

Supplementary material for

**Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size**

Elizabeth Skippington, Todd J. Barkman, Danny W. Rice and Jeffrey D. Palmer\*

\*Corresponding author. E-mail: [jpalmer@indiana.edu](mailto:jpalmer@indiana.edu)

**This PDF file includes:**

Tables S1 to S2

Figures S1 to S8

**Table S1.** Gene content of the *Viscum album* mitogenome.

| Gene            | Product                            | Coordinates |        |
|-----------------|------------------------------------|-------------|--------|
| <i>atp1</i>     | ATPase subunit 1                   | 439103      | 437565 |
| <i>atp4</i>     | ATPase subunit 4                   | 509449      | 508976 |
| <i>atp6</i>     | ATPase subunit 6                   | 99644       | 100558 |
| <i>atp8</i>     | ATPase subunit 8                   | 52588       | 52094  |
| <i>atp9</i>     | ATPase subunit 9                   | 497342      | 497118 |
| <i>ccmB</i>     | cytochrome c biogenesis subunit B  | 273822      | 274385 |
| <i>ccmC</i>     | cytochrome c biogenesis subunit C  | 418365      | 419021 |
| <i>ccmFc</i>    | cytochrome c biogenesis subunit FC | 493580      | 492756 |
|                 |                                    | 491957      | 491475 |
| <i>ccmFn</i>    | cytochrome c biogenesis subunit FN | 512059      | 510281 |
| <i>cob</i>      | apocytochrome b                    | 95609       | 96820  |
| <i>cox1</i>     | cytochrome c oxidase subunit 1     | 51203       | 50493  |
|                 |                                    | 49500       | 48610  |
| <i>cox2</i>     | cytochrome c oxidase subunit 2     | 502063      | 501692 |
|                 |                                    | 43549       | 43860  |
|                 |                                    | 46039       | 46107  |
| <i>cox3</i>     | cytochrome c oxidase subunit 3     | 173779      | 174519 |
| <i>matR</i>     | maturase (splicing factor)         | 194269      | 192311 |
| <i>mttB</i>     | Tat protein translocator subunit 2 | 18014       | 18781  |
| <i>rpl16</i>    | ribosomal protein L16              | 518119      | 518496 |
| <i>rps3</i>     | ribosomal protein S3               | 516629      | 518086 |
| <i>rps4</i>     | ribosomal protein S4               | 413310      | 414317 |
| <i>rps10</i>    | ribosomal protein S10              | 423343      | 423068 |
| <i>rps12</i>    | ribosomal protein S12              | 177860      | 178279 |
| <i>sdh3</i>     | succinate dehydrogenase subunit 3  | 282757      | 282981 |
| <i>rrn5</i>     | 5S ribosomal RNA                   | 380044      | 380161 |
| <i>rrn18</i>    | 18S ribosomal RNA                  | 377450      | 379569 |
| <i>rrn26</i>    | 26S ribosomal RNA                  | 21049       | 25415  |
| <i>trnG-GCC</i> | tRNA-Gly                           | 92604       | 92533  |
| <i>trnK-TTT</i> | tRNA-Lys                           | 500844      | 500766 |
| <i>trnM-CAT</i> | tRNA-Met                           | 42653       | 42728  |
| <i>trnM-CAT</i> | tRNA-Met                           | 412252      | 412180 |
| <i>trnW-CCA</i> | tRNA-Trp                           | 6553        | 6628   |
| <i>trnW-CCA</i> | tRNA-Trp                           | 148987      | 149060 |

**Table S2.** Test for shifts in selective pressures using RELAX [1].

| Locus        | Selection intensity $k$ | $P$ Value |
|--------------|-------------------------|-----------|
| <i>atp1</i>  | 0.17                    | 1.84E-07  |
| <i>atp6</i>  | 0.20                    | 4.06E-08  |
| <i>atp9</i>  | 0.49                    | 6.40E-05  |
| <i>ccmC</i>  | 0.01                    | 5.91E-05  |
| <i>cox1</i>  | 0.27                    | 3.81E-05  |
| <i>cox2</i>  | 0.34                    | 2.22E-04  |
| <i>cox3</i>  | 0.22                    | 1.07E-05  |
| <i>cob</i>   | 0.70                    | 2.66E-03  |
| <i>rps12</i> | 3.40                    | 2.24E-02  |

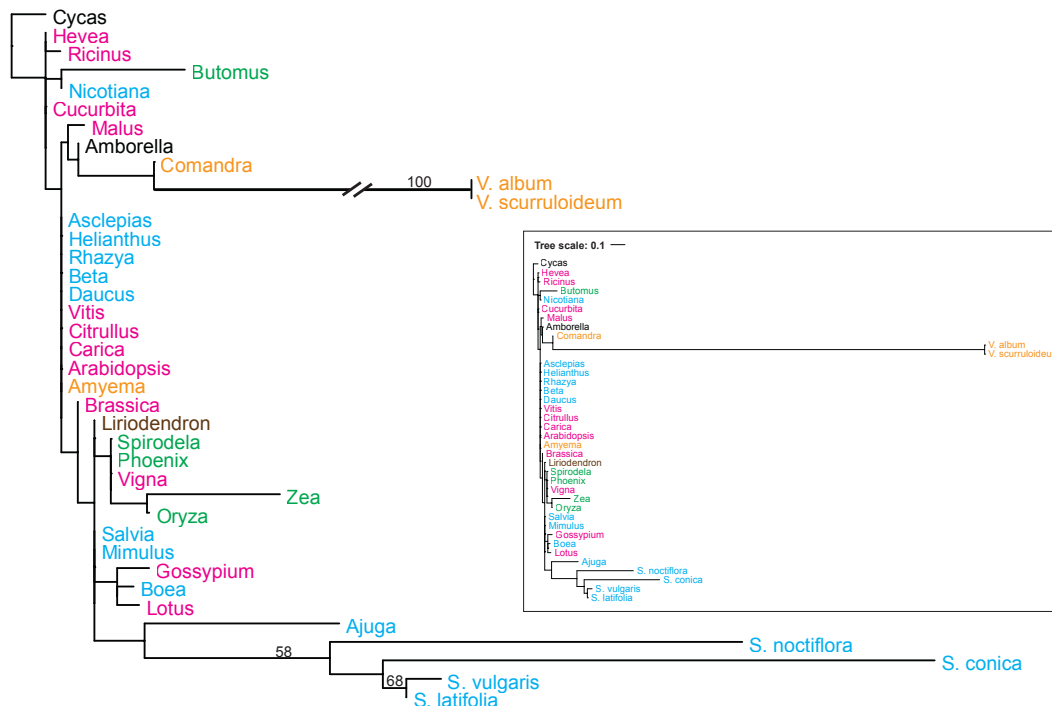
Eight of the nine best conserved protein genes in *Viscum* showed a significant signature of relaxed selection ( $k < 1$ ).



