Sub-functionalization and epigenetic regulation of a biosynthetic gene cluster in *Solanaceae*

Santiago Priego-Cubero[1,*], Eva Knoch[1,*], Zhidan Wang[2], Saleh Alseekh[3,4], Karl-Heinz Braun[1], Philipp Chapman[1], Alisdair Fernie[3,4], Chang Liu[2], Claude Becker[1,‡]

[1] Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany

[2] Department of Epigenetics, Institute of Biology, University of Hohenheim, 70599 Stuttgart, Germany

[3] Max-Planck-Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany

[4] Center of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria

[*] These authors contributed equally to this work.

[‡] To whom correspondence should be addressed: claude.becker@lmu.de

ORCIDs

Santiago Priego Cubero: 0009-0006-5209-7942

Eva Knoch: 0000-0001-5316-9174

Zhidan Wang: 0000-0001-8893-6732

Saleh Alseekh: 0000-0003-2067-5235

Chang Liu: 0000-0003-2859-4288

Claude Becker: 0000-0003-3406-4670

Abstract

Biosynthetic gene clusters (BGCs) are sets of often heterologous genes that are genetically and functionally linked. Among eukaryotes, BGCs are most common in plants and fungi, where they ensure the co-expression of the different enzymes coordinating the biosynthesis of specialised metabolites. Here, we report on a newly identified withanolide BGC in *Physalis grisea* (ground-cherry), a member of the nightshade family (*Solanaceae*). A combination of transcriptomic, epigenomic and metabolic analyses, revealed that, following a duplication event, this BGC evolved two tissue-specifically expressed sub-clusters, containing eight pairs of paralogs that contribute to related but distinct biochemical processes; this sub-functionalisation is tightly associated with epigenetic features and the local chromatin environment. The two sub-clusters appear strictly isolated from each other at the chromatin structural level, each forming a highly self-interacting chromatin domain with tissue-dependent levels of condensation. This correlates with gene expression exclusively in either above- or below-ground tissue, thus spatially separating the production of different withanolide compounds. By comparative phylogenomic analysis, we show that the withanolide BGC most likely evolved before the diversification of the *Solanaceae* family and underwent lineage-specific diversifications and losses. The tissue-specific sub-functionalisation is common to species of the *Physalideae* tribe but distinct from other, independent duplication events outside of this clade. In sum, our study reports on the first instance of an epigenetically modulated sub-functionalisation within a BGC and sheds light on the biosynthesis of withanolides, a highly diverse group of steroidal triterpenoids important in plant defence and amenable to pharmaceutical applications due to their anti-inflammatory, antibiotic and anti-cancer properties.

Introduction

Plants produce a plethora of specialised metabolites that typically serve as protectants, repellents, attractants, or signalling molecules in ecological interactions. Some have growth-stimulating or -inhibitory properties. Among the 200,000 to one million different compounds that plants are estimated to produce, triterpenoids are one of the largest and most diversified groups (Cárdenas et al. 2019). Withanolides, in turn, are highly diverse steroidal triterpenoids that are prominent in *Solanaceae* but have also been detected in some species outside the nightshade family (e.g. Yang et al. (2011)). The name-giving plant species, *Withania somnifera* (commonly referred to by its Sanskrit name “ashwagandha”), has been used as a medicinal plant in traditional Indian medicine for several thousand years; modern pharmaceutical research is investigating several withanolides for their anti-inflammatory properties and their potential use in treatments of cancer and autoimmune diseases (Huang et al. 2020; Zhang et al. 2023; Cigler et al. 2024). Besides in *Withania*, withanolide analogues occur in many but not all *Solanaceae*; they are present in different *Physalis* and *Petunia* species but are notably absent in the majority of *Solanum* species such as tomato and potato. Currently, more than 600 compounds belonging to this class are known, differing in the configuration of their side chains and the positioning of double bonds in their structure (Dhar et al. 2015; Xia et al. 2022). This diversity makes withanolides, which act as herbivore repellents and antibiotics in the plant, a promising source for novel pharmaceutical agents.

The biosynthesis of withanolides branches off the phytosterol pathway. The first dedicated step towards withanolides is mediated by a sterol Δ^{24}-isomerase (24ISO) that catalyses the isomerization from 24-methylenecholesterol to 24-methyldesmosterol (Knoch et al. 2018). 24ISO is a homolog of DWARF1 (DWF1), with two paralogs, SSR1 and SSR2 (for Sterol Side-chain Reductase), in *Solanum lycopersicum* (tomato) and *S. tuberosum* (potato). SSR1 catalyses the synthesis of campesterol from 24-methylenecholesterol, an intermediate in brassinolide biosynthesis, while SSR2 plays a role in the production of cholesterol and hence steroidal glycoalkaloids (Sawai et al. 2014; Akiyama et al. 2023).

