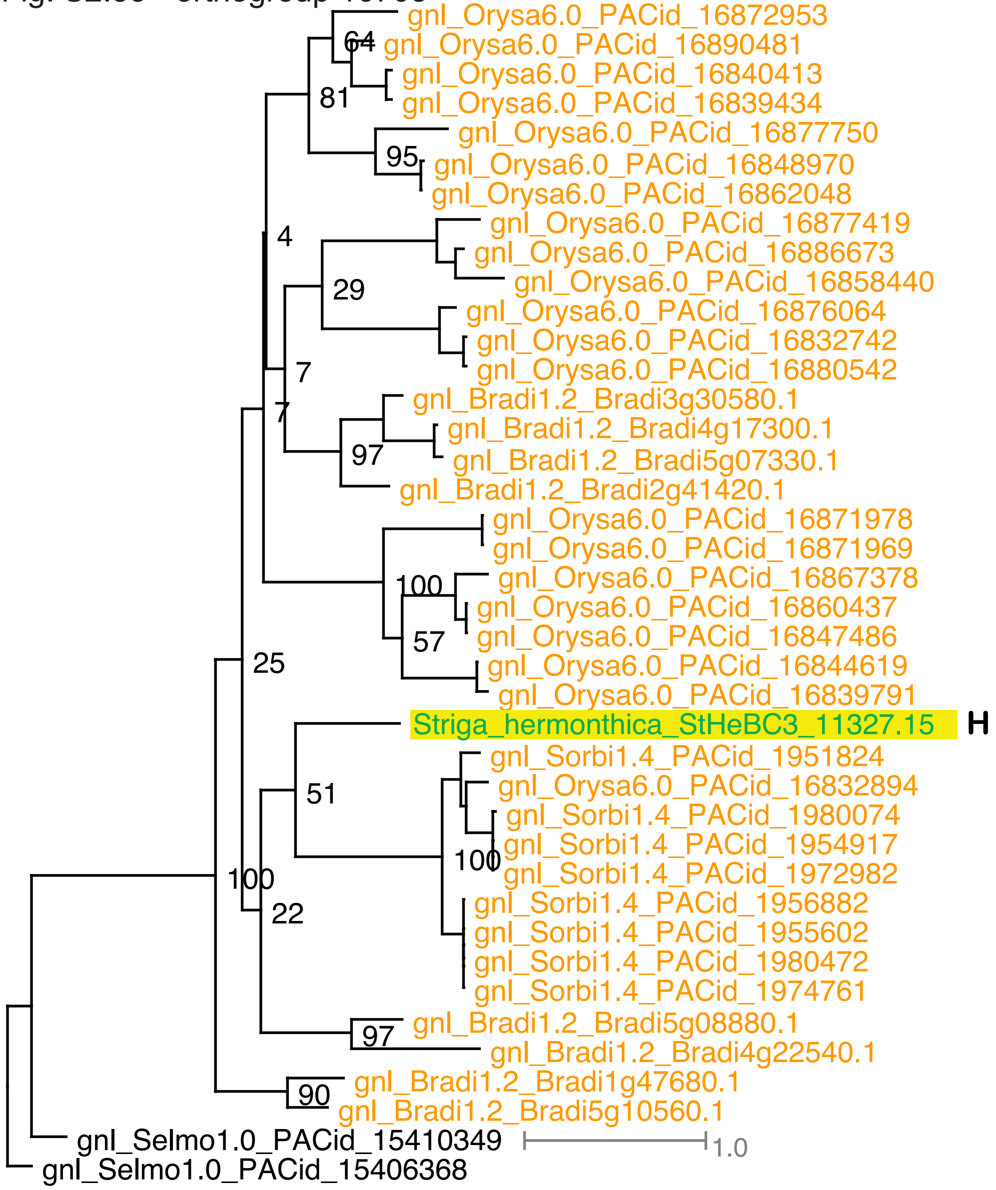


Fig. S2.34 - orthogroup 18354

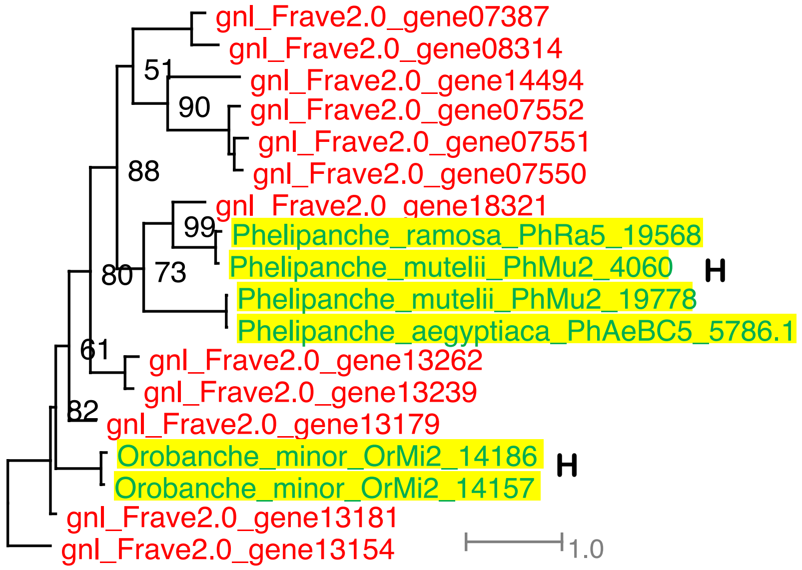


Fig. S2.35 - orthogroup 18709

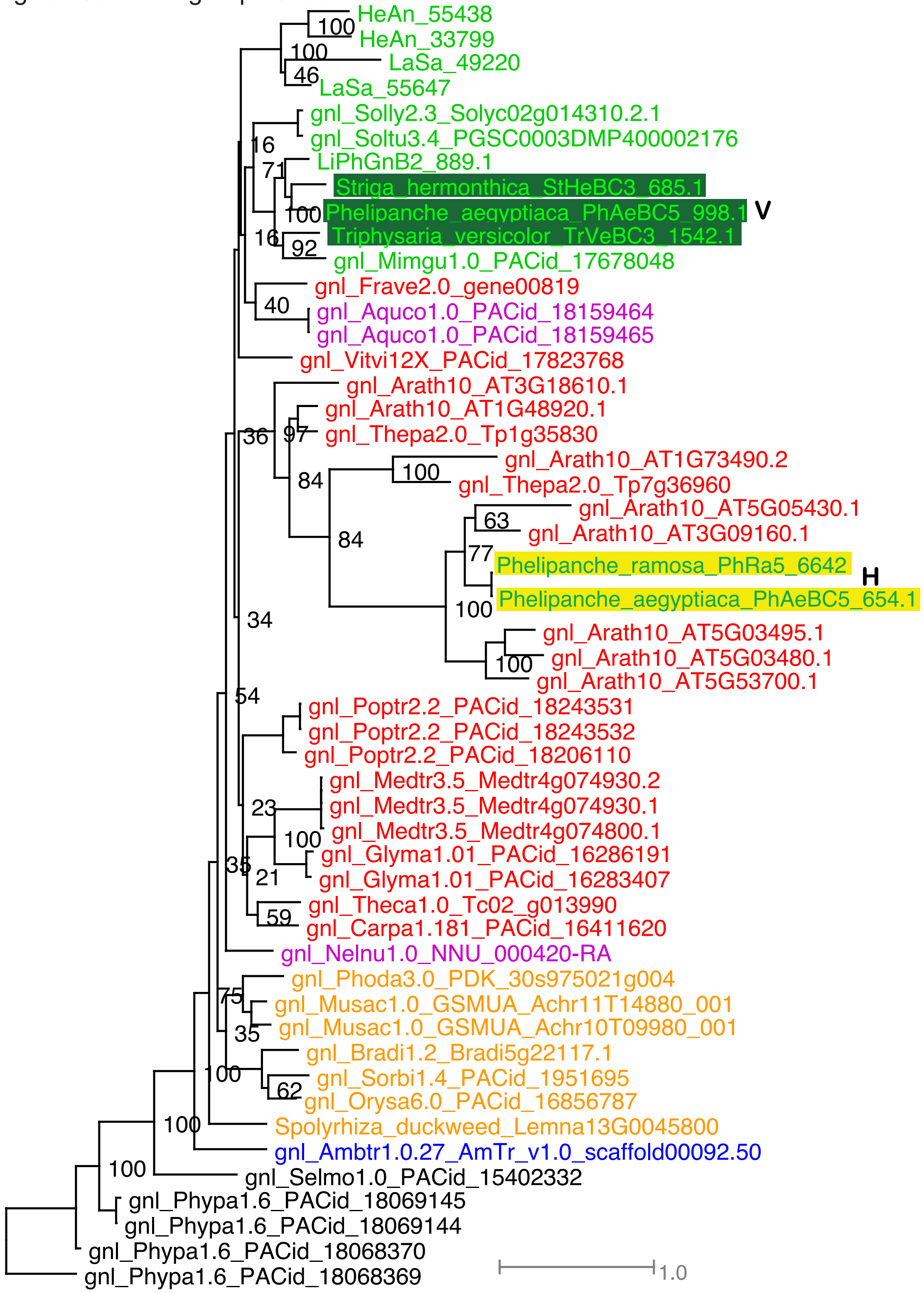


Fig. S2.36 - orthogroup 18774

Fig. S2.37 - orthogroup 19297

Fig. S2.38 - orthogroup 19696

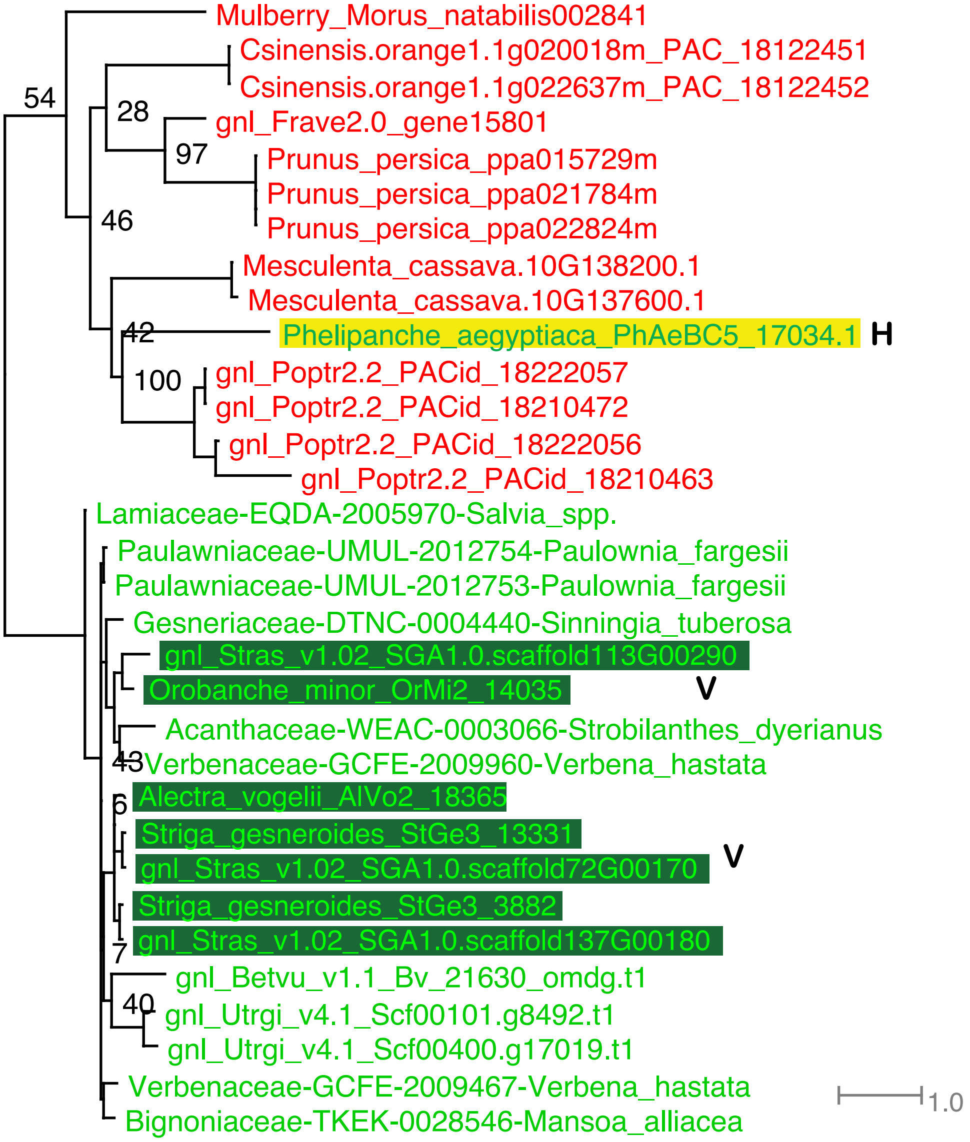


Fig. S2.39 - orthogroup 20188

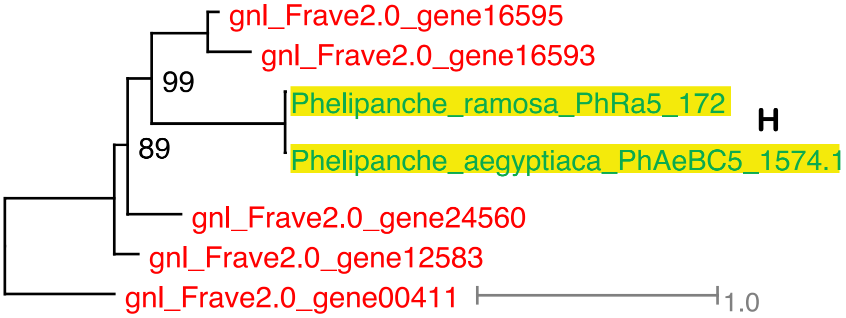


Fig. S2.40 - orthogroup 20190

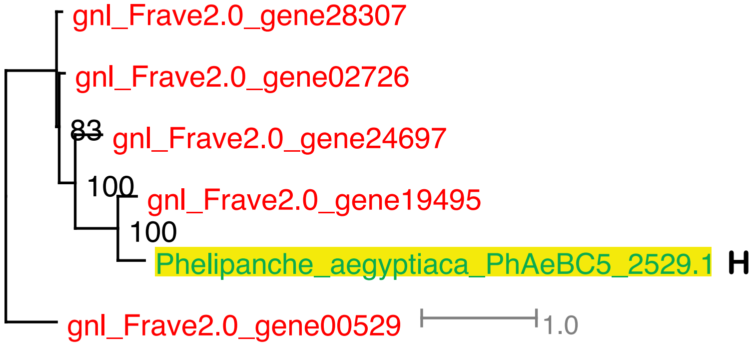


Fig. S2.41 - orthogroup 23343