rrn5

tree scale: 0.01 –  
alignment length: 130 nt

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

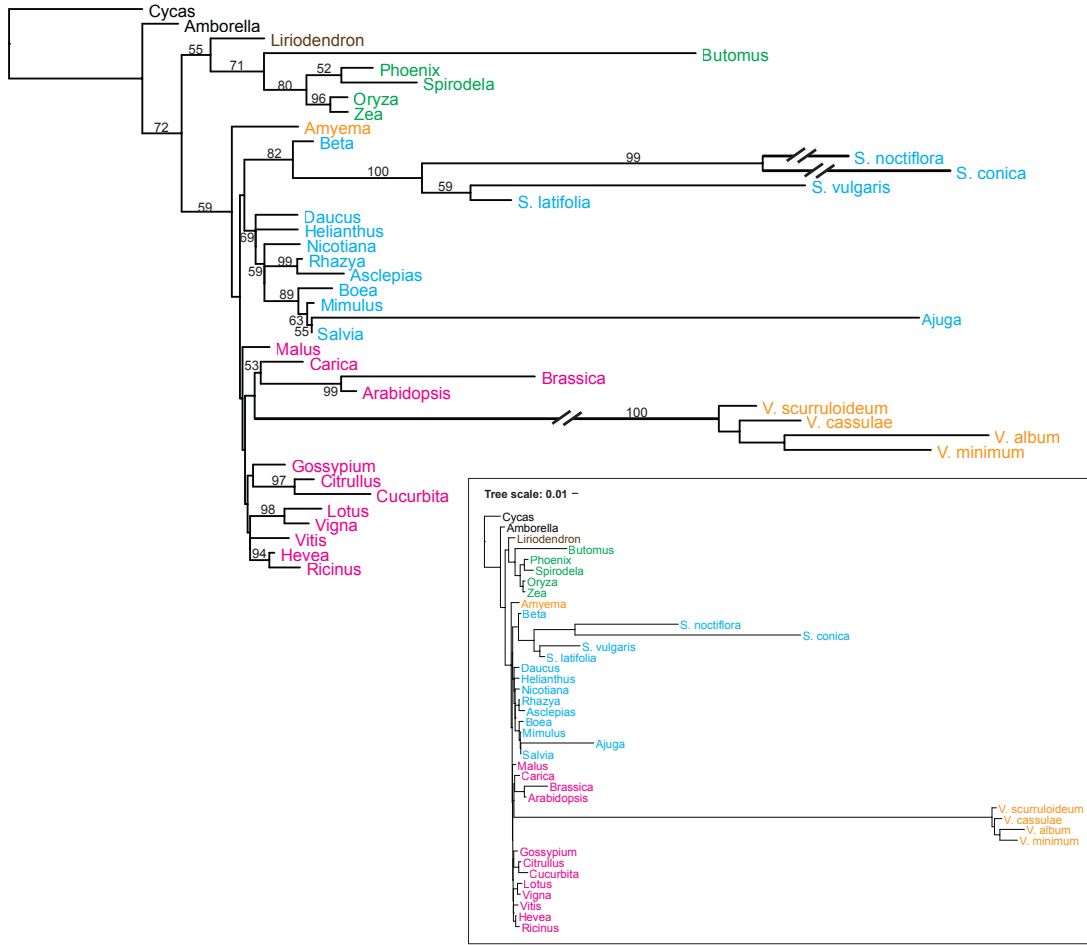


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

rrnS

tree scale: 0.01 —  
alignment length: 3347 nt

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

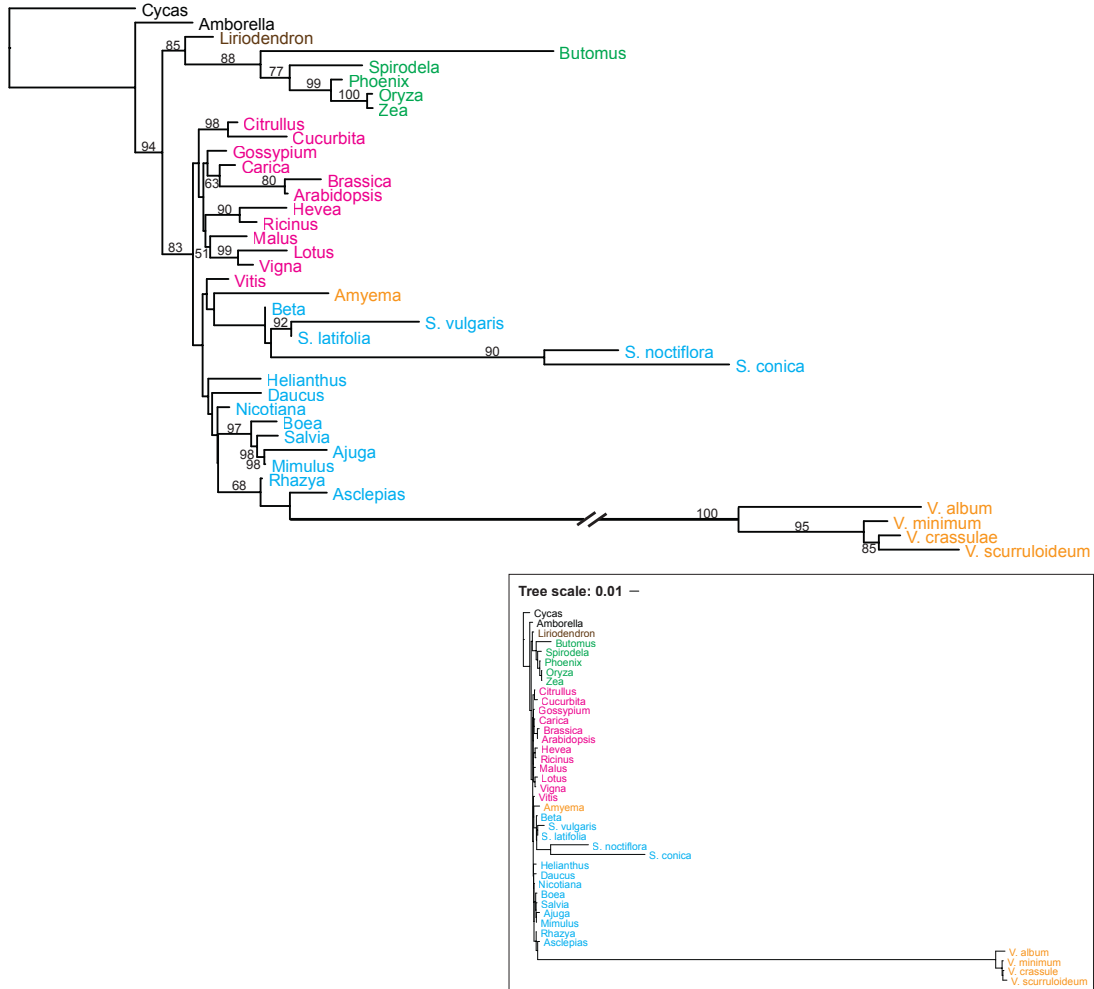


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

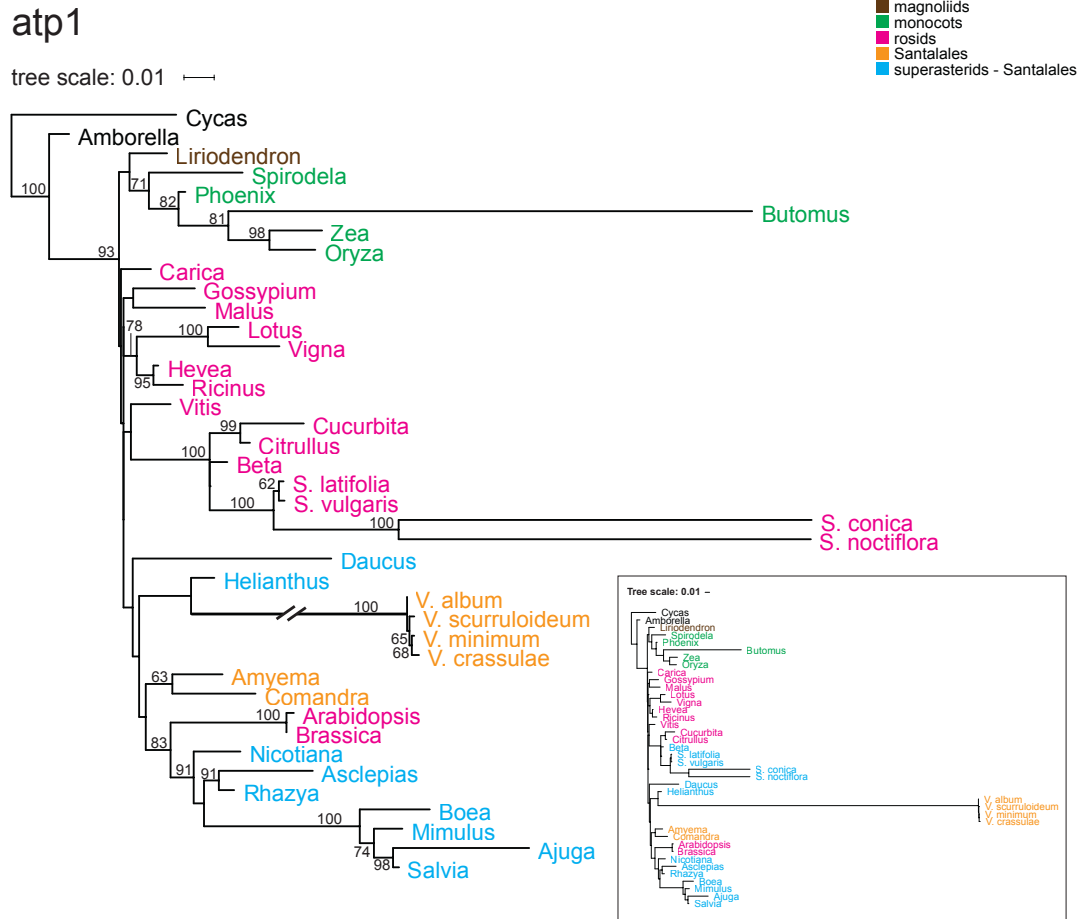
rrnL

tree scale: 0.01 —  
alignment length: 5716 nt

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales



**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

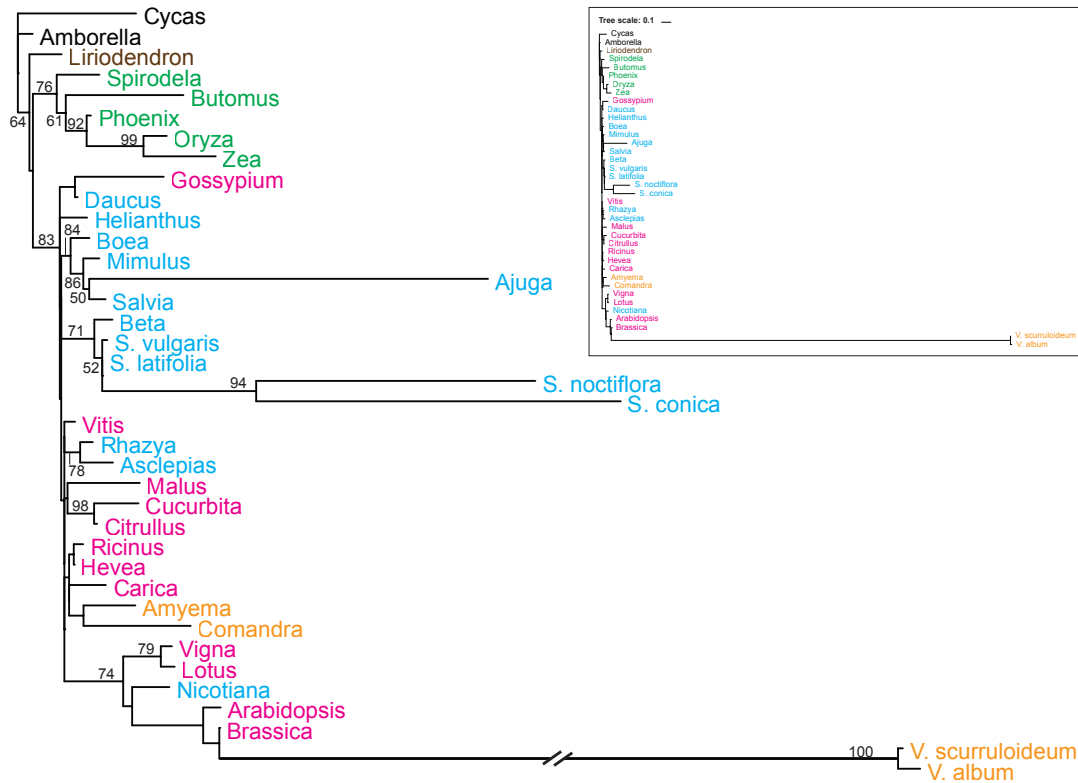


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

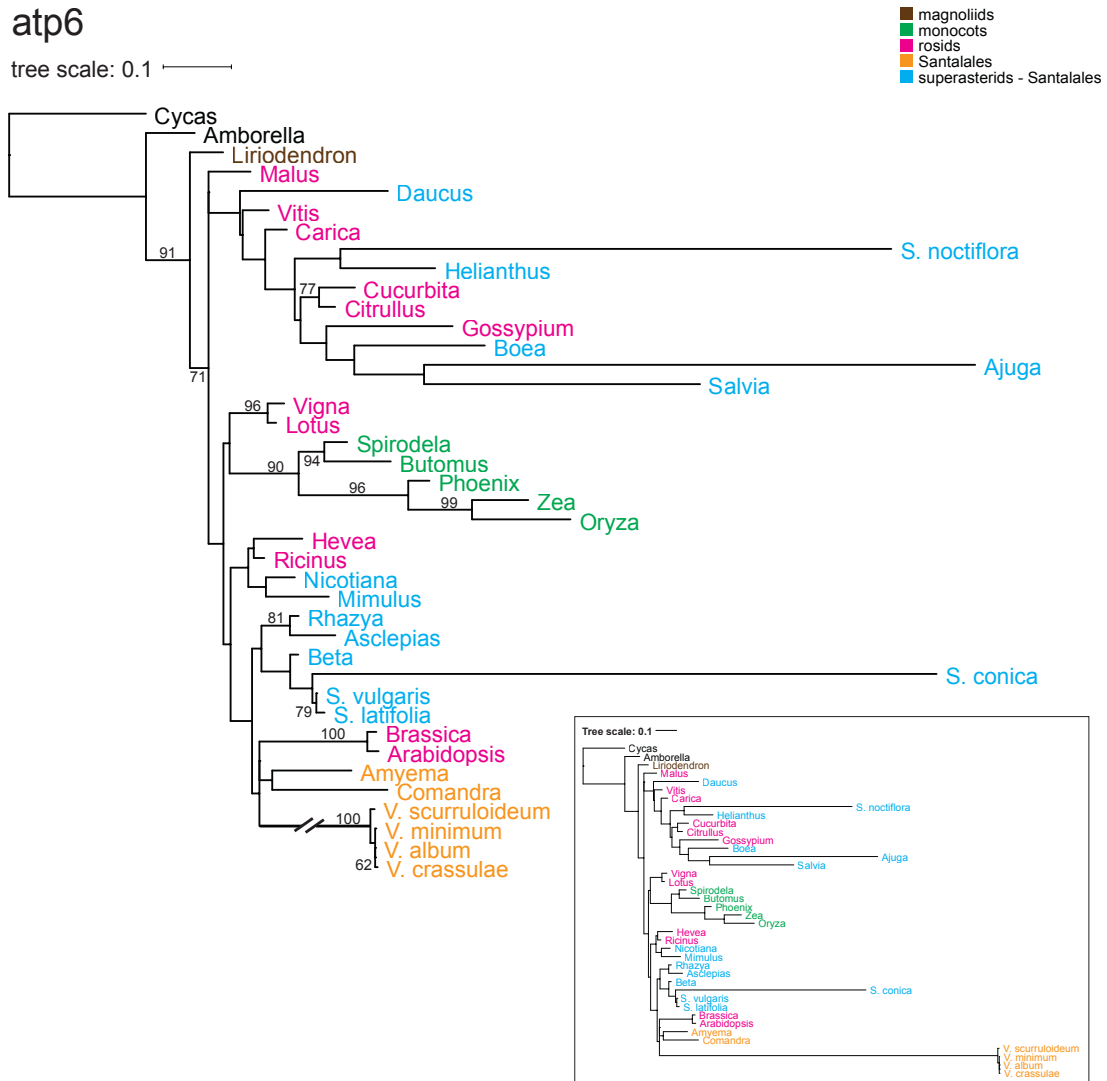
