The ABCG1 transporter facilitates sesquiterpene accumulation in *Marchantia polymorpha* oil bodies.

Edith C. F. Forestier, Paola Asprilla, Facundo Romani, Ignacy Bonter, Eftychios Frangedakis and Jim Haseloff.

ABSTRACT

Marchantia polymorpha oil bodies (OBs) are specialised cell structures housing a diverse array of C15-terpenes, called sesquiterpenes. These compounds are known for their roles as herbivore repellents, yet the enzymes responsible for the biosynthesis of their precursors (C5 isoprenoid units) remain poorly characterized. Discrepancies remain between enzyme localizations suggested by computational predictions and those observed in earlier experimental studies, complicating our understanding of terpene biosynthesis. We investigated the localization of isoprenoid biosynthetic enzymes using translational and transcriptional reporters, coupled with confocal microscopy. Most enzymes localized as predicted (*e.g.*, cytosol, chloroplast and the endoplasmic reticulum), and OB cells were identified as the primary sites of terpene biosynthesis.

To explore OBs as potential storage sites for terpenes, we attempted to produce exogenous but easily identifiable compounds in *Marchantia*, such as the diterpene taxadiene and the triterpene β-amyrin. Targeting to OB cells resulted in measurable amounts of these compounds, but their yields remained unaffected by the over-expression of key precursor genes, underscoring challenges in redirecting metabolic flux.

To further investigate terpene accumulation in OBs, we focused on ABCG1, an ABC transporter previously reported to localize at the OB membrane. Overexpression of ABCG1 in OB cells, alongside an exogenous sesquiterpene synthase, only increased the levels of endogenous sesquiterpenes, while CRISPR-mediated disruption of ABCG1 resulted in a dramatic reduction in sesquiterpene accumulation. These findings establish ABCG1 as a critical factor for sesquiterpene retention within OBs and provide new insights into the mechanisms governing terpene metabolism and storage in *Marchantia polymorpha*.

INTRODUCTION.

In recent years, *Marchantia polymorpha* has emerged as a useful model plant system[1] for studying terpene biosynthesis and accumulation within specialized storage structures known as oil bodies (OBs)[2]. The oil bodies in *Marchantia* are membrane-bound compartments that primarily house terpenes[3] and bisbibenzyls[4], a class of phenylpropanoid derivatives; the mixture confers anti-feedant properties to protect against herbivores[5]. Terpenes are a diverse class of natural products derived from the condensation of isoprene (C_5H_8) units[6], found across many organisms but predominantly in the plant kingdom. Beyond deterring herbivores as observed in *Marchantia*, they serve other vital roles in plants, such as preventing fungal infections or attracting pollinators. Additionally, a broad range of bioactive properties makes them valuable in medicine, agriculture, food and other industrial applications.

The sequestration of bisbibenzyls and terpenes into OBs is critical for protecting *Marchantia* from the potential toxicity of these metabolites when released into other tissues. The unique compartmentalization offered by OBs presents a promising avenue for metabolic engineering, as it provides a natural reservoir for accumulating metabolites that might otherwise be harmful to the plant.

Recent developmental studies have provided significant insights into the mechanisms governing OB cell fate determination and formation, highlighting key transcription factors (TFs) involved in these processes. Specific TFs, such as ERF13[7], C1HDZ[5], TGA[8] and MYB2[9,10] have been identified as crucial regulators of OB cell differentiation and maturation[11]. These discoveries have expanded our understanding of how OB cells develop and contribute to the plant's metabolic capacity.

Manipulating these TFs has further revealed the delicate balance between OB cell proliferation and plant fitness. For instance, CRISPR-mediated mutation of *TGA*[8] or gain-of-function mutants of *ERF13* led to a dramatic increase in OB cell numbers per plant[7]. However, these alterations were accompanied by reduced growth rates or morphological defects, suggesting that OB cell number is tightly regulated in *Marchantia* to avoid compromising overall plant health and fitness, or that these TFs have pleiotropic effects that influence additional developmental processes.

Early studies on *Marchantia's* terpene biosynthetic pathway, which were based on immunolocalization of isoprenoid enzymes, hypothesized that key enzymatic steps might occur at the oil body (OB) membrane[3]. Today, with the availability of whole-

genome sequencing and detailed gene annotations informed by sequence homology with other species, our understanding of the early steps of isoprenoid biosynthesis has significantly improved[12–14] (Table 1 and S1). The mevalonate (MVA) and methylerythritol phosphate (MEP)[15–17] pathways are the two primary metabolic routes responsible for producing the precursors — isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) — for various classes of terpenes (Table 1). The MVA pathway, comprising seven enzymatic steps[18] and typically operating in the cytosol[19], endoplasmic reticulum[20] (ER) and peroxysome[21], generally supplies precursors to sesquiterpenes and triterpenes synthesis[22,23], while the MEP pathway, with eight enzymatic steps and primarily localized in plastids[24], commonly provides precursors for monoterpenes and diterpenes[22,23] (Table 1 and S1). In *Marchantia*, sesquiterpenes are the dominant terpene class found in OBs[2], produced by several microbial- and fungal-type terpene synthases[25]. Additionally, *Marchantia* produces the monoterpene limonene, likely synthesized through the cis-prenyl precursor neryl pyrophosphate (NPP) rather than the trans form geranyl pyrophosphate (GPP), as demonstrated by Kumar et al[26]. While cell-type-specific metabolic analyses revealed that only sesquiterpenes are stored inside the OB compartment[2], plants defective in OB cell differentiation show reduced levels of both sesquiterpenes and monoterpenes[5], making it unclear whether limonene is specifically stored within OBs. Examination of genome data[12–14,27] along with computational predictive tools for enzyme localization such as DeepLoc 2.1[28], enabled us to identify putative precursor genes involved in prenyl phosphate synthesis for terpene production (Table 1). However, discrepancies remain between earlier theories suggesting that terpene biosynthetic enzymes are localized to OB membranes[3] and more recent predictions that place MVA enzymes in the cytosol, ER and peroxisome, and MEP enzymes in plastids (Table 1). These discrepancies prompted us to re-examine the localization of key enzymes using modern techniques such as confocal microscopy combined with translational reporters[7]. This approach allowed us to identify the specific cells expressing these genes and determine the subcellular localization of their corresponding proteins.

Building on our findings regarding terpene precursor gene expression and localization, we then explored the potential for metabolic engineering in *Marchantia polymorpha*'s OBs. Specifically, we aimed to produce two valuable terpenes within OBs: the

102 diterpene taxadiene[29], a precursor to the valuable anti-cancer compound Taxol®[30,31],
103 and the triterpene β-amyrin[32], a precursor to compounds such as glycyrrhizin[33], a
104 sweet-tasting constituent of liquorice. These compounds were selected due to their
105 distinct chemical signatures, which facilitate easy detection in chromatographic
106 analyses given the minimal background of these terpene sub-classes in *Marchantia*.
107 To further evaluate the metabolic capacity of *Marchantia* and investigate the transport
108 of sesquiterpenes, we also examined the production of the sesquiterpene amorpha-
109 4,11-diene[34] alongside ABCG1, an ATP-binding cassette (ABC) transporter highly and
110 exclusively expressed in OB cells[35]. Prior studies have demonstrated its clear
111 localization to the OB membrane using a translational reporter[7], suggesting a potential
112 role in metabolite transport. This study provides new insights into the involvement of
113 ABCG1 in terpene accumulation within OBs and contributes to a broader
114 understanding of the partitioning of terpene metabolism in *Marchantia polymorpha*.
115
116 **MAIN**
117
118 **1. Translational and transcriptional reporters for genes expression in the**
119 **terpenoid pathway.**
120 To determine the expression pattern and subcellular localization of key enzymes
121 involved in *Marchantia*'s terpene biosynthesis, we generated constructs of
122 translational reporters, in which each gene's native promoter drove its respective
123 coding sequence fused to the fluorescent tag *mVenus*[36]. In cases where translational
124 reporters were unavailable due to technical challenges, transcriptional reporters were
125 used instead to determine cell-type expression only. These consisted of *Marchantia*'s
126 native promoter driving mVenus with a nuclear localisation signal[37] (mVenus-N7),
127 enabling us to investigate the spatial and temporal expression of key genes. Promoter
128 lengths were selected based on Assay for Transposase-Accessible Chromatin with
129 Sequencing (ATAC-seq)[38] data from version 6.1 of the Tak accession in the
130 *Marchantia.info* database[13,25] (Table S2). ATAC-seq identifies regions of open
131 chromatin, which are more accessible to transcription factors and associated with
132 regulatory domains that influence gene expression. Since relaxed chromatin is
133 frequently found near the start of a gene, it serves as a useful indicator for selecting
134 promoter regions likely to drive gene expression in their native cellular contexts. To

135 delineate cell boundaries, each construct also included a fluorescent membrane
136 marker, *mScarlet-LTI6b* fusion[39–41], driven by the strong, constitutive *ubiquitin-*
137 *conjugating enzyme E2* gene promoter and 5'UTR from *Marchantia polymorpha*
138 ($_{Prom5}$*UBE2*)[41].
139 For the identification of different subcellular compartments, such as
140 chloroplasts/plastids, ER, cytosol, Golgi, and mitochondria, we interpreted our findings
141 by comparing the localization patterns observed in our images to those of previously
142 characterised standard reporters in *Marchantia*[41,42], though distinguishing between the
143 ER and cytosol remained challenging under the resolution constraints of our
144 experimental setup.
145 For this study, we selected key genes from the MVA and MEP pathways on the basis
146 of their predicted roles as rate-limiting enzymes, their redundancy, or their involvement
147 in the final steps of the pathway. Additionally, we included genes responsible for the
148 formation of linear scaffold precursors for sesquiterpenes, diterpenes, and triterpenes
149 to capture a more comprehensive view of terpene biosynthesis. Among these, several
150 genes were annotated as isoprenyl diphosphate synthases (*IDS*)[27] (Table 1). To
151 ensure accurate functional assignments, we compared their protein sequences with
152 *Arabidopsis* homologs and tentatively re-annotated those likely involved in terpene
153 precursor formation for our study.
154 We first analysed the two annotated *1-deoxy-D-xylulose-5-phosphate synthases*
155 (*MpDXS1 and MpDXS2*) (Table 1), which encode the first committed and rate-limiting
156 enzymes of the MEP pathway[43]. For *MpDXS1*, tagged with mVenus and driven by its
157 native promoter and 5'UTR ($_{Prom5}$*MpDXS1:MpDXS1-mVenus*), mVenus fluorescence
158 was observed in all cell types throughout the plant, with a stronger signal in OB cells
159 (Figure 1A). The protein localized to plastids, displaying a punctate pattern in non-OB
160 cells rather than a homogeneous distribution (Figure 1A). By contrast,
161 $_{Prom5}$*DXS2:DXS2-mVenus* showed a restricted localization pattern, with fluorescence
162 observed in oil body plastids only from Day 0 gemmae to 14-days-old plants (Figure
163 1B and S1D).
164 We examined the final step of the MEP pathway by generating a translational reporter
165 for *(E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase* (*MpHDR*)[44,45] (Table
166 1). We also generated a translational reporter for *(E)-4-hydroxy-3-methylbut-2-enyl*
167 *diphosphate synthase* (*MpHDS*)[44,46] (Table 1). *MpHDS* is adjacent to *MpHDR* and
168 oriented in the opposite direction on the same chromosome, suggesting they share a

169 bidirectional promoter (Figure S2A). Previous studies have highlighted the potentially
170 toxic nature of the product of HDS[47,48], which is detoxified in the subsequent enzymatic
171 step by HDR[47,48], making it particularly interesting to study both enzymes together. To
172 our knowledge, this is the first report suggesting that *HDS* and *HDR* genes may be
173 regulated by a shared promoter region. Initial experiments revealed that the 2 kb
174 region upstream of *MpHDR* was insufficient to drive expression, whereas the 1.9 kb
175 region upstream of *MpHDS* was adequate (Figures S2B). Expression of a construct
176 driving *MpHDS-eGFP* and *MpHDR-mVenus* bidirectionally (Figure S2C)
177 demonstrated that, like other MEP enzymes, both proteins localized to plastids (Figure
178 1C and 1D). Their expression was stronger, if not exclusive, to OB cells in Day 0
179 gemmae (Figure 1C and 1D), before becoming more widespread across tissues at
180 later developmental stages (Figure S1A and S1B).
181 Finally, we analysed the sub-cellular localisation and cell-type expression of the
182 enzyme forming the linear diterpene precursor geranylgeranyl pyrophosphate
183 (GGPP)[49,50]. Among the IDS candidates[27], IDS2 and IDS3 share 68% and 62% protein
184 homology, respectively, with GERANYLGERANYL PYROPHOSPHATE SYNTHASE
185 11 (GGPPS11) from *Arabidopsis thaliana*, the most highly expressed and functionally
186 dominant GGPPS in *Arabidopsis*[51], which operates as a homodimer[52]. However,
187 transcriptomic data revealed that *IDS3* exhibited consistently low expression across
188 all tissues and developmental stages[12–14]. The other IDS candidates, IDS4 and IDS5,
189 share 71% and 57% protein homology with SOLANESYL PYROPHOSPHATE
190 SYNTHASE and GERANYL PYROPHOSPHATE SYNTHASE, respectively,
191 suggesting potential roles in ubiquinone biosynthesis[53] (IDS4) and trans-monoterpene
192 precursor formation (IDS5). IDS2 emerged as the strongest candidate for the primary
193 GGPPS in *Marchantia*, given its higher sequence homology and stronger expression.
194 We therefore re-annotated it as MpGGPPS (Table 1 and S1). The translational fusion
195 to *MpGGPPS* showed localization to plastids in all cell types, with particularly strong
196 expression in OB cells (Figure 1E).
197 For the MVA pathway and sesquiterpene synthesis, we examined the rate-limiting
198 enzyme 3-HYDROXY-3-METHYLGLUTARYL-COA REDUCTASE (MpHMGR)[54]
199 (Table 1). Expression of $_{Prom5}$*MpHMGR:MpHMGR-mVenus* revealed a restricted
200 localisation to oil body cells in gemmae (Figure 1G) and 14-day-old plants (Figure
201 S1D). The enzyme appeared to display a dual localization pattern: a punctate
202 distribution around the OB, indicative of Golgi, and a web-like pattern surrounding the

