

Title: **Over 100 million years of enzyme evolution underpinning the production of morphine in the Papaveraceae family of flowering plants**

General information for gene trees

The best-scoring maximum likelihood tree was inferred using RAxML (Stamatakis, 2006) All branches are drawn to scale as indicated by the scale bar (substitutions/site) and the bootstrap values from an analysis of 100 replicates are shown at the nodes. The solid diamonds indicate the root of the tree. The strongly supported clades are indicated with dark grey bars. Branch colours indicate species. Each subclade that contains sequences from species of the same family is highlighted in colour. Additional sequence information is in Supplemental Table S5.

Species information is embedded in the sequence identifiers: ACxG, EC, MC, MI, NN, OS or PS prefixes denote sequences identified from the genome annotation datasets of *A. coerulea*, *E. californica*, *M. cordata*, *M. integrifolia*, *N. nucifera*, *O. sativa* and *P. somniferum* respectively. NCBI Genebank accession number corresponding to these annotated protein sequences and/or gene names are appended to their sequence identifiers, separated by a slash. For the remaining Genebank sequences, a three letter code following an underscore symbol attached at the end of each sequence identifier indicates the species information of the protein sequence. These are: *_Aar* (*Actinidia arguta*), *_Aer* (*Actinidia eriantha*), *_Afi* (*Aristolochia fimbriata*), *_Ame* (*Argemone mexicana*), *_Ath* (*Arabidopsis thaliana*), *_Bly* (*Berberis lyceum*), *_Bst* (*Berberis stolonifera*), *_Cch* (*Coptis chinensis*), *_Cja* (*Coptis japonica*), *_Cma* (*Chelidonium majus*), *_Csa* (*Corydalis saxicola*), *_Cte* (*Coptis teeta*), *_Dve* (*Dysosma versipellis*), *_Eca* (*Eschscholzia californica*), *_Gfl* (*Glaucium flavum*), *_Mno* (*Morus notabilis*), *_Pan* (*Parasponia andersonii*), *_Pbr* (*Papaver bracteatum*), *_Por* (*Papaver orientale*), *_Ppe* (*Podophyllum peltatum*), *_Pps* (*Papaver pseudo-orientale*), *_Prh* (*Papaver rhoeas*), *_Qsu* (*Quercus suber*), *_She* (*Sinopodophyllum hexandrum*), *_Tfl* (*Thalictrum flavum*), *_Tco* (*Tinospora cordifolia*) and *_Tor* (*Trema orientale*). Gene names are attached where available, separated by a slash symbol. Detailed information on the sequence identifiers in the trees are given in Supplemental Table S5. The species and the families they belong to are highlighted in different colours.