atp4

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales



**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

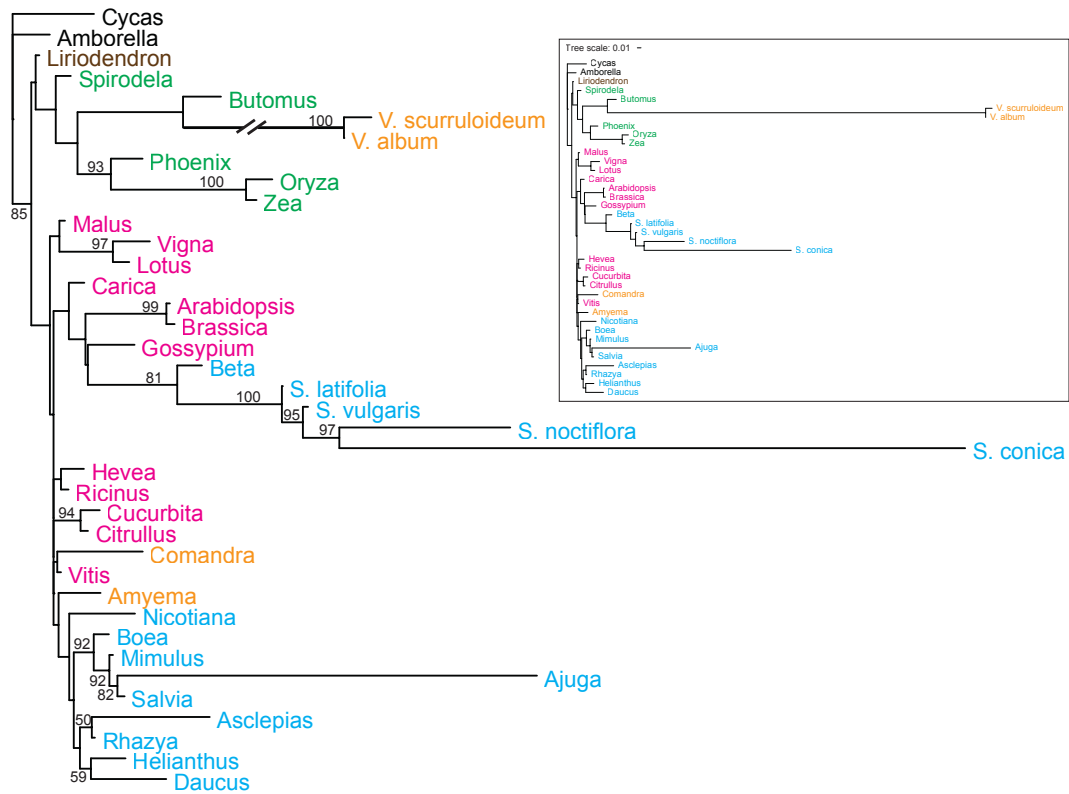


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

atp8

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

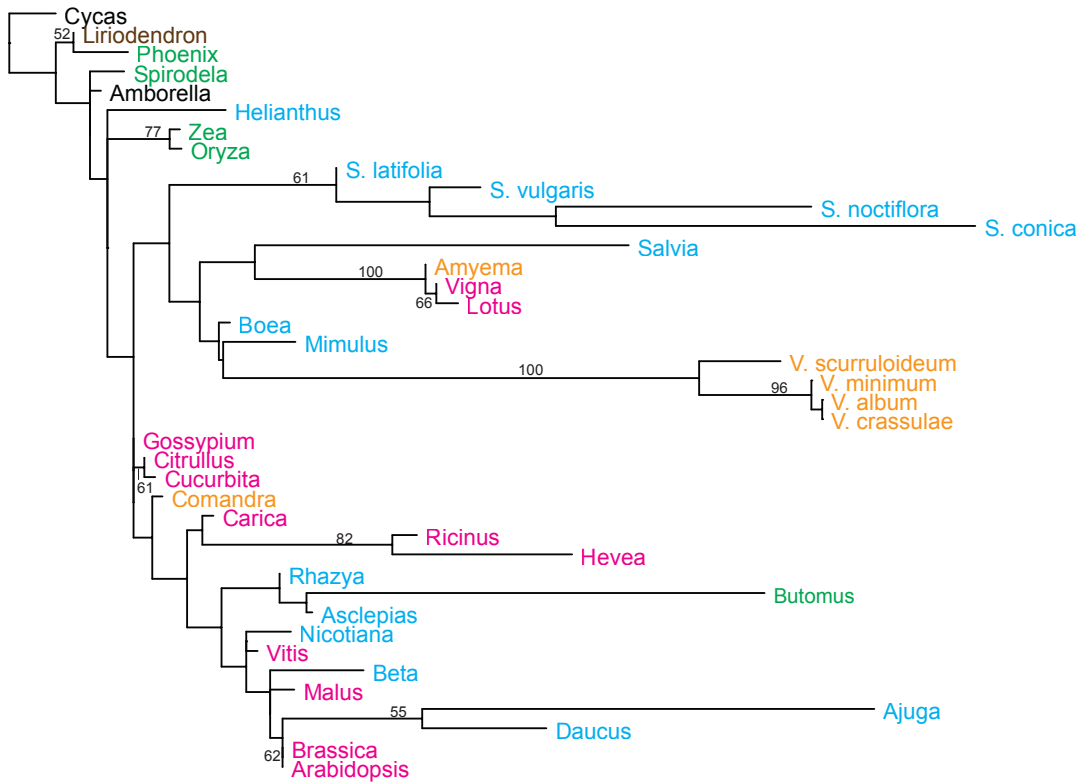


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

atp9

tree scale: 0.01

magnoliids  
 monocots  
 rosids  
 Santalales  
 superasterids - Santalales

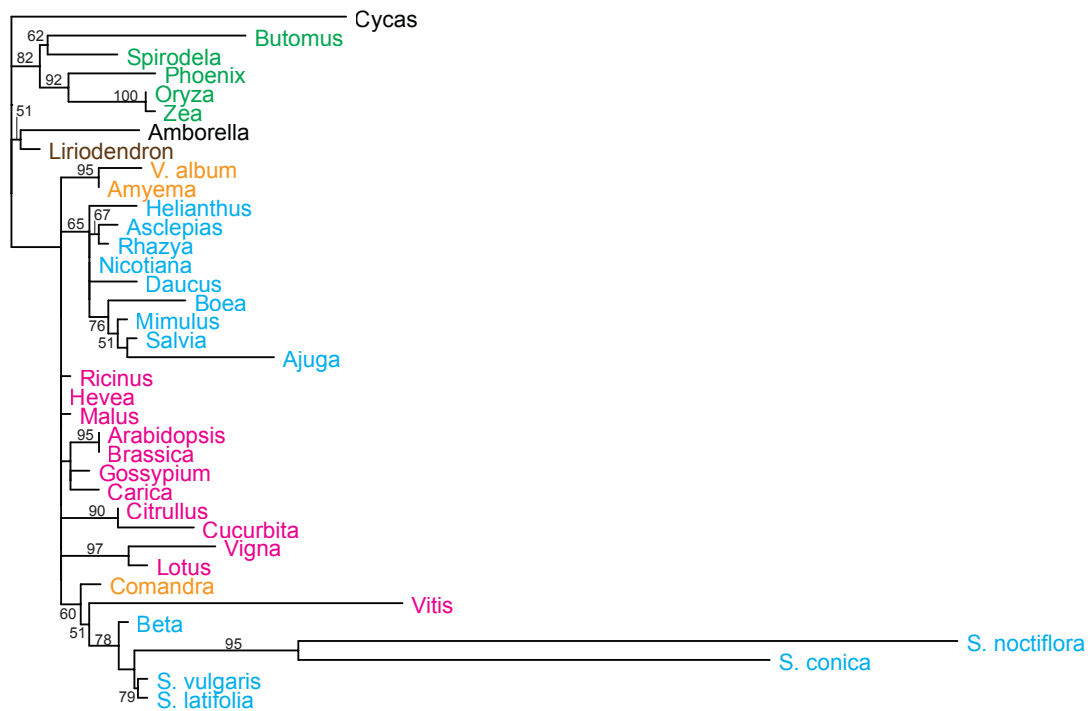


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

ccmB

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

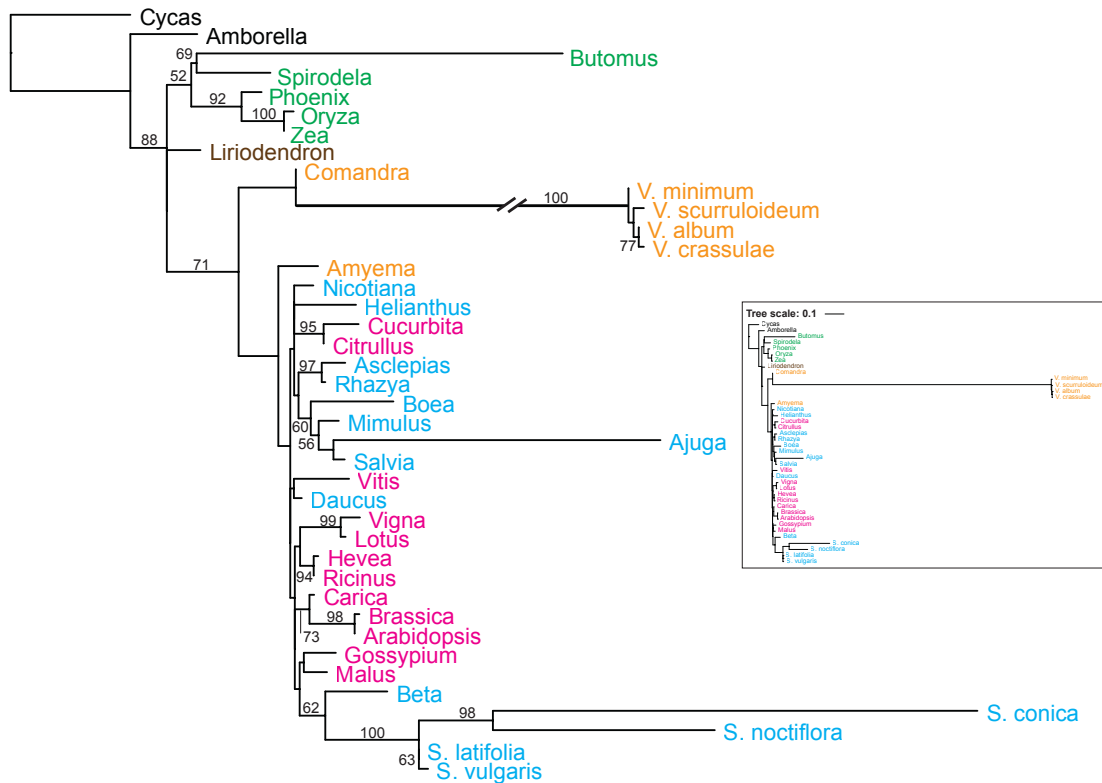


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

