**The polyketide pathway in sporopollenin biosynthesis is specific to land plants (Embryophyta)**

Dae-Yeon Suh*, Damanpreet K. Sraan and Neil W. Ashton

Department of Chemistry and Biochemistry, University of Regina, Regina, SK S4S 0A2, Canada

*Author for correspondence: email suhdaey@uregina.ca

Supporting Information

Table S1 Sequences used for 4CL family tree construction (Fig S1).

Table S2 Sequences used for CCR family tree construction (Fig S2).

Fig. S1 Maximum Likelihood tree of embryophyte ACOS and 4CL sequences and algal homologues.

Fig. S2 Maximum Likelihood tree of TKPR and other enzymes of the CCR family and algal homologues.

Fig. S3 Multiple sequence alignment of ACOSs, 4CLs and algal homologues.

Fig. S4 Multiple sequence alignment of representative enzymes of the embryophyte CCR family and algal homologues.

**Abstract**

Sporopollenin (SP) is a complex biopolymer in the outer cell wall of spores and pollen and provides protection from environmental stresses. Its extraordinary biological and chemical resistance has defied efforts to elucidate its chemical composition and structure. This led to the adoption of broad and imprecise definitions of SP, which encouraged claims for its widespread occurrence among diverse embryophyte and non-embryophyte taxa and also misguided research on the evolution of SP and its roles during terrestrialisation. We previously proposed a more precise biochemical definition that can be used as a diagnostic tool to distinguish genuine SP from other chemically resistant cell wall materials. The definition was centred on the possession of ASCL (Anther-Specific Chalcone synthase-Like), an embryophyte-specific enzyme of the polyketide pathway that provides precursors for SP biosynthesis. Herein, we demonstrate that all five enzymes (CYP703, CYP704, ACOS, ASCL and TKPR) of the polyketide pathway are present in embryophytes but absent in green algae and propose a new, more comprehensive definition of SP: "Sporopollenin is a chemically resistant complex heteropolymer present in the outer walls of spores and pollen grains and is composed partly of hydroxylated polyketides derived from the conserved polyketide pathway, which typically involves CYP703A, CYP704B, ACOS, ASCL and TKPR." Furthermore, we examine the evolution of all five enzymes of the polyketide pathway from ancestral enzymes of primary metabolism. We discuss the co-selection of ACOS, ASCL and TKPR and also provide evidence that gene dosage of all five genes in the pathway has been subjected to strict evolutionary control. This study further signifies SP as an evolutionary innovation unique to the embryophyte lineage and encourages research on possible evolutionary relationships between genuine embryophyte SP and other 'SP-like materials' in reproductive bodies including algal zygospores and dinoflagellate cysts.

**Key words:** sporopollenin, CYP450, acyl-CoA synthetase, anther-specific chalcone synthase-like, tetraketide α-pyrone reductase, plant evolution and terrestrialisation, polyketide pathway, gene copy number