Supplemental Figures

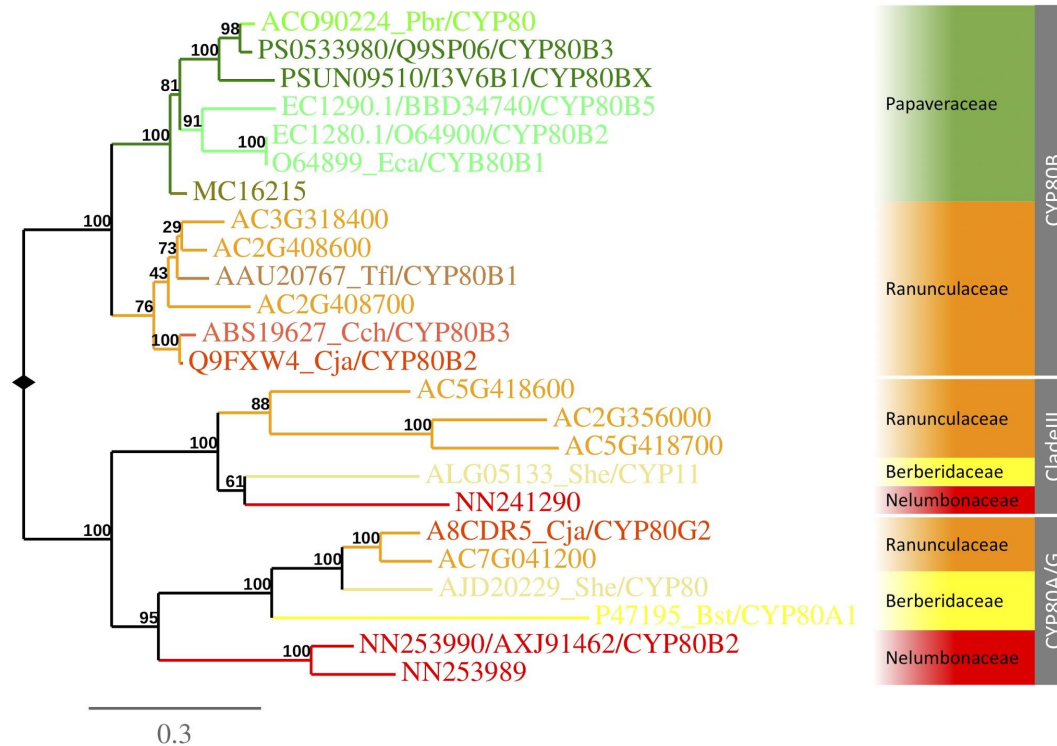


Figure S1. The CYP80 gene tree shows a monophyletic CYP80B clade in the order of Ranunculales. All functionally characterised (*S*)-*N*-methylcoclaurine 3'-hydroxylases are members of the CYP80B clade. Sequences from species of the Ranunculaceae or Papaveraceae family form two strongly supported subclades.

The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

Additional discussion on Figure S1.

The CYP80 gene tree shows that the CYP80A/G clade contains only Ranunculaceae and Berberidaceae sequences including a single sequence from *A. coerulea* genome. Two functionally characterized members of the CYP80A/G clade are CYP80G2 and CYP80A1. CYP80G2 (corytuberine synthase) were reported to catalyse the conversion of (*S*)-reticuline to (*S*)-corytuberine for the biosynthesis of aporphine alkaloid magnoflorine in *C. japonica* (Ikezawa *et al.*, 2008), whereas CYP80A1 is a berbaminine synthase found to catalyse the production of the bisbenzylisoquinoline alkaloid berbaminine in *Berberis stolonifera* (Kraus and Kutchan 1995). However, no orthologues of CYP80A1/G2 are present in the genome assemblies of the three Papaveraceae species. The absence of orthologous copies of CYP80G2/CYP80A1 in Papaveraceae species suggests that these genes were lost before the divergence of the three Papaveraceae species in this family lineage.