ccmC

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

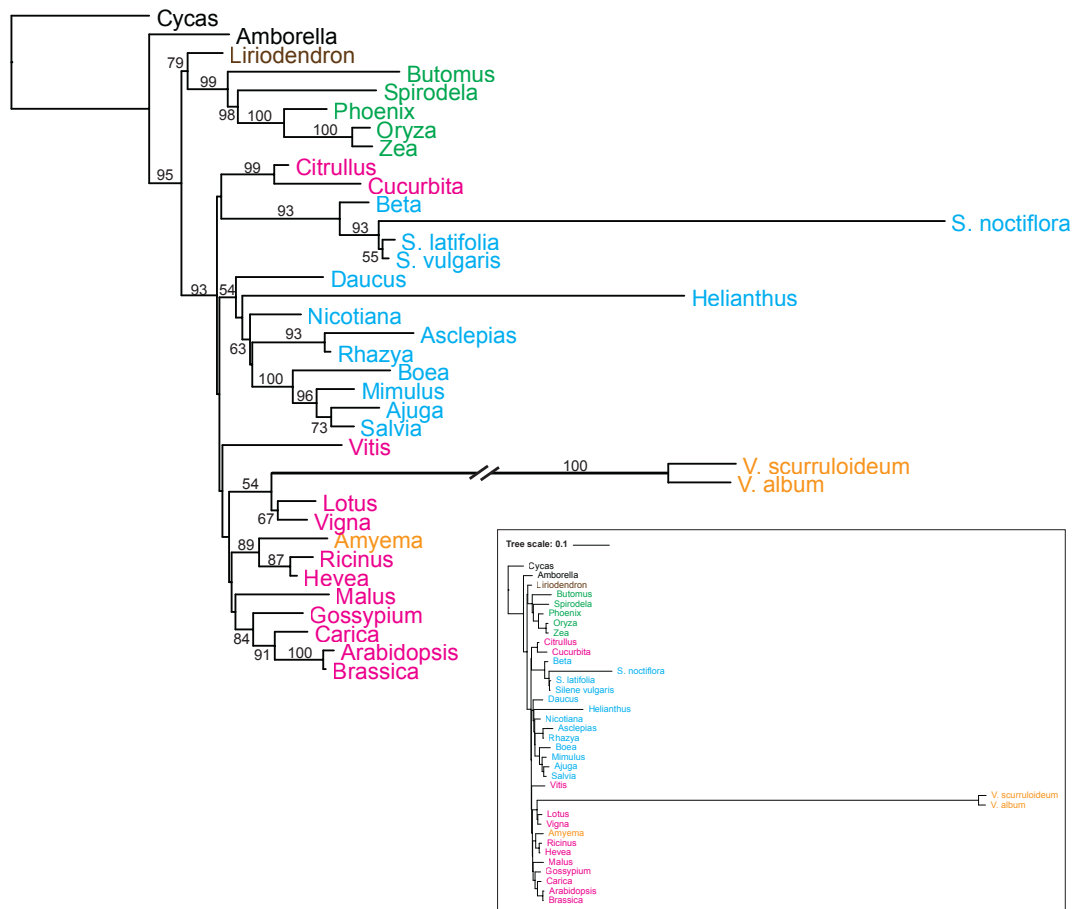


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

ccmFc

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

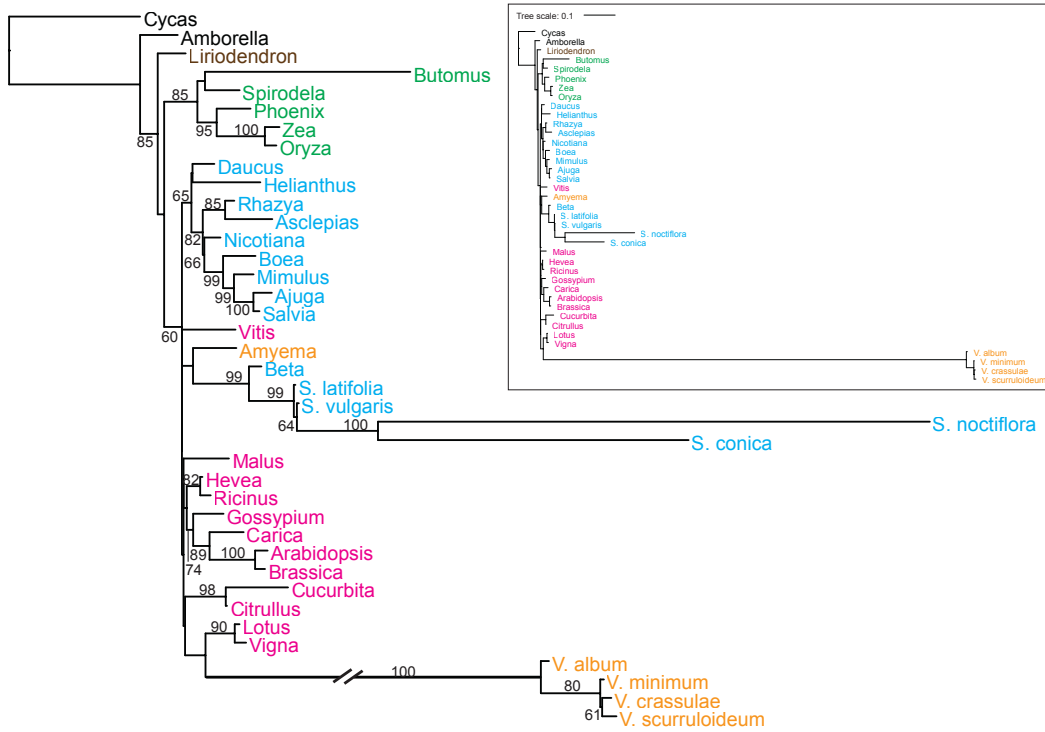


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

ccmFn

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales



**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

cob

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

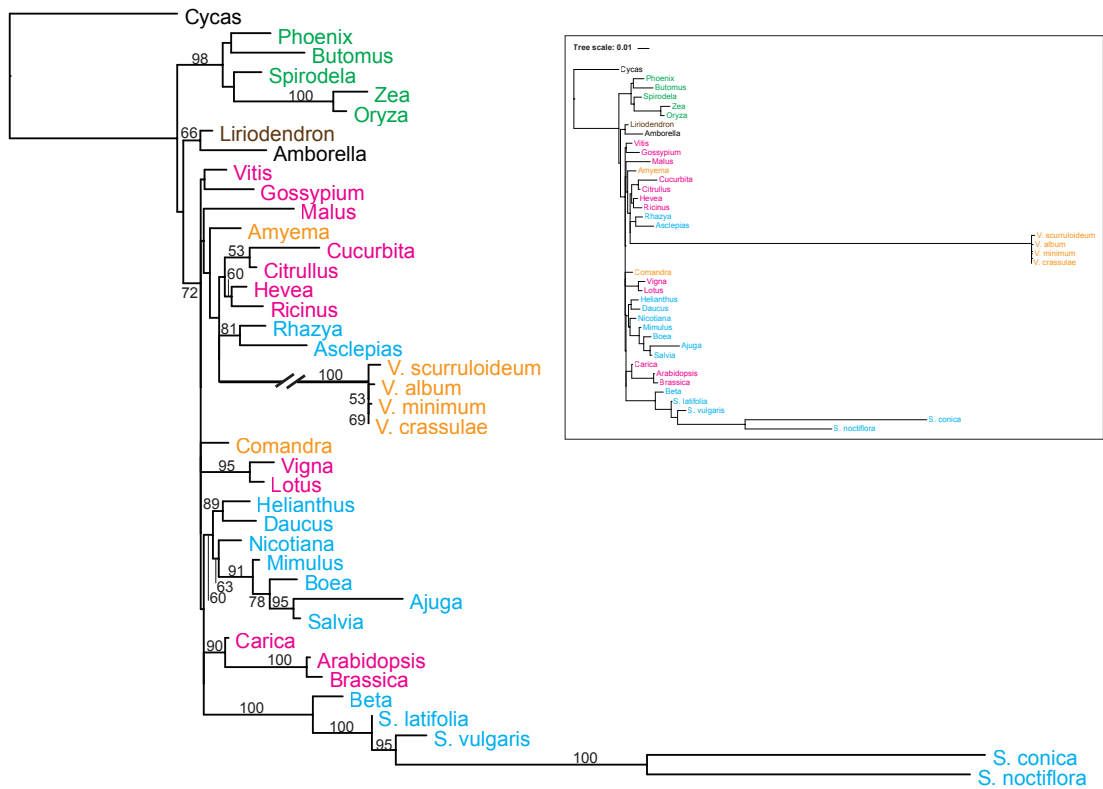


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

cox1

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

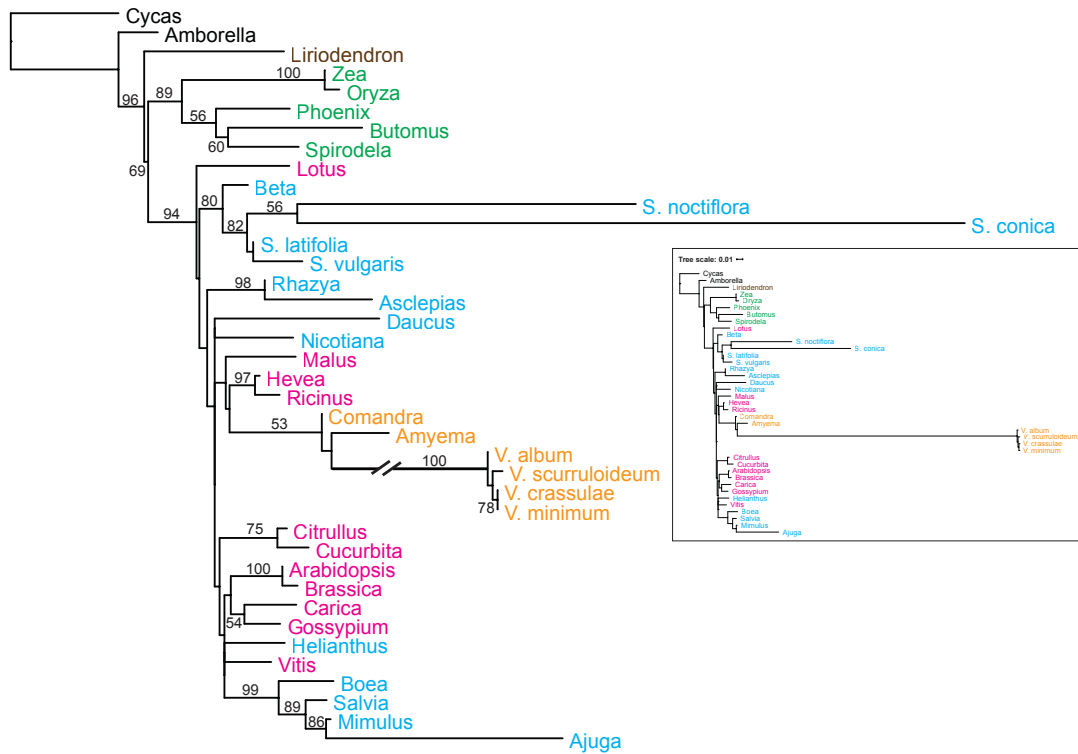


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