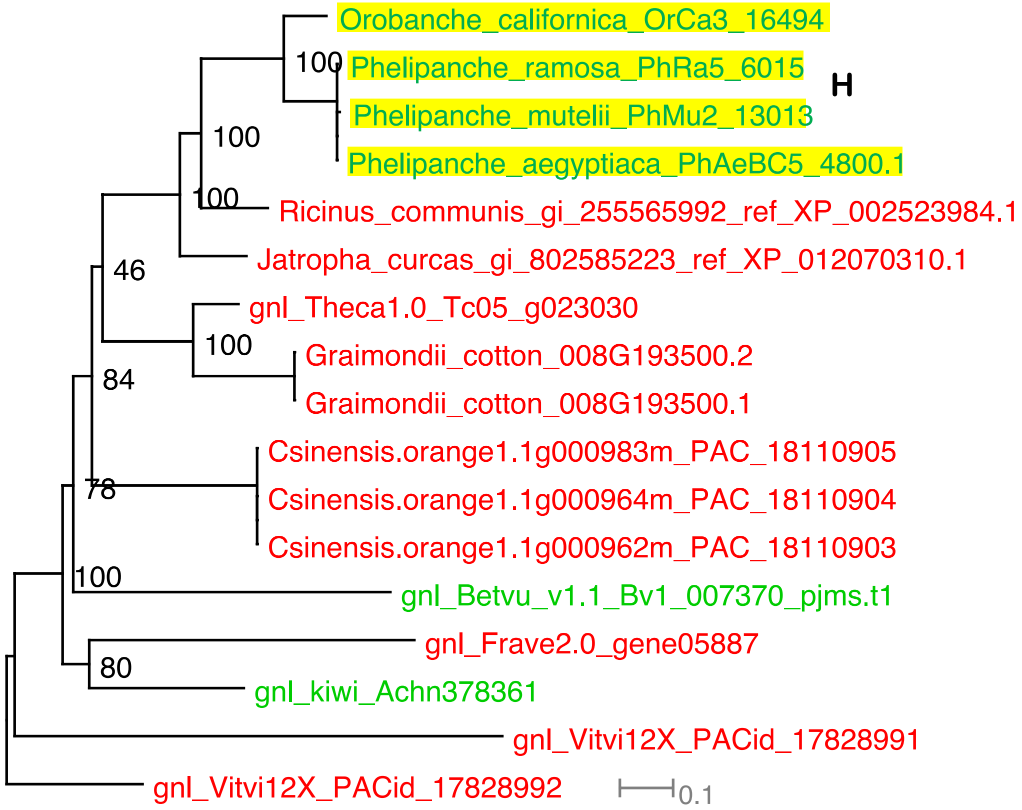


Fig. S2.42 - orthogroup 23480

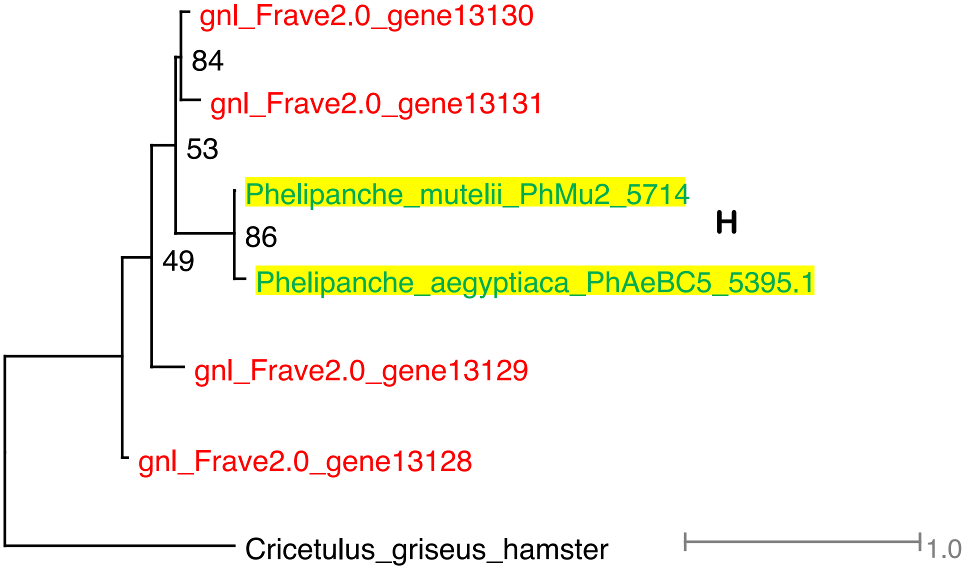


Figure S2. RAxML-based maximum likelihood trees for 42 HGT orthogroups. Yellow highlighting and "H" symbol refer to the HGT clade, green highlighting and "V" symbol refer to the vertical clade. The species are represented by "gnl_" followed by the first three letters of the genus and first two letters of the species name, which can also be referred to in Fig. S1.