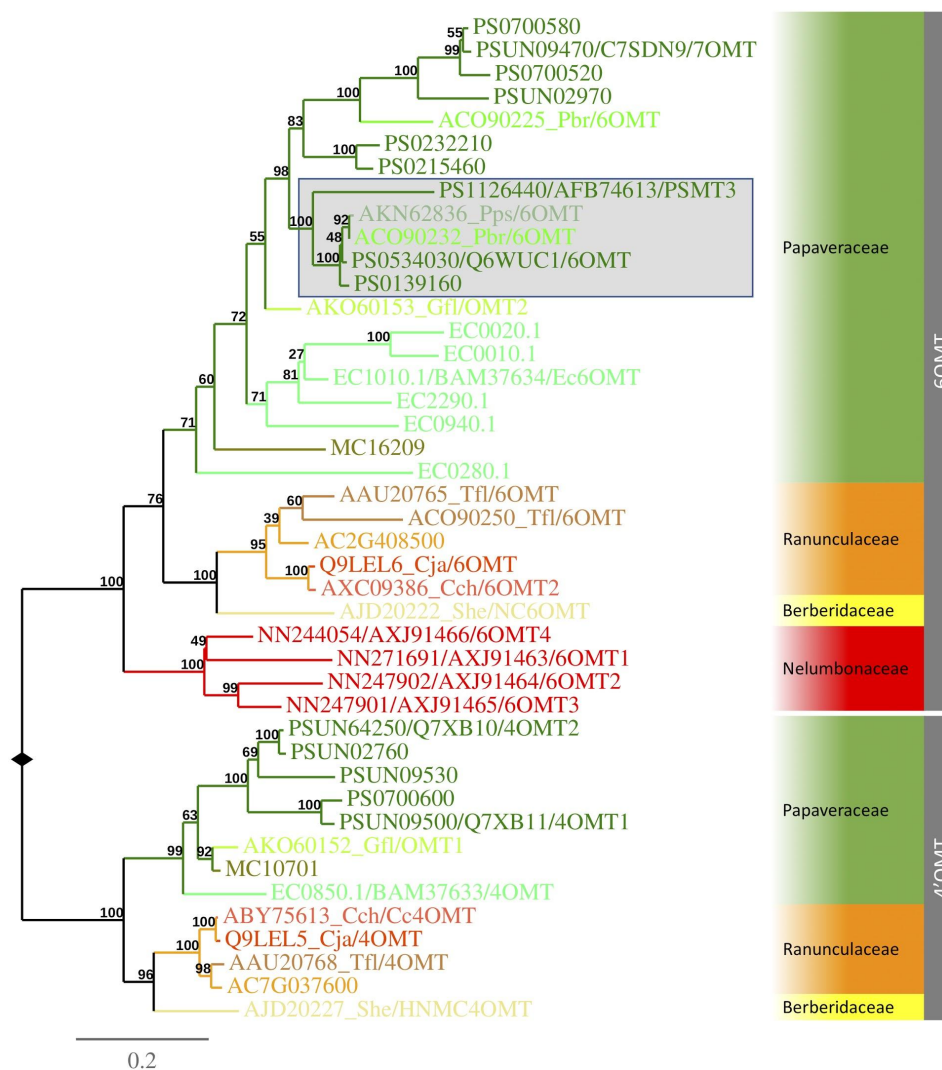


Figure S2. Phylogenetic analysis of the 4'OMT/6OMT subfamily. The subfamily is comprised of two strong supported 4'OMT and 6OMT sister clades. The 4'OMT clade contains all functionally characterised 4'OMTs, whereas all characterised 6OMTs are members of the 6OMT clade. Sequences from the species of each plant family form strongly supported subclades. PSMT3, the *O*-methyltransferase 3 involved in the noscapine biosynthesis, is a member of a robust opium poppy subgroup nested inside the 6OMT clade which is highlighted by the shaded box in the tree. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