While 24ISO is the starting point of the withanolide biosynthesis pathway, several enzymatic steps, particularly those typically involving cytochrome P450s (CYP450s) and 2-oxoglutarate-dependent dioxygenases (2ODDs), have been hypothesised as being necessary to complete the biosynthesis from 24-methyldesmosterol to the withanolides (Knoch et al. 2018). Potential gene clusters containing genes encoding CYP450s and 2ODDs and located in the same genomic region as *24ISO* were identified in related solanaceous plants, suggesting a larger withanolide biosynthetic gene cluster (BGC) (Knoch et al. 2018). Several BGCs of plant specialised metabolites have been identified in recent years, e.g. encoding the biosynthesis

of diterpenes in rice and barley, or benzoxazinoids in maize and other grasses (Polturak and Osbourn 2021; Polturak et al. 2022). A BGC encoding withanolide production, although hypothesised in the past, has not been identified to date.

Here, we present the discovery of a withanolide BGC in the *Solanaceae* species *Physalis grisea*, a close relative of *W. somnifera*. Our data show that the genes in the cluster play an essential role in the synthesis of withanolides. The cluster architecture revealed a tandem duplication, resulting in a sub-functionalization into two distinct sub-clusters, each containing a set of core genes required for withanolide biosynthesis. Through the combined analysis of transcriptomic, epigenomic and metabolic datasets, we show that sub-cluster expression and withanolide formation are tissue-specific, and that sub-cluster differentiation is associated with tissue-specific epigenetic configuration and higher-order chromatin organisation. Through phylogenomic analysis, we show that this sub-functionalisation is not specific to *P. grisea* but is common to the *Physalideae* tribe. In sum, our study uncovers how biosynthesis of plant specialised metabolites can diversify via duplication and differential organisation of the chromatin environment.

Results

The putative withanolide biosynthetic gene cluster in P. grisea underwent tandem duplication

At the time Knoch *et al.* identified 24ISO as the enzyme catalysing the first dedicated biosynthesis step in withanolide production, only a few *Solanaceae* genomes had been assembled and annotated. Nonetheless, the authors had already noticed that the gene encoding the 24ISO ortholog in *Capsicum annuum*, *Solanum melongena* (in which it appeared pseudogenic) and *Petunia inflata* was flanked by genes encoding CYP450s (CYP88 subfamily, among others) and 2-oxoglutarate dependent (2-ODDs) dioxygenases that might be involved in the metabolization of the withanolide-intermediate 24-methyldesmosterol (Knoch et al. 2018). We set out to study the possibility that withanolide biosynthesis was organised in a BGC. In absence of a *W. somnifera* genome assembly, we referred to the most closely related available assembly and gene annotation, that of *P. grisea* (He et al. 2023; Huang et al. 2023). Like *W. somnifera*, *Physalis* spp. belong to the *Physalideae* tribe. Its members produce withanolide analogs, known as physalins (Huang et al. 2020). Given this close phylogenetic and biochemical relationship between the genera *Physalis* and *Withania*, we hypothesised that the *P. grisea* genome could harbour orthologs of *Ws24ISO* and could serve as a proxy to study putative BGCs, particularly since *P. grisea* is currently being established as a new genetic model in the *Solanaceae* family (Van Eck 2022; Dale et al. 2024).