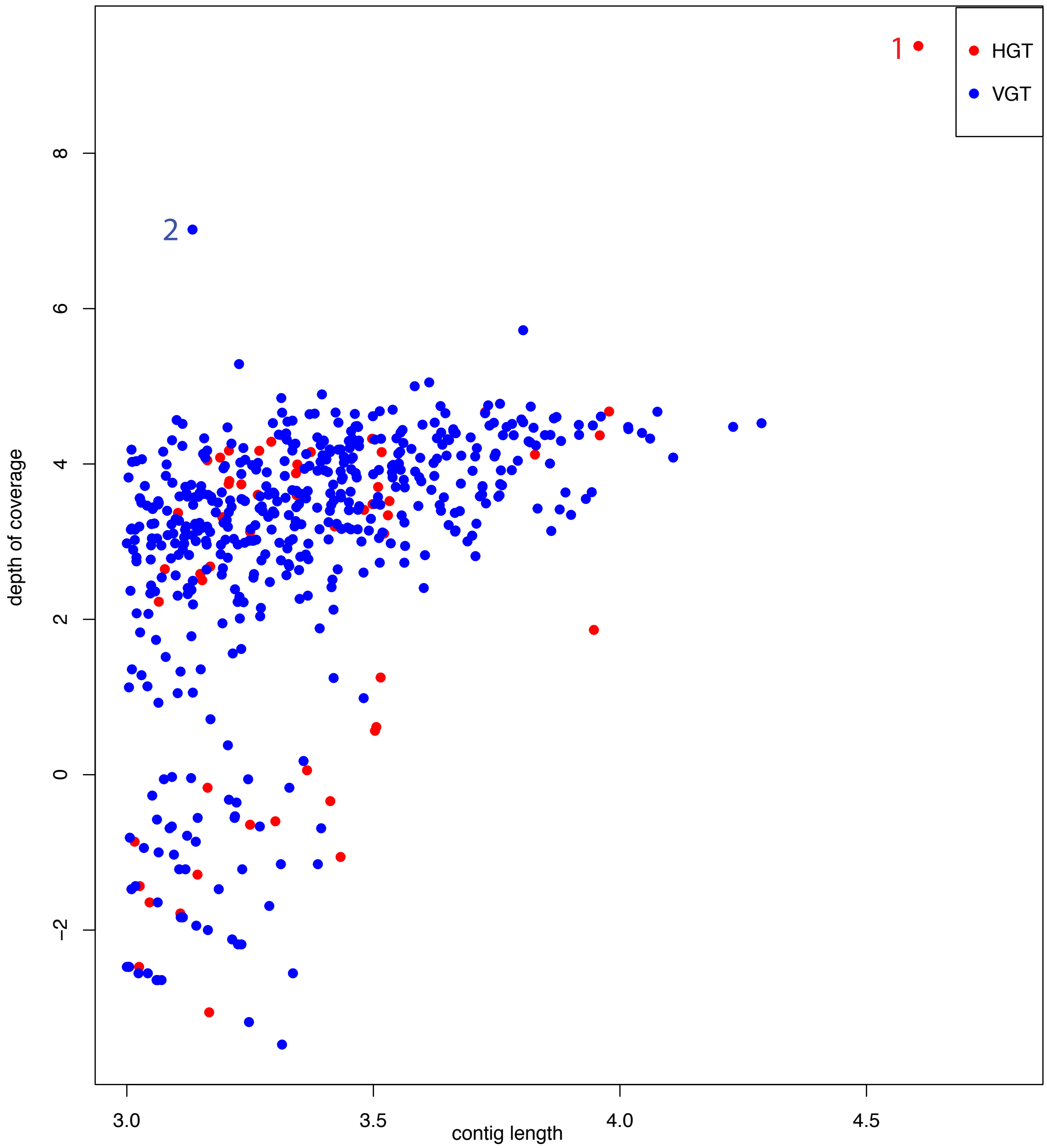


Figure S3. A stoichiometry plot showing the coverage of HGT genomic contigs and vertically transmitted genomic contigs. The X-axis is the log10 transformed contig length (to remove noise, only contigs longer than 1kb were included), and the Y-axis is the log2 transformed read deapth of coverage. This analysis includes 56 HGT-derived genomic contigs, and 493 vertically transmitted gneomic contigs from all three parasitic plants. Two high-coverage genomic contigs are labeled to show outliers: 1 is a mitochondrial genomic contig with much greater sequence depth than nuclear contigs, 2 is a highly-repetitive sequence.

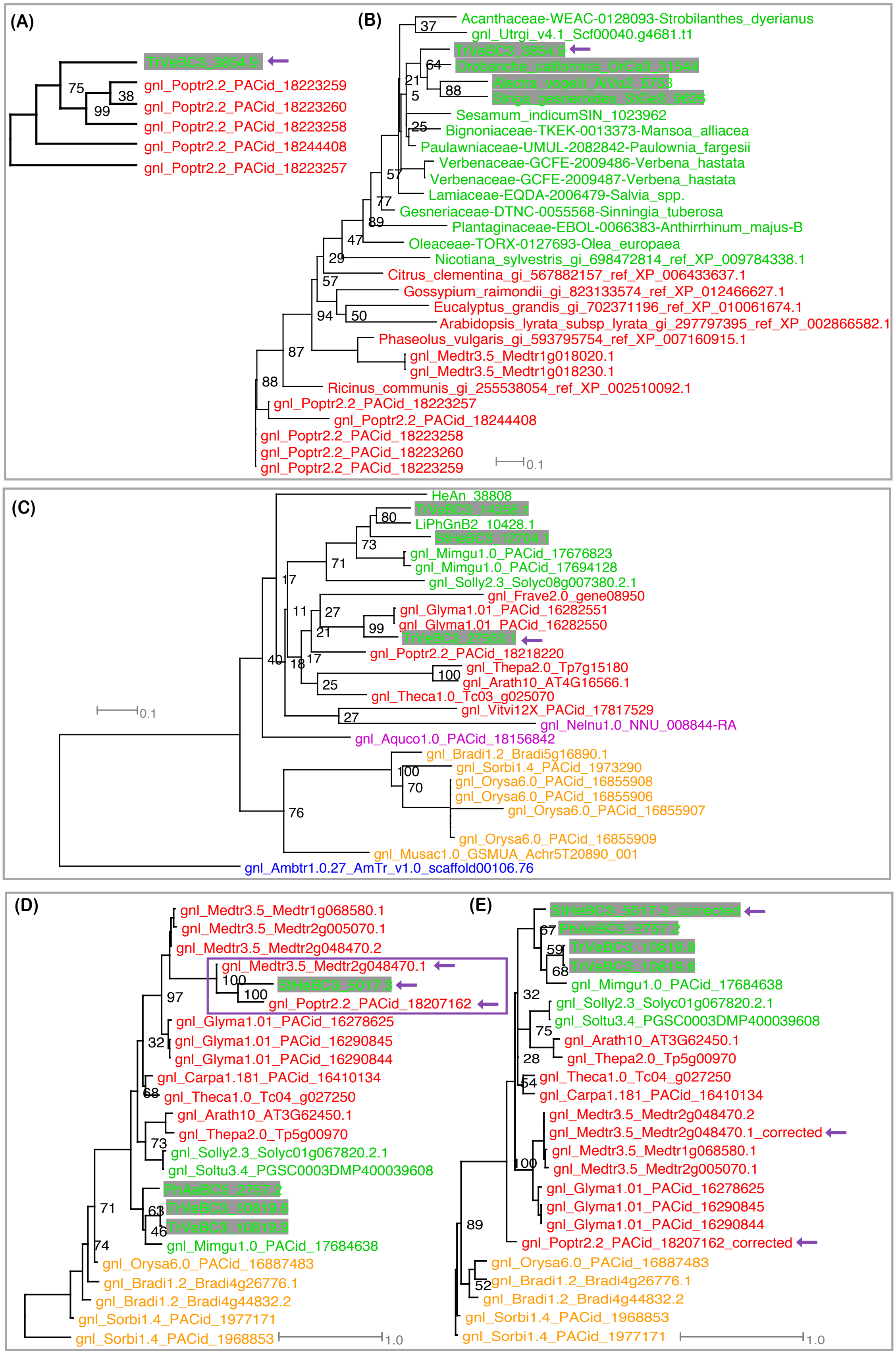
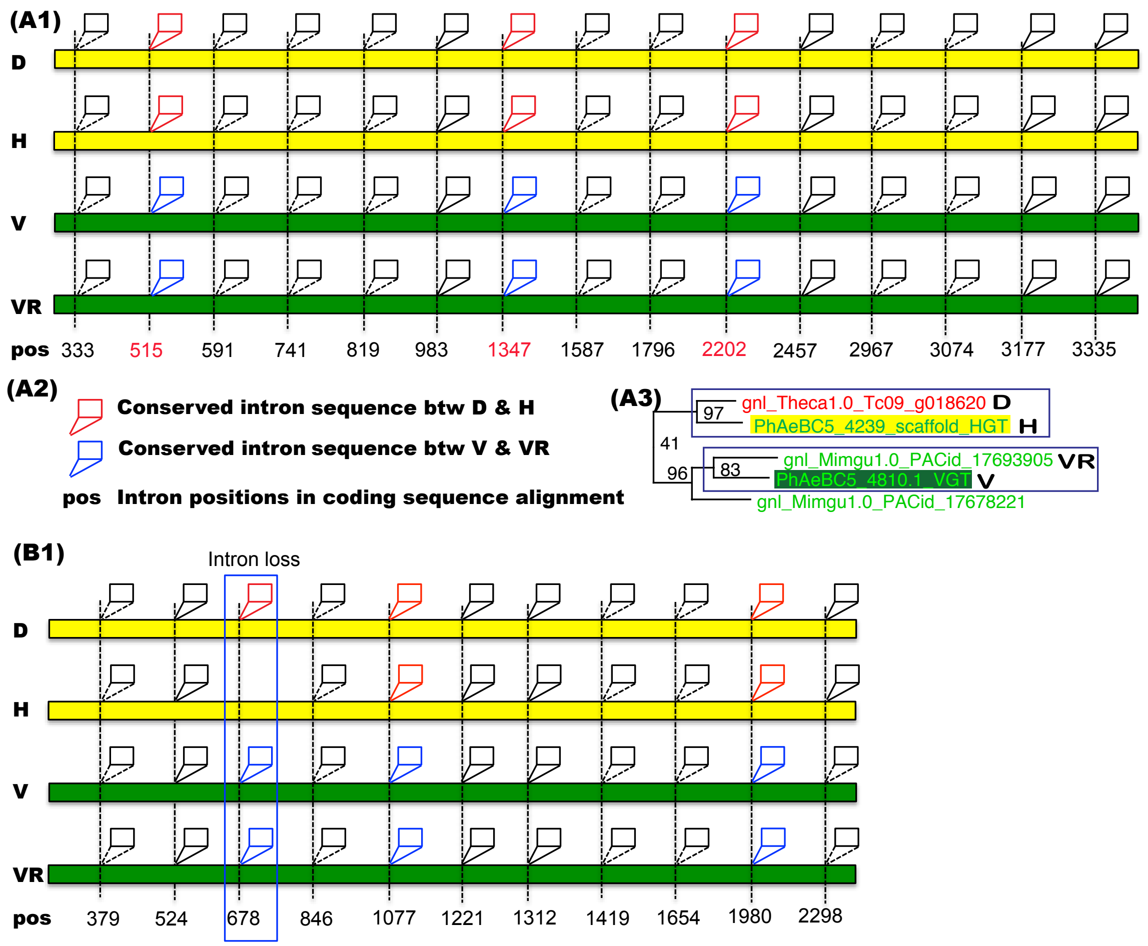
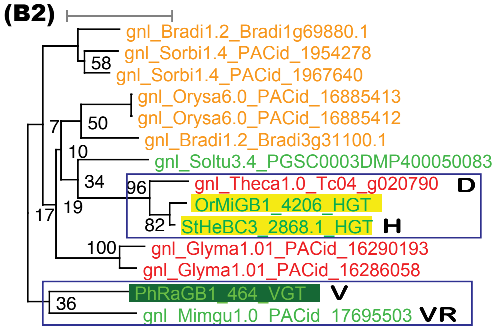