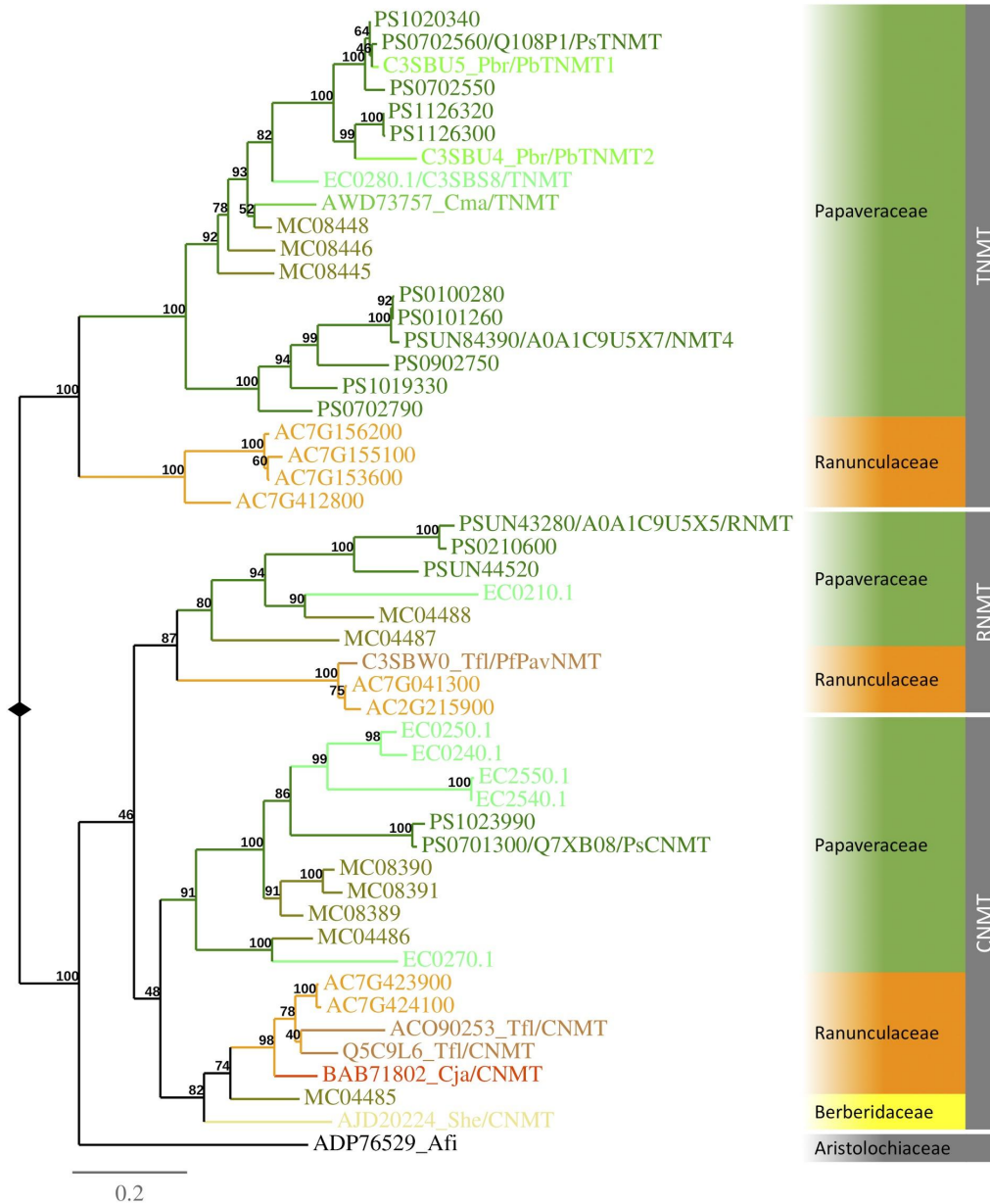


Figure S3. Phylogenetic analysis of the *N*-methyltransferase subfamily. The *N*-methyltransferase subfamily consists of three clades, including CNMT, RNMT and TNMT clades. Each clade contains sequences exclusively from the Ranunculales and the subclades within these clades sequences are comprised overwhelmingly of members from the same family. MC04485 of *M. cordata*, being the only anomalous case, falls outside the Papaveraceae subclade within the CNMT clade. All four Ranunculales genomes are represented in each of these three clades. The CNMT clade contains all functionally characterised CNMTs, whereas all characterised TNMTs are members of the TNMT clade. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