115 Using the *W. somnifera* 24ISO amino acid sequence as a query for BLAST, we identified two
116 *24ISO* copies in the *P. grisea* genome, localised at the end of chromosome 1 and separated
117 by ~350 kb (**Figure 1A**). Closer inspection of this genomic region revealed aggregation of
118 several genes encoding enzymes putatively involved in the biosynthesis of secondary
119 metabolites (**Figure 1A**). In line with the observations by Knoch et al. (2018) in *C. annuum*, *S.*
120 *melongena* and *P. inflata*, we found nine genes that encoded CYP450s belonging to different
121 families and subfamilies (*CYP87G1*, *CYP749B2*, *CYP88C8*, *CYP88C7* and *CYP88C10*), and
122 six *2-ODD* genes surrounding the *24ISO* orthologs, suggesting the existence of a putative
123 withanolide BGC in *P. grisea*. Additionally, we also identified two genes (one of them
124 appearing pseudogenic) encoding Short-Chain Dehydrogenase/Reductases (SDRs), two for
125 Sulfotransferases (Sulf), and one encoding an acylsugar acyltransferase (ASAT) (**Figure 1A**).
126 We noticed that eight genes appeared to be duplicated within a 376 kb region, such that
127 overall, the locus exhibited signals of a tandem duplication that had resulted in two groups of
128 genes. Remarkably, each group encoded core paralogs of a putative withanolide biosynthetic
129 pathway (24ISOa and b, respectively, CYP87G1a/b, CYP749B2a/b, CYP88C7a/b,
130 CYP88C10a/b, 2ODD3a/b, Sulfa/b, and SDRa/b), and was separated from the other group by
131 an intermediate gene-depleted region of approximately 50 kb in length (**Figure 1A**). Sequence
132 analysis of the respective gene and protein copies corroborated the notion that these gene
133 pairs were paralogs that most likely had arisen from a duplication, with pairwise nucleotide
134 and amino acid sequence identities ranging between 84.4-94.6% and 80-95%, respectively
135 (**Supplementary Table 1**).

136 In sum, our data showed that *P. grisea* has two *24ISO* paralogs that are embedded in a
137 putative withanolide BGC and which appeared to have undergone a partial tandem
138 duplication. This raised the questions whether this cluster architecture was specific to *P. grisea*
139 or widespread in the *Solanaceae* family, and whether the tandem duplication had any
140 functional consequences.

141 *Withanolide production is organised by two differentially regulated sub-clusters*

142 A key feature of BGCs is the co-regulation of expression of the genes that form the cluster,
143 which is thought to ensure the faithful spatio-temporal production of a given metabolite, also
144 in response to different elicitors (Nützmann and Osbourn 2015). To test whether the members
145 of the putative withanolide BGC were co-regulated, we performed whole-genome
146 transcriptomic analysis via mRNA sequencing (RNA-seq) across two sets of samples
147 (**Supplementary Dataset 1**). The first consisted of nine different tissues and developmental
148 stages: shoot and root tissue of two-week-old seedlings; young leaf, mature leaf, flower, stem,
149 petiole, calyx, and fruit (**Supplementary Figure 1**). For the second set, we extracted RNA
150 from roots and shoots of two-week-old seedlings treated with methyl jasmonate (MeJA), which

is known to trigger withanolide production in *W. somnifera* (Rana et al. 2014). We further included treatments with the stress hormone abscisic acid (ABA), which induces expression of the withanolide analogue transport protein PDR2 in *Petunia* (Sasse et al. 2016), and the bacterial elicitor flagellin 22 (Flg22) to explore their respective impact on the gene cluster expression.