(A)
TrVeBC3_3854.9
gnl_Poptr2.2_PACid_18223259
gnl_Poptr2.2_PACid_18223260
gnl_Poptr2.2_PACid_18223258
gnl_Poptr2.2_PACid_18244408
gnl_Poptr2.2_PACid_18223257
75
38
99
(B)
Acanthaceae-WEAC-0128093-Strobilanthes_dyerianus
gnl_Utrgi_v4.1_Scf00040.g4681.t1
TrVeBC3_3854.9
Orobanche_californica_OrCa3_31544
Alectra_vogelii_AlVo2_5753
Striga_gesneroides_StGe3_9626
Sesamum_indicumSIN_1023962
Bignoniaceae-TKEK-0013373-Mansoa_alliacea
Paulawniaceae-UMUL-2082842-Paulownia_fargesii
Verbenaceae-GCFE-2009486-Verbena_hastata
Verbenaceae-GCFE-2009487-Verbena_hastata
Lamiaceae-EQDA-2006479-Salvia_spp.
Gesneriaceae-DTNC-0055568-Sinningia_tuberosa
Plantaginaceae-EBOL-0066383-Anthirrhinum_majus-B
Oleaceae-TORX-0127693-Olea_europaea
Nicotiana_sylvestris_gi_698472814_ref_XP_009784338.1
Citrus_clementina_gi_567882157_ref_XP_006433637.1
Gossypium_raimondii_gi_823133574_ref_XP_012466627.1
Eucalyptus_grandis_gi_702371196_ref_XP_010061674.1
Arabidopsis_lyrata_subsp_lyrata_gi_297797395_ref_XP_002866582.1
Phaseolus_vulgaris_gi_593795754_ref_XP_007160915.1
gnl_Medtr3.5_Medtr1g018020.1
gnl_Medtr3.5_Medtr1g018230.1
Ricinus_communis_gi_255538054_ref_XP_002510092.1
gnl_Poptr2.2_PACid_18223257
gnl_Poptr2.2_PACid_18244408
gnl_Poptr2.2_PACid_18223258
gnl_Poptr2.2_PACid_18223260
gnl_Poptr2.2_PACid_18223259
37
64
21
5
88
25
57
77
89
47
29
57
94
50
87
88
0.1
(C)
HeAn_38808
TrVeBC3_14358.1
LiPhGnB2_10428.1
StHeBC3_12704.1
gnl_Mimgu1.0_PACid_17676823
gnl_Mimgu1.0_PACid_17694128
gnl_Solly2.3_Solyc08g007380.2.1
gnl_Frave2.0_gene08950
gnl_Glyma1.01_PACid_16282551
gnl_Glyma1.01_PACid_16282550
TrVeBC3_27563.1
gnl_Poptr2.2_PACid_18218220
gnl_Thepa2.0_Tp7g15180
gnl_Arath10_AT4G16566.1
gnl_Theca1.0_Tc03_g025070
gnl_Vitvi12X_PACid_17817529
gnl_Nelnu1.0_NNU_008844-RA
gnl_Aquco1.0_PACid_18156842
gnl_Bradi1.2_Bradi5g16890.1
gnl_Sorbi1.4_PACid_1973290
gnl_Orysa6.0_PACid_16855908
gnl_Orysa6.0_PACid_16855906
gnl_Orysa6.0_PACid_16855907
gnl_Orysa6.0_PACid_16855909
gnl_Musac1.0_GSMUA_Achr5T20890_001
gnl_Ambtr1.0.27_AmTr_v1.0_scaffold00106.76
80
73
71
17
11
27
99
21
40
18
17
100
25
27
100
70
76
0.1
(D)
gnl_Medtr3.5_Medtr1g068580.1
gnl_Medtr3.5_Medtr2g005070.1
gnl_Medtr3.5_Medtr2g048470.2
gnl_Medtr3.5_Medtr2g048470.1
StHeBC3_5017.3
gnl_Poptr2.2_PACid_18207162
gnl_Glyma1.01_PACid_16278625
gnl_Glyma1.01_PACid_16290845
gnl_Glyma1.01_PACid_16290844
gnl_Carpa1.181_PACid_16410134
gnl_Theca1.0_Tc04_g027250
gnl_Arath10_AT3G62450.1
gnl_Thepa2.0_Tp5g00970
gnl_Solly2.3_Solyc01g067820.2.1
gnl_Soltu3.4_PGSC0003DMP400039608
PhAeBC5_2757.2
TrVeBC3_10819.6
TrVeBC3_10819.9
gnl_Mimgu1.0_PACid_17684638
gnl_Orysa6.0_PACid_16887483
gnl_Bradi1.2_Bradi4g26776.1
gnl_Bradi1.2_Bradi4g44832.2
gnl_Sorbi1.4_PACid_1977171
gnl_Sorbi1.4_PACid_1968853
100
100
97
32
68
73
71
63
46
74
1.0
(E)
StHeBC3_5017.3_corrected
PhAeBC5_2757.2
TrVeBC3_10819.9
TrVeBC3_10819.6
gnl_Mimgu1.0_PACid_17684638
gnl_Solly2.3_Solyc01g067820.2.1
gnl_Soltu3.4_PGSC0003DMP400039608
gnl_Arath10_AT3G62450.1
gnl_Thepa2.0_Tp5g00970
gnl_Theca1.0_Tc04_g027250
gnl_Carpa1.181_PACid_16410134
gnl_Medtr3.5_Medtr2g048470.2
gnl_Medtr3.5_Medtr2g048470.1_corrected
gnl_Medtr3.5_Medtr1g068580.1
gnl_Medtr3.5_Medtr2g005070.1
gnl_Glyma1.01_PACid_16278625
gnl_Glyma1.01_PACid_16290845
gnl_Glyma1.01_PACid_16290844
gnl_Poptr2.2_PACid_18207162_corrected
gnl_Orysa6.0_PACid_16887483
gnl_Bradi1.2_Bradi4g26776.1
gnl_Bradi1.2_Bradi4g44832.2
gnl_Sorbi1.4_PACid_1968853
gnl_Sorbi1.4_PACid_1977171
67
59
68
32
75
28
54
100
89
52
1.0

Figure S4. HGT artifacts due to insufficient taxon sampling (A-B), host plant contamination (C), and frame-shift errors (D-E). RAxML-based maximum likelihood tree for orthogroup 20348 from the initial automated pipeline (A) shows the placement of a *Triphysaria* gene (purple arrow; TrVeBC3_3854.9) as a member of a rosid clade composed of many *Populus* sequences (Poptr), indicative of a potential HGT. (B) Increase of taxon sampling in the Lamiales order (green non-shaded) and the rosid groups (red) identifies the *Triphysaria* sequence (TrVeBC3_3854.9) as a vertically transmitted gene which groups with its closely related parasitic taxa (green foreground, grey shading) in Lamiales. (C) Discovery of an unannotated host plant gene, rather than HGT (orthogroup 8224), explains the placement of the *Triphysaria* gene (TrVeBC3_27563.1) with *Glycine max* (Glyma) sequences, a close relative of its experimental host – *Medicago truncatula*. (D-E) show an HGT artifact due to frame-shift errors in orthogroup 20348. (D) Purple branch shows a clade with long branch composed of three sequences – StHeBC3_5017.3, gnl_Medtr3.5_Medtr2g048470.1, and gnl_Poptr2.2_PACid_18207162. *Striga* sequence was identified in preliminary screening as a potential HGT sequence because of strong support for placement with a rosid donor. Careful examination of the alignment revealed frame-shift errors in all these three sequences, and correction of this error resolved the *Striga* gene as a vertical gene (E). The two rosid sequences from *Poplar* and *Medicago* also went to their expected positions within the rosid clades without long branches (E).