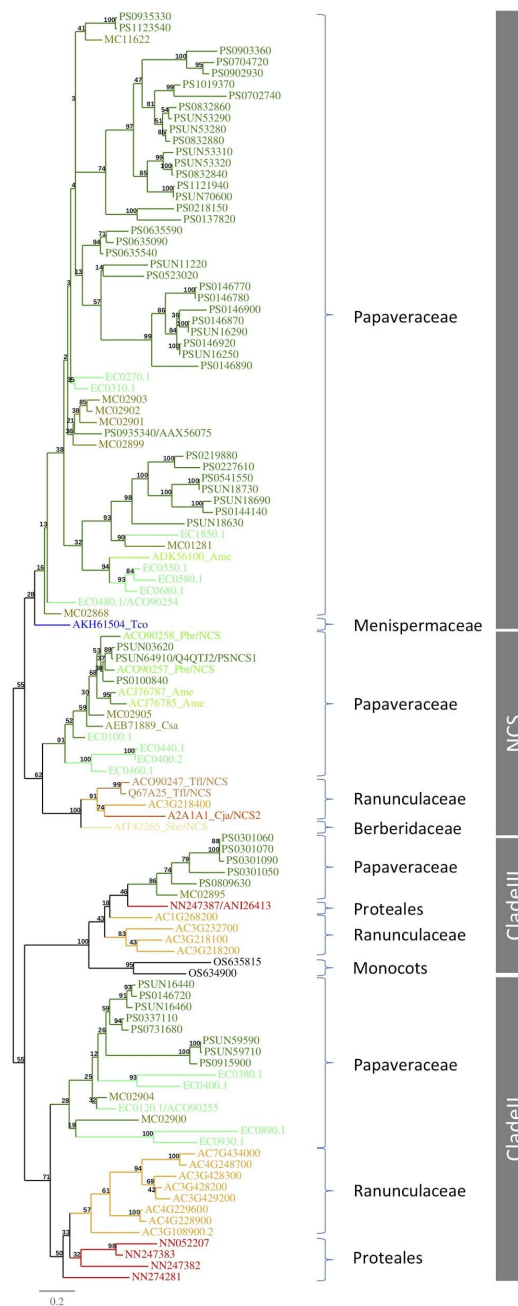


Figure S4. Phylogenetic analyses suggests paralogous relationship of the sacred lotus NCS like sequences to the Ranunculales NCS. Phylogenetic analysis of NCS related sequences identifies three distinct clades among them. The NCS clade consists of sequences exclusively from the Ranunculales, and relationships within the NCS clade is further refined in Figure 3. The other two clades, II and III, contain five sequences from *N. nucifera*, with NN247387 in clade III and the other four in clade II. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees. Curly brackets indicate plant family of the species for the group of sequences.

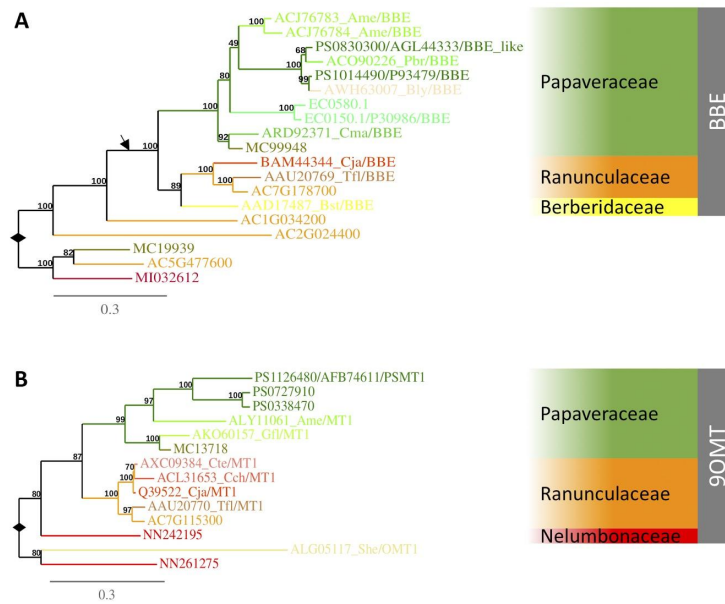


Figure S5. Monophyletic origin of the protoberberine biosynthetic enzymes BBE and 9OMT. Gene tree analyses identify robust clades for functionally characterised BBE and 9OMT clades respectively. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees. **(A).** The root of BBE clade is indicated by a black arrow in the BBE tree, nested in a group containing exclusively Ranunculales sequences. The BBE clade consists of all functionally characterised BBE enzymes. Orthologous sequences from all four Ranunculales genomes are present. Sequences from species of the same family form strongly supported subclades with a single exception. The anomalous unpublished sequence, AWH63007, was downloaded from the Genbank NR database and entered as from *Berberis lyceum* of the Berberidaceae family. **(B).** The 9OMT clade contains a strongly supported Ranunculales group which is sister to a single sacred lotus sequence, NN242195, in the 9OMT tree. Unlike the BBE clade, no orthologous sequence was found in the genome of *E. californica* but the three Ranunculales genomes are represented. Sequences from species of the same family form strongly supported subclades.