cox2

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

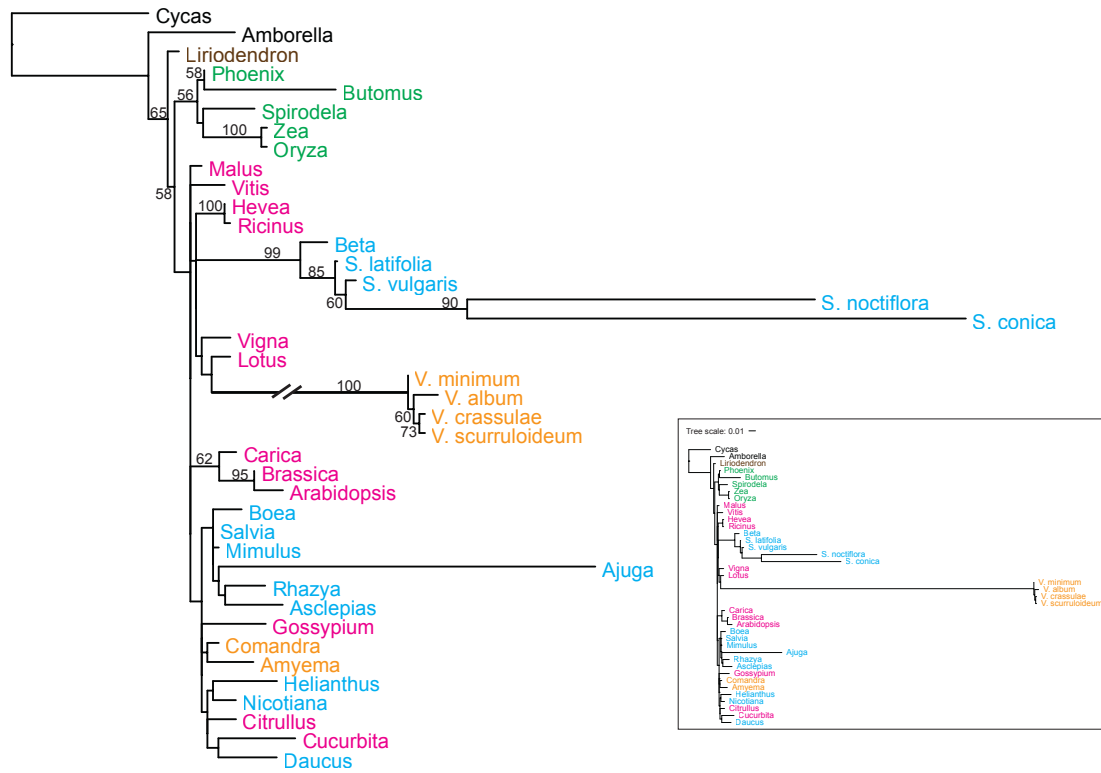


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

# cox3

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

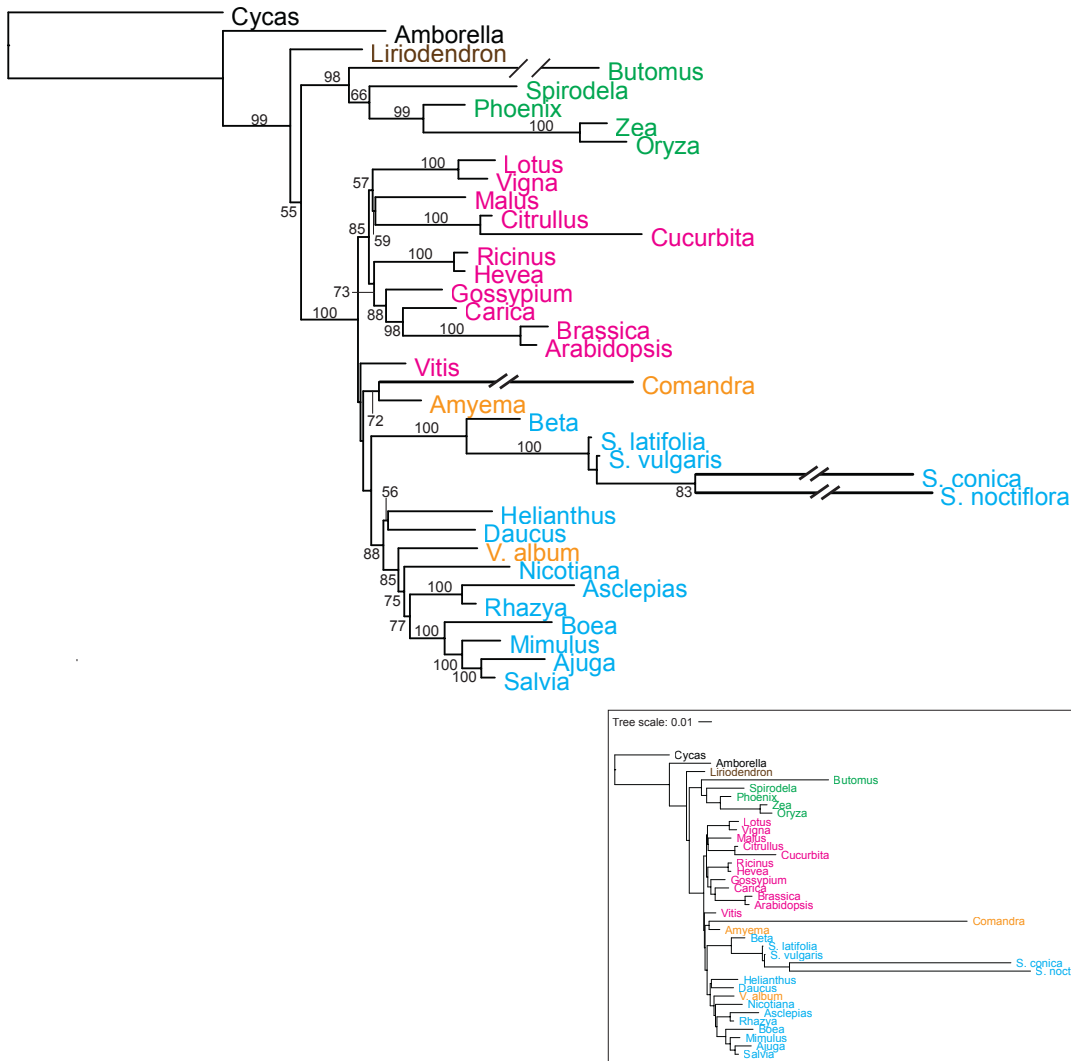


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

matR

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

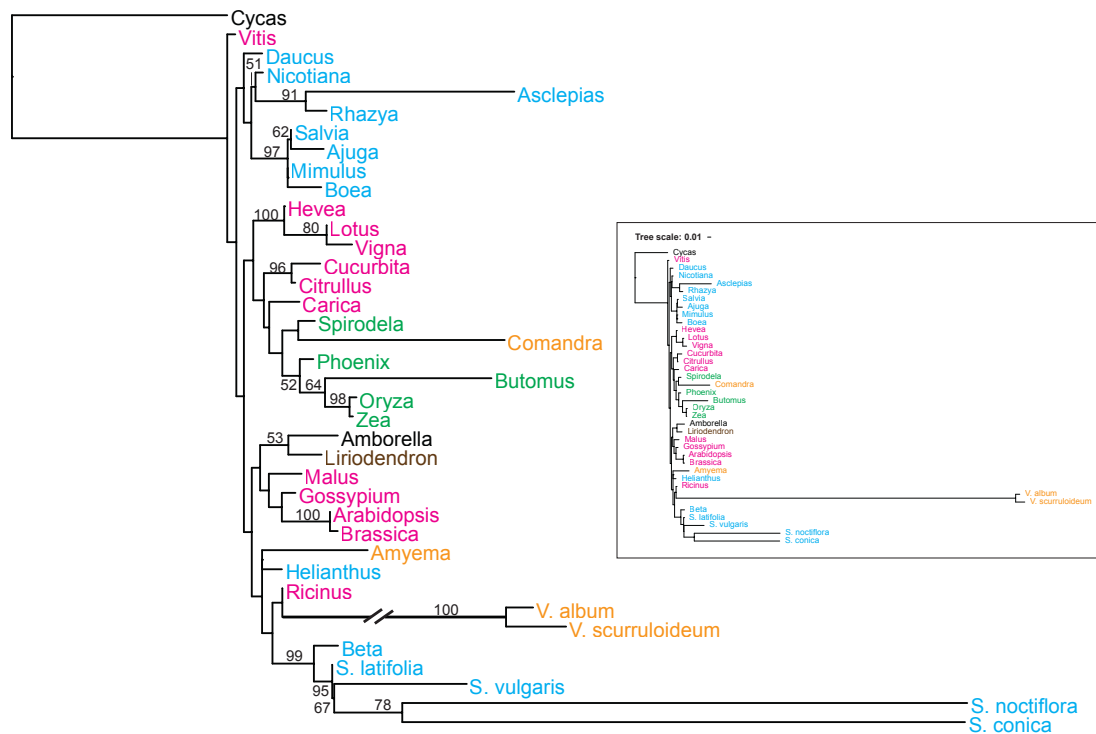


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

mttB

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

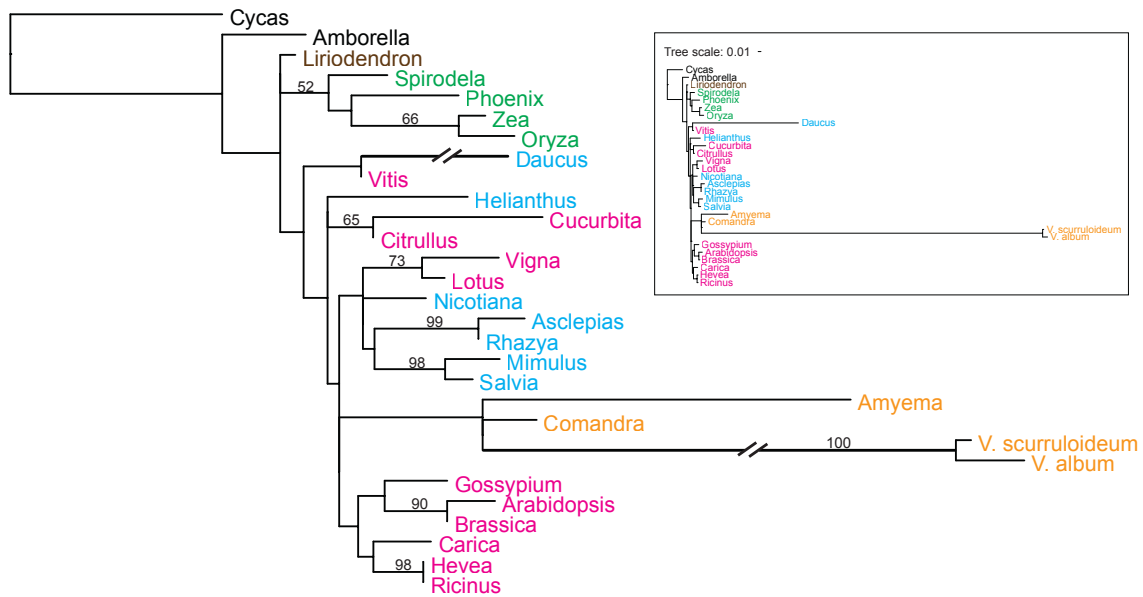


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

rpl16

tree scale: 0.01

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