28 We previously proposed a new definition of sporopollenin (SP) as follows:
29 "Sporopollenin is a chemically resistant complex heteropolymer present in the outer walls of
30 spores and pollen grains and is composed partly of hydroxylated polyketides derived from the
31 conserved polyketide pathway, which involves ASCL." (Suh & Ashton, 2022).
32 As a protectant of progeny, SP is extraordinarily resistant to biological and chemical treatments,
33 and it has been challenging to determine its chemical composition and structure. Hence, prior to
34 our definition, resistance to acetolysis had been the sole criterion for the identification of SP
35 (Rowley & Southworth, 1967; Heslop-Harrison, 1969). This imprecise definition resulted in
36 claims for the existence of SP in algae and various microorganisms and also misguided research
37 on the evolution of SP and its roles during terrestrialisation (Delwiche *et al*., 1989; Feofilova, 2010;
38 de Vries & Archibald, 2018).
39 Recently, genetic and biochemical research has uncovered the biosynthesis pathways for
40 various SP precursors (Grienenberger & Quilichini, 2021; Xue *et al*., 2023). The polyketide
41 pathway that comprises five enzymes, CYP703A, CYP704B, ACOS (acyl-CoA synthetase),
42 ASCL (anther-specific chalcone synthase-like) and TKPR (tetraketide α-pyrone reductase)
43 provides polyketide SP precursors and is conserved in embryophytes. ASCL belongs to the type
44 III polyketide synthase family, and previous phylogenetic studies of the family indicated that
45 ASCL is specific to embryophytes (Jiang *et al*., 2008; Shimizu *et al*., 2017). Indeed, we discovered
46 at least one *ASCL* gene in representative embryophytes from all lineages that produce spores or
47 pollen with sporopolleninous exine. In sharp contrast, we failed to find any *ASCL* in the genomes
48 of all algae species for which data are available, including Charophytes that were considered to
49 produce SP-like substances. This finding led us to propose the above-mentioned definition for SP
50 to distinguish genuine SP from other similarly resistant substances and to use *ASCL* as an
51 identifying marker for the presence of SP in spore and pollen walls. We recognized that there are
52 other enzymes than ASCL in the polyketide pathway and exploring algal genomes for orthologues
53 encoding these enzymes has the potential for augmenting our SP definition (Suh & Ashton, 2022).
54 In this study, we investigated whether CYP703A, CYP704B, ACOS and TKPR are conserved
55 in embryophytes and whether they are absent in green algae. The cytochrome P450 family had
56 already been extensively studied and meticulously curated in the P450 database
57 (https://drnelson.uthsc.edu/, Nelson, 2009). CYP703A is present only in embryophytes, while
58 CYP704B is found in embryophytes, protists and red algae, but, importantly, not in green algae

59 (Morant *et al*., 2007; Li *et al*., 2010; Hansen *et al*., 2021). It had been suggested previously that
60 ACOS and TKPR are embryophyte-specific (de Azevedo Souza *et al*., 2008; Grienenberger *et al*.,
61 2010). However, information about algal homologues of these enzymes was lacking. Taking
62 advantage of newly determined genome sequences, we expanded earlier studies and searched for
63 ACOS and TKPR in the genomes of thirty-five embryophytes representing all major lineages
64 (Tables S1, S2). Putative ACOS and TKPR sequences collated from BLASTp searches in
65 Phytozome 13 (https://phytozome-next.jgi.doe.gov/) and other databases were identified based on
66 phylogenetic clustering with known enzymes (Figs S1, S2). In parallel, for additional support for
67 enzyme identification, we searched for amino acid residues that are conserved in ACOS but not in
68 its paralogues. ACOS and 4-coumarate:CoA ligase (4CL), both of which belong to the 4CL family,
69 catalyze formation of CoA esters. Both enzymes use ATP and CoA as co-substrates, but they differ
70 in their acid substrates. While ACOS accepts hydroxyfatty acids (provided by CYP703A and
71 CYP704B) as substrate, 4CL accepts 4-coumaric acid derivatives. Cystal structures of several
72 4CLs pinpointed active site residues responsible for substrate specificity of the enzyme (Hu *et al*.,
73 2010; Fig. S3). Sequence alignment of representative enzymes of the 4CL family and comparison
74 of 4CL crystal structures (e.g., tobacco 4CL2, PDB id, 5BST) with ACOS structure models
75 (https://swissmodel.expasy.org/; https://alphafold.ebi.ac.uk/) indicated that $Gly^{241}$ and
76 $S^{339}$C(V/I)Tx(S/T) (numbering of AtACOS5) are uniquely conserved in ACOS and could serve as
77 a determinant of ACOS (Fig. S3). We then performed a BLASTp search of all algal genomes in
78 the PhycoCosm database (https://phycocosm.jgi.doe.gov/phycocosm/home, accessed in February
79 2024) for algal homologues (E values lower than $e^{-20}$) of ACOS (Table S1) and constructed a
80 phylogenetic tree with representative embryophyte enzymes of the 4CL family. As shown in Fig.
81 S1, none of the algal homologues, including 13 from charophytes, is found in the ACOS clade and
82 none of them possesses the above-mentioned ACOS signature residues (Fig. S3). We concluded
83 that none of the algae examined has an ACOS.
84 Similarly, sequences and modeled structures of TKPRs were compared with those of
85 paralogous DFR (dihydroflavonol 4-reductase) and CCR (cinnamoyl-CoA reductase) to identify
86 two substrate binding site residues $Arg^{131}$ and $(Ser/Thr)^{188}$ (numbering of AtTKPR2) as
87 determinants of TKPR (Fig. S4). BLASTp searches yielded a total of 29 algal homologues,
88 including four from charophytes (Table S2). None of them resides in the TKPR clade (Fig. S2)
89 and, furthermore, no charophyte homologue has the TKPR signature residues, Arg and (Ser/Thr).