203 nucleus, consistent with ER localization (Figure 1G). As a key enzyme downstream
204 the MVA pathway, we analysed FARNESYL PYROPHOSPHATE SYNTHASE
205 (MpFPS) (Table 1), which forms farnesyl pyrophosphate (FPP)[55], the linear precursor
206 to all sesquiterpenes. Annotated as IDS1[27] in the Marchantia.info database, MpFPS
207 shares 66% sequence similarity with both FPS homologs from *Arabidopsis thaliana*.
208 Expression analysis revealed that MpFPS is specifically expressed in OB cells,
209 persisting even after 14 days (Figure 1F and S1E). Contrary to the prediction of
210 mitochondrial localization by DeepLoc 2.1[28] (Table 1), the enzyme appeared
211 predominantly cytosolic, as indicated by its uniform signal surrounding the OB and
212 delineating the plastids (Figure 1F).
213 We further investigated the expression pattern[56] of the promoters driving the genes
214 encoding *1-deoxy-D-xylulose 5-phosphate reductoisomerases*[57,58] (*MpDXR1* and
215 *MpDXR2*) and *mevalonate diphosphate decarboxylase* (MpMVD)[59] using
216 transcriptional reporters. The corresponding enzymes, MpDXRs and MpMVD,
217 catalyze the second enzymatic step of the MEP pathway and the last enzymatic step
218 of the MVA pathway, respectively (Table 1). $_{Prom5}$*MpDXR1:mVenus-N7* showed
219 ubiquitous expression (Figure 2A), whereas $_{Prom5}$*DXR2:mVenus-N7* showed specificity
220 to OB cells in Day 0 gemma (Figure 2B). Similarly, $_{Prom5}$*MpMVD:mVenus-N7* exhibited
221 specificity to OB cells, even in 14-day-old thalli (Figure 2C).
222 Finally, given that the expression of some MVA biosynthetic genes and *MpFPS*
223 appears specific to OB cells, we investigated whether this pattern extends to
224 phytosterol biosynthesis. Phytosterols are essential structural components of plant
225 membranes[60] and are initially formed through the fusion of two FPP molecules to
226 produce squalene, a reaction catalyzed by SQUALENE SYNTHASE[61] (SQS). We
227 therefore examined the transcriptional reporter of the *MpSQS* promoter. Among the
228 three squalene synthase-like genes annotated in the *Marchantia* genome (*MpSQS1-*
229 *3*) (Table 1 and S1), we selected *MpSQS3* based on its higher protein sequence
230 homology (60%) with *Arabidopsis thaliana* SQUALENE SYNTHASE. In contrast,
231 MpSQS1 showed greater similarity (73%) to *Arabidopsis* PHYTOENE SYNTHASE,
232 suggesting a role in carotenoid biosynthesis, while MpSQS2 shared 52% homology
233 with an uncharacterized terpenoid synthase. Plants expressing the
234 $_{Prom5}$*SQS3:mVenus-N7* constructs show a mVenus signal in meristematic regions in
235 Day 0 gemmae, followed by ubiquitous expression in later developmental stages

236 (Figure 2D). These findings indicate that sterol biosynthesis is not confined to OB cells
237 but occurs broadly throughout the plant.
238 Overall, our findings provide a comprehensive map of the subcellular localization and
239 expression patterns of key terpene biosynthetic enzymes, offering insights into the
240 spatial organization of biosynthetic intermediates and terpene scaffolds in *Marchantia*
241 *polymorpha*.
242
243 **2. Exogenous production of taxadiene and β-amyrin in *Marchantia polymorpha***
244 **whole plants *vs*. oil body cells.**
245 Given the observed expression of terpene precursor enzymes predominantly in OB
246 cells, we aimed to harness the metabolic potential of these specialised structures. As
247 proof of concept, we attempted to produce the diterpene taxadiene[29], and the
248 triterpene β-amyrin[32].
249 We generated constructs to express *Taxus baccata taxadiene synthase* (*TXS*)[62] or
250 *Talinum paniculatum β-amyrin synthase*[63] (*β-AS*), driven by either the constitutive
251 $_{Prom5}$*UBE2* or the oil body-specific Mp*R2R3-MYB2* promoter ($_{Prom5}$*MYB2*)[37], enabling
252 expression throughout the entire plant or specifically within OB cells, respectively. TXS
253 catalyzes the cyclization of the linear diterpene precursor GGPP into taxadiene, while
254 β-AS converts the linear triterpene precursor 2,3-oxidosqualene into β-amyrin.
255 In plants expressing $_{Prom5}$*UBE2:TXS*, we detected a prominent new peak in hexane
256 extracts (Figure 3A), corresponding to taxadiene based on its mass spectrum (MS),
257 as described in prior publications[64,65]. While no commercial taxadiene standards were
258 available, we identified the compound by its characteristic ion at m/z 122 and the
259 molecular ion [M+] at m/z 272 (Figure S3A). Quantification of taxadiene in four
260 independent transformants expressing either $_{Prom5}$*UBE2:TXS* or $_{Prom5}$*MYB2:TXS*
261 revealed significantly higher levels of taxadiene in whole plants compared to those
262 with expression restricted to OB cells (Figure 3D). Specifically, $_{Prom5}$*UBE2:TXS* lines
263 produced approximately 21 µg of taxadiene per g of fresh weight (FW), whereas
264 $_{Prom5}$*MYB2:TXS* lines produced only 0.7 µg/g FW, dodecane equivalent. The correct
265 targeting of TXS protein in OB cells was confirmed using a $_{Prom5}$*MYB2:TXS-mVenus*
266 fusion construct, which showed plastid localization (Figure 3C).
267 Similarly, plants expressing $_{Prom5}$*UBE2:β-AS* produced a new compound, identified as
268 β-amyrin through its MS profile and comparison with an authentic standard (Figure 3B
269 and S3B). The $_{Prom5}$*UBE2:β-AS* lines produced an average of 56 µg of β-amyrin per g

270 of fresh weight (coprostanol equivalent), whereas lines expressing the enzyme
271 specifically in OB cells yielded an average of 1.4 µg/g FW (Figure 3D). In MYB2-driven
272 *β-AS-mVenus* lines, the protein localized specifically in the cytosol of the OB cells
273 (Figure 3C).
274 To investigate whether precursor flux limitations were constraining the production of
275 taxadiene and β-amyrin in oil bodies, we co-expressed *TXS* and *β-AS* with key rate-
276 limiting enzymes of the MEP and MVA pathways (*i.e DXS* and *HMGR*, respectively),
277 as our work demonstrated that many of these enzymes appeared to be highly or
278 specifically expressed to OB cells.
279 For taxadiene production via the MEP pathway, we co-expressed *MpDXS1* or
280 *MpDXS2* with *TXS*. To mitigate potential homology-dependent gene silencing
281 associated with overexpressing native *Marchantia* genes[66], we also expressed the
282 homologous *PpDXS1* and *PpDXS2* genes from *Physcomitrium patens,* which have
283 recently been re-annotated as *PpDXS1A* and *PpDXS1D*[67]. To further enhance
284 precursor flux, we introduced *MpHDR* and *MpGGPPS*, two key enzymes previously
285 shown to increase GGPP flux for diterpene production in *Nicotiana benthamiana*[68,69].
286 For β-amyrin production through the MVA pathway[70], we overexpressed the *HMGR*
287 genes from *Marchantia* (*MpHMGR*) and *Physcomitrium* (*PpHMGR*), alongside
288 truncated versions designed to remove their negative feedback regulatory domain[71,72],
289 thereby increasing the pool of terpene precursors and allowing unrestrained
290 production of β-amyrin. These truncated variants (*Mp-tHMGR* and *Pp-tHMGR*) lack
291 the first 148 and 146 amino acids, respectively. All genes were expressed under the
292 MYB2 promoter to specifically target OB cells.
293 Quantitative analysis of four independent transformants showed no significant
294 increase in taxadiene or β-amyrin levels upon over-expression of these precursor
295 supply genes (Figure 3E). We confirmed the correct targeting to OB cells and
296 subcellular localization of *Marchantia DXS*, *HDR*, *GGPPS*, *HMGR*, and *tHMGR* by
297 fusing these genes with mVenus under the control of the MYB2 promoter (Figure S4).
298 DXSs, HDR, and GGPPS enzymes specifically localized to plastids (Figure S4A, S4B,
299 S4C and S4D). However, despite co-localization with TXS, this did not result in
300 increased taxadiene levels. Over-expression of enzymes for precursor biosynthesis in
301 the OB cells didn't increase either the level of main endogenous sesquiterpenes
302 (Dataset 1 and Dataset 2), which were tentatively identified using Kovats retention
303 indices[73] (Figure S5 and Table S3). While tHMGR-mVenus appeared to localize in the

304 ER and Golgi apparatus of OB cells (Figure S4E), the native version driven by the
305 MYB2 promoter remained undetectable. To address this, we re-cloned *MpHMGR*
306 under the 2x35S promoter and successfully detected the protein, which accumulated
307 exclusively in the ER of OB cells (Figure S4F). Interestingly, when the truncated
308 version (tHMGR) was expressed under the 2x35S promoter, the fusion protein showed
309 expression across all cells (Figure S4G), suggesting that the negative feedback
310 domain removed in this construct may play a role in restricting localization to OB cells.
311 Overall, these results demonstrate that over-expression of selected rate-limiting
312 enzymes was insufficient to significantly enhance yields of exogenous terpenes in
313 *Marchantia* OB cells, suggesting that additional regulatory mechanisms may play
314 critical roles in terpene biosynthesis and accumulation in these specialized cells.
315
316 **3. Accumulation of terpenes in oil bodies may require a specific transporter.**
317 Despite the high specificity and/or strong expression of terpene biosynthetic pathway
318 genes in OB cells, the over-expression of key enzymes did not lead to increased levels
319 of exogenous terpenes. Moreover, these compounds may not naturally accumulate in
320 OBs: taxadiene may primarily localize to plastids, while β-amyrin could accumulate in
321 the cytosol. We hypothesized that precursor flux in *Marchantia* predominantly supplies
322 sesquiterpene biosynthesis and that a specific transporter may be required for their
323 accumulation inside OBs. Previous studies demonstrated high and exclusive
324 expression of the *ABCG1* gene in oil body cells[35], with its protein localized to the OB
325 membrane when expressed under its native promoter, and to the plasma membrane
326 of other cells when driven by a non-OB-specific promoter[7].
327 To investigate whether ABCG1 could facilitate the accumulation of any sesquiterpenes
328 in *Marchantia*'s OBs, we expressed *Artemisia annua amorphadiene synthase* (*AMS*),
329 a sesquiterpene synthase that converts FPP into amorpha-4,11-diene[34]. Cyclisation of
330 FPP into this sesquiterpene likely occurs via the bisabolyl cation[74], similar to the
331 endogenous β-chamigrene[75], therefore we expected this structural similarity to enable
332 amorpha-4,11-diene transport into OBs. We targeted *AMS* expression to either the
333 whole plant ($_{Prom5}$*UBE2*) or specifically in OB cells ($_{Prom5}$*MYB2*). Additionally, we co-
334 expressed *AMS* and *ABCG1* in OB cells, alongside *HMGR*, *DXS2*, and *FPS*, with the
335 FPS enzyme recognized as rate-limiting in sesquiterpene biosynthesis[76]. By
336 introducing ABCG1 and FPS, we aimed to overcome potential transport limitations and

337 to test whether this combination could further increase FPP availability for
338 sesquiterpene production. Amorpha-4,11-diene was detected only in transgenic lines
339 expressing the construct $_{Prom5}$*UBE2:AMS* (Figure S6), with an average yield of 0.75
340 µg/g FW in six out of eight independent transformants (Figure 4A, Dataset 3). No
341 amorpha-4,11-diene was detected when expression was targeted to OB cells, either
342 with or without precursor supply genes and *ABCG1* (Figure S6). Localization studies
343 of *AMS-mVenus* and *MpFPS-mVenus* driven by the MYB2 promoter showed that both
344 proteins localized to the cytosol of OB cells (Figure 4B).
345 We measured endogenous sesquiterpene levels in eight independent transformants
346 for each gene combination. Lines expressing $_{Prom5}$*MYB2:AMS* alone produced
347 relatively low yields, but co-expression of precursor genes significantly increased
348 sesquiterpene levels (Figure 4C, Dataset 3). Among combinations without *ABCG1*,
349 *FPS* + *HMGR* boosted sesquiterpene levels the most, with a 2.8-fold average increase
350 for the major sesquiterpenes compared to a 1.8-fold increase with *FPS* and *DXS2*.
351 The over-expression of *ABCG1* alongside all precursor genes led to the most
352 significant increases in endogenous sesquiterpene levels, with a 3-fold increase in cis-
353 thujopsene, a 3.2-fold increase in β-chamigrene and γ-cuprene, and a 3.6-fold
354 increase in thujopsan-2α-ol (Figure 4C, Dataset 3). The observed increase in
355 endogenous sesquiterpene levels when *HMGR* was expressed under the MYB2
356 promoter also suggests that HMGR protein is likely localized to OB cells, despite its
357 apparent absence in localization studies using the mVenus tag.
358 To further validate the transport role of ABCG1, we generated CRISPR lines to disrupt
359 its function. We obtained six independent transformants with early stop codons in the
360 *ABCG1* DNA sequence (Figure 5A). Terpene analysis revealed a dramatic reduction
361 or complete absence of sesquiterpenes in *abcg1* mutants compared to Cas9-only
362 controls (Figures 5B and 5C). Notably, this reduction was specific to sesquiterpenes,
363 as the levels of phytosterols remained unaffected (Figure 5B and 5C). Given the
364 apparent differences in peak abundance observed in the overlaid chromatograms
365 (Figure 5B), we quantified additional compounds alongside neophytadiene and phytol,
366 such as (*Z*)-1,3-phytadiene and (*E*)-1,3-phytadiene, phytol derivatives identified in
367 prior studies[77,78] (Figure 5B). Quantification revealed no significant changes in these
368 compounds when comparing the five Cas9 controls and six *abcg1* mutant lines (Figure
369 S7). These findings suggest that ABCG1 plays a critical role in the accumulation of