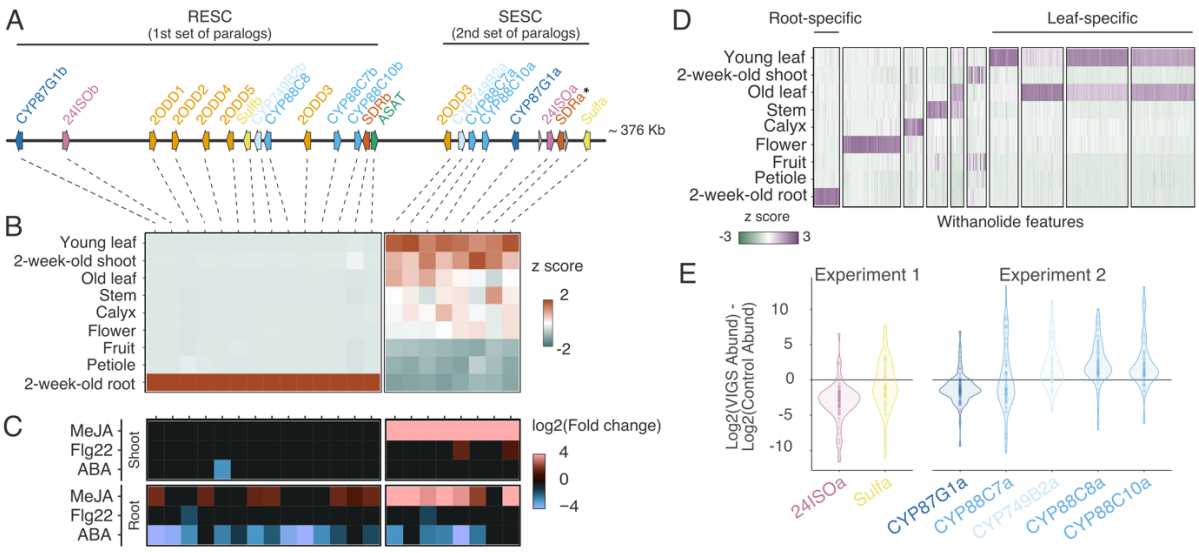


Figure 1. The withanolide biosynthetic gene cluster is organised into two sub-clusters. A) Genomic organisation of the withanolide BGC on *P. grisea* chromosome 1; distances between genes are drawn to scale; arrows indicate the sense of transcription. RESC and SESC: root- and shoot-expressed sub-clusters, respectively. Asterisk indicates a paralog that appears to be pseudogenic. **B)** Heatmap showing the relative expression level of each gene within the cluster across different tissues. Expression levels are presented as z-scores of transcripts per million (TPM), relative to the mean for each gene. **C)** Heatmap showing the expression of each gene within the cluster in response to the putative elicitors methyl jasmonate (MeJA), flagellin 22 (Flg22), and abscisic acid (ABA). Differentially expressed genes were defined as those with $\log_2$(fold-change) >1 or <(-1) and an adjusted p-value <0.01. Log_2(fold-change) values between -1 and 1 were set to black in the colour key. Differential expression was calculated independently for each tissue using the respective mock treatment as the reference. **D)** Relative abundance of withanolides across the same tissues as in (B), expressed as the z-score of the intensity values relative to the mean for each compound. **E**) Relative abundance of the leaf-specific withanolides in (D) after VIGS of different genes from the shoot-expressed sub-cluster.

We combined the data from both RNA-seq experiments in a co-expression analysis using weighted gene correlation network analysis (WGCNA) (Zhang and Horvath 2005; Langfelder and Horvath 2008; Russo et al. 2018) (**Supplementary Figure 2**). Instead of the expected co-regulation of all the genes in the putative BGC, the two groups of paralogs showed distinct and contrasting tissue-specific expression: the first 14 collinear genes, containing the first set of core paralogs in the BGC (**Figure 1A**), were exclusively expressed in root (expression module M2) (**Supplementary Figure 2D**). In contrast, the second set of eight paralogs were all placed in module M1, which grouped genes with high expression in shoot and aerial tissues (**Figure 1B; Supplementary Figure 2D**). Based on these expression profiles, we named these sub-clusters 'root-expressed sub-cluster' (RESC) and 'shoot-expressed sub-cluster' (SESC), respectively.

181 In the treatment data, the expression of each sub-cluster was elicited by MeJA. Genes from
182 the SESC responded in both shoot and root tissue, while the RESC genes responded
183 exclusively in root tissue, indicating that the expression of the RESC is tightly repressed in
184 shoot tissue. In contrast, ABA repressed the expression of both the RESC and the SESC in
185 roots (**Figure 1C; Supplementary Figure 2F**); neither sub-cluster responded to flg22. Our
186 data indicate that the tandem duplication resulted in contrasting expression patterns of the two
187 sub-clusters, which seem to be regulated tissue-specifically not only in their respective
188 constitutive expression but also in their response to external stimuli.

189 Duplication of enzymes in specialised metabolism often leads to neofunctionalization and, in
190 turn, the production of novel metabolites (Smit and Lichman 2022). We hypothesised that the
191 sub-division and tissue-specific regulation of expression could indicate the sub-
192 functionalization of the putative withanolide BGC and result in qualitative differences in
193 withanolide production in the respective tissues. To test this hypothesis, we performed
194 untargeted metabolomics on the tissue samples used for the RNA-seq. After several filtering
195 steps to select features potentially derived from withanolides in the MS/MS spectrum, we
196 compiled a list of withanolide-associated features (see *Material and Methods* for details). We
197 observed a distinct pattern of tissue-specific accumulation of different withanolide-related
198 features in roots and diverse aerial tissues (**Figure 1D**). To further confirm the role of the BGC
199 in withanolide biosynthesis, we employed virus-induced gene silencing (VIGS). Since VIGS
200 can only be applied in leaves but not in roots, we targeted genes from the SESC. The silencing
201 of *24ISOa*, *CYP87G1a* and *Sulfa* led to an overall decrease in the abundance of leaf-specific
202 withanolide features (**Figure 1E; Supplemental Figure 3**). Silencing of *CYP88C7a* had
203 similar effects, but we also noted higher levels for some features, which potentially reflects an
204 accumulation of certain intermediates (**Figure 1E**). Lastly, silencing of *CYP749B2a*,
205 *CYP88C8a* and *CYP88C10a* led to the reduction of few withanolide features and the general
206 accumulation of other, potentially intermediate, products (**Figure 1E**).