90 These findings led us to conclude that all of the five enzymes in the polyketide pathway are present
91 in embryophytes but absent in green algae, and produce polyketide precursors of genuine SP.
92 Therefore, we propose expanding our definition of SP as follows:
93 "Sporopollenin is a chemically resistant complex heteropolymer present in the outer walls of
94 spores and pollen grains and is composed partly of hydroxylated polyketides derived from the
95 conserved polyketide pathway, which typically involves CYP703A, CYP704B, ACOS, ASCL and
96 TKPR."
97
98 **Evolutionary origin of the polyketide pathway**
99 Next we searched the literature for studies on the evolutionary origins of the enzymes in the
100 polyketide pathway. It is generally accepted that enzymes of secondary metabolism have been
101 derived from pre-existing enzymes in primary metabolism via gene duplication and
102 neofunctionalisation (Firn & Jones, 2000).
103 (1) ACOS
104 The 4CL family belongs to the adenylate-forming enzyme superfamily present in all organisms.
105 The superfamily shares a common reaction of the formation of an adenylated intermediate as in
106 the fatty acid activation by fatty acyl-CoA synthetase. 4CL produces CoA esters of
107 hydroxycinnamic acids which are intermediates in the biosynthesis of a vast array of phenolic
108 secondary metabolites, including monolignols and flavonoids (Hahlbrock & Scheel, 1989). Based
109 on kinetic and phylogenetic analysis, it has been suggested that the members of the superfamily
110 share a common ancestor similar to extant acyl-CoA synthetase in the primary fatty acid
111 metabolism (Fig. 1) (Linne *et al*., 2007; de Azevedo Souza *et al*., 2008).
112 (2) ASCL
113 ASCL belongs to the family of type III polyketide synthases (PKS). Type III PKSs condense a
114 starter CoA substrate with a number of (mostly three) malonyl-CoA molecules and produce a
115 variety of polyketide secondary metabolites with different biological functions (Abe & Morita,
116 2010). There are more than twenty type III PKSs in plants and the most well-known is chalcone
117 synthase (CHS) that catalyzes the first committed step in flavonoid biosynthesis. Type III PKSs,
118 along with several condensing enzymes in fatty acid metabolism, belong to the thiolase
119 superfamily. An earlier phylogenetic study showed that type III PKSs have evolved, along with β-

120 ketoacyl synthases in fatty acid synthesis, from an ancestral thiolase-like condensing enzyme in
121 lipid metabolism (Fig. 1) (Jiang et al. 2008).
122 (3) TKPR
123 TKPR and its paralogues DFR and CCR all belong to the short-chain dehydrogenase/reductase
124 (SDR) superfamily present in all living organisms. DFR plays a role in the biosynthesis of
125 anthocyanins while CCR catalyses the first committed step of lignin biosynthesis (Lacombe *et al*.,
126 1997; Petit *et al*., 2007). Enzymes in the SDR family display similarities to sugar-metabolizing
127 enzymes in bacteria and it has been suggested that an ancestor of the superfamily could have been
128 UDP-galactose-4-epimerase in primary sugar metabolism (Fig. 1) (Baker & Blasco, 1992;
129 Kallberg *et al*., 2002).
130 (4) CYP703A and CYP704B
131 The CYP family is one of the largest enzyme families and present in all kingdoms of life. The
132 family is made of many clans and each clan contains individual CYP enzymes. The CYP 71 and
133 86 clans, to which CYP 703A and 704B belong respectively, are mainly involved in fatty acid
134 metabolism. All extant CYPs are monophyletic, and CYP51 (sterol 14α-demethylase) in sterol
135 biosynthesis is widely accepted as the ancestral progenitor (Yoshida et al. 2000; Lamb et al., 2021).
136 To summarize, it appears that ACOS, ASCL and TKPR all evolved from ancestral enzymes in
137 primary metabolism in the common ancestor of embryophytes. It could be more than a coincidence
138 that their paralogues play important roles in phenylpropanoid pathways (Fig. 1). It may reflect
139 critical contributions that sporopollenin and phenylpropanoids made during early evolution of land
140 plants.
141
142 **ACOS, ASCL and TKPR were co-selected during evolution**
143 Earlier, we found that the genome of *Zostera marina* (a seagrass) that produces exine-less pollen
144 does not contain *ASCL*, which we attributed to secondary gene loss (Suh & Ashton, 2022). The *Z.*
145 *marina* genome also lacks an *ACOS*, and it contains one copy of a *TKPR*-like sequence that would
146 produce a truncated TKPR-like protein because of nonsense mutation and pseudogenisation. Also,
147 *Zingiber officinale* produces pollen with a reduced exine with regular discontinuities (Subbarayudu
148 *et al*., 2014). A tBLASTn search against the genome of *Zingiber officinale* Roscoe (taxid:94328)
149 (https://blast.ncbi.nlm.nih.gov/, accessed in February 2024) failed to find *ACOS*, *ASCL* or *TKPR*.
150 Therefore, in line with our new and expanded definition, neither *Z. marina* nor *Z. officinale* makes