370 endogenous sesquiterpenes while having no effect on other terpene sub-classes or
371 on the accumulation of exogenous sesquiterpenes.
372
373 **DISCUSSION**
374
375 **Mapping of cell-type specificity and sub cellular localisation of isoprenoid**
376 **biosynthetic enzymes.**
377 We used translational and trancriptional reporters to map both the cell-type specificity
378 and subcellular localization of enzymes involved in the early steps of terpene
379 biosynthesis. These findings corroborate earlier transcriptomic analyses in OB-
380 defective plants[5,7], as well as co-expression network analysis[14] and single-cell RNA
381 sequencing data[35], all of which indicate that these enzymes are OB-cell specific.
382 Enzymatic steps of the MEP pathway and diterpene biosynthesis, including DXS,
383 HDS, HDR, and GGPPS, were localized to chloroplasts/plastids – referred to as
384 chloro-amyloplasts in OB cells by Suire et al[3] – consistent with their predicted roles in
385 producing isoprenoid precursors for plastid-derived terpenes. In contrast, steps of the
386 MVA pathway and sesquiterpene biosynthesis, represented by HMGR and FPS, were
387 localized to the ER, Golgi, and cytosol, supporting their involvement in cytosolic
388 terpene precursor production. With the exception of FPS, all studied enzymes
389 localized as predicted by DeepLoc2.1[28]. However, our experimental conditions
390 allowed us to localise steady-state accumulation of proteins and did not account for
391 any dynamic behavior due to protein trafficking[79]. Our findings corroborate earlier
392 studies[3] that reported high or exclusive expression of isoprenoid biosynthetic enzymes
393 in OB cells. Given that OBs are formed through redirection of the secretory pathway[7],
394 some enzymes like HMGR – which localizes to both the ER and Golgi depending on
395 the OB cell or plant being imaged – may transit to other compartments during OB
396 development, supporting the possibility of their transient presence around the OB
397 membrane[3].
398 For each translational or expression pattern reporter, we selected promoter lengths
399 based on a single ATAC-seq[38] peak, ensuring coverage of the putative core promoter
400 and upstream regulatory regions. In some cases, promoters (including the 5'UTR)
401 were shorter than the typically advised 2.5 kb length[37,41], balancing the need for
402 regulatory coverage with constraints related to sequence synthesis. For instance, a
403 promoter as short as 970 bp like $_{Prom5}$*GGPPS* was sufficient to drive strong and

404 ubiquitous signal in chloro-amyloplasts. However, in other instances such as
405 $_{Prom5}$*DXR2*, promoter functionality required extending the length from 1.7 kb to 2.7 kb
406 to achieve detectable expression. Altering promoter lengths could include or exclude
407 positive or negative regulatory elements affecting cell-type specificity, therefore further
408 systematic studies would be needed to draw concrete conclusions about the
409 relationship between promoter structure and regulatory outcomes.
410 Nevertheless, our results confirm the specialized role of OB cells in terpene
411 biosynthesis and highlight the importance of carefully selecting promoters to ensure
412 accurate expression and localization of biosynthetic enzymes.
413
414 **Pathway specificity in terpene synthesis: MVA and MEP contributions.**
415 Our study suggests that the MVA pathway predominantly provides precursors for
416 sesquiterpene synthesis in *Marchantia polymorpha*, supported by the apparent
417 exclusive expression of HMGR and MVD in OB cells and the significant increases in
418 endogenous sesquiterpene levels observed when *HMGR* and *FPS* were co-
419 expressed. Notably, FPS – also a potential OB-specific enzyme – emerged as a critical
420 limiting step in sesquiterpene biosynthesis, consistent with previous findings in
421 *Nicotiana tabacum*[80].
422 The MEP pathway's contribution to sesquiterpene synthesis cannot be ruled out, given
423 the specific expression of *DXS2* and *DXR2* in OB cells and the fact that a previous
424 study using GC-MS analysis of physically extracted OB content detected
425 sesquiterpenes[2] as the only sub-class of terpenes. If OBs also store monoterpenes
426 like limonene[26] or specific diterpenes, their levels remained undetectable or
427 unchanged despite the targeted expression of precursor supply genes. Furthermore,
428 over-expression of key MEP genes and *GGPPS* had no effect on taxadiene yield,
429 despite co-localization of these enzymes with TXS in plastids. This suggests that a
430 significant portion of the precursors is redirected toward sesquiterpene synthesis
431 instead. Supporting this, co-expression of *DXS2* with *FPS*, *HMGR*, and *ABCG1* mildly
432 increased the sesquiterpene levels compared to the same combination of genes
433 without *DXS2*, demonstrating a modest effect of *DXS2* on sesquiterpene yield. Dual
434 contribution from the MVA and MEP pathways in *Marchantia polymorpha* would
435 parallel findings in *Artemisia annua*, where inhibitor assays demonstrated a mixed
436 origin of precursors[81].

437 In contrast, *DXS1* and *DXR1* are ubiquitously expressed and likely play broader roles
438 in synthesizing essential metabolites, such as phytyl diphosphate for chlorophyll
439 production. Supporting this, ChIP-seq studies in *Marchantia* have shown that the
440 chloroplast biogenesis regulator GLK binds to the promoter of *DXR1*[82], further
441 highlighting its role in primary metabolism rather than specialized metabolite
442 production.
443 The specific localisation of HMGR, MVD, and FPS protein fusions in OB cells raises
444 intriguing questions about how essential metabolites, such as sterols, are produced in
445 non-OB cells. Despite the ubiquitous expression of squalene synthase *SQS3*, which
446 catalyzes the formation of squalene as a precursor to sterols, the source of precursors
447 for sterol synthesis in non-OB tissues remains unclear. Notably, sterol levels in *abcg1*
448 mutants were unaffected, suggesting that OBs are not the primary site of sterol
449 synthesis or storage for the rest of the plant. This is further supported by the
450 observation that sterol levels remain unchanged in plants defective in OB cell
451 differentiation[5]. This raises the possibility that other, as-yet-unannotated MVA
452 enzymes may supply precursors for sterol biosynthesis in non-OB cells. Alternatively,
453 it is plausible that the genes studied here are expressed ubiquitously but accumulate
454 at levels too low to be detected in other cell types with our confocal microscopy
455 studies.
456 These findings highlight the complexity of precursor flux and compartmentalization in
457 *Marchantia*, accentuating the need to further investigate how primary metabolites
458 synthesis is maintained across tissues.
459
460 **Putative role of ABCG1 transporter and implications for metabolic engineering**
461 **in OB cells.**
462 Our study demonstrated that exogenous terpenes, such as amorpha-4,11-diene,
463 taxadiene, and β-amyrin, could be produced throughout the whole plant. However,
464 achieving useful yields within OB cells proved less straightforward, even with the
465 overexpression of key biosynthetic genes. This led us to consider whether transport
466 into OBs might be a limiting factor for the accumulation of exogenous compounds.
467 Supporting this, CRISPR-mediated disruption of *ABCG1* resulted in a sharp reduction
468 in endogenous sesquiterpene levels, demonstrating that ABCG1 plays a crucial role
469 in terpene accumulation. However, the specificity of its transport function remains an
470 open question.

471 Some ABCG transporters exhibit broad substrate specificity, while others are highly
472 selective. For example, *Nicotiana tabacum* NtNPR1 shows increased ATPase activity
473 when exposed to structurally distinct terpene substrates, including sesquiterpenes like
474 sclareol and capsidiol and the diterpene cembrene[83], suggesting that it can
475 accommodate chemically diverse terpenes. In contrast, *Arabidopsis thaliana* ABCG29
476 is a highly specific transporter of *p*-coumaryl alcohol (a lignin precursor)[84], while
477 several transport assay studies reveal that multiple ABCG transporters exclusively
478 carry the phytohormone abscisic acid (ABA), as reviewed by Do et al[85]. These
479 examples raise the question: does MpABCG1 transport all the sesquiterpenes
480 produced in *Marchantia* OBs, or is it more selective?
481 Our chromatographic analysis identified over 35 sesquiterpenes in *Marchantia*, though
482 the true number may be even higher, as our GC-MS approach — like previous studies[2]
483 — primarily detects compounds with limited hydroxylation. If ABCG1 were responsible
484 for transporting all these metabolites into OBs, it would need to accommodate a
485 remarkably high number of substrates. Yet, MpABCG1 did not facilitate the
486 accumulation of exogenous amorpha-4,11-diene, indicating that its function may be
487 more selective than initially expected.
488 We hypothesize that MpABCG1 could be more exclusive, transporting the
489 sesquiterpene precursor FPP rather than the numerous end-products. To date, no
490 ABCG transporter has been shown to facilitate the movement of a phosphate-
491 containing molecule, with the closest example being *Arabidopsis* ABCC5, an ABCC
492 sub-family transporter that carries inositol hexakisphosphate[86]. However, FPP shares
493 structural features with ABA — not only in its amphiphilic nature but also in its
494 sesquiterpenoid backbone — raising the possibility that an ABCG transporter could
495 accommodate FPP as a substrate.
496 A serendipitous observation in our study lends further support to the hypothesis that
497 FPP, rather than sesquiterpenes, is transported. A translational reporter of the *fungal*
498 *terpene synthase-like 2* (*FTPSL2*), an endogenous sesquiterpene synthase
499 characterized by Kumar et al.[26], was detected within the OB lumen (Figure S8). This
500 finding aligns with earlier theories suggesting that the OB lumen is not merely a
501 storage site but also a catalytic compartment for terpene biosynthesis[3]. However, the
502 conditions under which FTPSL2-mVenus was observed inside OBs remain unclear,
503 as this localization pattern was detected in only one lobe out of four in plants no
504 younger than 14 days old (Figure S8).

If this theory holds true, it implies that precursor flux from both the MVA and MEP pathways is primarily channeled toward FPP production, which is then transported into OBs for conversion by endogenous sesquiterpene synthases. This would explain why co-expression of different *HMGR* versions with *β-AS* did not increase β-amyrin yield: little FPP may remain in the cytosol, limiting its availability for β-amyrin biosynthesis. To determine the precise substrate specificity and transport mechanism of ABCG1, direct biochemical validation of its role in FPP or sesquiterpene transport will be essential. If ABCG1 is indeed specialized in transporting terpene precursors, optimizing taxadiene biosynthesis within OBs would require (1) redirecting GGPPS activity outside plastids, (2) identifying a transporter to mediate GGPP entry into OBs, and (3) engineering TXS for OB localization. Similarly, producing amorpha-4,11-diene or β-amyrin in OBs would necessitate targeting their respective biosynthetic enzymes – and SQS3 in the case of β-amyrin – to this compartment.

We present a preliminary map of terpene biosynthesis in *Marchantia* OB cells (Figure 6). This map includes key steps in primary metabolism, emphasizing the role of certain terpene precursors in vital plant functions, such as sterols, chlorophyll and gibberellin derivative GA_{12}[87], although the presence of the latter in OB cells has yet to be confirmed (Figure 6).

Overall, our work provides a new comprehensive view of terpene synthesis in *Marchantia polymorpha* OB cells by redefining the subcellular localization of key isoprenoid biosynthetic enzymes. Our study also highlights the critical role of ABCG1 in sesquiterpene accumulation and contributes to the broader understanding of metabolic compartmentalization. These findings provide a foundation for engineering terpene biosynthesis with greater precision in specialized plant cells. Beyond *Marchantia*, this work may help uncover similar transport mechanisms in other plant species, offering new strategies for metabolic engineering and plant synthetic biology.

MATERIAL AND METHODS.

Plant material and growth conditions.

Marchantia polymorpha subspecies *ruderalis* accessions Cam-1 (male) and Cam-2 (female) were used in this study[88]. Plants were cultivated on solid 0.5× Gamborg B-5 basal medium (#G398, PhytoTech Labs, Lenexa, Kansas, USA) adjusted to pH 5.7–5.8 and solidified with 1.2% (w/v) agar micropropagation grade (#A296, PhytoTech

539 Labs). They were maintained under continuous light at 22°C with a light intensity of
540 150 µmol/m²/s. For general propagation of the lines, plants were grown in 94 × 16 mm
541 Petri dishes (#633181, Greiner Bio-One, Kremsmünster, Austria). For imaging
542 purposes, plants were cultivated on gridded 65 × 14.5 mm Gosselin™ Petri dishes
543 (#BB64-01, Corning, Corning, NY, USA). Several gemmae, originating from four
544 independent transformants, were grown on these plates and imaged at developmental
545 stages corresponding to Day 3, Day 5, Day 7, and Day 14 after germination.
546 For gemma cup production and the collection of material for terpene analysis, the
547 media were supplemented with 0.5% (w/v) sucrose. Plants destined for terpene
548 extraction and quantification were grown in 100 × 25 mm Petri dishes (#D943,
549 Phytotech Labs) containing 50 mL of medium.
550 Spore production was carried out in Microbox micropropagation containers (SacO2)
551 under long-day conditions (16 h light/8 h dark) with light supplemented by far-red
552 illumination, following previously established protocol[41].
553
554 **Protein sequence analysis.**
555 Protein sequences of the selected IDS and SQS candidates from *Marchantia*
556 *polymorpha* were analyzed using NCBI BLASTP (Basic Local Alignment Search Tool
557 for Proteins, https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the *Arabidopsis thaliana*
558 protein database. Sequence similarity was assessed based on percentage identity
559 and coverage to infer functional homology.
560
561 **Plasmid construction for over-expression studies and CRISPR-based mutant**
562 **generation.**
563 DNA fragments used to generate mVenus and metabolic constructs were synthesized
564 by Genewiz (Azenta, Burlington, Massachusetts, USA) either as linear fragments or
565 in pUAP1 vector[89]. Linear fragments included overhangs compatible with LguI
566 (#ER1931, ThermoFisher Scientific, Waltham, Massachusetts, USA) cloning into L0
567 acceptor vectors following the Loop protocol[41]. Linear fragments and sequences
568 synthesized in pUAP1 vectors, contained overhangs enabling direct cloning into L1 or
569 pBy01[37] vectors using BsaI-HF ®v2 enzyme (#R3733S, New England Biolabs,
570 Ipswitch, Massachusetts, USA).
571 *MpFPS* and *MpHDS* sequences were cloned from the *Marchantia* transcriptome as
572 they didn't require to remove the internal restriction sites of BsaI and LguI,