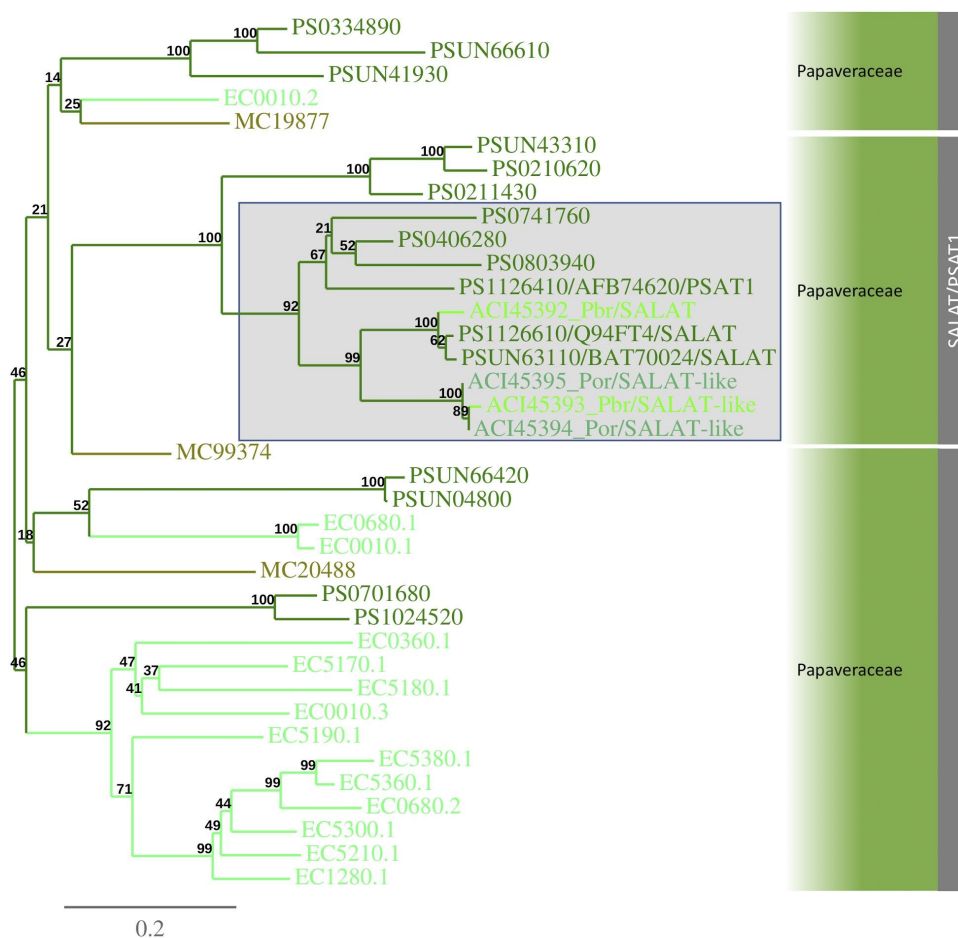


Figure S6. Phylogenetic analysis of the acetyltransferase subfamily reveals close relationship of SALAT and PSAT1. This acetyltransferase subfamily is comprised of exclusively Papaveraceae sequences. However, the acetyltransferase gene tree shows that PSAT1 and SALAT, involved in the biosynthesis of phthalideisoquinoline alkaloid noscapine and morphinan alkaloids in opium poppy, are two members of the same clade in the opium poppy lineage. A subgroup containing both sequences is highlighted by the shaded box in SALAT/PSAT1 clade. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