151 genuine SP. Based on these data, we also concluded that the three genes of the polyketide pathway
152 were co-selected during embryophyte evolution; loss of one of these genes would have alleviated
153 the selective pressure for retention of the remaining genes in the pathway. *CYP703A* and *CYP704B*
154 genes were found in the *Z. marina* and *Z. officinale* genomes unlike the other three genes in the
155 polyketide pathway. That they did not suffer gene loss is probably because the CYP enzymes they
156 encode are required for the biosynthesis of cutin monomers in addition to SP biosynthesis (Li *et*
157 *al*., 2010).
158
159 **Gene dosage: Copy number of the genes in the polyketide pathway**
160 Lastly, we examined whether the copy number of the five genes in the pathway is controlled. The
161 *CYP* family underwent diversification bursts (Nelson, 2019) and clans 71 and 86 are the two largest
162 clans in embryophytes. For instance, the genome of *Physcomitrium patens* has 71 CYPs and 41 of
163 them are in clan 71. Nonetheless, *CYP703A* is often the only single gene family in the *CYP71* clan
164 (Nelson, 2019). Also, *CYP704* is often maintained as a single copy or at low copy numbers
165 (Hansen *et al*., 2021). In the 4CL family, while bona fide *4CL* genes have undergone gene family
166 expansion, *ACOS* genes are encoded by single-copy genes in most plants (de Azevedo Souza *et*
167 *al*., 2009). Out of 35 plants we examined, only the genomes of *P. patens* and *Glycine max* have
168 two copies of *ACOS* and the rest have a single copy (Table S1). Likewise, whereas multiple copies
169 of *CHS* are found in most plants, only one (bryophytes, lycophytes, gymnosperms) or two copies
170 of *ASCL* (angiosperms) were found in plant species that we examined (Suh & Ashton, 2022). As
171 for *TKPR*, we found two copies of *TKPR* in most plant species and a single copy in a few species
172 including *Ginkgo biloba* and *Spirodela polyrhiza* (Table S2). These data indicate that gene dosage
173 of all five genes in the pathway has been subjected to strict control during evolution. Indeed, the
174 genomes of *A. thaliana* and rice contain one copy each of *CYP703*, *CYP704* and *ACOS* and two
175 copies each of *ASCL* and *TKPR*. The genome of *P. patens* has one copy each of *ACOS* and *ASCL*
176 and two copies of *TKPR* despite undergoing two whole genome duplications. We argue that the
177 evolutionary benefits of gene dosage control were maintenance of steady metabolic flux and
178 avoidance of the accumulation of intermediate(s) that could be prematurely incorporated into SP,
179 harming its structural integrity.
180
181 **Concluding remarks**

182 All evidence indicates that the polyketide pathway and therefore SP is embryophyte-specific. The
183 addition of the polyketide pathway in the definition of genuine SP will allow separation of SP from
184 algaenans and other chemically resistant 'SP-like' algal spore wall substances. This new, expanded
185 definition of SP should help establish SP as a major evolutionary innovation unique to the
186 embryophyte lineage that may have contributed to plant terrestrialisation. This study should also
187 spur research on the genetic and biochemical nature of the algal 'SP-like' spore wall substances
188 and the possible evolutionary relationships between algal spore wall substances and embryophyte
189 SP.
190