573 incompatible with the Loop cloning system. Total RNA was first extracted from
574 *Marchantia* tissue using the RNeasy Plant Mini Kit (#74904, Qiagen, Hilden, Germany)
575 and reverse-transcribed into cDNA with Superscript IV (#18090010, Invitrogen,
576 Waltham, Massachusetts, USA) using random hexamer oligos (#N8080127,
577 Invitrogen) following the protocol from the manufacturer. Target sequences were
578 subsequently amplified from the cDNA using VeriFi® Polymerase Mix (#PB10.43-01,
579 PCR Biosystem, London, UK). PCR products were purified using the QIAquick PCR
580 Purification Kit (#28104, Qiagen) before being used in the cloning steps.
581 DNA sequences exogenous to *Marchantia* were domesticated when necessary and
582 codon-optimised using the Genewiz codon-optimisation tool prior to ordering. NCBI
583 accession numbers for each exogenous genes are as follows: *TXS* (AY424738), *β-AS*
584 (MG492000), *PpDXS1* (XM_024524883), *PpDXS2* (XM_024533934), *PpHMGR*
585 (XM_024507706) and *AMS* (AY006482). Certain sequences were re-amplified to
586 remove stop codons and introduce the appropriate overhangs for fusion with mVenus
587 or eGFP tags. Primers used in this study were ordered from IDT (Integrated DNA
588 Technology, Coralville, Iowa, USA) (Table S4).
589 Vectors, promoters, terminators, tags, and pre-assembled cassettes used to generate
590 the constructs were sourced from the OpenPlant toolkit[41] or adapted from Romani et
591 al[37]. Final plasmids (L2, L3, or pBy01) were sequenced at Plasmidsaurus (Eugene,
592 Oregon, USA) to confirm proper assembly and integration of all transcription units.
593 Details of the sequences and vectors syntax are available in Dataset 5.
594 The guide RNA sequence used to generate *abcg1* mutants was designed using the
595 CasFinder tool[90]. This guide was ordered as primers with appropriate overhangs, and
596 cloned into a single-guide acceptor vector following the protocol of Sauret-Guëto et
597 al[23].
598
599 ***Agrobacterium*-mediated transformation of *Marchantia* spores and selection of**
600 **independent transformants.**
601 *Marchantia* spores were sterilized and transformed following previously described
602 protocols, with slight modifications. Briefly, a single archegoniophore, dried and stored
603 in silica beads, was mixed in 1.5 mL of a chlorinated solution consisting of one Milton
604 sterilizing tablet (troclosene sodium; Boots, UK) dissolved in 10 mL of sterile water.
605 Spores were released into the solution, filtered through a 40 µm cell strainer (#542040,

606 Greiner Bio-One) and incubated in the sterilizing solution for 30 minutes. After
607 sterilization, spores were harvested by centrifugation, resuspended in sterile water,
608 and spread on Gamborg medium agar plates (without sucrose).
609 After 5 days of germination, the sporelings were co-cultured with *Agrobacterium*
610 *tumefaciens* GV3101 carrying the constructs of interest. Transformation of
611 *Agrobacterium* with the constructs was performed using the freeze thaw method[91].
612 Sporelings and *Agrobacterium* were co-cultured for 2 days in a 4 mL solution of 0.5×
613 Gamborg B-5 plus supplements like previously described[41], and incubated at 22°C
614 under continuous shaking and light. Following co-culture, sporellings were collected
615 on 70 µm cell strainers (#542070, Greiner Bio-One), rinsed thoroughly with sterile
616 water, and plated on 90 mm Petri dishes containing 0.5× Gamborg B5 medium
617 supplemented with 100 µg/mL cefotaxime (#BIC0111, Apollo Scientific, Bredbury, UK)
618 to eliminate *Agrobacterium* and 20 µg/mL hygromycin (#10687010, Invitrogen) to
619 select for transformants.
620 After two weeks on the first selective plates, eight transformants were typically
621 transferred to a second plate containing the same selection agents to confirm
622 successful transformation and eliminate residual *Agrobacterium*. For lines expressing
623 fluorescent proteins (mScarlet, mVenus or eGFP), transformants were pre-selected
624 under a Leica stereo microscope (#MDG41, Leica, Wetzlar, Germany) using suitable
625 filter sets to confirm the presence of fluorescence signals. For lines expressing
626 exogenous terpenes, eight plants were randomly chosen, grown for one month on the
627 second selective plates, and subsequently extracted to confirm the presence of the
628 terpene of interest. The four best lines producing the desired terpene were then
629 transferred to sucrose-supplemented plates to quantify endogenous and exogenous
630 terpene levels.
631 For lines co-expressing *AMS* and precursor supply genes, a single vector carrying up
632 to six transcription units could not be constructed due to cloning or vector size
633 limitations. Attempts to co-transform using two *Agrobacterium* strains carrying distinct
634 plasmids — one with hygromycin B as the selection agent and the other with
635 chlorsulfuron — resulted in an insufficient number of transformants. Therefore,
636 transformants were selected based on hygromycin B resistance for *AMS* (with or
637 without *ABCG1*) on one vector, and the presence of an mScarlet-Lti6b fluorescent
638 signal from the second vector carrying *MpFPS*, *MpHMGR*, and/or *MpDXS2*. All eight

639 independent lines were retained to measure amorphadiene and endogenous terpene
640 level.
641 To genotype CRISPR-generated *abcg1* mutants and Cas9-only control lines, genomic
642 DNA was extracted using the InnuPrep Plant DNA kit (#845-KS-1060050, IST
643 Innuscreen, Berlin, Germany). For *abcg1* lines, a 500 bp DNA fragment spanning the
644 gRNA target site was amplified using Q5® High-Fidelity DNA Polymerase (#M0491S,
645 New England Biolabs). The resulting PCR products were purified and sequenced by
646 Genewiz to confirm the disruption or not of the *ABCG1* coding sequence. The
647 presence of the Cas9 cassette in control lines was confirmed by amplification with
648 specific primers (listed in Table S4) and subsequent sequencing. A total of 16 lines
649 were genotyped for the abcg1 ones, and eight for the Cas9 controls.
650
651 **Laser scanning confocal microscopy.**
652 For transgenic lines expressing fluorescent proteins, images were acquired on an
653 upright Leica SP8X confocal microscope equipped with a 460–670 nm
654 supercontinuum white light laser, two continuous wavelength laser lines of 405 nm and
655 442 nm and a five-channel spectral scanhead (four hybrid detectors and one
656 photomultiplier). Imaging was conducted using either a 25× water immersion objective
657 (Fluotar VISIR 25×/0.95 WATER) for whole-gemmae or meristematic area imaging, or
658 a 40× water immersion objective (HC PL APO CS2 40×/1.10 WATER) for higher
659 magnification images, with an additional digital zoom applied up to a factor of 5× to
660 enhance visualization of subcellular localizations. Excitation laser wavelength and
661 fluorescence emission bandwidth windows were as follows: 515 nm and 525–550 nm
662 (for mVenus); 570 nm and 591-621 nm (for mScarlet); 442 nm and 645–664 nm (for
663 chlorophyll autofluorescence). Each channel was acquired separately using a hybrid
664 detector, with a sequential scan for each channel performed on the Leica LAS X
665 software. To image sequentially the mVenus, eGFP and mScarlet fluorescences on
666 the translational reporter lines carrying the bidirectional construct of *HDS-eGFP* and
667 *HDR-mVenus*, excitation and emission bandwidth windows were adjusted as follow to
668 prevent signal crosstalk: 487 nm and 497-513 nm (for eGFP); 521 nm and 530-550
669 nm (for mVenus); 570 nm and 588-598 nm (for mScarlet).
670 Day 0 gemmae were imaged by mounting them on a glass slide with perfluorodecalin[92]
671 (#130040250, ThermoFisher) and covering them with a glass coverslip. For plants
672 aged 3 to 14 days, a 15 x 16 mm gene frame (#AB0577, ThermoFisher) was placed

673 on the glass slide to prevent compression of the plant material under the coverslip.
674 Whole gemmae and overviews of the meristematic region in older plants were imaged
675 using Z-stack scans, which were processed in Fiji[93] to generate maximum intensity
676 projections of the Z-stacks. High magnification images were acquired as single scans
677 to better display subcellular features.
678
679 **Extraction of terpenes and analysis by gas chromatography-mass spectrometry**
680 **(GC-MS).**
681 Approximately 200 mg of frozen tissue from 1-month-old *Marchantia polymorpha*
682 plants was extracted with 1 mL of cold methanol containing 5 mM NaCl to quench
683 enzymatic activity, following the method of Kumar et al[26]. To enable quantification, 5
684 µg of dodecane (#297879, Sigma-Aldrich) was included as an internal standard. The
685 tissue was disrupted in the solvent using a 3 mm stainless steel ball (#2205, Durston,
686 High Wycombe, UK) and homogenized with a TissueLyser II (Qiagen). Samples were
687 subsequently agitated on a Vibrax® shaker (#0002819002, IKA, Staufen, Germany)
688 at 2000 rpm for two hours, and the resulting extracts were centrifuged to remove plant
689 debris. The methanolic phase was then extracted once with hexane to isolate non-
690 polar and medium-polar terpenes from the upper organic layer.
691 Hexane extracts (200 µL) were analyzed on a Trace 1300 gas chromatograph
692 (ThermoFisher) coupled to an ISQ 700 mass spectrometer (ThermoFisher) with a
693 Zebron CD-5MS column (30 m × 0.25 mm × 0.25 µm; ThermoFisher). A splitless
694 injection (1 µL) was conducted at 230°C, and the GC-MS oven parameters were
695 adapted from Kumar's work[26] as follows: the oven temperature was initially set to 70°C
696 and held for 3 min, followed by a ramp of 20°C/min to 90°C, a second ramp of 3°C/min
697 to 180°C, a third ramp of 5°C/min to 240°C, and a final ramp of 20°C/min to 300°C,
698 with a 6-min hold at this temperature. The MS began data acquisition after a 5.5-
699 minute solvent delay, with the transfer line and ion source temperatures set at 250°C
700 and 270°C, respectively. Scanning was conducted in full-scan mode (scan time: 0.17
701 s) over a mass range of 40–600 atomic mass units. Helium was used as the carrier
702 gas, with a flow rate of 1.2 mL/min for sesquiterpenes in the taxadiene and β-amyrin
703 datasets. For amorphadiene detection, the carrier flow rate was reduced to 0.9
704 mL/min. Chromatograms were processed using Chromeleon™ software
705 (ThermoFisher), and terpene quantification was performed relative to the internal
706 standard dodecane. Tentative identification of the major sesquiterpenes was achieved

707 by comparing mass spectra to published data and by calculating retention indices
708 using a single sample run with a C8–C40 alkane standard mixture (#40147-U,
709 Superlco, Bellefonte, Pennsylvania, USA) as described by Adams[73]. Note that the
710 endogenous terpene composition in CAM accessions may slightly differ from that
711 reported for other accessions, such as Tak, in previous studies. The authentic standard
712 of amorpha-4,11-diene was provided by Tomasz Czechowski and Ian Graham
713 (University of York).
714 To achieve more accurate quantification of β-amyrin and sterols, the extraction
715 protocol and GC-MS method were modified. The same amount of frozen material was
716 extracted with cold methanol supplemented with 10 µg of coprostan-3-ol (#C7578,
717 Sigma-Aldrich) as an internal standard. The samples were extracted twice with hexane
718 to obtain a total hexane volume of 1.5 mL, which was then dried under nitrogen flow
719 in a Genevac™ concentrator EZ-2 (#EZ3P-23050-NN0, Genevac Ltd, Ipswich, UK).
720 Extracts were derivatized with 50 µL 1-trimethylsilyl-imidazole (TMS; #92718,
721 Superlco) and resuspended in 300 µL hexane. For GC-MS analysis, 200 µL of the
722 derivatized samples was transferred into vials for direct injection. The method used for
723 sterol/triterpene analysis was as follows: 130°C hold for 2 min; ramp of 30°C/min until
724 220°C, then 2°C/min until 300°C, hold for 10 min. Solvent delay of 10 minutes before
725 acquiring the MS data. The authentic standard of β-amyrin was provided by James
726 Reed and Anne Osbourne (John Innes Center, Norwich, UK).
727 For terpene quantification in *abcg1* CRISPR and Cas9 control lines, extractions were
728 performed on genotyped plants from the second selection stage, as loss of the
729 CRISPR genotype was observed during propagation from gemmae. Given this
730 observation, we considered the possibility of chimerism[94] and extracted as much
731 biomass as was reasonably available from individual plants. This approach aimed to
732 minimize variability, homogenize the material, and ensure accurate results.
733 Box plots showing individual data points for each quantified terpene were created
734 using PRISM software version 10 (GraphPad, La Jolla, California, USA), and the
735 results of statistical comparisons were integrated into the visualizations. For lines
736 expressing exogenous terpene synthases, statistical analyses were performed using
737 one-way analysis of variance (ANOVA) to assess differences across gene
738 combinations. Homogeneity of variances was evaluated using the Brown-Forsythe
739 and Bartlett's tests, with Dunnett's multiple comparison test applied post-hoc to
740 compare each gene combination to the respective control ($_{Prom5}$*MYB2:β-AS*, *TXS*, or

741 *AMS*). The significance threshold (alpha) was set at 0.05. Four biological replicates
742 expressing the constructs were analyzed to quantify both exogenous (β-amyrin and
743 taxadiene) and endogenous compounds (n=4), while eight biological replicates
744 expressing the constructs were analyzed for amorphadiene lines, focusing on the
745 quantification of endogenous terpenes since amorphadiene was not detected (n=8).
746 For comparisons of terpene levels between *abcg1* mutants (n = 6) and Cas9 controls
747 (n = 5), Welch's t-test was used when variances were unequal, and unpaired t-test
748 was used when variances were equal, as determined by PRISM. Both tests accounted
749 for differences in sample size between the two groups. Asterisks were added to the
750 box plots to indicate statistically significant differences (P<0.05) between *abcg1*
751 mutants and Cas9 control lines, as determined by Welch's or unpaired t-tests.
752
753 **REFERENCES**
754
755 1. Bowman, J. L. *et al.* The renaissance and enlightenment of Marchantia as a
756 model system. *Plant Cell* **34**, 3512–3542 (2022).
757 2. Tanaka, M. *et al.* Direct evidence of specific localization of sesquiterpenes and
758 marchantin A in oil body cells of Marchantia polymorpha L. *Phytochemistry* **130**, 77–
759 84 (2016).
760 3. Suire, C. *et al.* Cellular Localization of Isoprenoid Biosynthetic Enzymes in
761 *Marchantia polymorpha* . Uncovering a New Role of Oil Bodies. *Plant Physiology* **124**,
762 971–978 (2000).
763 4. Jensen, S. *et al.* Marchantin A, a macrocyclic bisbibenzyl ether, isolated from
764 the liverwort Marchantia polymorpha, inhibits protozoal growth in vitro. *Phytomedicine*
765 **19**, 1191–1195 (2012).
766 5. Romani, F. *et al.* Oil Body Formation in Marchantia polymorpha Is Controlled
767 by MpC1HDZ and Serves as a Defense against Arthropod Herbivores. *Current Biology*
768 **30**, 2815-2828.e8 (2020).
769 6. Ruzicka, L. The isoprene rule and the biogenesis of terpenic compounds.
770 *Experientia* **9**, 357–367 (1953).
771 7. Kanazawa, T. *et al.* The liverwort oil body is formed by redirection of the
772 secretory pathway. *Nat Commun* **11**, 6152 (2020).
773 8. Gutsche, N. *et al.* Mp *TGA* , together with Mp *NPR* , regulates sexual
774 reproduction and independently affects oil body formation in *Marchantia polymorpha*.