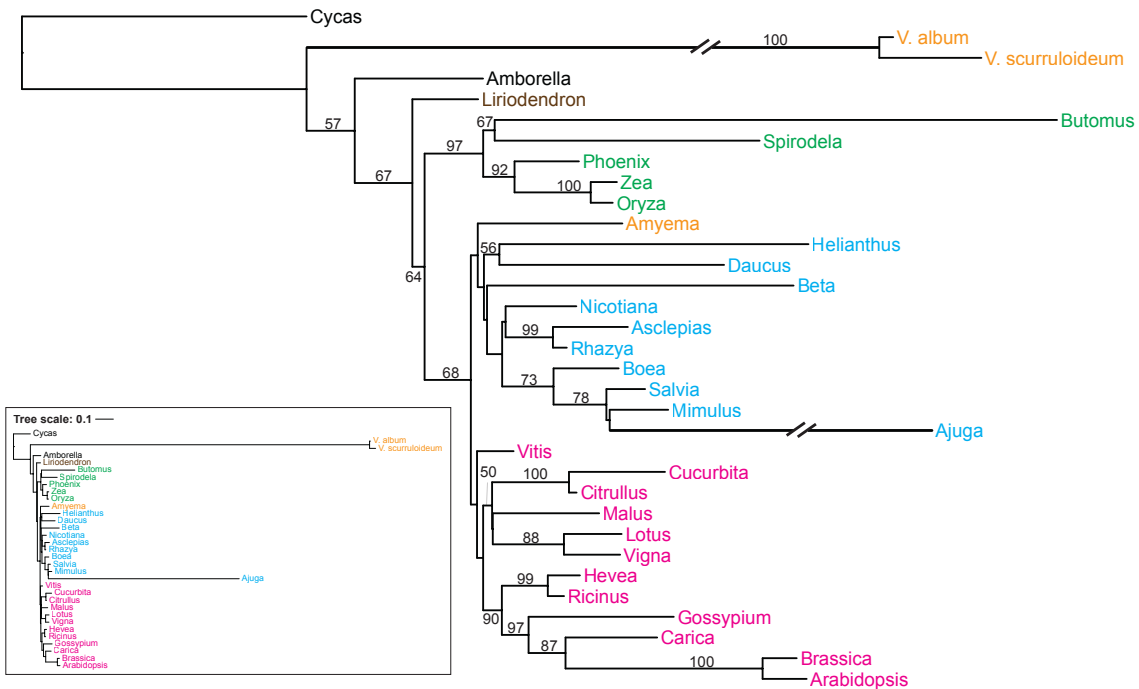


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

rps3

tree scale: 0.01 —

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

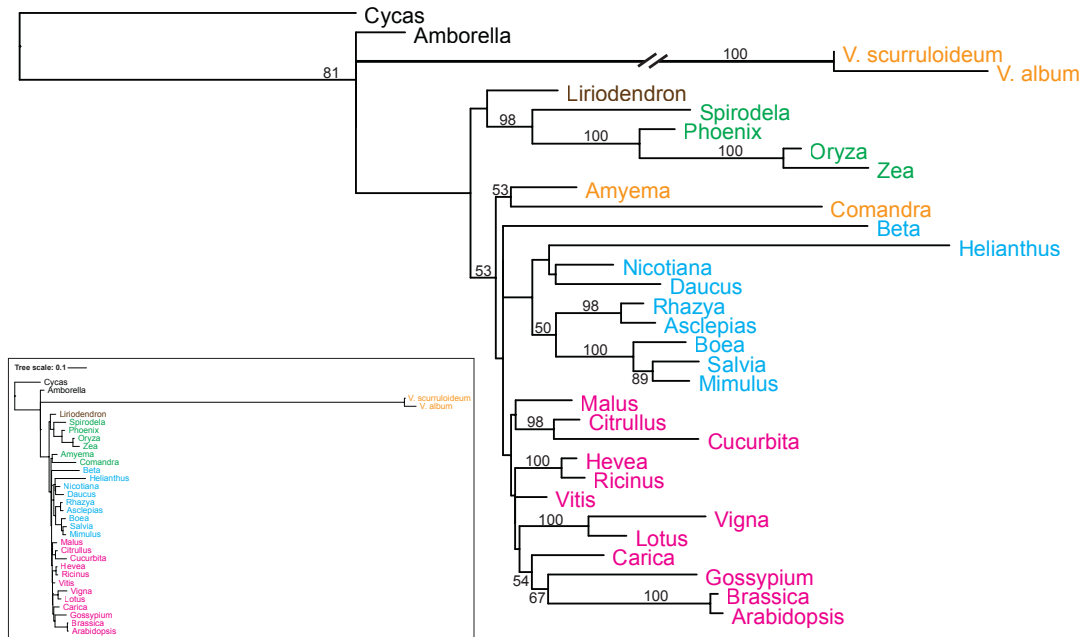


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

rps4

tree scale: 0.01 —

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

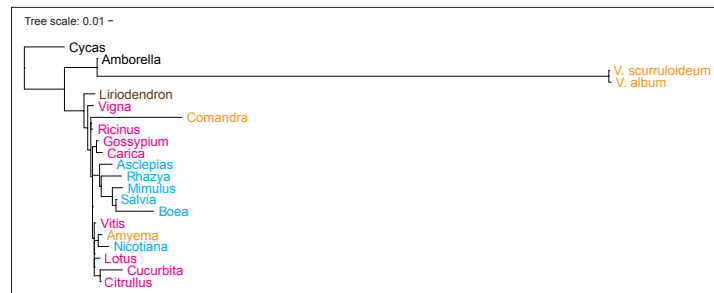
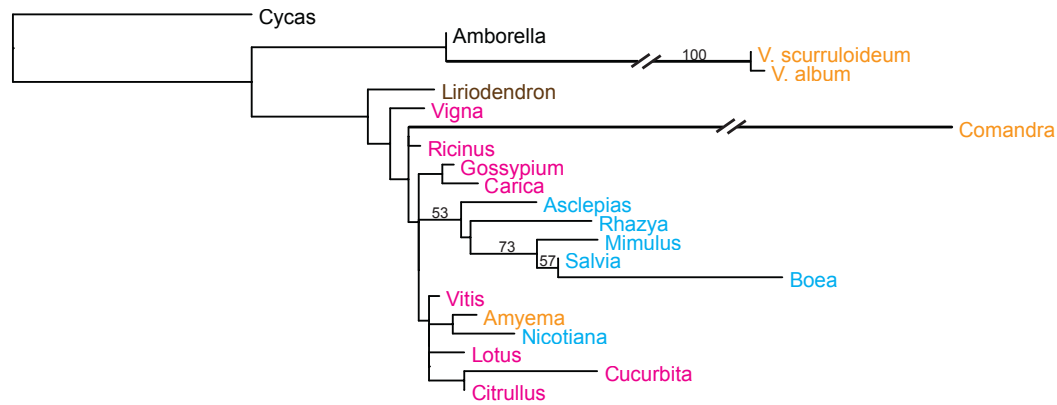


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

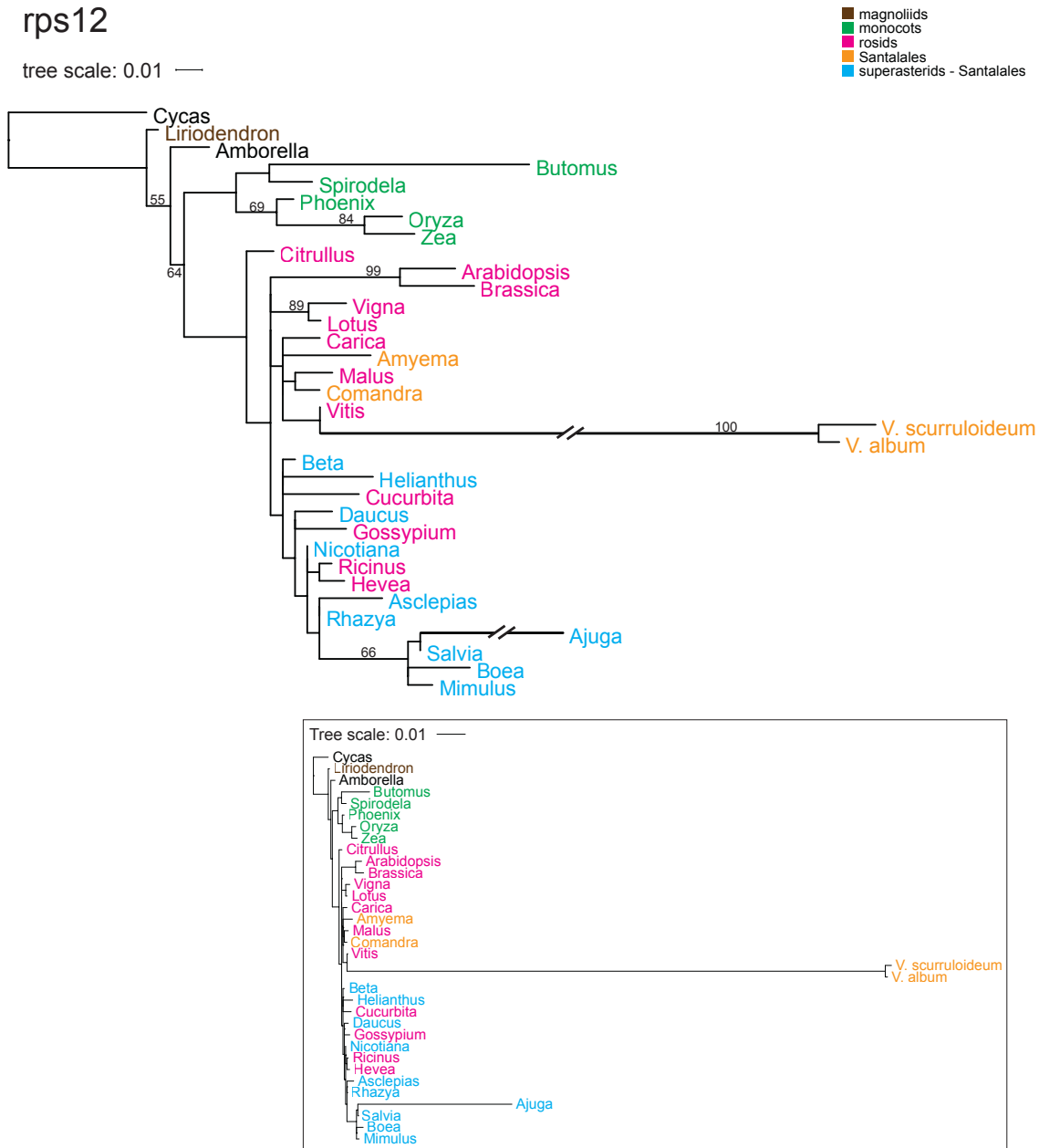
rps10

tree scale: 0.01 —

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales



**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

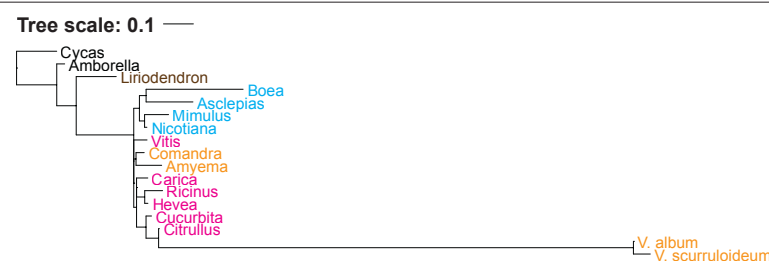
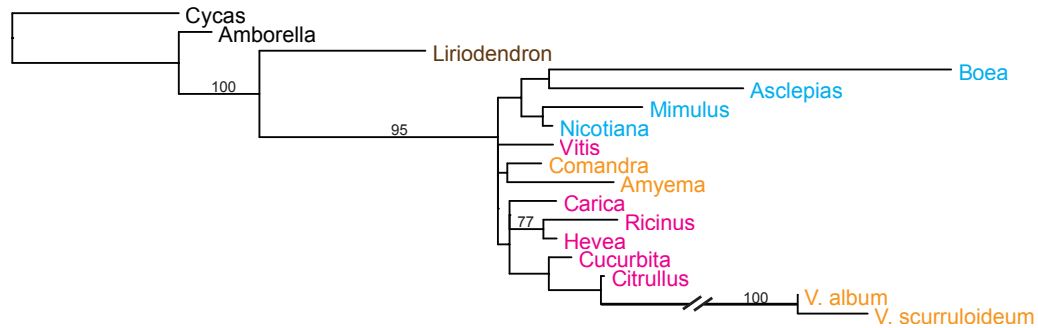