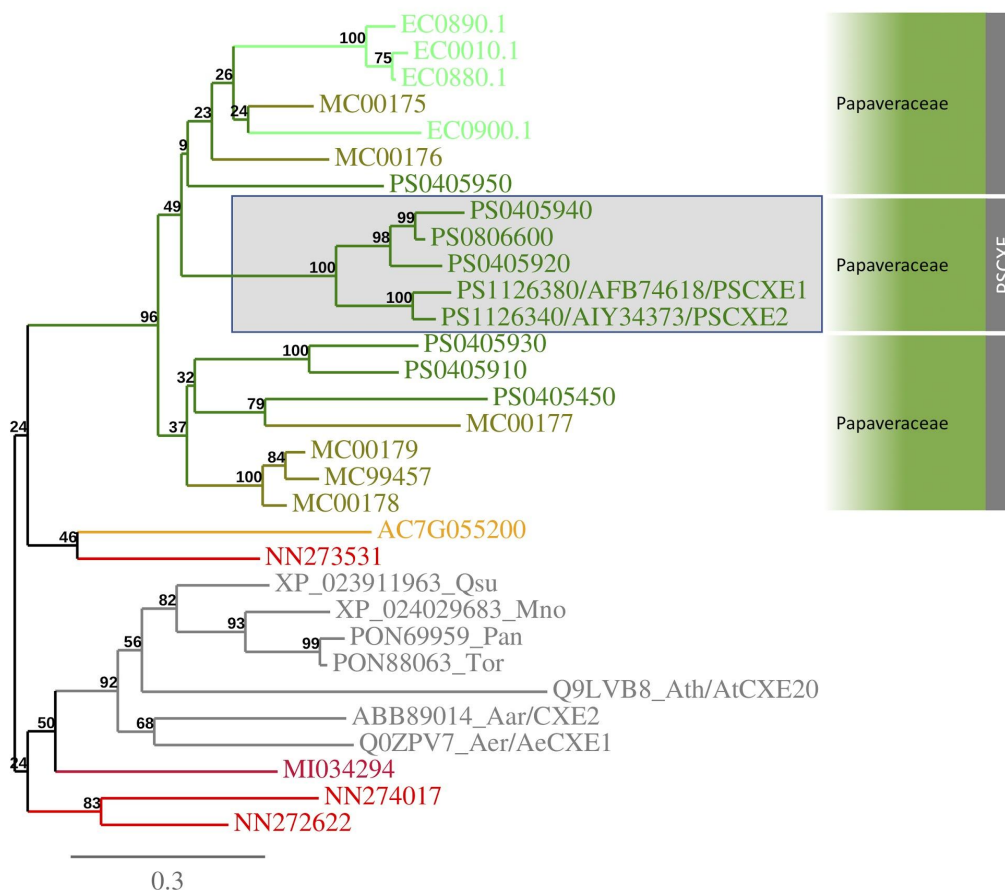


Figure S7. PXCXE1 is a member of the CXE clade specific to the opium poppy lineage. The carboxylesterase gene tree shows that PSCXE1, involved in the biosynthesis of phthalideisoquinoline alkaloid noscapine in opium poppy, is a member of the PSCXE clade that is nested in a larger group that contains sequences exclusively from the Papaveraceae family. However, the PSCXE clade is comprised of only opium poppy sequences and is highlighted by the shaded box in the tree. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

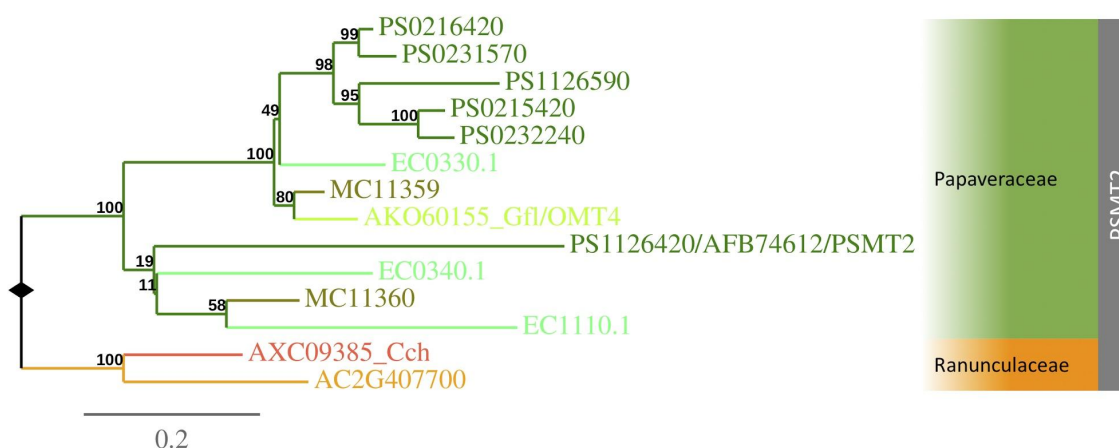


Figure S8. Phylogenetic analysis identifies a PSMT2 clade in Ranunculales. The PSMT2 clade consists of two sister subclades. One contains two sequences from the Ranunculaceae family, and the other has sequences exclusively from the Papaveraceae family. All four Ranunculales genomes are represented. PSMT2 is the only member that has been functionally characterised in this clade so far. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