207 Taken together, these results provide support for the existence of two functionally distinct sub-
208 clusters that produce a distinct array of withanolides in a tissue-specific manner.

209 *The withanolide sub-clusters are functionally isolated from each other and display*
210 *distinct levels of tissue-specific chromatin compaction*

211 The presumed tandem duplication resulted in two neighbouring gene clusters with highly
212 distinct expression profiles, raising the question how such a functional division could be
213 organised and maintained at the regulatory level. Given the size of the sub-clusters and the
214 number of genes (14 and 8, respectively) that were coordinately expressed, we hypothesised
215 that canonical transcription-factor-mediated regulation might insufficiently explain the

216 observed pattern. Several recent studies have highlighted that local chromatin accessibility
217 and higher-order organisation at the three-dimensional (3D) level can play an important role
218 in the coordinated regulation of gene expression of BGCs in plants (Nützmann et al. 2020;
219 Yang et al. 2021; Li et al. 2023). BGCs are located within topologically associated domains
220 (TADs), self-interacting chromatin regions with suppressed interaction with neighbouring
221 chromatin (Doğan and Liu 2018), or within regions with TAD-like features (Nützmann et al.
222 2020). This topological organisation facilitates precise spatio-temporal co-expression of the
223 genes within the cluster, enabling long-range interactions between genes and between genes
224 and their *cis*-regulatory elements (Zhao et al. 2022; Li et al. 2023).

225 To test whether the distinct expression patterns of the RESC and SESC are regulated at the
226 chromatin level, we first performed Assay for Transposase-Accessible Chromatin using
227 sequencing (ATAC-seq) of *P. grisea* root and leaf tissue, in four biological replicates per tissue
228 (**Supplementary Figures 4-6**; **Supplementary Dataset 2**). Our data clearly showed that
229 local chromatin accessibility in the RESC was high in root tissue, but that the chromatin in the
230 same sub-cluster was inaccessible in leaves (**Figure 2A and 2B**). We observed the opposite
231 for the SESC: while the ATAC-seq showed many open chromatin regions in leaf tissue, the
232 same region was less accessible in roots. Together, these results corroborated our hypothesis
233 that the chromatin landscape differed between tissues; in general, each sub-cluster was more
234 accessible in the tissue in which it was expressed. Additionally, we observed that most of the
235 peaks in both tissues and sub-clusters were located in intergenic regions, suggesting a role
236 of distal *cis*-regulatory elements and long-range chromatin contacts to the tissue-specific
237 regulation of each sub-cluster (**Figure 2C**).

238 To gain further insights into the chromatin configuration at this genomic locus, we captured
239 whole-genome chromatin-chromatin interactions in *P. grisea* roots and leaves using Hi-C
240 (Lieberman-Aiden et al. 2009) (**Supplementary Dataset 3**). We hypothesised that each
241 subcluster would exhibit distinct patterns of chromatin interactions in a tissue-specific manner.
242 We obtained a total of ~103 million valid unique read pairs, combining data from two
243 independent biological replicates for each tissue (see *Material and Methods* for details;
244 **Supplementary Figure 7A**). We first examined the genome packaging of *P. grisea* at the
245 chromosomal level with a Hi-C contact map at 1 Mb resolution (**Supplementary Figure 7B**).