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

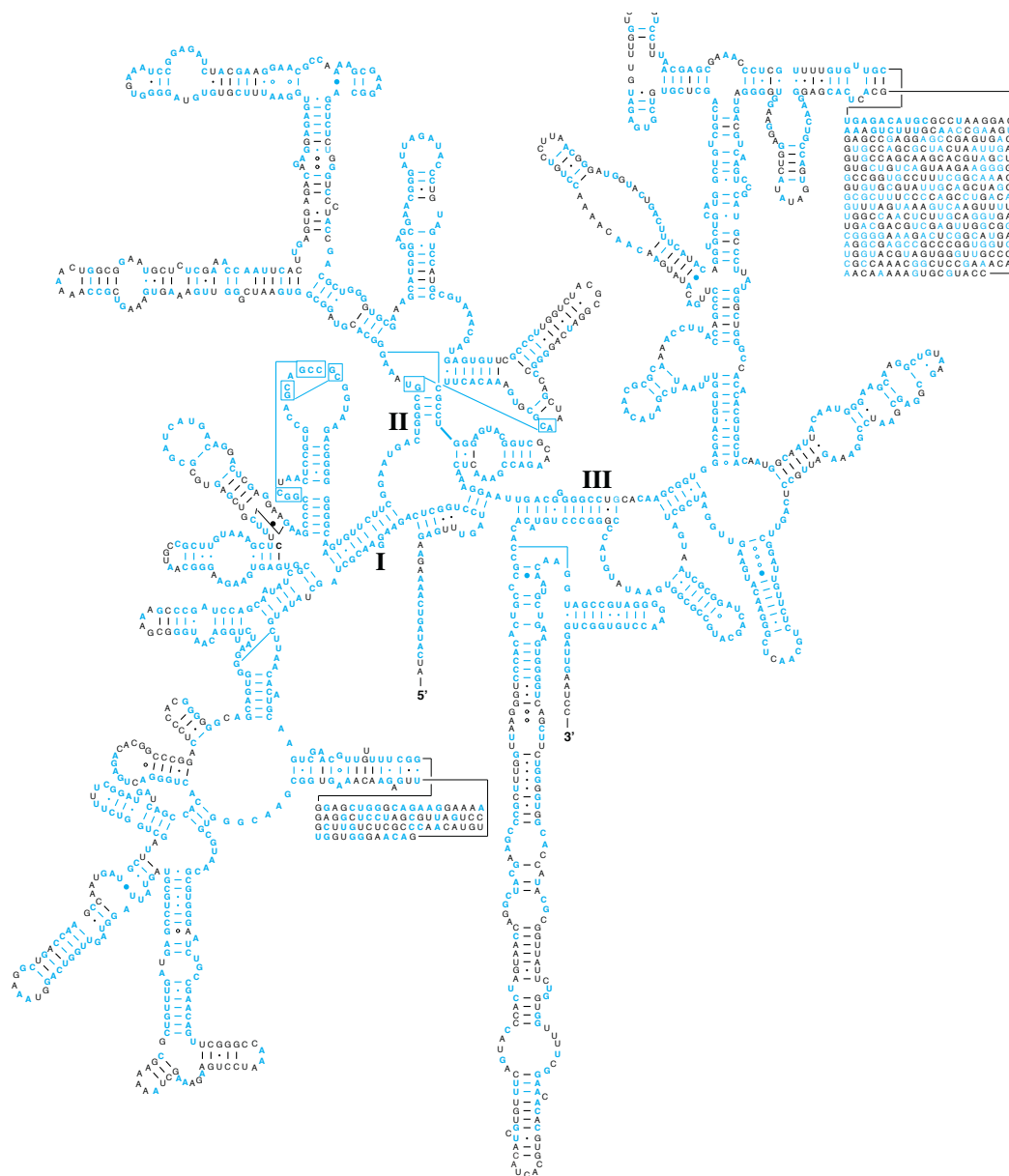
sdh3

tree scale: 0.01 —

■ magnoliids  
■ monocots  
■ rosids  
■ Santalales  
■ superasterids - Santalales

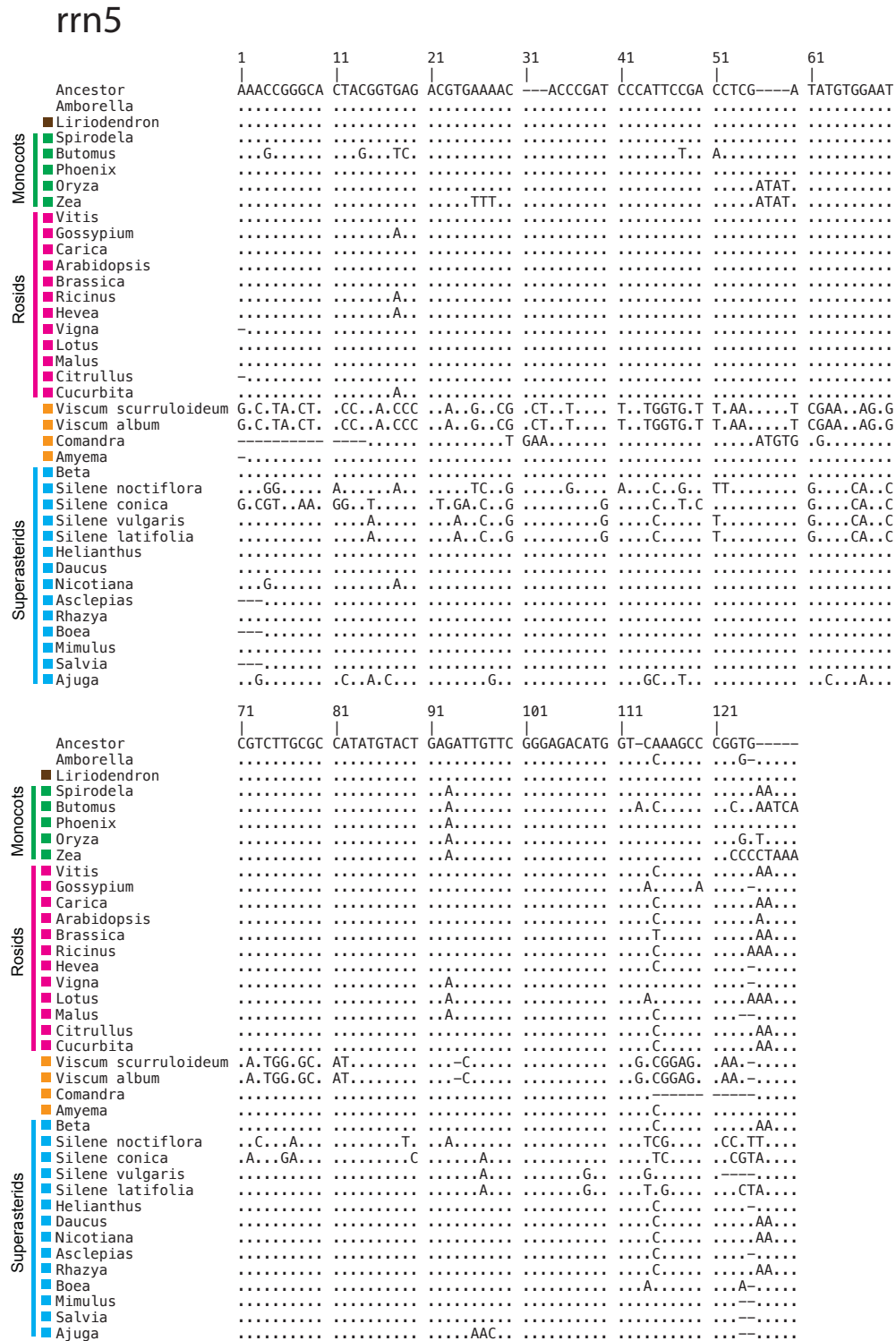


**Fig. S2** Maximum likelihood phylogenies based on all-position nucleotide alignments for three mitochondrial rRNA genes and 21 mitochondrial protein genes. Bootstrap support values > 50% from 1,000 replicates are shown. For most genes, one or a few branch lengths (usually only the branch leading to *Viscum*) are shown at a reduced scale in the main tree, with all branches shown to the same scale in the accompanying inset tree.

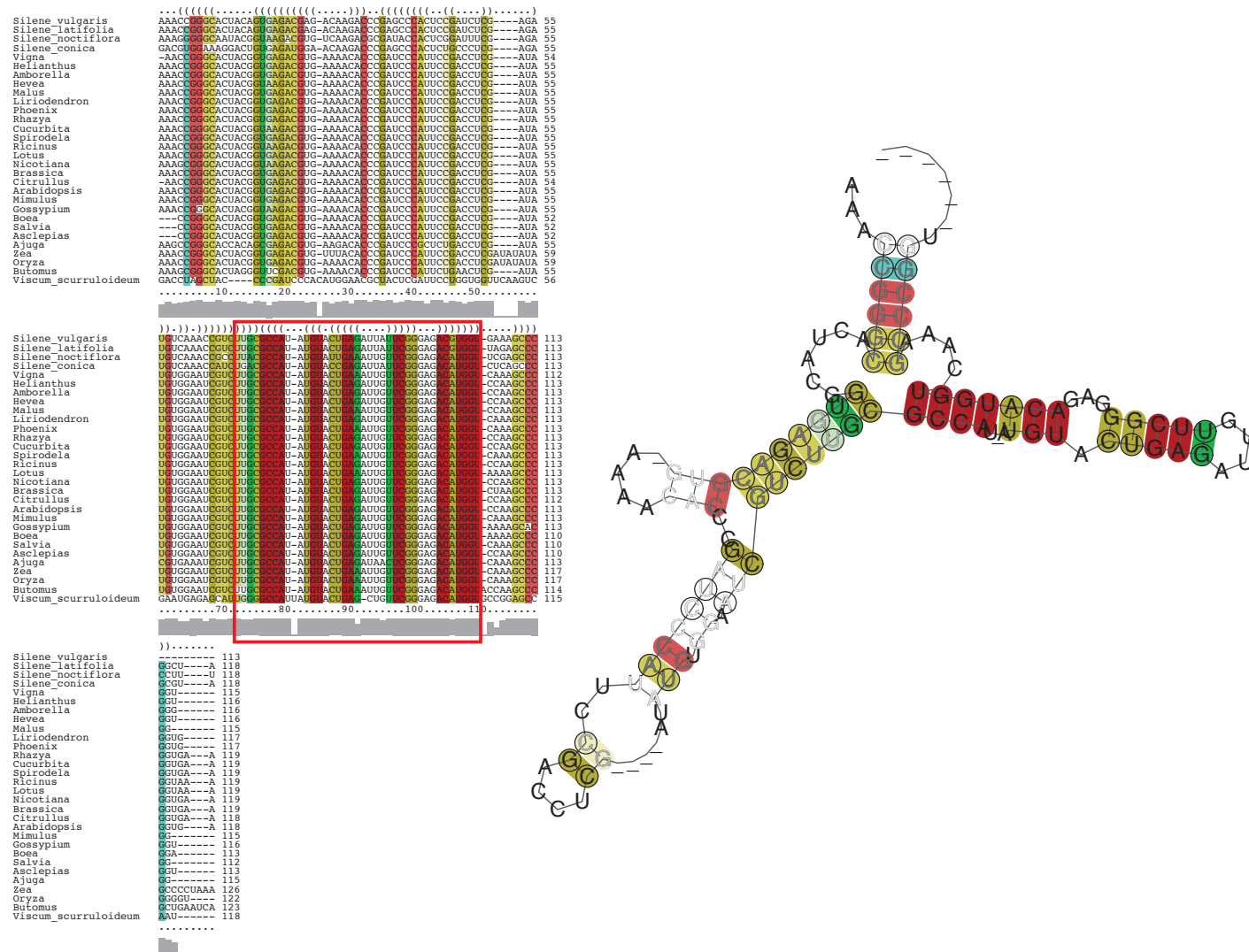


**Fig. S3** Secondary structure model of *Oenothera berteriana* SSU mitochondrial rRNA. Regions of sequence identity between the SSU rRNAs of *O. berteriana* and *V. album* are indicated in blue. This secondary structure was downloaded from the Comparative RNA Website: <http://www.rna.ccbb.utexas.edu> [2].



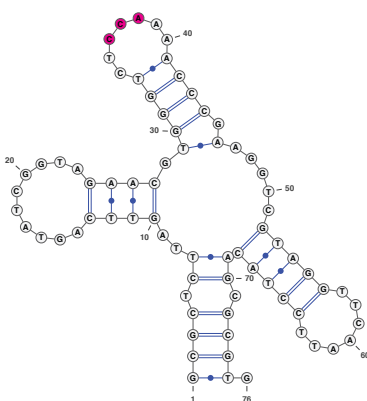


**Fig. S5** Nucleotide alignment of the mitochondrial 5S rRNA gene. Taxa are shown in phylogenetic order. The ancestral sequence was reconstructed using baseml in the PAML package [3].

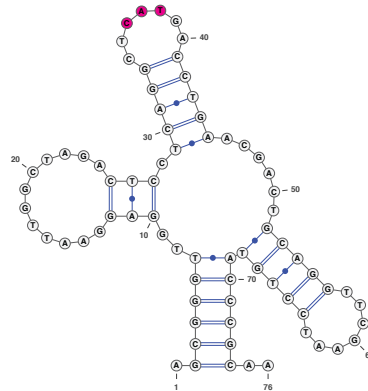


**Fig. S6** Angiosperm 5S rRNA alignment within annotated consensus structure constructed using the LocARNA [4] tool for multiple alignment of RNA molecules. The red box indicates the original *V. scurruloideum* 5S rRNA annotation.