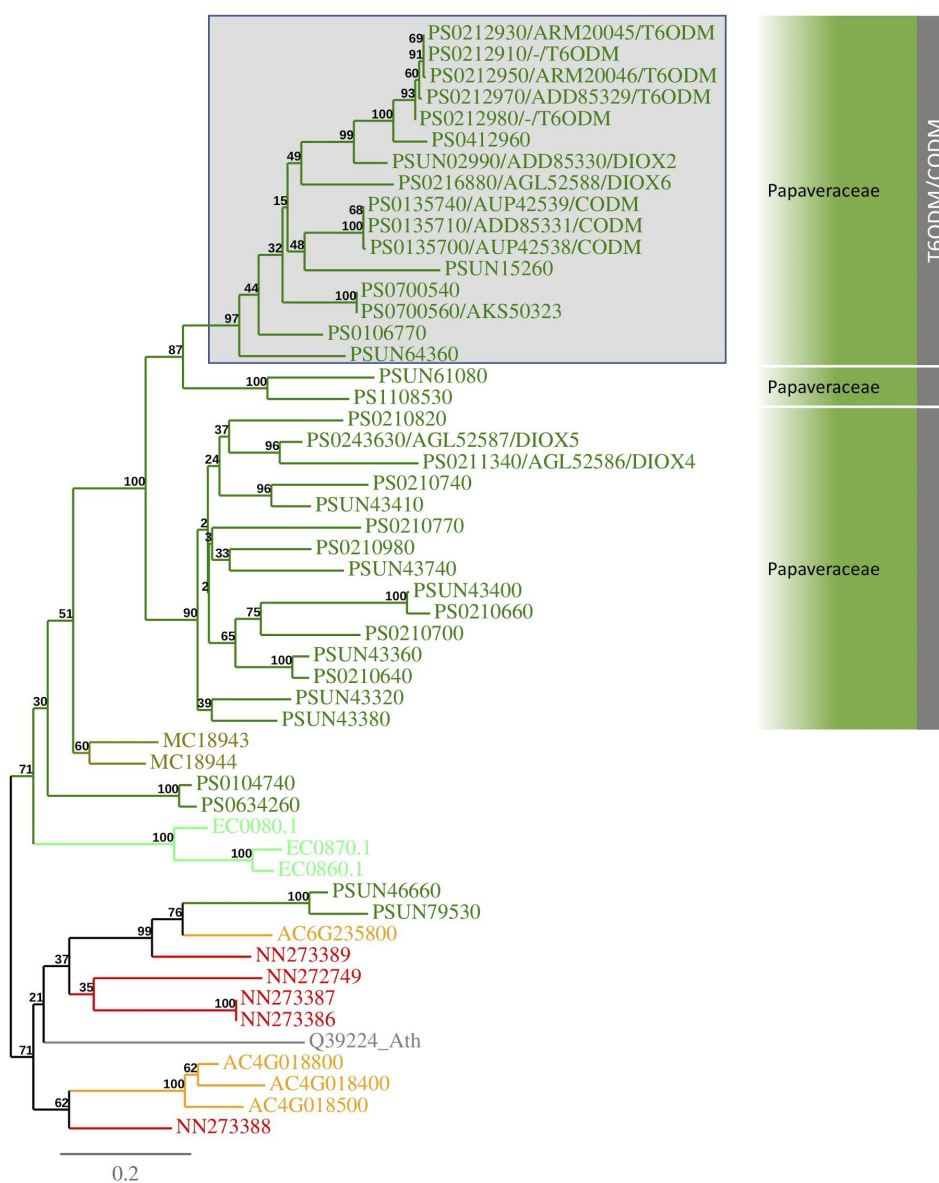


Figure S9. Lineage specific expansion of the 2-oxoglutarate/Fe(II)-dependent dioxygenase subfamily in the opium poppy lineage. The phylogenetic analysis of the 2-oxoglutarate/Fe(II)-dependent dioxygenase subfamily identifies a strong 33 membered opium poppy group of three clades. One of these is the CODM/T6ODM clade which is highlighted by the shaded box in the tree. The figure descriptors and legends are in the same format as described in the above section: General information for gene trees.

Supplemental References

- Ikezawa, N., Iwasa, K., Sato, F. (2008).** Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (*S*)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* cells. *J Biol Chem.* **283(14)**:8810-21. doi: 10.1074/jbc.M705082200.
- Kraus, P.F., Kutchan, T.M. (1995).** Molecular cloning and heterologous expression of a cDNA encoding berbaminine synthase, a C--O phenol-coupling cytochrome P450 from the higher plant *Berberis stolonifera*. *Proc Natl Acad Sci U S A.* **92(6)**:2071-5.