191 **Acknowledgements**

192 The authors thank William B. Browing (Southern Illinois Univ., USA) for informing us of the
193 absence of *ASCL* in the *Zingiber officinale* genome. This work was funded by the Natural Sciences
194 and Engineering Research Council of Canada Discovery grant to D-YS. (RGPIN-2018-04286).
195 DKS was supported in part by University of Regina Graduate Scholarships.
196

197 **Author contributions**

198 D-YS conceived the concept, DKS collected sequences and constructed phylogenetic trees. D-YS
199 and NWA wrote the manuscript.
200

201 **References**

202 **Abe I, Morita H. 2010.** Structure and function of the chalcone synthase superfamily of plant type
203 III polyketide synthases. *Natural Product Reports* **27:** 809–838.
204 **de Azevedo Souza C, Barbazuk B, Ralph SG, Bohlmann J, Hamberger B, Douglas CJ. 2008.**
205 Genome-wide analysis of a land plant-specific *acyl:coenzymeA synthetase* (*ACS*) gene family
206 in *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytologist* **179**: 987–1003.
207 **de Azevedo Souza C, Kim SS, Koch S, Kienow L, Schneider K, McKim SM, Haughn GW,**
208 **Kombrink E, Douglas CJ. 2009.** A novel fatty acyl-CoA synthetase is required for pollen
209 development and sporopollenin biosynthesis in *Arabidopsis*. *Plant Cell* **21**: 507–525.
210 **Baker ME, Blasco R. 1992.** Expansion of the mammalian 3β-hydroxysteroid dehydrogenase/plant
211 dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a