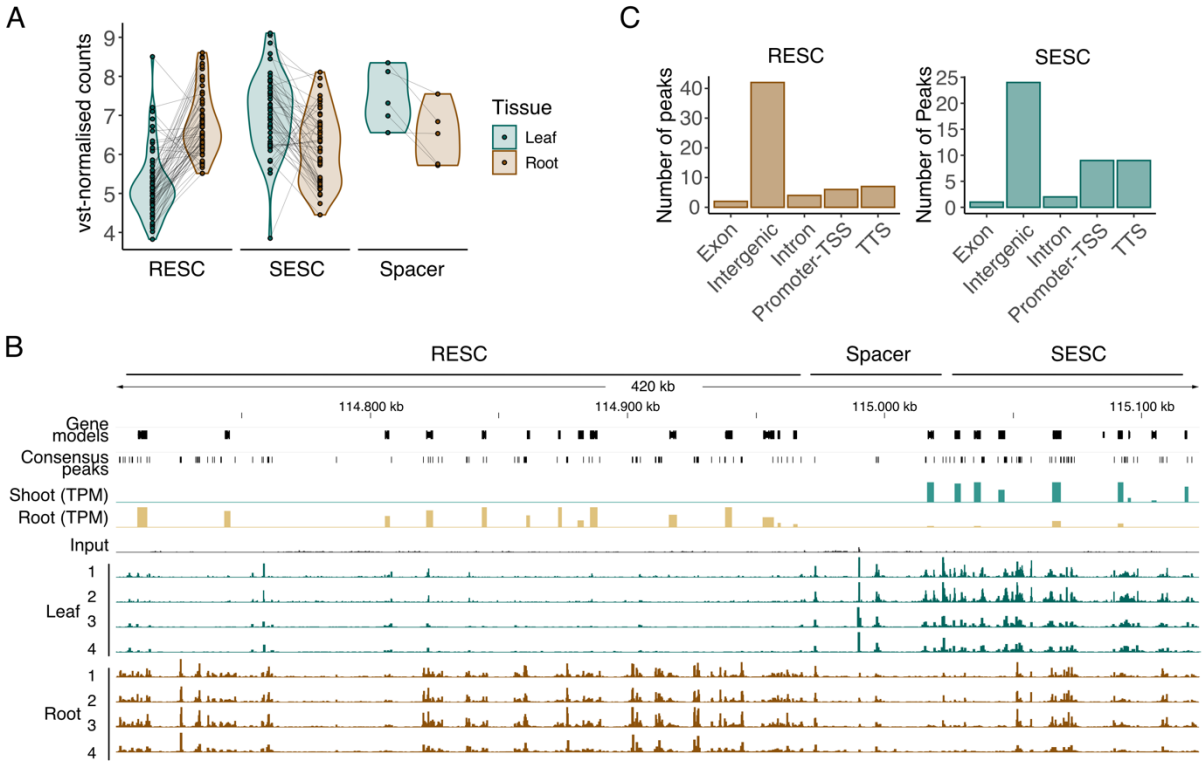


Figure 2. Tissue-specific local chromatin accessibility at the locus of the withanolide BGC by ATAC-seq. **A)** Normalised read counts after variance-stabilising transformation using the DESeq2 package (Love et al. 2014) of the consensus narrow peaks called on the Tn5-sensitive regions within each subcluster for a combination of four replicates per tissue. Black lines connect the value of each narrow peak in leaf and root for each subcluster. **B)** Tn5 integration footprint in the withanolide BGC region. From top to bottom: gene models, consensus narrow peaks, expression (TPM: transcript per million; scale was set to 0-400 for both tracks) of 2-week-old shoot and root tissue. ATAC-seq profiles are represented as reads per million (RPM; scale was set to 0-1 for all tracks) for input, leaf (four replicates) and root (four replicates) samples. **C)** Number of narrow peaks in the root and shoot subclusters, separately called for root and leaf samples, respectively, and their location with respect to the genes on each subcluster. Promoter-TSS = promoter and transcription start site, defined from -1 kb to +100 bp; TTS = transcription termination site, defined from -100 bp to +1 kb.

The Hi-C map revealed prominent interactions along the main diagonal, attributed to the increased likelihood of interactions among neighbouring linear genome segments (**Supplementary Figure 7B**). We also detected intense intra- and interchromosomal interactions at the chromosomal ends (**Supplementary Figure 7B**), where gene density is highest in the *P. grisea* genome (**Supplementary Figure 7C**). In an A/B compartment annotation of each chromosome, which generally classifies open, euchromatic vs. more compacted, heterochromatic regions, those gene-rich regions at the distal chromosome ends were distinct from the large central region, where transposon density is highest (**Supplementary Figure 7C**) (He et al. 2023). Consequently, we annotated the distal ends as A (accessible) and the central part as B (inaccessible) compartments. This chromosomal organisation was highly similar to that observed in tomato (*S. lycopersicum*) (Dong et al. 2017).