Figure S5. The conserved 5'-UTR intron shared between the donor represented by *Sorghum bicolor* gene Sb10g026740 and the HGT *Striga hermonthica* gene StHeGnB1_80049 (orthogroup 14624). Geneious alignment view of the sequence alignment between the donor and recipient gene for 5'-UTR region (A), 5'-UTR intron (B), CDS (C), and 3'-UTR (D). The 5'-UTR intron was revealed by read mapping data.

Figure S6. The conserved 3’-UTR intron shared between the donor represented by *Sorghum bicolor* gene Sb01g013250 and the HGT *Striga hermonthica* gene ShContig9483 (orthogroup 19297). Geneious alignment view of the sequence alignment between the donor and recipient gene for 5’-UTR region (A), CDS (B), 3’-UTR (C), and 3’-UTR intron (D). A shortened intron revealed in the parasitic *Striga* HGT gene, as shown in (C) where the N-terminal intron of the host gene resulted in unspliced transcribed mRNA in the 3’-UTR region. The 3’-UTR intron was revealed by read mapping data.





(C)

orthogroup	annotation	# of CIPs
806	valyl-tRNA synthetase	15
2270	methionyl-tRNA synthetase	11
4067	tRNAHis guanylyl-transferase	9

Figure S7. Intron phylogeny and intron-positions for three HGT orthogroups that encode tRNA synthetase/transferase. (A) and (B) show the conserved intron positions and intron sequence comparisons between two pairs of sequences: H (the HGT gene) and D (the represented donor), V (the vertical gene) and VR (the vertical sequence relative) (*Mimulus* sequence) (corresponding coding sequence phylogeny are shown in Fig. S1. (A1) (orthogroup 806) and (B1) (orthogroup 2270) show coding sequence alignment with introns marked with boxes. (A3) and (B2) show the intron sequence phylogenies supporting strong sequence similarity of two pairs, the HGT gene and its donor, the vertical gene and its vertical relative in *Mimulus*. Intron sequence phylogeny were made by the conserved introns colored in red and blue of (A1) and (B1). The intron origins match the coding sequence origin, supporting genomic transfers. (C) Annotation of three HGT orthogroups, all encoding tRNA synthetase/tranferases involved in the attachment of the codon and tRNA anticodon. Also, the number of conserved intron positions (CIPs) in the forced CDS alignment were also shown, 4067 is the orthogroup in main Figure 3 that also encodes tRNA^{His} transferase.

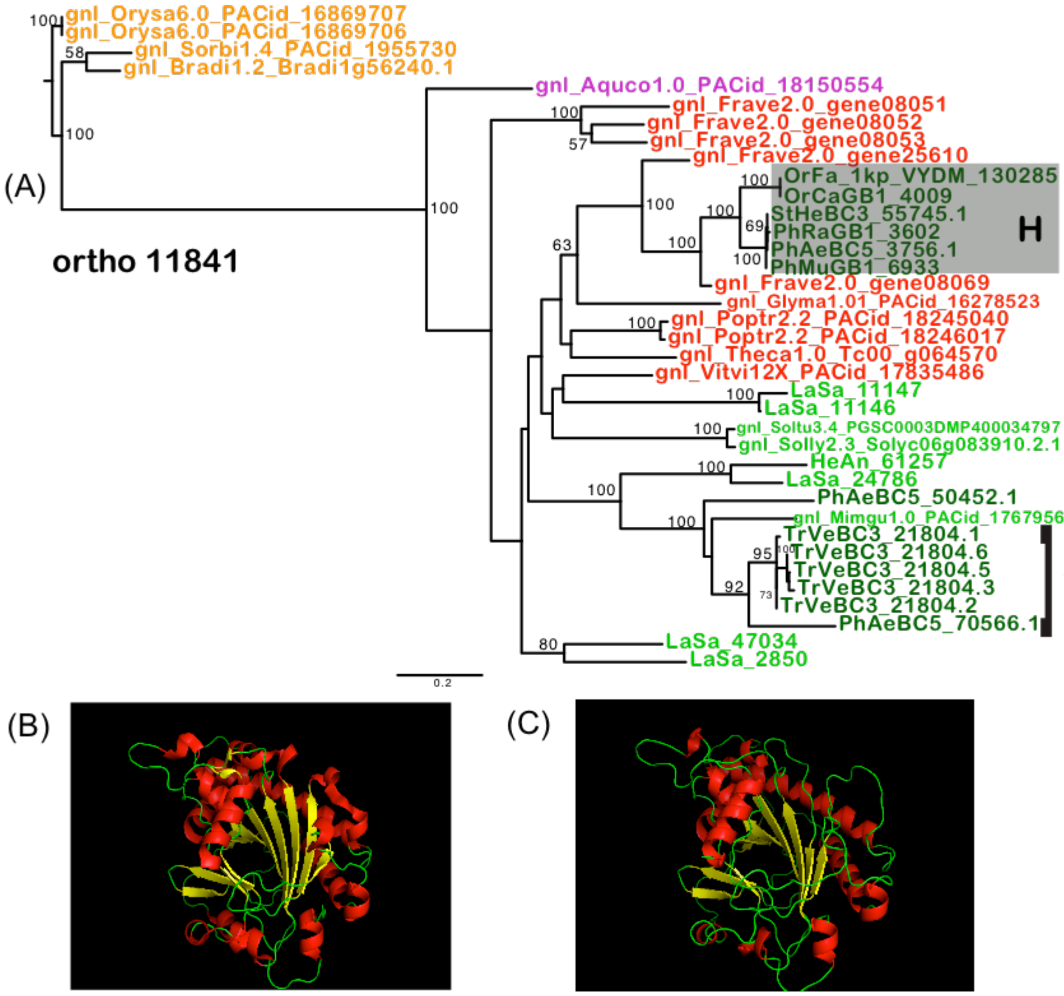


Figure S8. Phylogeny of orthogroup 11841 and predicted 3D structure.
The HGT genes identified are PhAeBC5_3756.1 in *Phelipanche aegyptiaca* and StHeBC3_55745.1 in *Striga hermonthica*. For this particular orthogroup, the HGT gene was also identified in *Orobanche fasciculata* (OrFa_1kp_VYDM_130285) from the 1KP database. Additional Orobanchaceae orthologs were also found through shallow transcriptome sequencing of additional species, including *Myzorrhiza californica* (OrCaGB1_4009), *Phelipanche ramosa* (PhRaGB1_3602) and *Phelipanche mutelii* (PhMuGB1_6933). HGT and vertical clade was labeled with "H" and "V" respectively on the phylogeny. (A). phylogeny of orthogroup 11841 supporting HGT; (B). Protein 3D structure of Anthocyanidin synthase from *Arabidopsis thaliana* (PDB ID: 1GP4); (C). Predicted protein 3D structure of PhAeBC5_3756.1. Color scheme: alpha helixes shown in red; beta sheets shown in yellow; loops shown in green.

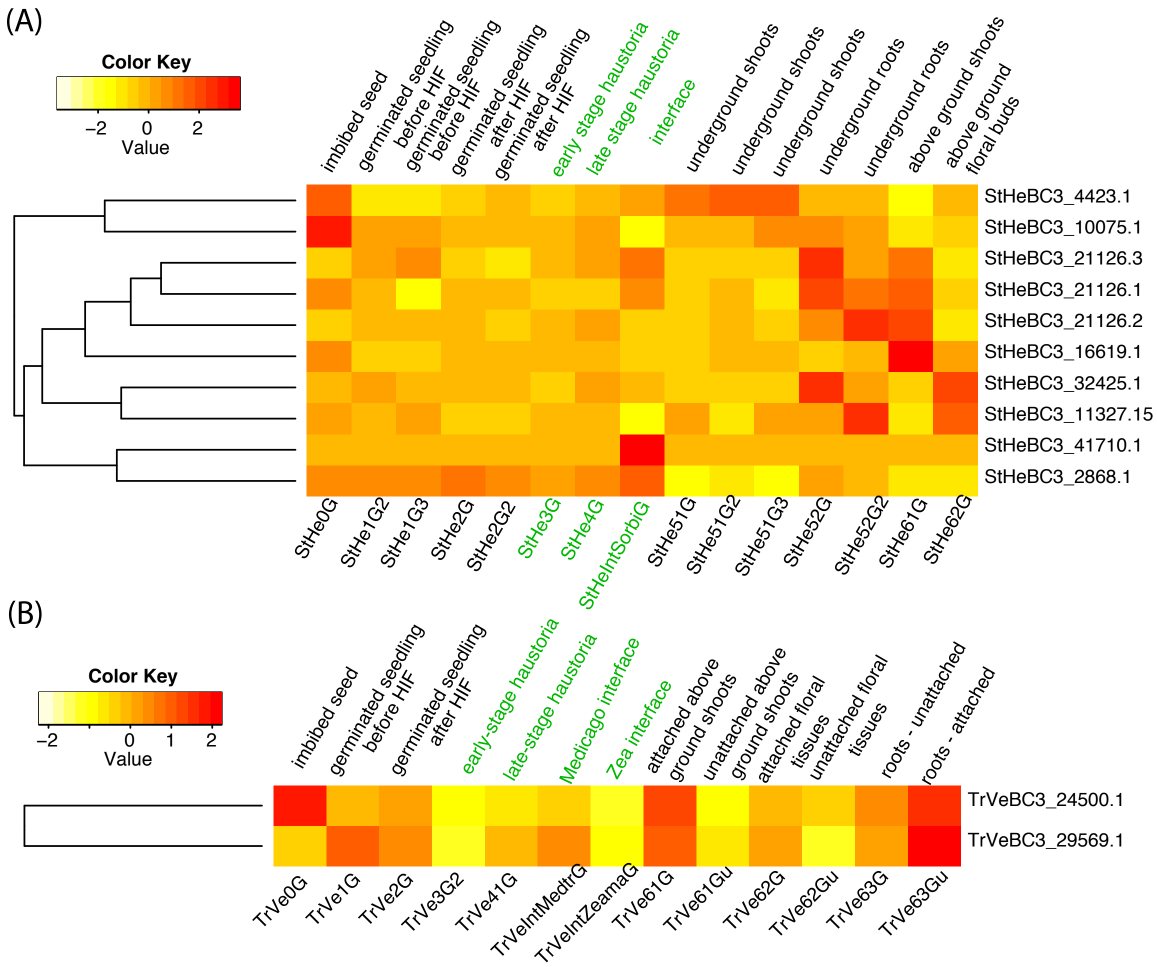


Figure S9. Heatmaps of HGT transgenes in *Striga hermonthica* (A), and *Triphysaria versicolor* (B). These genes were identified from the 42 HGT orthogroups. Heatmap intensity represents the FPKM-transformed z-score.