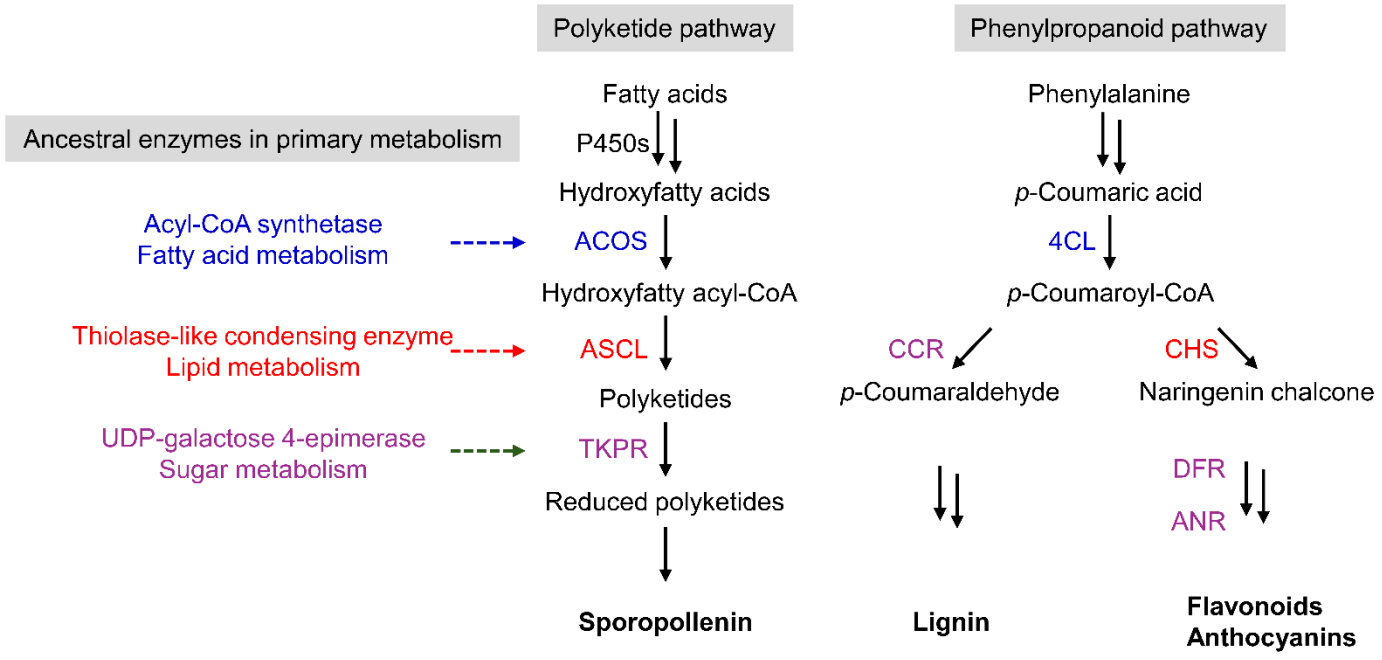
212 bacterial UDP-galactose 4-epimerase, and open reading frames in vaccinia virus and fish
213 lymphocystis disease virus. *FEBS Letters* **301**: 89–93.
214 **Delwiche CF, Graham LE, Thomson N. 1989.** Lignin-like compounds and sporopollenin in
215 Coleochaete, an algal model for land plant ancestry. *Science* **245**: 399–401.
216 **Feofilova EP. 2010.** The fungal cell wall: modern concepts of its composition and biological
217 function. *Microbiology* **79**: 711–720.
218 **Ferrer J-L, Austin MB, Sterwart Jr C, Noel JP. 2008.** Structure and function of enzymes
219 involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry* **46**:
220 356–370.
221 **Firn RD, Jones CG. 2000.** The evolution of secondary metabolism—a unifying model. *Molecular*
222 *Microbiology* **37:** 989–994.
223 **Grienenberger E, Kim SS, Lallemand B, Geoffroy P, Heintz D, de Azevedo Souza C, Heitz**
224 **T, Douglas CJ, Legrand M. 2010.** Analysis of *TETRAKETIDE α-PYRONE REDUCTASE*
225 function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical
226 pathway in sporopollenin monomer biosynthesis. *Plant Cell* **22**: 4067–4083.
227 **Grienenberger E, Quilichini TD. 2021.** The toughest material in the plant kingdom: An update
228 on sporopollenin. *Frontiers in Plant Science* **12**: 703864.
229 **Hahlbrock K, Scheel D. 1989.** Physiology and molecular biology of phenylpropanoid metabolism.
230 *Annual Review of Plant Physiology and Plant Molecular Biology* **40:** 347–369.
231 **Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021.** Plant cytochrome P450
232 plasticity and evolution. *Molecular Plant* **14:** 1244–1265.
233 **Heslop-Harrison J. 1969.** An acetolysis-resistant membrane investing tapetum and sporogenous
234 tissue in the anthers of certain Compositae. *Canadian Journal of Botany* **47:** 541–542.
235 **Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C. 2010.** Crystal
236 structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic
237 mechanisms. *Plant Cell* **22:** 3093–3104.
238 **Jiang C, Kim SY, Suh D-Y. 2008.** Divergent evolution of the thiolase superfamily and chalcone
239 synthase family. *Molecular Phylogenetics and Evolution* **49**: 691–701.
240 **Kallberg U, Oppermann U, Jörnvall H, Persson B. 2002.** Short-chain dehydrogenase/reductase
241 (SDR) relationships: A large family with eight clusters common to human, animal, and plant
242 genomes. *Protein Science* **11**: 636–641.