We next explored the withanolide BGC locus, located at the end of chromosome 1 (114.71-115.12 Mb), by generating normalised Hi-C maps at 20 kb resolution. In both root and shoot tissue, chromatin interactions between the RESC and SESC were rare, indicating that both

subclusters were isolated from each other at the chromatin level (**Figure 3A**). Such a spatial separation may contribute to the regulation of tissue-specific expression, allowing each sub-cluster to interact with specific cis-elements. In contrast, the RESC and, to a lesser extent, the SESC exhibited frequent intra-sub-cluster chromatin interactions (**Figure 3A**). We further noticed that the RESC exhibited more frequent interactions with itself in the shoot compared to the root; a weaker, opposite trend applied to the SESC (**Figure 3A)**. This became more obvious when considering the differential chromatin interactions between tissues (**Figure 3B**).

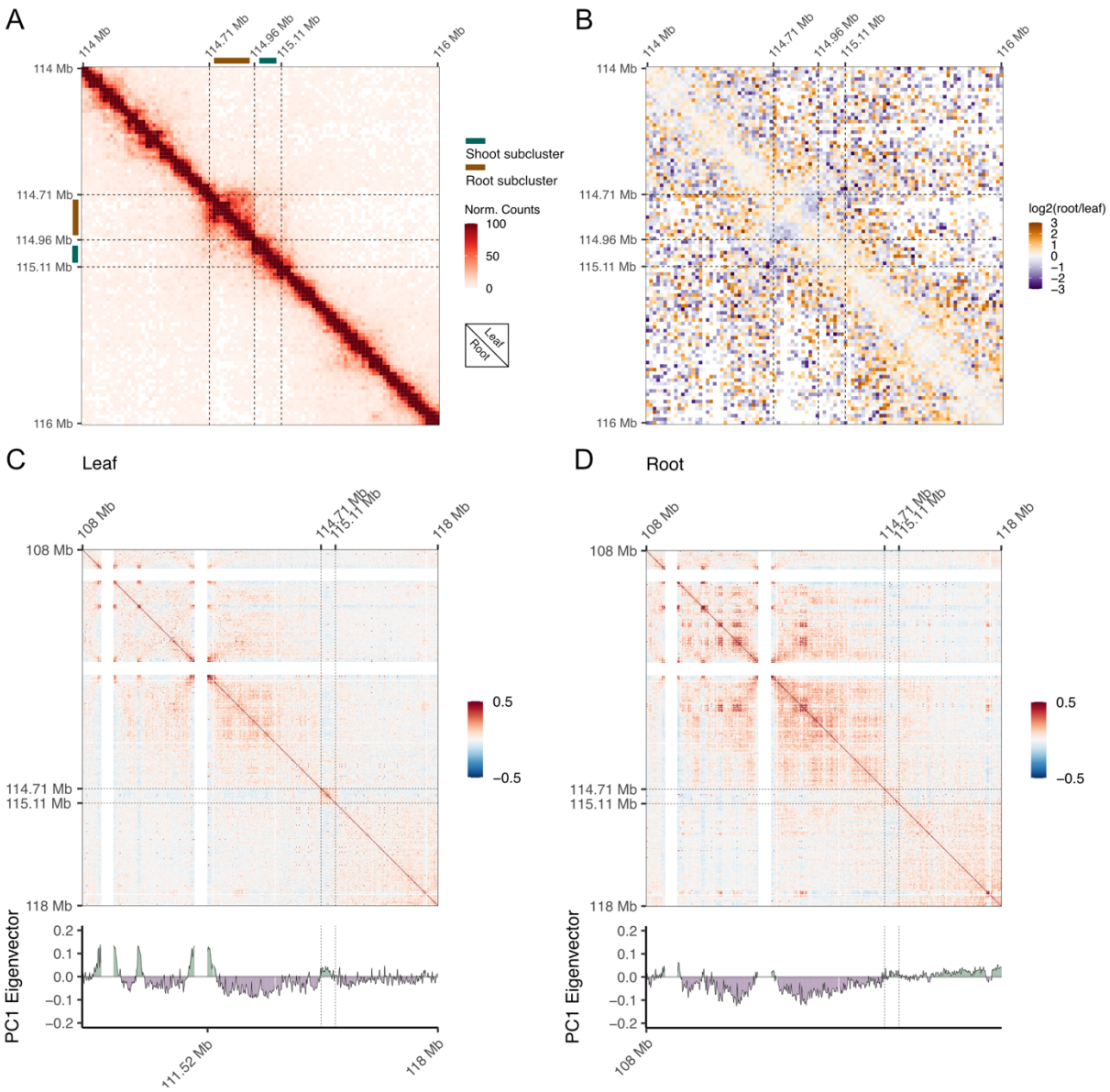


Figure 3. Higher-order chromatin configuration and chromatin-chromatin interactions at the withanolide BGC locus. A) Normalised Hi-C contact matrices at 20-kb resolution covering the locus of the withanolide BGC of leaf (above the diagonal) and root (below the diagonal). **B)** Binary logarithm (log2) of the ratio between the root and shoot normalised Hi-C contact matrices at 20-kb resolution. Positive and negative values indicate more interaction frequency in root and leaf tissue, respectively. **C)** and **D)** Pearson's correlation matrices of chromosome 1 right arm and sub-A/B compartment annotation in **C)** leaf and **D)** root. In the four panels, vertical and horizontal dashed black lines indicate the withanolide BGC locus.