Table S1. Scoring scheme used to score each phylogenetic tree based on bootstrap support, depth of donor clades, and long branches.

Criterion1		**Criterion2**		**Criterion3**		**Scoring**	
Bootstrap support	score	Sampling depth	score	Branch length	score	Summed score	confidence
x=100	10	rich sampling in the donor clade and strong support	5	branch length not long	3	x≤9	low
90≤x<100	8						
80≤x<90	6	rich sampling in the donor clade and low support	3	branch length < 2 times of the average	2	10≤x≤14	medium
70≤x<80	4						
60≤x<70	3						
50≤x<60	2	mixed samples in the ancestral node	1	branch length >=2 times	1	15≤x≤18	high
x<50	1						

Table S2. SH test to evaluate number of transfers in HGT trees by constraining multiple HGT genes to one monophyletic clade.

orthogroup	**logL (original)**	**logL (constrained)**	**Significant worse (0.01)?**	**# of transfers**
218	-106049.78	-106490.58	Y	≥ 2

1021	-40178.73	-40281.56	Y	≥ 2
3861	-42295.88	-42296.95	N	1
5002	-95429.05	-95694.82	Y	≥ 2
8888	-35555.79	-35813.86	Y	≥ 2
11437	-121967.75	-122293.12	Y	≥ 2
13512	113878.47	-114045.44	Y	≥ 2
14233	-88970.83	-89052.98	Y	≥ 2
18354	-11889.45	-11941.91	Y	≥ 2
18774	-48200.19	-48661.65	Y	≥ 2
19297	-48200.19	-48661.65	Y	≥ 2

logL: log likelihood value, Y means the constrained tree is significant worse than the original tree, N means the constrained tree is not significant worse than the original tree.

Table S3. Verification of HGT sequences by multiple lines of evidences including RT-PCR (RT), genomic PCR (GP), covered by genomic sequencing (GS), present in transcriptomes of multiple parasitic species (MP), or by existing studies.

Orthogroup (genes)	Verification	Orthogroup	Verification
1886 (PhAeBC5_15496.1)	RT, GP, GS	12835 (StHeBC3_10075.1)	GS
14624 (StHeGnB1_80049)	GP, GS	14230 (PhAeBC5_34854.1)	GS
1685 (PhAeBC5_6791.2/9731.1/15355.1)	RT, GP	17 (PhAeBC5_8480.1)	GS
10143 (PhAeBC5_14056.1)	RT	226 (PhAeBC5_14888.1)	GS
8888 (2) (PhAeBC5_15086.1)	RT	9613 (PhAeBC5_16890.17)	GS
10050 (PhAeBC5_4284.2)	RT, GP (I)	19696 (PhAeBC5_17034.1)	GS
13656 (StHeBC3_41710.1)	RT	16703 (StHeBC3_11327.15)	GS
14233 (2) (StHeBC3_16619.1)	RT	18709 (PhAeBC5_654.1)	GS
18774 (2) (StHeBC3_4423.1)	RT	13892 (PhAeBC5_4266.1)	GS
20190 (PhAeBC5_2529.1)	RT	1021 (2) (PhAeBC5_17559.1)	GS
2376 (PhAeBC5_270.1 StHeBC3_48088.1)	RT, GS	5002 (2) (PhAeBC5_9781.2)	GS
11841 (PhAeBC5_3756.1 StHeBC3_55745.1)	RT, GS, GP (I-Pa_3756.1)	15149 (PhAeBC5_8832.4)	GS
1226 (PhAeBC5_11914.4)	RT, GS	20188 (PhAeBC5_1574.1)	GS
3861 (PhAeBC5_ 9914.1 /22555.2)	RT, GS, GP (I)	4572 (PhAeBC5_8623.1)	MP (3)
5896 (StHeBC3_21126.1)	RT, GS	4598 (PhAeBC5_13756.1)	MP(3)
806 (PhAeBC5_4239.1)	RT, GS, GP (I)	11437 (2) (PhAeBC5_13226.1/6799.1)	MP (4)
2270 (StHeBC3_2868.1)	RT, GS	13512 (2) (PhAeBC5_1584.1/3740.1)	MP (4)
4067 (PhAeBC5_9762.1)	RT, GS	14675 (PhAeBC5_12620.5)	MP (4)
218 (2) (PhAeBC5_20413.1)	RT, GS	18354 (2) (PhAeBC5_5786.1)	MP (4)
19297 (2) (StGe3_4270)	Reference – Yoshida et al. 2010 (26)	23343 (PhAeBC5_4800.1)	MP(4)
15246 (PhAeBC5_2525.1)	Reference – Zhang et al. 2013 (28)	23480 (PhAeBC5_5395.1)	MP (2)

Number in the parenthesis following orthogroup and transcriptome stands for the number of events for that orthogroup, and the number of parasite transcriptomes containing the HGT gene, respectively. “I” followed by GP (genomic PCR) represents introns in the sequenced genomic PCR product.

Table S4. An example of orthogroup 806 showing conserved intron positions in coding sequence (CDS) multiple sequence alignment (MSA).

Sequence_id	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos
gnl_Ambtr1.0.27_AmTr_v1.0_scaffold00038.43	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Sorbi1.4_PACid_1975126	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Orysa6.0_PACid_16846845	333	515	591	657	741	819	903	983	1347	1587	1796		2510				
gnl_Orysa6.0_PACid_16844493	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	
gnl_Orysa6.0_PACid_16844494	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	
gnl_Orysa6.0_PACid_16849430	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Orysa6.0_PACid_16849429	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Bradi1.2_Bradi2g04960.1	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Bradi1.2_Bradi3g30350.1	333	504		657	741	819	903	987	1347	1587	1796		2457	2967	3074	3177	3335
gnl_Orysa6.0_PACid_16886435	333	515	591	657	741	819	903	983	1347	1545	1796	2202	2457	2967	3074	3177	3335
gnl_Aquco1.0_PACid_18164810	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Aquco1.0_PACid_18157961	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Aquco1.0_PACid_18144325	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Poptr2.2_PACid_18207203					741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Poptr2.2_PACid_18244025		515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Arath10_AT1G14610.1	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Thepa2.0_Tp1g12970	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	
gnl_Theca1.0_Tc09_g018620 (HGT donor)	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
PhAeBC5_4239.1 (HGT gene)	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Medtr3.5_Medtr1g101620.1		515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Glyma1.01_PACid_16281403		515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Glyma1.01_PACid_16281404		515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Vitvi12X_PACid_17822742	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Solly2.3_Solyc09g007540.2.1	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Mimgu1.0_PACid_17678221 (Vertical relative)	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	
gnl_Mimgu1.0_PACid_17684099 (Vertical relative)	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	3335
gnl_Mimgu1.0_PACid_17693905 (Vertical relative)	333	515	591	657	741	819	903	983	1347	1587	1796	2202	2457	2967	3074	3177	
PhAeBC5_4810.1 (vertical gene)	333	515	591	-	741	819	-	983	1347	1587	1796	2202	2457	2967	3074	3177	3335

pos: intron position in MSA; "-" means not available due to low coverage of genomic DNA sequencing data.

Table S5. Summary of conserved intron positions (CIPs) in HGT clade and the overall MSA.

orthogroup	annotation	HGT recipient genus	Donor representative	# of CIPs in HGT clade	scorable # of CIPs in HGT clade	# of CIPs in MSA
2376	proteasome subunit	P, S	Poptr	8	8	10
11841	Oxidoreductase, 2OG-Fe oxygenase	P, S, O	Frave	1	3	3
4067	tRNAHis	P	Frave	8	9	12

	guanylyltransferase					
8888	tRNA adenylyltransferase activity	P	Frave	7	7	14
10050	tRNA synthetase class II core domain containing protein	P	Frave	4	4	6
806	Valyl-tRNA synthetase	P, O	Theca	15	15	19
2270	methionyl-tRNA synthetase	P, S, O	Theca	11	13	16
1685	Gnk2-homologous domain	P	Medtr, Glyma	7	7	7
1886	Ankyrin repeat family protein	P	Frave	0	2	2
9613	HAD-superfamily hydrolase	P	Glyma	2	3	12
3861	Poly (ADP-ribose) glycohydrolase (PARG)	P, L	Glyma	7	7	9
5896	uroporphyrinogen-III synthase	S	Glyma	8	8	8
18774	Domain of unknown function	S	Orysa	0	0	0
Total				78	79	118

HGT recipient genus: P – *Phelipanche*, S – *Striga*, O – *Orobanche*, L – *Lindenbergia*. Nomenclature of the donor follows the five-letter rule (The first three letters of the genus, and first two letters of the species).

Table S6. PAML analyses with branch test on codon alignment of 42 HGT orthogroups, testing mode of selection in HGT sequences (foreground) compared to VGT sequences (background).