775 *New Phytologist* **241**, 1559–1573 (2024).
776 9. Kubo, H. *et al.* Biosynthesis of riccionidins and marchantins is regulated by
777 R2R3-MYB transcription factors in Marchantia polymorpha. *J Plant Res* **131**, 849–864
778 (2018).
779 10. Kubo, H., Sunhwa, K., Teramori, H. & Takanashi, K. MpR2R3-MYB2 is a key
780 regulator of oil body formation in Marchantia polymorpha. *Planta* **260**, 68 (2024).
781 11. Romani, F. *et al.* Liverwort oil bodies: diversity, biochemistry, and molecular cell
782 biology of the earliest secretory structure of land plants. *Journal of Experimental*
783 *Botany* **73**, 4427–4439 (2022).
784 12. Ishizaki, K., Nishihama, R., Yamato, K. T. & Kohchi, T. Molecular Genetic Tools
785 and Techniques for Marchantia polymorpha Research. *Plant and Cell Physiology* **57**,
786 262–270 (2016).
787 13. Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals
788 an Ancestral Association between H3K27me3, Transposons, and Constitutive
789 Heterochromatin. *Current Biology* **30**, 573-588.e7 (2020).
790 14. Kawamura, S. *et al.* MarpolBase Expression: A Web-Based, Comprehensive
791 Platform for Visualization and Analysis of Transcriptomes in the Liverwort Marchantia
792 polymorpha. *Plant and Cell Physiology* **63**, 1745–1755 (2022).
793 15. Rohmer, M., Knani, M., Simonin, P., Sutter, B. & Sahm, H. Isoprenoid
794 biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl
795 diphosphate. *Biochemical Journal* **295**, 517–524 (1993).
796 16. Lichtenthaler, H. K., Schwender, J., Disch, A. & Rohmer, M. Biosynthesis of
797 isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent
798 pathway. *Febs Letters* **400**, 271–274 (1997).
799 17. Rohmer, M. The discovery of a mevalonate-independent pathway for
800 isoprenoid biosynthesis in bacteria, algae and higher plants†. *Nat. Prod. Rep.* **16**, 565–
801 574 (1999).
802 18. Bouvier, F., Rahier, A. & Camara, B. Biogenesis, molecular regulation and
803 function of plant isoprenoids. *Progress in Lipid Research* **44**, 357–429 (2005).
804 19. Rodríguez-Concepción, M. & Boronat, A. Breaking new ground in the regulation
805 of the early steps of plant isoprenoid biosynthesis. *Current Opinion in Plant Biology*
806 **25**, 17–22 (2015).
807 20. Leivar, P. *et al.* Subcellular Localization of Arabidopsis 3-Hydroxy-3-
808 Methylglutaryl-Coenzyme A Reductase. *Plant Physiology* **137**, 57–69 (2005).

809 21. Simkin, A. J. *et al.* Peroxisomal localisation of the final steps of the mevalonic
810 acid pathway in planta. *Planta* **234**, 903–914 (2011).
811 22. Chappell, J. Biochemistry and Molecular Biology of the Isoprenoid Biosynthetic
812 Pathway in Plants. *Annual Review of Plant Biology* **46**, 521–547 (1995).
813 23. KUZUYAMA, T. Mevalonate and Nonmevalonate Pathways for the
814 Biosynthesis of Isoprene Units. *Bioscience, Biotechnology, and Biochemistry* **66**,
815 1619–1627 (2002).
816 24. Hsieh, M.-H., Chang, C.-Y., Hsu, S.-J. & Chen, J.-J. Chloroplast localization of
817 methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes
818 in ispD and ispE albino mutants in Arabidopsis. *Plant Mol Biol* **66**, 663–673 (2008).
819 25. Bowman, J. L. *et al.* Insights into Land Plant Evolution Garnered from the
820 Marchantia polymorpha Genome. *Cell* **171**, 287-304.e15 (2017).
821 26. Kumar, S. *et al.* Molecular Diversity of Terpene Synthases in the Liverwort
822 Marchantia polymorpha. *Plant Cell* tpc.00062.2016 (2016) doi:10.1105/tpc.16.00062.
823 27. Chen, F. *et al.* Terpenoid Secondary Metabolites in Bryophytes: Chemical
824 Diversity, Biosynthesis and Biological Functions. *Critical Reviews in Plant Sciences*
825 **37**, 210–231 (2018).
826 28. Ødum, M. T. *et al.* DeepLoc 2.1: multi-label membrane protein type prediction
827 using protein language models. *Nucleic Acids Research* **52**, W215–W220 (2024).
828 29. Koepp, A. E. *et al.* Cyclization of Geranylgeranyl Diphosphate to Taxa-
829 4(5),11(12)-diene Is the Committed Step of Taxol Biosynthesis in Pacific Yew *.
830 *Journal of Biological Chemistry* **270**, 8686–8690 (1995).
831 30. Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. Plant
832 antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and
833 antitumor agent from Taxus brevifolia. *J. Am. Chem. Soc.* **93**, 2325–2327 (1971).
834 31. Suffness, M. *TAXOL®: Science and Applications*. (CRC Press, London, 2021).
835 doi:10.1201/9780138737368.
836 32. Kushiro, T., Shibuya, M. & Ebizuka, Y. β-Amyrin synthase. *European Journal*
837 *of Biochemistry* **256**, 238–244 (1998).
838 33. Fiore, C., Eisenhut, M., Ragazzi, E., Zanchin, G. & Armanini, D. A history of the
839 therapeutic use of liquorice in Europe. *J Ethnopharmacol* **99**, 317–324 (2005).
840 34. Bouwmeester, H. J. *et al.* Amorpha-4,11-diene synthase catalyses the first
841 probable step in artemisinin biosynthesis. *Phytochemistry* **52**, 843–854 (1999).
842 35. Wang, L. *et al.* The maturation and aging trajectory of Marchantia polymorpha

843 at single-cell resolution. *Developmental Cell* **58**, 1429-1444.e6 (2023).
844 36. Kremers, G.-J., Goedhart, J., van Munster, E. B. & Gadella, T. W. J. Cyan and
845 Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and
846 FRET Förster Radius,. *Biochemistry* **45**, 6570–6580 (2006).
847 37. Romani, F. *et al.* The landscape of transcription factor promoter activity during
848 vegetative development in Marchantia. *Plant Cell* **36**, 2140–2159 (2024).
849 38. Karaaslan, E. S. *et al.* Marchantia TCP transcription factor activity correlates
850 with three-dimensional chromatin structure. *Nat. Plants* **6**, 1250–1261 (2020).
851 39. Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for
852 cellular imaging. *Nat Methods* **14**, 53–56 (2017).
853 40. Cutler, S. R., Ehrhardt, D. W., Griffitts, J. S. & Somerville, C. R. Random
854 GFP::cDNA fusions enable visualization of subcellular structures in cells of
855 *Arabidopsis* at a high frequency. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3718–3723 (2000).
856 41. Sauret-Güeto, S. *et al.* Systematic Tools for Reprogramming Plant Gene
857 Expression in a Simple Model, *Marchantia polymorpha*. *ACS Synth. Biol.* **9**, 864–882
858 (2020).
859 42. Minamino, N. *et al.* RAB GTPases in the Basal Land Plant Marchantia
860 polymorpha. *Plant and Cell Physiology* **59**, 850–861 (2018).
861 43. Lichtenthaler, H. K. THE 1-DEOXY-D-XYLULOSE-5-PHOSPHATE PATHWAY
862 OF ISOPRENOID BIOSYNTHESIS IN PLANTS. *Annual Review of Plant Biology* **50**,
863 47–65 (1999).
864 44. Rodríguez-Concepción, M. & Boronat, A. Elucidation of the Methylerythritol
865 Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic
866 Milestone Achieved through Genomics. *Plant Physiology* **130**, 1079–1089 (2002).
867 45. Hsieh, M.-H. The Arabidopsis IspH Homolog Is Involved in the Plastid
868 Nonmevalonate Pathway of Isoprenoid Biosynthesis. *PLANT PHYSIOLOGY* **138**,
869 641–653 (2005).
870 46. Seemann, M., Tse Sum Bui, B., Wolff, M., Miginiac-Maslow, M. & Rohmer, M.
871 Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: Direct
872 thylakoid/ferredoxin-dependent photoreduction of GcpE/IspG. *FEBS Letters* **580**,
873 1547–1552 (2006).
874 47. Li, Q. *et al.* Balanced activation of IspG and IspH to eliminate MEP intermediate
875 accumulation and improve isoprenoids production in Escherichia coli. *Metabolic*
876 *Engineering* **44**, 13–21 (2017).

877 48. Forestier, E. C. F., Brown, G. D., Harvey, D., Larson, T. R. & Graham, I. A.
878 Engineering Production of a Novel Diterpene Synthase Precursor in Nicotiana
879 benthamiana. *Frontiers in Plant Science* **12**, 8 (2021).
880 49. Laferrière, A. & Beyer, P. Purification of geranylgeranyl diphosphate synthase
881 from *Sinapis alba* etioplasts. *Biochimica et Biophysica Acta (BBA) - Protein Structure*
882 *and Molecular Enzymology* **1077**, 167–172 (1991).
883 50. Kuntz, M. *et al.* Identification of a cDNA for the plastid-located geranylgeranyl
884 pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme
885 activity and transcript level during fruit ripening. *The Plant Journal* **2**, 25–34 (1992).
886 51. Ruiz-Sola, M. Á. *et al. Arabidopsis* GERANYLGERANYL DIPHOSPHATE
887 SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-
888 related isoprenoids. *New Phytologist* **209**, 252–264 (2016).
889 52. Wang, C. *et al.* Structural Analyses of Short-Chain Prenyltransferases Identify
890 an Evolutionarily Conserved GFPPS Clade in Brassicaceae Plants. *Molecular Plant* **9**,
891 195–204 (2016).
892 53. Lai, D.-H. *et al.* Solanesyl Diphosphate Synthase, an Enzyme of the Ubiquinone
893 Synthetic Pathway, Is Required throughout the Life Cycle of Trypanosoma brucei.
894 *Eukaryot Cell* **13**, 320–328 (2014).
895 54. Bach, T. J. Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol
896 synthesis? *Lipids* **21**, 82–88 (1986).
897 55. Poulter, C. D., Argyle, J. C. & Mash, E. A. Farnesyl pyrophosphate synthetase.
898 Mechanistic studies of the 1'-4 coupling reaction with 2-fluorogeranyl pyrophosphate.
899 *Journal of Biological Chemistry* **253**, 7227–7233 (1978).
900 56. Tse, S. W. *et al.* Optimizing Promoters and Subcellular Localization for
901 Constitutive Transgene Expression in Marchantia polymorpha. *Plant and Cell*
902 *Physiology* **65**, 1298 (2024).
903 57. Kuzuyama, T., Takahashi, S., Watanabe, H. & Seto, H. Direct formation of 2-
904 C-methyl-d-erythritol 4-phosphate from 1-deoxy-d-xylulose 5-phosphate by 1-deoxy-
905 d-xylulose 5-phosphate reductoisomerase, a new enzyme in the non-mevalonate
906 pathway to isopentenyl diphosphate. *Tetrahedron Letters* **39**, 4509–4512 (1998).
907 58. Takahashi, S., Kuzuyama, T., Watanabe, H. & Seto, H. A 1-deoxy-D-xylulose
908 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-
909 phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis.
910 *Proceedings of the National Academy of Sciences of the United States of America* **95**,