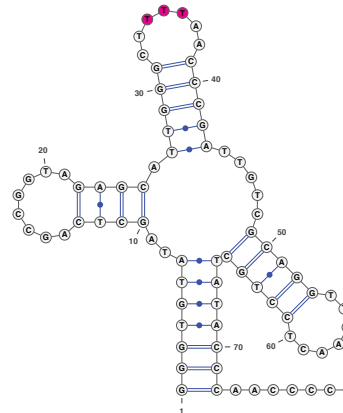
Tryptophan  
(W)(CCA)  
coordinates: 6553-6628  
cove score: 52.48



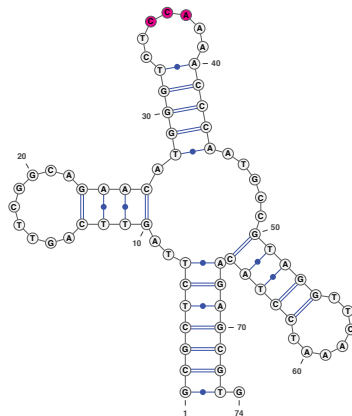
Methionine  
(M)(CAT)  
coordinates: 42653-42728  
cove score: 40.43



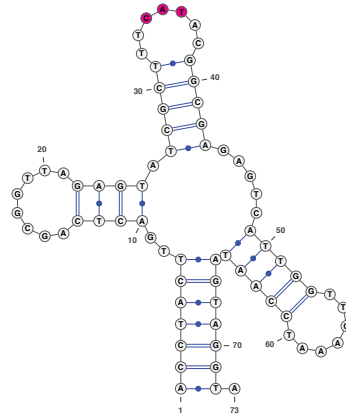
Lysine  
(K)(TTT)  
coordinates: complement(500766-500844)  
cove score: 78.53



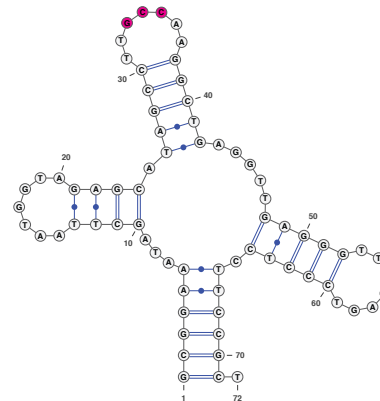
Tryptophan  
(W)(CCA)  
coordinates: 148987-149060  
cove score: 70.10



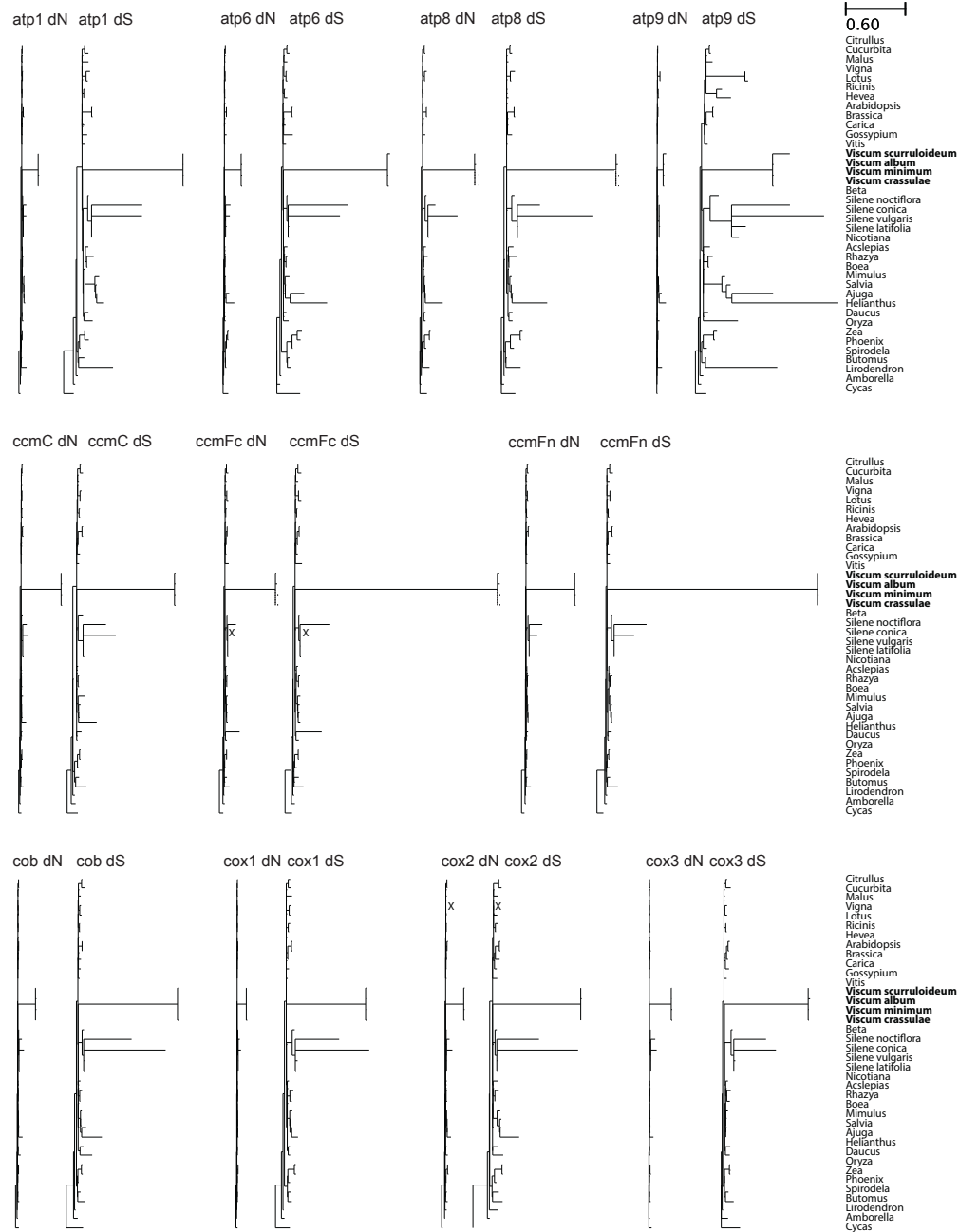
Methionine  
(M)(CAT)  
coordinates: complement(412180-412252)  
cove score: 70.15



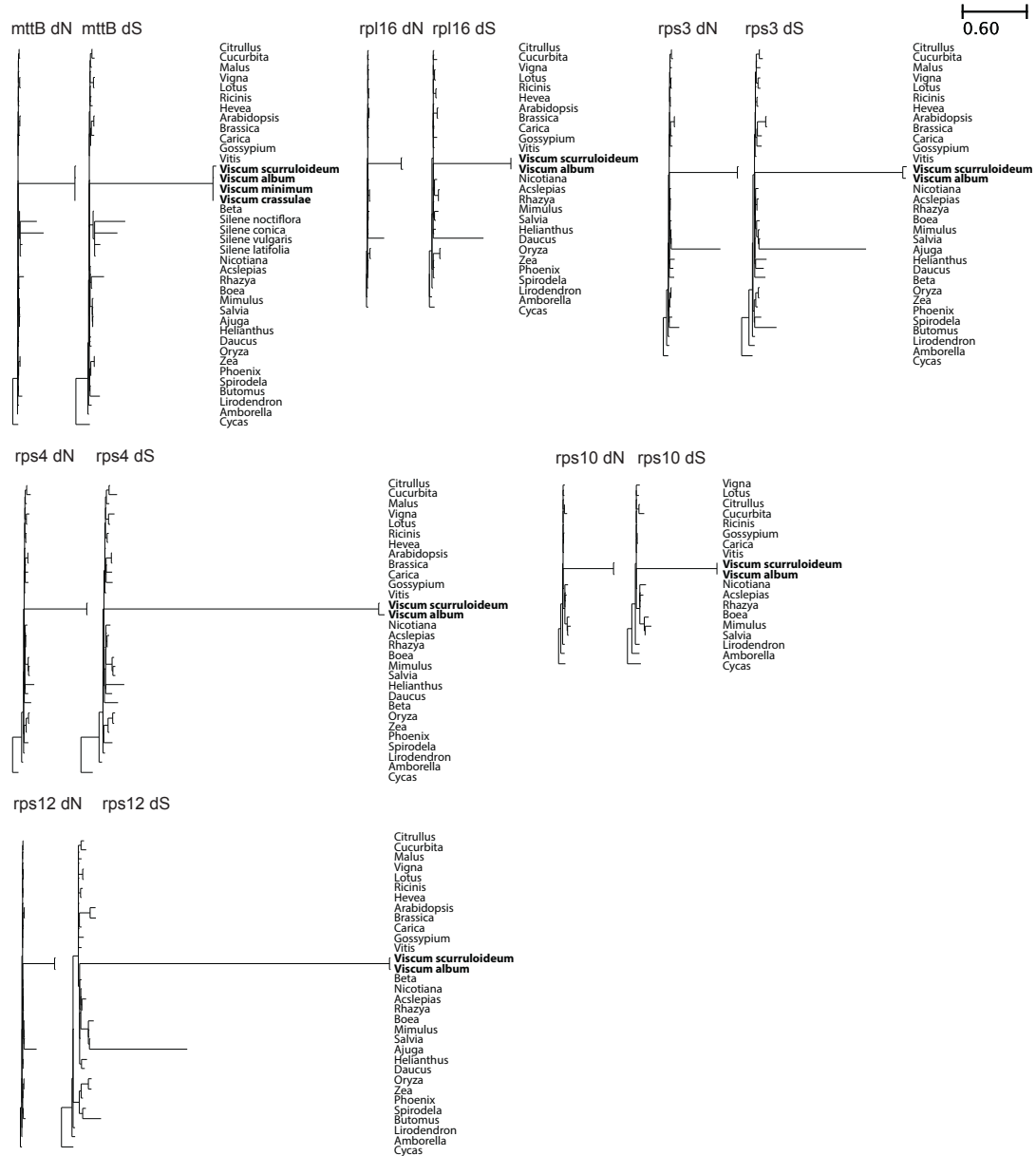
Glycine  
(G)(GCC)  
coordinates: complement(92533-92604)  
cove score: 67.59



**Fig. S7** *Viscum album* tRNAs. Anticodons are marked in pink. Secondary structures and anticodons were predicted using tRNAscan-SE [5]. Prediction scores are shown for each tRNA. Secondary structures were visualized using the VARNA java web start applet (<http://varna.lri.fr/>) [6].



**Fig. S8** Mitochondrial sequence divergence among 37 angiosperms and outgroup *Cycas*. Shown are topologically constrained phylograms of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) divergence for 17 of the 19 mitochondrial protein genes present in both *V. scurruloideum* and *V. album*. All trees are shown to the same scale. For “core” genes (*atp1-mttB*), an “X” indicates gene loss, and a dotted line indicates that a gene has not been annotated in *V. crassulae* and *V. minimum* by Petersen et al. (2015), but is present in both *V. scurruloideum* and *V. album*, and is probably present in the other two. For “variable” genes (*rpl16-rps12*), only species that contain the gene are shown. Phylograms for *atp4* and *sdh3* are not shown due to exceptionally high amino acid divergence.



**Fig. S8** Mitochondrial sequence divergence among 37 angiosperms and outgroup *Cycas*. Shown are topologically constrained phylograms of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) divergence for 17 of the 19 mitochondrial protein genes present in both *V. scurruloideum* and *V. album*. All trees are shown to the same scale. For “core” genes (*atp1*-*mttB*), an “X” indicates gene loss, and a dotted line indicates that a gene has not been annotated in *V. crassulae* and *V. minimum* by Petersen et al. (2015), but is present in both *V. scurruloideum* and *V. album*, and is probably present in the other two. For “variable” genes (*rpl16*-*rps12*), only species that contain the gene are shown. Phylograms for *atp4* and *sdh3* are not shown due to exceptionally high amino acid divergence.

## References

1. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 2015;32:820-32.
2. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 2012;3:2.
3. Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586-91.
4. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* 2012;18:900-14.
5. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955-64.
6. Darty K, Denise A, Ponty Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 2009;25:1974-75.