Ortho ID	D_n	D_s	Foreground ω	Background ω	P-value	Significant at 0.05	Selection
806	0.09	0.33	0.27	0.14	< 0.00001	S	RP
1886	0.48	0.82	0.58	0.16	< 0.00001	S	RP
2270	0.09	0.28	0.31	0.14	< 0.00001	S	RP
2376	0.03	0.15	0.18	0.04	0.000108	S	RP
4067	0.10	0.24	0.43	0.20	0.000187	S	RP
8888	0.12	0.28	0.43	0.23	0.001553	S	RP
9613	0.51	1.03	0.50	0.14	< 0.00001	S	RP
10050	0.13	0.26	0.51	0.20	< 0.00001	S	RP
10143	0.10	0.21	0.46	0.20	< 0.00001	S	RP
1226	0.22	0.41	0.55	0.27	0.002847	S	RP
4598	0.22	0.56	0.39	0.27	0.016327	S	RP
11437	0.40	0.53	0.74	0.36	< 0.00001	S	RP
13512	0.29	0.17	1.65	0.34	< 0.00001	S	POS
3861	0.14	0.10	1.48	0.28	< 0.00001	S	POS
18709	0.19	0.17	1.16	0.22	0.000495	S	POS
19696	0.56	73.08	0.01	0.29	0.022212	S	SP
11841	0.03	0.20	0.14	0.26	0.03422	S	SP
13656	0.01	3.08	0.00	0.30	< 0.00001	S	SP
14233	0.32	1.16	0.27	0.48	0.005556	S	SP
5896	0.06	0.21	0.29	0.30	0.943628	NS	P
1685	0.06	0.21	0.29	0.27	0.811464	NS	P
17	0.21	0.78	0.27	0.23	0.054126	NS	P
218	0.07	0.54	0.12	0.20	0.431345	NS	P
226	0.12	0.51	0.24	0.27	0.449545	NS	P
1021	0.13	0.72	0.18	0.19	0.943886	NS	P
4572	0.13	0.42	0.31	0.47	0.054707	NS	P
5002	0.17	1.00	0.17	0.23	0.172164	NS	P
12835	0.11	0.68	0.16	0.18	0.731762	NS	P
13892	0.02	0.09	0.26	0.24	0.917998	NS	P
14230	0.28	2.22	0.13	0.16	0.331354	NS	P
14624	0.22	1.02	0.22	0.26	0.420936	NS	P
14675	0.49	1.22	0.40	0.47	0.705675	NS	P
15149	0.31	1.24	0.25	0.28	0.734189	NS	P
15246	0.06	0.10	0.65	0.24	0.139996	NS	P
16703	0.40	70.39	0.01	0.33	0.09305	NS	P
18354	0.54	1.01	0.54	0.62	0.908563	NS	P

18774	0.20	0.32	0.61	0.51	0.26167	NS	P
19297	0.08	0.14	0.61	0.51	0.26167	NS	P
20188	0.55	0.66	0.83	0.82	0.971238	NS	P
20190	0.19	0.22	0.88	0.91	0.946516	NS	P
23343	0.08	0.32	0.26	0.34	0.05642	NS	P
23480	0.22	0.26	0.56	0.84	0.218423	NS	P

Significant branch tests (at $p<0.05$) reveal relaxed constraint (RP), positive selection (POS), or stronger purifying selection (SP) in the HGT foreground compared to background. Non-significant branch tests show a common level of purifying selection (P) in the foreground HGT sequences compared to the VGT background.

Table S7. PAML analyses with the branch-site model on codon alignment of 15 HGT orthogroups with greater D_n/D_s on HGT genes compared to the background, identifying the presence of sites under positive selection.

HGT transgene IDs	Orthogroup ID	Sites identified having positive selection (P<=0.05)
PhAeBC5_4239.1 OrMi2_4015 PhRa5_26371 PhMu2_16115	806*	556 D;
PhAeBC5_11914.4 PhRa5_49596 PhRa5_27280 PhMu2_10869 PhMu2_12766	1226*	135 E; 250 V;
PhAeBC5_15496.1 PhRa5_103413 PhRa5_58721	1886*	1 M; 39 E; 83 F; 84 S; 86 E; 92 K; 146 L; 153 S; 170 T; 173 D; 176 C; 210 T; 214 L; 230 T; 236 T; 247 R; 249 Q; 251 C; 254 D; 256 V; 317 G; 318 K; 322 E; 330 P; 347 S; 363 A; 403 G; 404 F ; 406 S; 416 C; 421 G; 426 F; 428 G; 429 P; 438 C; 444 W; 472 A; 485 A; 488 V; 493 A; 500 G; 530 S; 533 E; 535 V; 581 D; 582 L;
StHeBC3_2868.1 PhAeBC5_3356.1 PhRa5_6157 PhMu2_12619 PhMu2_44156 PhMu2_20937 PhMu2_5516 PhMu2_14031 OrMi2_4206 StGe3_7730 StGe3_20099	2270	NA
PhAeBC5_270.3 StHeBC3_48088.1 PhRa5_4383 PhMu2_6326 PhRa5_4382	2376*	20 V;
PhAeBC5_9914.1OrMi2_2808 PhRa5_12205 PhMu2_12820 PhMu2_11052	3861*	23 S; 45 E; 68 A; 71 D; 93 L; 96 D; 97 D; 254 D; 289 K; 301 V; 486 C; 543 W; 568 P;
PhAeBC5_9762.1 PhRa5_10084 PhMu2_15393	4067*	218 K;
PhAeBC5_13756.1 PhMu2_15647 PhRa5_33568	4598	31 R; 32 P; 156 S; 159 G; 160 L; 169 K; 187 L; 188 S; 233 Q; 235 K; 240 K; 244 T; 247 E; 248 A; 249 M; 251 L; 253 P; 282 L; 284 S; 379 K; 481 N; 495 E; 500 G; 535 A; 536 T; 539 V; 540 A; 543 C; 545 P; 547 N;
Phelipanche_ramosa_PhRa5_6603 Phelipanche_mutelii_PhMu2_37370 Phelipanche_ramosa_PhRa5_6605 PhAeBC5_15086.1 Phelipanche_ramosa_PhRa5_6604	8888*	209 Q; 538 D; 559 P;
PhAeBC5_16890.17 PhRa5_14643 PhRa5_14643	9613*	3 P; 4 S; 22 S; 26 R; 33 F; 42 K; 50 K; 53 T; 55 N; 73 L; 74 P; 77 D; 78 A; 81 I; 82 G; 85 L; 86 Q; 87 I; 90 E; 95 V; 96 E; 99 F; 100 V; 101 H; 102 L; 104 F; 106 C; 107 E; 109 K; 110 P; 112 H; 114 V; 116 S; 121 S; 122 K; 123 P; 126 K; 127 F; 162 T;
PhAeBC5_4284.2 PhMu2_5340 PhRa5_11783 PhMu2_31112	10050*	472 S; 666 S; 709 G; 710 S; 763 F;
PhAeBC5_14056.1 OrMi2_39134 OrMi2_14052	10143*	521 H; 599 G;
PhAeBC5_1584.1 PhAeBC5_3740.1	11437*	1201 Y; 1242 E; 1287 T; 1417 A; 1637 G;
PhAeBC5_1584.1 OrCa3_3600 PhRa5_1522 PhMu2_40952	13512*	367 R; 479 C; 480 R; 497 S; 520 S; 533 F; 536 E; 547 L; 768 G; 831 N; 844 N; 886 E; 1005 P; 1107 Y; 1129 L; 1174 K; 1391 S;

PhAeBC5_654.1 PhRa5_6642	18709*	664 C; 676 Y; 710 K; 715 S; 724 R;

NA indicates no adaptive sites were identified. “*” after each orthogroup indicates presence of adaptive sites were identified only in parasitic lineages, not in a nonparasitic genome, such as *Mimulus*.

Table S8. RELAX analyses on codon alignment of 42 HGT orthogroups testing stronger selection or relaxed constraint in HGT sequences compared to the background.

Ortho ID	K	P-value	Significant at 0.05	Selection: HGT vs background
806	1.01	0.838	NS	same
9613	0.75	0.090	NS	same
14230	0.66	0.064	NS	same
10050	1.01	0.970	NS	same
11841	1.35	0.237	NS	same
19696	0.65	0.193	NS	same
18709	0.85	0.987	NS	same
11437	1.03	0.747	NS	same
13656	0.78	0.584	NS	same
17	1.03	0.720	NS	same
1021	1.14	0.325	NS	same
5002	1.05	0.538	NS	same
12835	1.78	0.103	NS	same
14624	1.02	0.914	NS	same
15246	1	0.995	NS	same
16703	1.02	0.914	NS	same
18354	1	0.970	NS	same
18774	0.92	0.436	NS	same
20188	0	0.067	NS	same
20190	0.73	0.294	NS	same
23343	1.21	0.227	NS	same
23480	1.02	0.960	NS	same
4572	50	0.004	S	stronger
5896	50	0.026	S	stronger
19297	49.44	0.013	S	stronger
1886	49.42	0.001	S	stronger
4598	35.1	0.000	S	stronger
226	2.22	0.000	S	stronger
14233	2.16	0.000	S	stronger
4067	1.99	0.000	S	stronger
218	1.93	0.000	S	stronger
13512	1.9	0.000	S	stronger
1226	1.61	0.000	S	stronger
14675	1.14	0.004	S	stronger
1685	0.82	0.004	S	relaxed
2270	0.68	0.000	S	relaxed
15149	0.66	0.038	S	relaxed
8888	0.61	0.013	S	relaxed
3861	0.55	0.000	S	relaxed
13892	0.54	0.030	S	relaxed
2376	0.21	0.000	S	relaxed
10143	0.08	0.001	S	relaxed

Significant RELAX results (at $p<0.05$) reveal stronger selection (stronger), relaxed constraint (relaxed). Non-significant RELAX results show the same level of selection in the foreground HGT sequences compared to background (same).