911 9879–9884 (1998).
912 59. Bloch, K., Chaykin, S., Phillips, A. H. & De Waard, A. Mevalonic Acid
913 Pyrophosphate and Isopentenylpyrophosphate. *Journal of Biological Chemistry* **234**,
914 2595–2604 (1959).
915 60. Schaller, H. The role of sterols in plant growth and development. *Prog Lipid Res*
916 **42**, 163–175 (2003).
917 61. Mookhtiar, K. A., Kalinowski, S. S., Zhang, D. & Poulter, C. D. Yeast squalene
918 synthase. A mechanism for addition of substrates and activation by NADPH. *Journal*
919 *of Biological Chemistry* **269**, 11201–11207 (1994).
920 62. Besumbes, Ó. *et al.* Metabolic engineering of isoprenoid biosynthesis in
921 Arabidopsis for the production of taxadiene, the first committed precursor of Taxol.
922 *Biotechnology and Bioengineering* **88**, 168–175 (2004).
923 63. Afifah, I. Q., Wibowo, I. & Faizal, A. A newly identified β-amyrin synthase gene
924 hypothetically involved in oleanane-saponin biosynthesis from Talinum paniculatum
925 (Jacq.) Gaertn. *Heliyon* **9**, e17707 (2023).
926 64. Hasan, M. M. *et al.* Metabolic engineering of Nicotiana benthamiana for the
927 increased production of taxadiene. *Plant Cell Rep* **33**, 895–904 (2014).
928 65. Li, J. *et al.* Chloroplastic metabolic engineering coupled with isoprenoid pool
929 enhancement for committed taxanes biosynthesis in Nicotiana benthamiana. *Nat*
930 *Commun* **10**, 4850 (2019).
931 66. Park, Y.-D. *et al.* Gene silencing mediated by promoter homology occurs at the
932 level of transcription and results in meiotically heritable alterations in methylation and
933 gene activity. *The Plant Journal* **9**, 183–194 (1996).
934 67. Horn, A., Lu, Y., Astorga Ríos, F. J., Toft Simonsen, H. & Becker, J. D.
935 Transcriptional and functional characterization in the terpenoid precursor pathway of
936 the early land plant *Physcomitrium patens*. *Plant Biol J* plb.13741 (2024)
937 doi:10.1111/plb.13741.
938 68. Brückner, K. & Tissier, A. High-level diterpene production by transient
939 expression in Nicotiana benthamiana. *Plant methods* **9**, 46 (2013).
940 69. Forestier, E. C. F. *et al.* Developing a *Nicotiana benthamiana* transgenic
941 platform for high-value diterpene production and candidate gene evaluation. *Plant*
942 *Biotechnol J* **19**, 1614–1623 (2021).
943 70. Schaller, H. *et al.* Expression of the Hevea brasiliensis (H.B.K.) Mull. Arg. 3-
944 Hydroxy-3-Methylglutaryl-Coenzyme A Reductase 1 in Tobacco Results in Sterol

945 Overproduction. *Plant Physiology* **109**, 761–770 (1995).
946 71. Chin, D. J. *et al.* Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme
947 A reductase, a glycoprotein of endoplasmic reticulum. *Nature* **308**, 613–617 (1984).
948 72. Chappell, J., Wolf, F., Proulx, J., Cuellar, R. & Saunders, C. Is the Reaction
949 Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting
950 Step for Isoprenoid Biosynthesis in Plants? *Plant Physiology* **109**, 1337–1343 (1995).
951 73. Adams, R. P. *Identification of Essential Oil Components by Gas*
952 *Chromatography Mass Spectroscopy*. (Allured Publishing Corporation, Carol Stream,
953 Ill, 2007).
954 74. Picaud, S. *et al.* Amorpha-4,11-diene synthase: mechanism and
955 stereochemistry of the enzymatic cyclization of farnesyl diphosphate. *Arch Biochem*
956 *Biophys* **448**, 150–155 (2006).
957 75. Cane, D. E. 2.06 - Sesquiterpene Biosynthesis: Cyclization Mechanisms. in
958 *Comprehensive Natural Products Chemistry* (eds. Barton, S. D., Nakanishi, K. & Meth-
959 Cohn, O.) 155–200 (Pergamon, Oxford, 1999). doi:10.1016/B978-0-08-091283-
960 7.00039-4.
961 76. Cunillera, N. *et al. Arabidopsis thaliana* Contains Two Differentially Expressed
962 Farnesyl-Diphosphate Synthase Genes (∗). *Journal of Biological Chemistry* **271**,
963 7774–7780 (1996).
964 77. Grossi, V. & Rontani, J.-F. Photosensitized oxygenation of phytadienes.
965 *Tetrahedron Letters* **36**, 3141–3144 (1995).
966 78. Grossi, V. *et al.* Formation of phytadienes in the water column: myth or reality?
967 *Organic Geochemistry* **24**, 833–839 (1996).
968 79. Crook, O. M. *et al.* Inferring differential subcellular localisation in comparative
969 spatial proteomics using BANDLE. *Nat Commun* **13**, 5948 (2022).
970 80. Daudonnet, S., Karst, F. & Tourte, Y. Expression of the farnesyldiphosphate
971 synthase gene of Saccharomyces cerevisiae in tobacco. *Molecular Breeding* **3**, 137–
972 145 (1997).
973 81. Nair, P. *et al.* Differentially Expressed Genes during Contrasting Growth Stages
974 of Artemisia annua for Artemisinin Content. *PLoS ONE* **8**, e60375 (2013).
975 82. Yelina, N. E. *et al.* Streamlined regulation of chloroplast development in the
976 liverwort Marchantia polymorpha. 2023.01.23.525199 Preprint at
977 https://doi.org/10.1101/2023.01.23.525199 (2024).
978 83. Pierman, B. *et al.* Activity of the purified plant ABC transporter NtPDR1 is

979 stimulated by diterpenes and sesquiterpenes involved in constitutive and induced
980 defenses. *Journal of Biological Chemistry* **292**, 19491–19502 (2017).
981 84. Alejandro, S. *et al.* AtABCG29 Is a Monolignol Transporter Involved in Lignin
982 Biosynthesis. *Current Biology* **22**, 1207–1212 (2012).
983 85. Do, T. H. T., Martinoia, E., Lee, Y. & Hwang, J.-U. 2021 update on ATP-binding
984 cassette (ABC) transporters: how they meet the needs of plants. *Plant Physiology* **187**,
985 1876–1892 (2021).
986 86. Nagy, R. *et al.* The Arabidopsis ATP-binding cassette protein
987 AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in
988 guard cell signaling and phytate storage. *J Biol Chem* **284**, 33614–33622 (2009).
989 87. Sun, R. Biosynthesis of gibberellin-related compounds modulates far-red light
990 responses in the liverwort Marchantia polymorpha.
991 88. Delmans, M., Pollak, B. & Haseloff, J. MarpoDB: An Open Registry for
992 Marchantia Polymorpha Genetic Parts. *Plant and Cell Physiology* **58**, e5 (2017).
993 89. Patron, N. J. *et al.* Standards for plant synthetic biology: a common syntax for
994 exchange of DNA parts. *New Phytologist* **208**, 13–19 (2015).
995 90. Aach, J., Mali, P. & Church, G. M. CasFinder: Flexible algorithm for identifying
996 specific Cas9 targets in genomes. 005074 Preprint at https://doi.org/10.1101/005074
997 (2014).
998 91. Höfgen, R. & Willmitzer, L. Storage of competent cells for Agrobacterium
999 transformation. *Nucleic Acids Res* **16**, 9877 (1988).
1000 92. Littlejohn, G. R., Gouveia, J. D., Edner, C., Smirnoff, N. & Love, J.
1001 Perfluorodecalin enhances in vivo confocal microscopy resolution of Arabidopsis
1002 thaliana mesophyll. *New Phytologist* **186**, 1018–1025 (2010).
1003 93. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis.
1004 *Nat Methods* **9**, 676–682 (2012).
1005 94. Song, G.-Q., Urban, G., Ryner, J. T. & Zhong, G.-Y. Gene Editing Profiles in 94
1006 CRISPR-Cas9 Expressing T0 Transgenic Tobacco Lines Reveal High Frequencies of
1007 Chimeric Editing of the Target Gene. *Plants* **11**, 3494 (2022).
1008
1009
1010
1011
1012

Acknowledgements

This work was funded by the BBSRC NEBP Transition Award BB/W014173/1, with part support from the BBSRC/EPSRC OpenPlant Synthetic Biology Research Centre Grant BB/ L014130/1 to J.H. I.B was funded by a Herchel Smith studentship.

The authors thank Davide Annesse and Connor Tansley (University of Cambridge) for their insightful discussions and support throughout this project. We also thank Anne Osbourn and James Reed (John Innes Centre, Norwich, UK) for providing a β-amyrin standard, as well as Ian Graham and Tomasz Czechowski (University of York) for providing an amorpha-4,11-diene standard.

Author information

AUTHORS AND AFFILIATIONS

Department of Plant Sciences, University of Cambridge, Cambridge, UK.

Edith C. F. Forestier, Paola Asprilla, Facundo Romani, Ignacy Bonter, Eftychios Frangedakis and Jim Haseloff.

CONTRIBUTIONS

E.C.F.F. designed the work. F.R. provided DNA parts and early insights. E.C.F.F., P.A., I.B. and E.F. carried out the work. E.C.F.F. wrote the manuscript with input from all authors. All authors approved the final version of the manuscript.

CORRESPONDING AUTHOR.

Correspondence to Jim Haseloff.

Ethics declarations

Competing interests

The authors declare no competing interests.

	Name	Abbreviation	Putative localisation (DeepLoc 2.1)
MEP	1-Deoxy-D-xylulose-5-phosphate synthase 1	*MpDXS1*	Plastids/soluble
	1-Deoxy-D-xylulose-5-phosphate synthase 2	*MpDXS2*	
	1-Deoxy-D-xylulose 5-phosphate reductoisomerase 1	*MpDXR1*	
	1-Deoxy-D-xylulose 5-phosphate reductoisomerase 2	*MpDXR2*	
	4-Diphosphocytidyl-2-*C*-methyl-*D*-erythritol synthase	*MpMCT*	
	4-Diphosphocytidyl-2-*C*-methyl-*D*-erythritol 2-phosphate kinase	*MpCMK*	
	2-*C*-methyl-*D*-erythritol 2,4-cyclodiphosphate synthase	*MpMDS*	
	1-Hydroxy-2-methyl-2-*(E)*-butenyl 4-diphosphate synthase	*MpHDS*	
	1-Hydroxy-2-methyl-2-*(E)*-butenyl 4-diphosphate reductase	*MpHDR*	
	Isopentenyl diphosphate isomerase 1	*MpIDI1*	
	Isoprenyl (geranylgeranyl) diphosphate synthase 2	*MpIDS2/MpGGPPS*	Plastids/soluble
	Isoprenyl diphosphate synthase 3	*MpIDS3*	Plastids/soluble
	Isoprenyl (solanesyl) diphosphate synthase 4	*MpIDS4*	Plastids/soluble
	Isoprenyl diphosphate synthase 5	*MpIDS5*	Mitochondria/soluble
MVA	Acetoacetyl-coA thiolase 2	*MpACT2*	Peroxisome/soluble
	3-Hydroxy-3-methylglutaryl-coA synthase	*MpHMGS*	Cytoplasm
	3-Hydroxy-3-methylglutaryl-coA reductase	*MpHMG/MpHMGR*	ER/transmembrane
	Mevalonate-5-kinase	*MpMK*	Cytoplasm
	Phosphomevanolate kinase	*MpPMK*	Cytoplasm
	Mevanolate diphosphate decarboxylase.	*MpMVD*	Cytoplasm
	Isopentenyl diphosphate isomerase 2	*MpIDI2*	Cytoplasm
	Farnesyl pyrophosphate synthase	*MpIDS1MpFPS*	Mitochondria/soluble
	Squalene synthase 3	*MpSQS3*	ER/transmembrane
	Squalene/phytoene synthase 1	*MpSQS1*	Plastids/peripheral
	Squalene/phytoene synthase 2	*MpSQS2*	Mitochondria/peripheral

Table 1. Overview of MVA, MEP, and terpene scaffold biosynthetic genes in *Marchantia polymorpha*.
This table lists key genes studied or referenced in this work, including those involved in the methylerythritol phosphate (MEP) and mevalonate (MVA) pathways, as well as enzymes responsible for synthesizing terpene scaffolds for different sub-classes. Abbreviations correspond to gene names, and their putative protein localizations were predicted using DeepLoc 2.1. Detailed annotations and functional descriptions for each gene are provided in Table S1.