243 **Lacombe E, Hawkins S, Van Doorsselaere J, Piquemal J, Goffner D, Poeydomenge O,**
244 **Boudet AM, Grima-Pettenati J. 1997.** Cinnamoyl CoA reductase, the first committed enzyme
245 of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships.
246 *Plant Journal* **11:** 429–441.
247 **Lamb DC, Hargrove TY, Zhao B, Wawrzak Z, Goldstone JV, Nes WD, Kelly SL, Waterman**
248 **MR, Stegeman JJ, Lepesheva GI. 2021.** Concerning P450 evolution: Structural analyses
249 support bacterial origin of sterol 14α-demethylases. *Molecular Biology and Evolution* **38:** 952–
250 967.
251 **Li H, Pinot F, Sauveplane V, Werck-Reichhart D, Diehl P, Schreiber L, Franke R, Zhang P,**
252 **Chen L, Gao Y *et al.* 2010.** Cytochrome P450 family member CYP704B2 catalyzes the ω-
253 hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine
254 formation in rice. *Plant Cell*, **22**: 173–190.
255 **Linne U, Schäfer A, Stubbs MT, Marahiel MA. 2007.** Aminoacyl-coenzyme A synthesis
256 catalyzed by adenylation domains. *FEBS Letters* **581:** 905–910.
257 **Morant M, Jørgensen K, Schaller H, Pinot F, Møller BL, Werck-Reichhart D, Bak S. 2007.**
258 CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of
259 lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* **19**: 1473–
260 1487.
261 **Nelson DR. 2009.** The Cytochrome P450 Homepage. *Human Genomics* **4**: 59–65.
262 **Nelson DR. 2019.** Cytochrome P450s in the sugarcane *Saccharum spontaneum*. *Tropical Plant*
263 *Biology* **12:** 150–157.
264 **Petit P, Granier T, d'Estaintot BL, Manigand C, Bathany K, Schmitter J-M, Lauvergeat V,**
265 **Hamdi S, Gallois B. 2007.** Crystal structure of grape dihydroflavonol 4-reductase, a key
266 enzyme in flavonoid biosynthesis. *Journal of Molecular Biology* **368:** 1345–1357.
267 **Rowley JR, Southworth D. 1967.** Deposition of sporopollenin on lamellae of unit membrane
268 dimensions. *Nature* **213:** 703–704.
269 **Subbarayudu S, Shankar Naik B, Sunitibala Devi H, Bhau BS, Shaik Sha Valli Khan P. 2014.**
270 Microsporogenesis and pollen formation in *Zingiber officinale* Roscoe. *Plant Systematics and*
271 *Evolution* **300:** 619–632.
272 **Shimizu Y, Ogata H, Goto S. 2017.** Type III polyketide synthases: Functional classification and
273 phylogenomics. *Chembiochem* **18:** 50–65.

274 **Suh D-Y, Ashton NW. 2022.** A sporopollenin definition for the genomics ages. *New Phytologist*
275 **236**: 2009–2013.
276 **de Vries J, Archibald JM. 2018.** Plant evolution: landmarks on the path to terrestrial life. *New*
277 *Phytologist* **217**: 1428–1434.
278 **Xue JS, Qiu S, Jia X-L, Shen S-Y, Shen C-W, Wang S, Xu P, Tong Q, Lou Y-X, Yang N-Y**
279 ***et al*. 2023.** Stepwise changes in flavonoids in spores/pollen contributed to terrestrial adaptation
280 of plants. *Plant Physiology* **193:** 627–642.
281 **Yoshida Y, Aoyama Y, Noshiro M, Gotoh O. 2000.** Sterol 14-demethylase P450 (CYP51)
282 provides a breakthrough for the discussion on the evolution of cytochrome P450 gene
283 superfamily. *Biochemical and Biophysical Research Communications* **273:** 799–804.
284

285 **Supporting Information**

286 **Fig. S1** Maximum Likelihood tree of embryophyte ACOS and 4CL sequences and algal
287 homologues.
288 **Fig. S2** Maximum Likelihood tree of TKPR and other enzymes of the CCR family and algal
289 homologues.
290 **Fig. S3** Multiple sequence alignment of ACOSs, 4CLs and algal homologues.
291 **Fig. S4** Multiple sequence alignment of representative enzymes of the embryophyte CCR family
292 and algal homologues.

293 **Figure legend**

294 **Fig. 1.** The origin and evolution of the polyketide pathway in sporopollenin biosynthesis.
295 The three enzymes (ACOS, ASCL/PKS and TKPR) in the polyketide pathway are shown with
296 their evolutionarily related enzymes in the phenylpropanoid pathway. Also shown are the enzymes
297 of primary metabolism proposed to share the same ancestors with the related enzymes in the
298 polyketide and phenylpropanoid pathways. Related enzymes are indicated by text in the same
299 colour. For instance, ACOS and 4CL are thought to have evolved from the ancestral form of acyl-
300 CoA synthetase in fatty acid metabolism and are presented using blue text. 4CL, 4-coumarate:CoA
301 ligase; ACOS, acyl-CoA synthetase; ANR, anthocyanidin reductase; ASCL, anther-specific
302 chalcone synthase-like; CCR, cinnamoyl-CoA reductase; CHS, chalcone synthase; DFR,
303 dihydroflavonol 4-reductase. For further information about the enzymes in the pathways, readers
304 are referred to Grienenberger & Quilichini (2021) and Ferrer et al. (2008).
305