Table S9. Inferred donor lineage and recipient taxa from the gene trees for the final 42 HGT orthogroups.

orthogroup ID	recipient taxa	donor family	taxa or common name
17	Phe	Amaranthaceae	*Beta vulgaris* (beet)
218	Tri, Ale, Phe, Str	Rosaceae	*Prunus* (peach)
226	Phe	Euphorbiaceae	*Morchella* (Cassava)
806	Phe, Oro	Malvaceae	*Theobroma cacao* (cocoa)
1021	Phe, Tri	Rosaceae	*Fragaria* (strawberry), *Malus* (apple)
1226	Phe	Euphorbiaceae	*Ricinus communis* (castor bean)
1685	Phe	Fabaceae	*Medicago*, *Glycine*
1886	Phe	Rosaceae	*Fragaria vesca* (strawberry)
2270	Phe, Oro, Str	Malvaceae	*Theobroma cacao* (cocoa)
2376	Phe, Str	Rosaceae	*Fragaria vesca* (strawberry)

3861	LiPh, Phe, Oro	Fabaceae	*Glycine* (soybean)
4067	Phe	Rosaceae	*Fragaria vesca* (strawberry)
4572	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
4598	Phe	Fabaceae	*Medicago* (alfalfa, barrel millet)
5002	Phe, Oro, Ale	Fabaceae	*Fragaria* (strawberry), *Prunus* (peach)
5896	Str	Fabaceae	*Glycine* (sobyean)
8888	Phe, Oro	Rosaceae, Fabaceae	*Fragaria* (strawberry), *Glycine* (soybean)
9613	Phe	Fabaceae	*Medicago* (alfalfa, barrel millet)
10050	Phe	Rosaceae	*Fragaria vesca* (strawberry)
10143	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
11437	Phe	Rosaceae	*Fragaria vesca* (strawberry)
11841	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
12835	Str	Poaceae	*Sorghum bicolor*
13512	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
13656	Str	Poaceae	*Sorghum bicolor*
13892	Phe, Oro	Fabaceae	*Medicago* (alfalfa, barrel millet)
14230	Phe	Rosaceae	*Prunus* (peach)
14233	Str	Poaceae	*Oryza sativa* (rice)
14624	Str	Poaceae	*Sorghum bicolor*, maize
14675	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
15149	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
15246	Phe, Oro, Ale, Tri	Fabaceae	*Medicago* (alfalfa, barrel millet)
16703	Str	Poaceae	*Sorghum bicolor*
18354	Phe, Oro	Rosaceae	*Fragaria vesca* (strawberry)
18709	Phe	Brassicaceae	*Arabidopsis* thaliana (thale cress)
18774	Str	Poaceae	*Setaria italica* (foxtail millet)
19297	Str	Poaceae	*Sorghum bicolor*
19696	Phe	Salicaceae	*Populus trichocarpa (poplar)*
20188	Phe, Str	Rosaceae	*Fragaria vesca* (strawberry)
20190	Phe	Rosaceae	*Fragaria vesca* (strawberry)
23343	Phe, Oro	Euphorbiaceae	*Ricinus communis (castor bean)*
23480	Phe	Rosaceae	*Fragaria vesca* (strawberry)

Phe: *Phelipanche*, Oro: *Orobanche*, Str: *Striga*, Tri: *Triphysaria*, Ale: *Alectra*, Lin: *Lindebergia*

Table S10. Inferred PFam domains for the HGT genes in the 42 orthogroups.

orthogroup	HGT unigene	Pfam domain
226	PhAeBC5_14888.1	[Cytochrome P450(31-483)]
1685	PhAeBC5_15353.2 PhAeBC5_9731.1 PhAeBC5_6791.2	[Bacteriophage coat protein B(47-66), ABC1 family(109-154), Beta-amyloid peptide (beta-APP)(154-162), Phosphotransferase enzyme family(215-253)] [Salt stress response/antifungal(62-151), Protein of unknown function (DUF2970)(177-195), ABC1 family(238-281)] [Salt stress response/antifungal(148-244), ABC1 family(327-376)]
2376	PhAeBC5_270.3	[Proteasome subunit A N-terminal signature(142-147)]
14624	StHeGnB1_80049	[BTB/POZ domain(272-300)]
23343	PhAeBC5_4800.1	[NB-ARC domain(174-451), CENP-Q, a CENPA-CAD centromere

		complex subunit(703-758), AAA ATPase domain(761-805), Leucine rich repeat(862-872)]
11841	PhAeBC5_3756.1	[non-haem dioxygenase in morphine synthesis N-terminal(39-150), 2OG-Fe(II) oxygenase superfamily(200-294)]
1886	PhAeBC5_15496.1	[PUL domain(281-350)]
11437	PhAeBC5_13226.1 PhAeBC5_6799.1	[Kelch motif(207-212), Kelch motif(228-244)] [Kelch motif(155-190), Kelch motif(257-283)]
8888	PhAeBC5_15086.1	[Poly A polymerase head domain(38-175), Probable RNA and SrmB-binding site of polymerase A(221-262)]
18709	PhAeBC5_654.1	[Nup53/35/40-type RNA recognition motif(116-150)]
806	PhAeBC5_4239.1	[Leucyl-tRNA synthetase, Domain 2(347-405), tRNA synthetases class I (C) catalytic domain(632-673), Anticodon-binding domain of tRNA(774-922), KRI1-like family(995-1025)]
2270	PhAeBC5_3356.1 StHeBC3_2868.1	[Adenylate kinase, active site lid(171-198)] [tRNA synthetases class I (I, L, M and V)(435-474), Anticodon-binding domain of tRNA(537-621)]
4067	PhAeBC5_9762.1	[tRNAHis guanylyltransferase(265-391)]
10050	PhAeBC5_4284.2	[Histidyl-tRNA synthetase(109-430), Anticodon binding domain(462-538)]
13892	PhAeBC5_4266.1	[Ribosomal protein S13/S18(60-132)]
17	PhAeBC5_8480.1	[ABC transporter transmembrane region(725-970), AAA domain(1076-1091), RNA helicase(1247-1282)]
9613	PhAeBC5_16890.17	[3-beta hydroxysteroid dehydrogenase/isomerase family(10-73), Dimerisation domain of Zinc Transporter(333-357)]
15246	PhAeBC5_2525.1	[Albumin I chain a(71-112)]
1226	PhAeBC5_11914.4	[DNA polymerase III beta subunit, central domain(12-79), BAAT / Acyl-CoA thioester hydrolase C terminal(135-162)]
10143	PhAeBC5_14056.1	[HEAT-like repeat(383-391), Tubulin folding cofactor D C terminal(427-614), HEAT repeat(640-657)]
3861	PhAeBC5_22555.2 PhAeBC5_9914.1	[Poly (ADP-ribose) glycohydrolase (PARG)(614-632)] [Poly (ADP-ribose) glycohydrolase (PARG)(78-432)]
4598	PhAeBC5_13756.1	[Colicin D(303-330)]
19696	PhAeBC5_17034.1	[Peptidase C39 family(300-311)]
16703	StHeBC3_11327.15	[GRF zinc finger(56-86)]
4572	PhAeBC5_8623.1	[Leucine Rich Repeat(182-192), Leucine Rich repeat(206-214), Leucine Rich repeats (2 copies)(316-325)]
5896	StHeBC3_21126.2	[Proline-rich nuclear receptor coactivator(17-65), Uncharacterised protein family (UPF0180)(206-237), PCI domain(299-377)]
218	PhAeBC5_31016.1 StHeBC3_32425.1 PhAeBC5_20413.1 TrVeBC3_29569.1	[Protein of unknown function (DUF 659)(16-145), Domain of unknown function (DUF4413)(278-387), hAT family C-terminal dimerisation region(456-535)] [BED zinc finger(57-115), Protein of unknown function (DUF 659)(199-334), Domain of unknown function (DUF4413)(492-590)] [Protein of unknown function (DUF 659)(371-392)] [Domain of unknown function (DUF4413)(250-355), Protein of unknown function (DUF 659)(393-413)]
1021	PhAeBC5_17559.1 TrVeBC3_24500.1	[DDE superfamily endonuclease(66-145), hAT family C-terminal dimerisation region(419-489)] [BED zinc finger(69-113), Protein of unknown function (DUF 659)(247-393)]
5002	PhAeBC5_9781.2	[BED zinc finger(316-368), Domain of unknown function (DUF4413)(414-512), hAT family C-terminal dimerisation region(568-650)]
12835	StHeBC3_10075.1	[Plant transposon protein(167-274)]
14230	PhAeBC5_34854.1	[RhoGEF domain(297-379), MULE transposase domain(433-464), Domain of Unknown Function with PDB structure (DUF3862)(495-

		547)]
15149	PhAeBC5_8832.4	[hAT family C-terminal dimerisation region(142-224)]
13512	PhAeBC5_1584.1 PhAeBC5_3740.1	[Kelch motif(200-213), Domain of unknown function (DUF4585)(252-270), Galactose oxidase, central domain(279-310)] [Kelch motif(144-157), Protein of unknown function (DUF1668)(222-244)]
14233	StHeBC3_16619.1	[Kelch motif(339-357)]
14675	PhAeBC5_12620.5	[Kelch motif(198-213)]
18354	PhAeBC5_5786.1	[Zinc-finger of C2H2 type(255-261), C2H2-type zinc finger(311-314)]
20190	PhAeBC5_2529.1	[DNA polymerase phi(10-104)]
23480	PhAeBC5_5395.1	[GAG-polyprotein viral zinc-finger(137-151)]
18774	StHeBC3_4423.1	[Protein of unknown function (DUF1618)(174-337)]
13656	StHeBC3_41710.1	NO HITS
19297	ShContig9483_gi295389393_19297	NO HITS
20188	PhAeBC5_1574.1	NO HITS

Genes from each orthogroup were annotated with a Pfam domain, with bracket separating the information for each gene. Domains for a gene are separated by comma, and each domain is followed by the corresponding region in the encoded protein sequence.