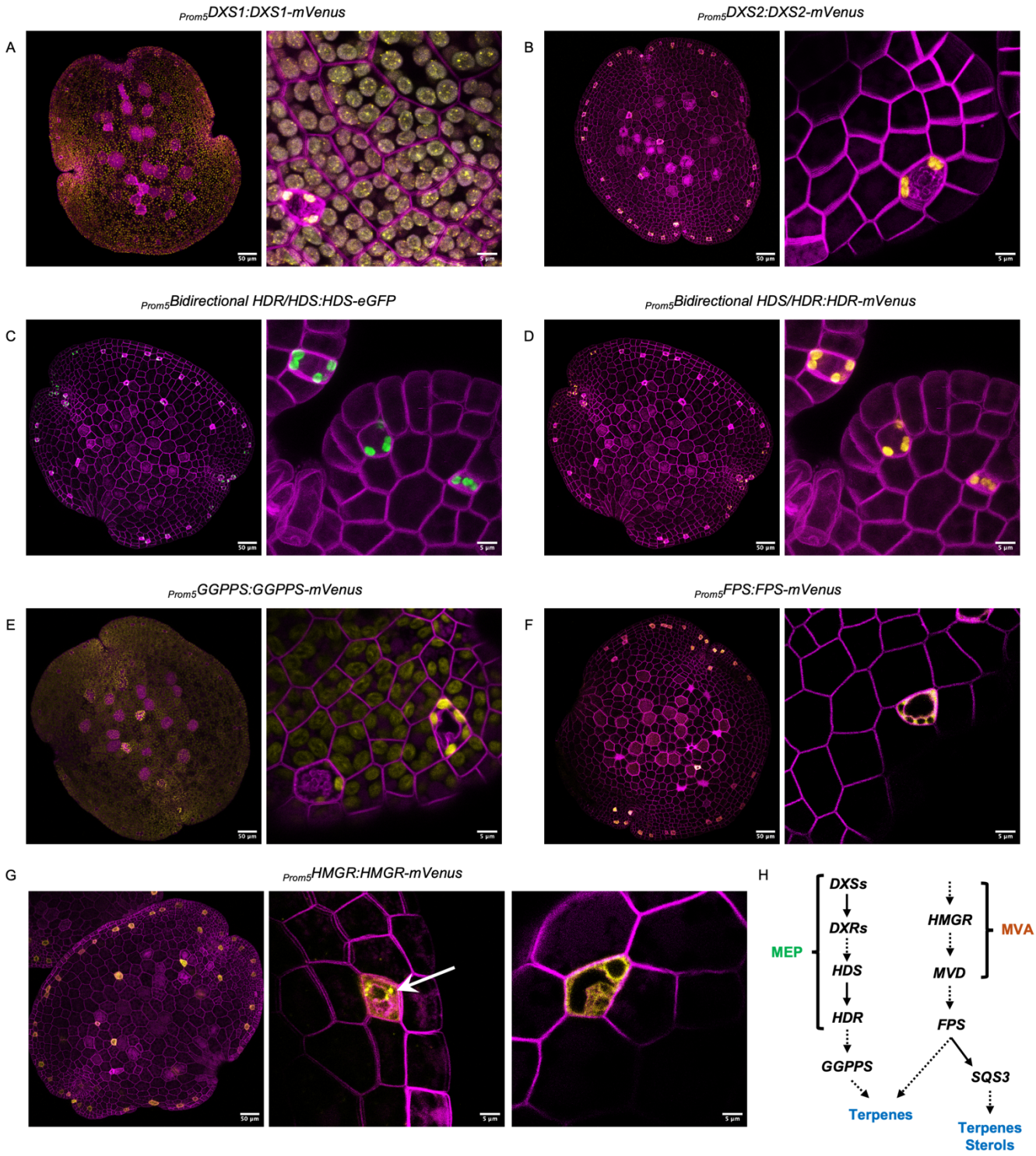


Figure 1. Confocal imaging of translational reporters for key *Marchantia polymorpha* isoprenoid biosynthetic genes.
(A) *Prom5DXS1:DXS1-mVenus*, (B) *Prom5DXS2:DXS2-mVenus*, (C) *Prom5Bidirectional HDR:HDS:HDS-eGFP*, (D) *Prom5Bidirectional HDS:HDR:HDR-mVenus*, (E) *Prom5GGPPS:GGPPS-mVenus*, (F) *Prom5FPS:FPS-mVenus*, and (G) *Prom5HMGR:HMGR-mVenus*. Each construct is represented by two images: the full gemmae at day 0 (left panels, scale bar: 50 µm) and a higher magnification view showing subcellular localization (right panels, scale bar: 5 µm). (H) Simplified biosynthetic pathway scheme highlighting enzymatic steps analyzed in this study. The *mVenus* signal (yellow) or eGFP (green) for HDS (C) indicates the subcellular localization of the corresponding enzymes, while *mScarlet* (purple) delineates the cellular boundaries. For the magnified image in (A), chloroplast/plastid autofluorescence (grey) is included to demonstrate sublocalization of DXS1 in the plastids. Panel (G) includes one whole-gemma image and two high magnification images to illustrate the localization of HMGR-mVenus to the ER and Golgi apparatus. White arrow shows the localizations to Golgi.

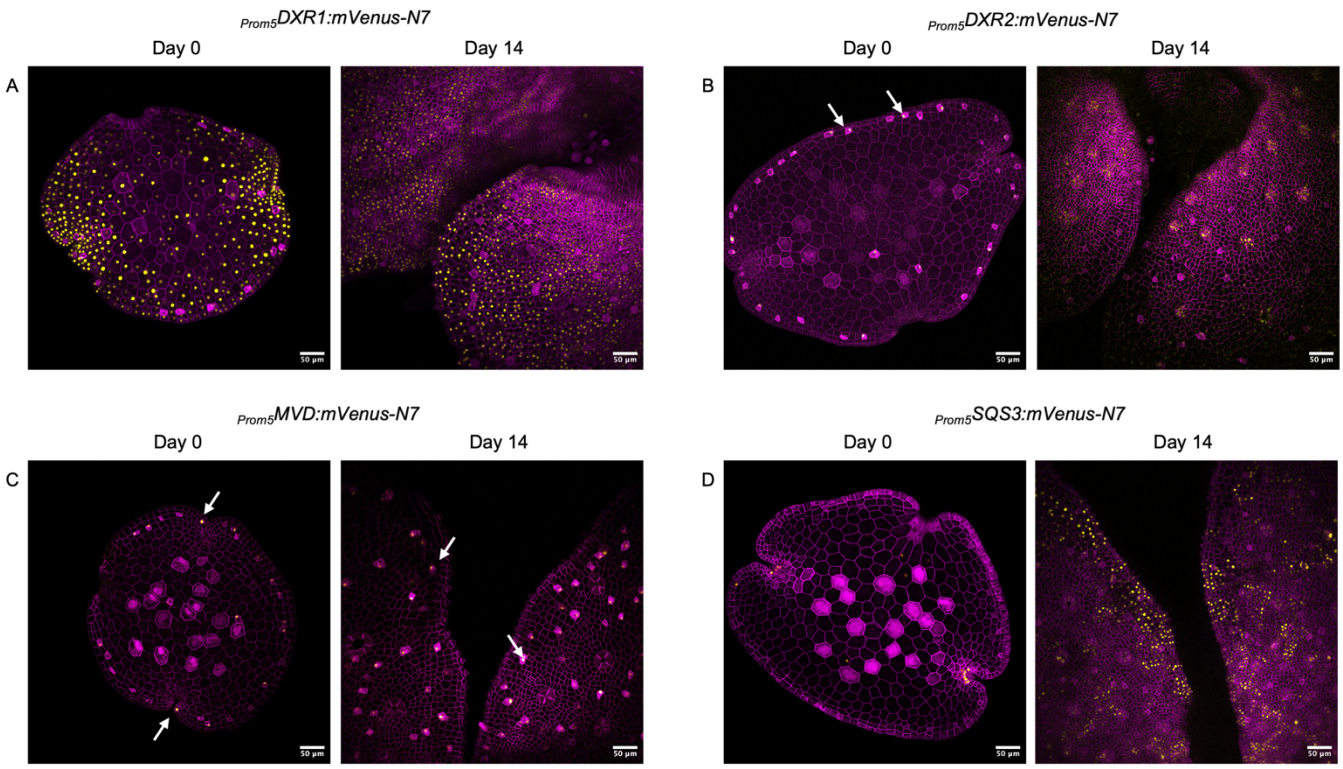


Figure 2. Confocal imaging of transcriptional reporters for additional isoprenoid biosynthetic genes in *Marchantia polymorpha.*

(A) $_{Prom5}$*DXR1:mVenus-N7*, (B) $_{Prom5}$*DXR2:mVenus-N7*, (C) $_{Prom5}$*MVD:mVenus-N7*, and (D) $_{Prom5}$*SQS3:mVenus-N7*. The mVenus signal (yellow) localized to the nucleus indicates the activation of these promoters in specific cell types, while mScarlet fluorescence (purple) delineates cellular boundaries. Each construct is represented by two panels: day 0 gemmae (left, scale bar: 50 µm) and the meristem area of 14-day-old plants (right, scale bar: 50 µm). White arrows indicate mVenus signals detected in oil body cells.

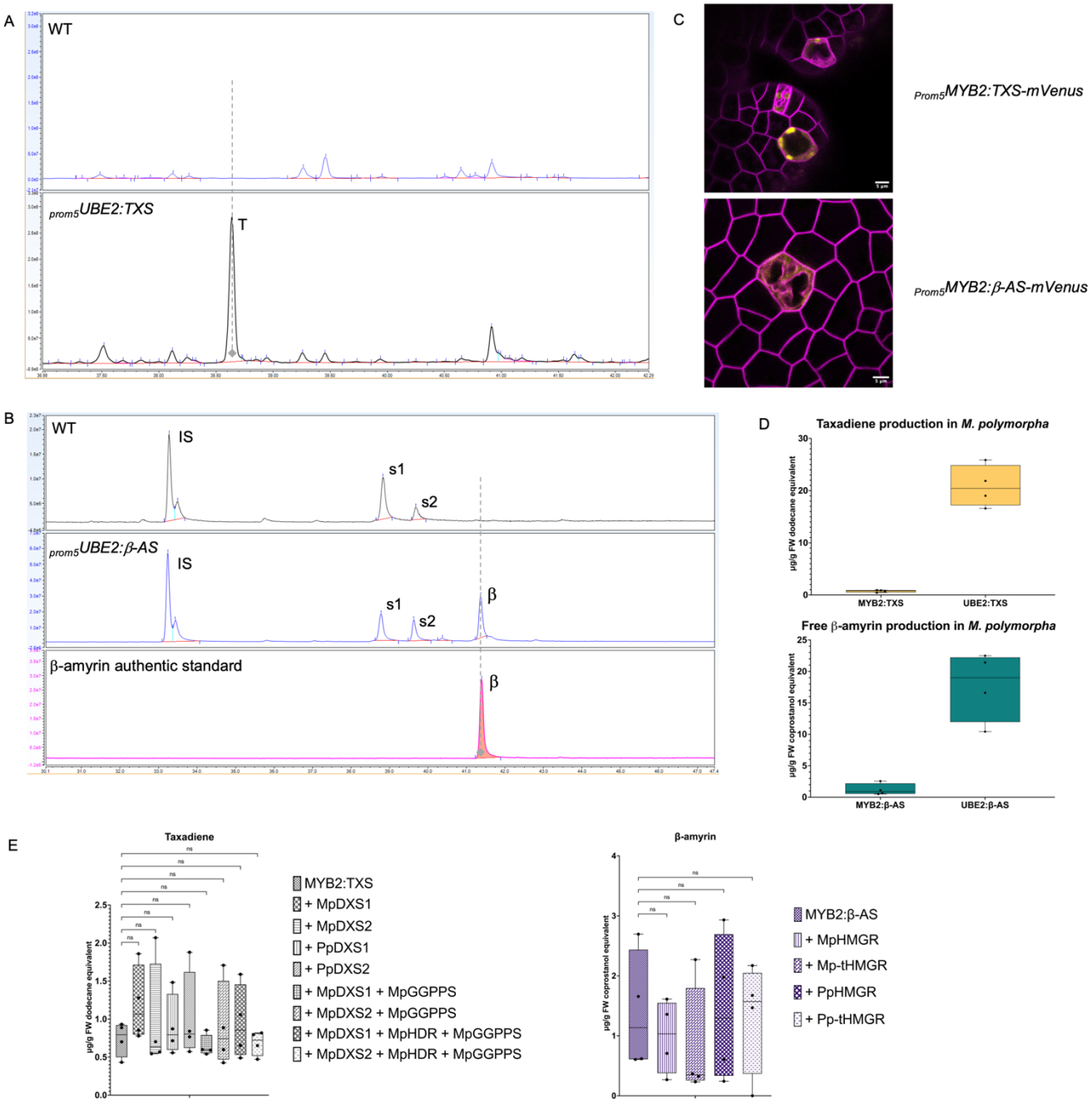


Figure 3. Production of taxadiene and β-amyrin in *Marchantia polymorpha* and subcellular localization of their synthases.
(A) Total ion chromatograms (TICs) of terpene extracts from wild-type (WT, upper panel) and transgenic plants expressing *Prom5UBE2:TXS* (*taxadiene synthase*, lower panel), highlighting the taxadiene peak (T). (B) TICs of trimethylsilyl-derivatized terpene extracts, showing WT (upper), a transgenic line expressing *Prom5UBE2:β-AS* (*β-amyrin synthase,* middle), and an authentic β-amyrin standard (lower). Peaks: IS (internal standard: coprostan-3-ol), s1 (campesterol), s2 (stigmasterol), and β (β-amyrin).
(C) Subcellular localization of TXS and β-AS fused to mVenus under the *Prom5MYB2* promoter, showing plastid localization for TXS (upper panel) and ER or cytosol localization for β-AS (lower panel). mScarlet fluorescence (purple) delineates cellular boundaries. Scale bar: 5 µm. (D) Quantification of taxadiene (upper graph) and β-amyrin (lower graph) in plants expressing *Prom5UBE2:TXS* and *Prom5UBE2:β-AS* respectively (whole plant expression) or *Prom5MYB2:TXS* and *Prom5MYB2:β-AS* respectively (oil body cell-specific expression). Data represent four independent lines per condition.
(E) Quantification of taxadiene (left) and β-amyrin (right) in transgenic lines over-expressing precursor genes. For taxadiene, combinations included *MpDXS1*, *MpDXS2*, *PpDXS1*, *PpDXS2*, *MpHDR*, and *MpGGPPS*. For β-amyrin, *MpHMGR*, *Mp-tHMGR*, *PpHMGR*, and *Pp-tHMGR* were tested. Box plots represent data points from independent transformants. One-way ANOVA with Dunnett's multiple comparison tests ($p < 0.05$) showed no significant differences across gene combinations.

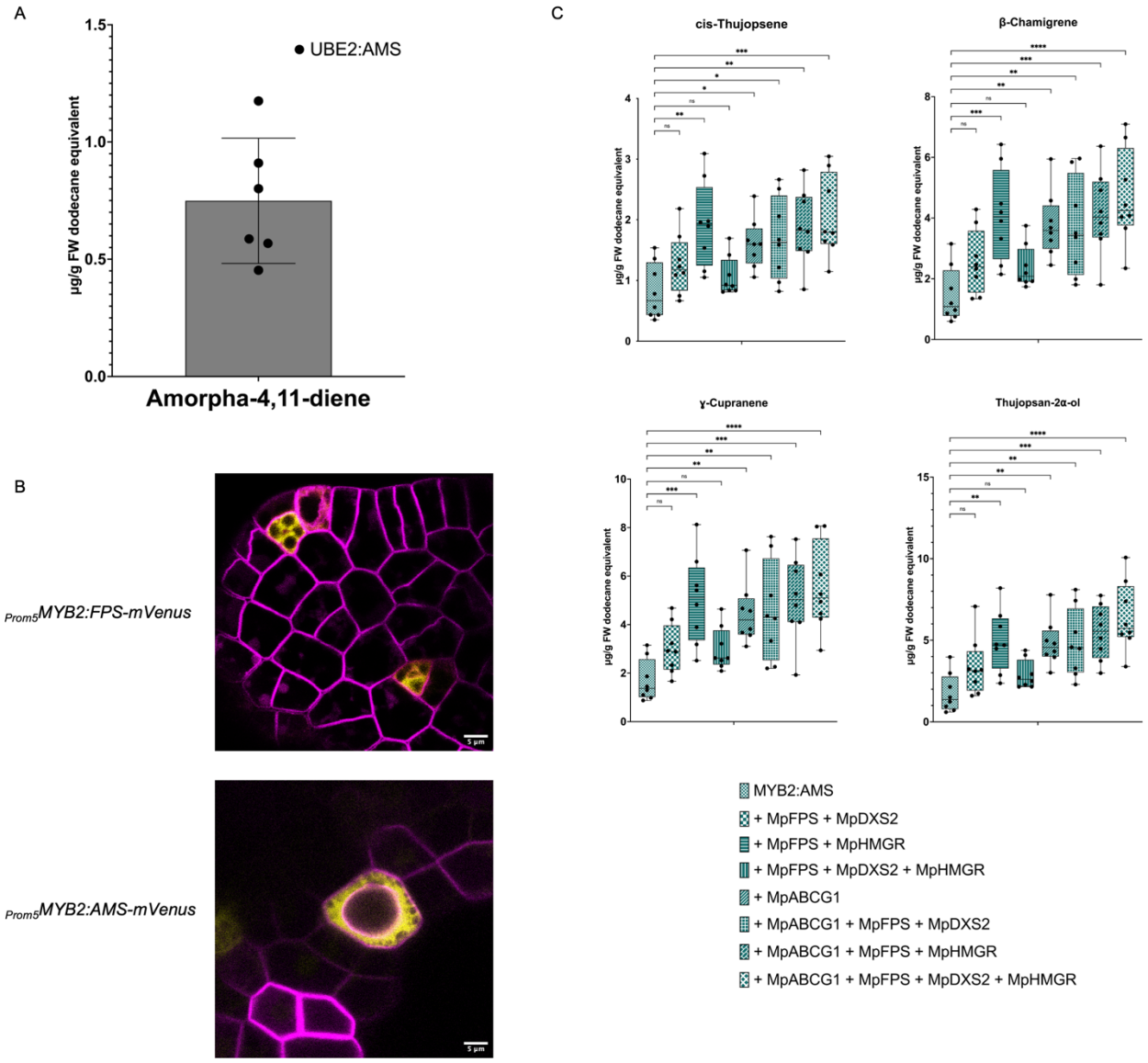
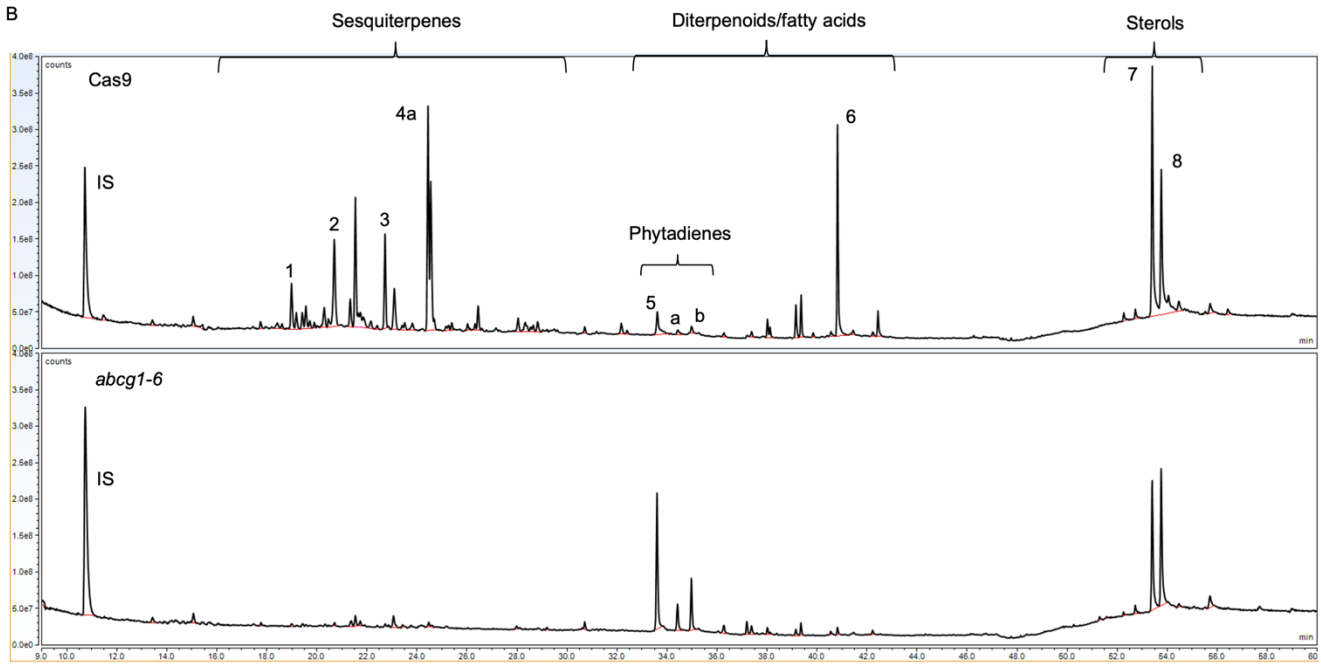


Figure 4. Expression of amorphadiene synthase (*AMS*) in *Marchantia polymorpha* with or without precursor supply genes and the ABCG1 transporter, and its effect on endogenous sesquiterpene levels.
(A) Quantification of amorpha-4,11-diene levels in plants expressing *Artemisia annua* AMS under the *UBE2* promoter. The compound was detected in six independent transformants (mean ± SD; n = 6). (B) Subcellular localization of FPS and AMS proteins when expressed under the oil-body-specific *MYB2* promoter. mVenus fluorescence (yellow) indicates protein localization, with signals observed in oil body cells, while mScarlet fluorescence (purple) marks cell boundaries. Scale bars: 5 µm. (C) Quantification of endogenous sesquiterpenes, including cis-thujopsene, β-chamigrene, γ-cuprenene, and thujopsan-2α-ol, in lines expressing *AMS* with or without precursor supply genes (*MpFPS*, *MpDXS2*, and/or *MpHMGR*) and with or without *ABCG1*, under $_{Prom5}$*MYB2*. Box plots represent data points from eight independent transformants (n = 8), with significance determined by one-way ANOVA followed by Dunnett's multiple comparison test ($p < 0.05$).

A

WT *ABCG1* 301- CAGATGAACCCAGAGGAGGATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_2 301- CAGATGAACCCAG __GAGGATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_5 301- CAGATGAACCCAGTCCCACTATGCAGATACGTGAGTATATGTCGGATATGTACCACTATGCAGAGGAGGATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_6 301- CAGATGAACCCAG __GAGGATAATGAGAAGTTTCTCGTAAGAGAACCTGCTTCATTATTTATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_7 301- CAGATGAACCTTCTCCATTTCAAGATACCTTTTTCATGTTATTGAATTGGTGAGTTCAATGAATTGGTTAATTTCAGGAGGATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_10 301- CAGATGAAC ____AGGAGGATAATGAGAAGTTTCTCGCCAAGCTAA
abcg1_11 301- CAGATGAACTGGGTTTCTCGCCAAGCTAAGATCCCATGATGAGAAGTTTCTCGCCAAGCTAA



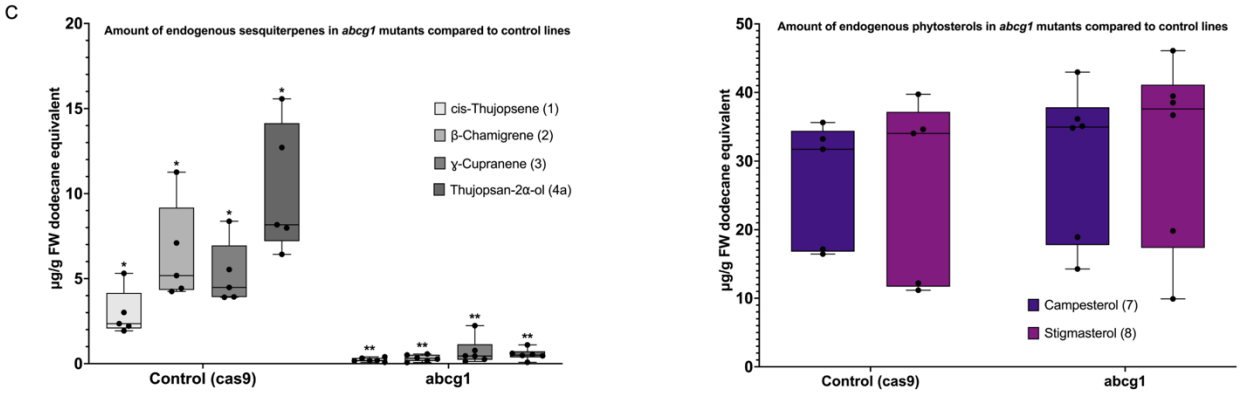


Figure 5. Analysis of endogenous terpene profiles in CRISPR-generated abcg1 mutants compared to Cas9 controls.
(A) DNA sequence alignment of the *abcg1* genomic region in wild-type (WT) and CRISPR-generated *abcg1* mutant lines. Dashes indicate deletions, red sequences represent insertions, and purple nucleotides correspond to premature stop codons introduced by the deletions and/or insertions. (B) TICs of terpene extracts from a Cas9 control (upper chromatogram) and abcg1 mutant line (lower chromatogram). Peaks corresponding to quantified compounds are labeled: sesquiterpenes (1: cis-thujopsene, 2: β-chamigrene, 3: γ-cuprenene, 4a: thujopsan-2α-ol), fatty acids/diterpenoids (5: neophytadiene, 6: phytol, a: (*Z*)-1,3-phytadiene, b: (*E*)-1,3-phytadiene), and phytosterols (7: campesterol, 8: stigmasterol). (C) Quantification of endogenous terpene levels in *abcg1* mutants (n=6) compared to Cas9 controls (n=5). The left graph shows the amounts of the four major sesquiterpenes (1 to 4a) labelled in panel B. The right graph displays the levels of phytosterol (7 and 8) labelled in panel B. Box plots display individual data points, with bars representing the median and interquartile range. Statistical significances are denoted by asterisks (one-way ANOVA with Dunnett's multiple comparison tests ($p < 0.05$)).

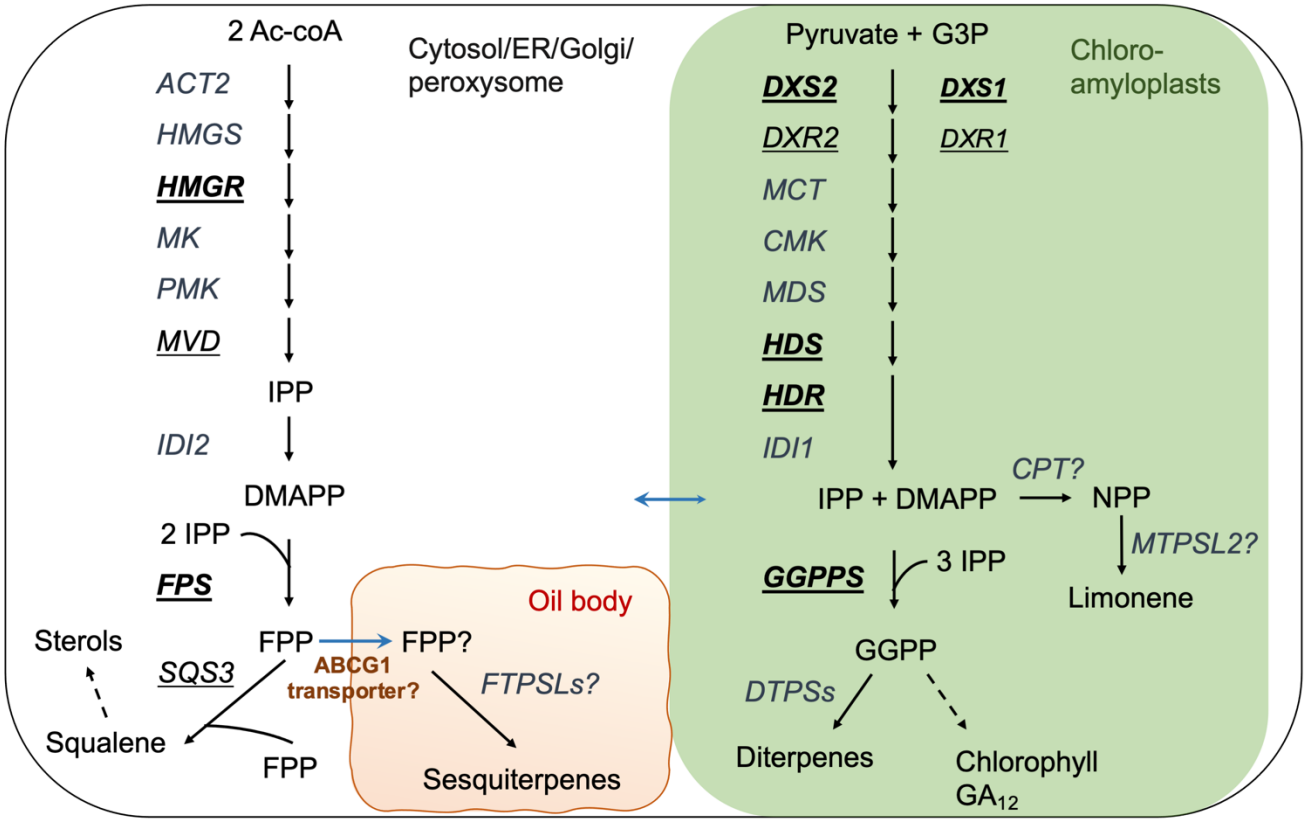


Figure 6. Proposed overview of the isoprenoid precursor pathways in *Marchantia polymorpha* oil body cells based on previous works and findings from this study.
Biosynthetic scheme illustrating the mevalonate (MVA) and methylerythritol phosphate (MEP) pathways in *Marchantia polymorpha*. Enzymes shown in bold and underlined were localized using translational reporters confirming their subcellular localization. Enzymes shown in underlined text only were analyzed with promoter expression constructs, indicating activity in oil body cells. Key precursor flows toward sterols, diterpenes, and other metabolites are also indicated. Details of enzymatic steps and annotations can be found in Table 1 and Table S1. Abbreviations: Ac-CoA, acetyl coenzyme A; G3P, glyceraldehyde-3-phosphate; CPT, cis-prenyl transferase; MTPSL2, microbial terpene synthase like 2, functionally characterised in yeast as a limonene synthase (Kumar et al.).