288 We noticed that the entire withanolide BGC locus displayed very few interactions with the
289 neighbouring euchromatic region, visible as a banded area in the Hi-C map, which was more
290 pronounced in the shoot than the root (**Supplementary Figure 7D,E**). This led us to
291 hypothesise that the local chromatin state of the whole BGC might differ from that of the
292 neighbouring regions. Global euchromatic regions (A compartments) and heterochromatic
293 regions (B compartments) can be further classified into local A/B compartments, i.e, local
294 regions within the A or B compartments that can be either euchromatic or heterochromatic at
295 a smaller scale and that correlate with activating or inactivating epigenetic marks (Dong et al.,
296 2017). Consequently, we annotated local A/B compartments within the euchromatic region
297 containing the cluster (**Figure 3C,D**). In leaf tissue, the BGC clearly stood out from the
298 neighbouring chromatin, suggesting that it preferentially interacted with heterochromatin
299 (**Figure 3C**). This was less prominent in root tissue, where the cluster seemed to interact with
300 the downstream euchromatic regions (**Figure 3D**). Collectively, the ATAC-seq and Hi-C
301 results suggested that the local 3D organisation of the chromatin contributes to the regulation
302 of the tissue-specific expression of each subcluster.

303 *The tissue-specific gene expression of each subcluster is associated with higher levels*
304 *of asymmetric DNA methylation*

305 Local epigenomic configurations are emerging as a key regulator of plant specialised
306 metabolism (Conneely et al. 2022; Méteignier et al. 2022). The tissue-specific chromatin
307 organisation at the withanolide BGC locus, revealed by our ATAC-seq and Hi-C data, led us
308 to hypothesise that distinct epigenetic states might be contributing to the regulation of the
309 bipartite sub-cluster-architecture of the withanolide BGC. Cytosine methylation at the 5th
310 carbon (5mC, or often more generally referred to as DNA methylation) is an important
311 epigenetic mark involved not only in gene expression regulation and the silencing of
312 transposable elements but also in modulating chromatin interactions (Zhang et al. 2018).
313 Moreover, recent studies showed that the expression of biosynthesis genes in several
314 specialised metabolism pathways are negatively correlated with DNA methylation (Li et al.
315 2019; de Bernonville et al. 2020; Kim et al. 2020). We therefore hypothesised that the
316 withanolide BGC sub-cluster architecture might be associated with distinct DNA methylation
317 patterns. We performed whole-genome bisulfite sequencing (WGBS) of shoot and root tissue
318 of 2-week-old seedlings, from the exact same biological samples as for the RNA-seq and
319 untargeted metabolomic analyses. We sequenced three biological replicates per tissue, with
320 a mean sequencing depth of 23-40X and bisulfite conversion rates >99.3% in every sample
321 (**Supplementary Dataset 4**; **Supplementary Figures 8 and 9**). We called differentially
322 methylated regions (DMRs) between shoot and root samples (Hüther et al. 2022). In a
323 principal component analysis (PCA) of mean methylation rates in context-specific DMRs,

samples clustered by tissue in all three contexts (**Supplementary Figure 8A**). In total, we identified 322 CG-, 183 CHG- and 9,522 CHH-context-specific DMRs. The withanolide BGC region contained 0 CG-, 1 CHG-, and 20 CHH-DMRs (**Figure 4A**). Contrary to our expectation, in each sub-cluster hypermethylated CHH-DMRs correlated with the active/expressed state (**Figure 4A**). Of the 20 CHH-DMRs in the withanolide BGC, 4 and 2 were located in the putative promoters (2 kb upstream) of RESC and SESC genes, respectively. Another CHH-DMR was located in the intron of the RESC gene *CYP87G1;* contrary to the trend, this gene was hyper-methylated in shoot tissue. The remaining 11 CHH-DMRs were located in intergenic regions.

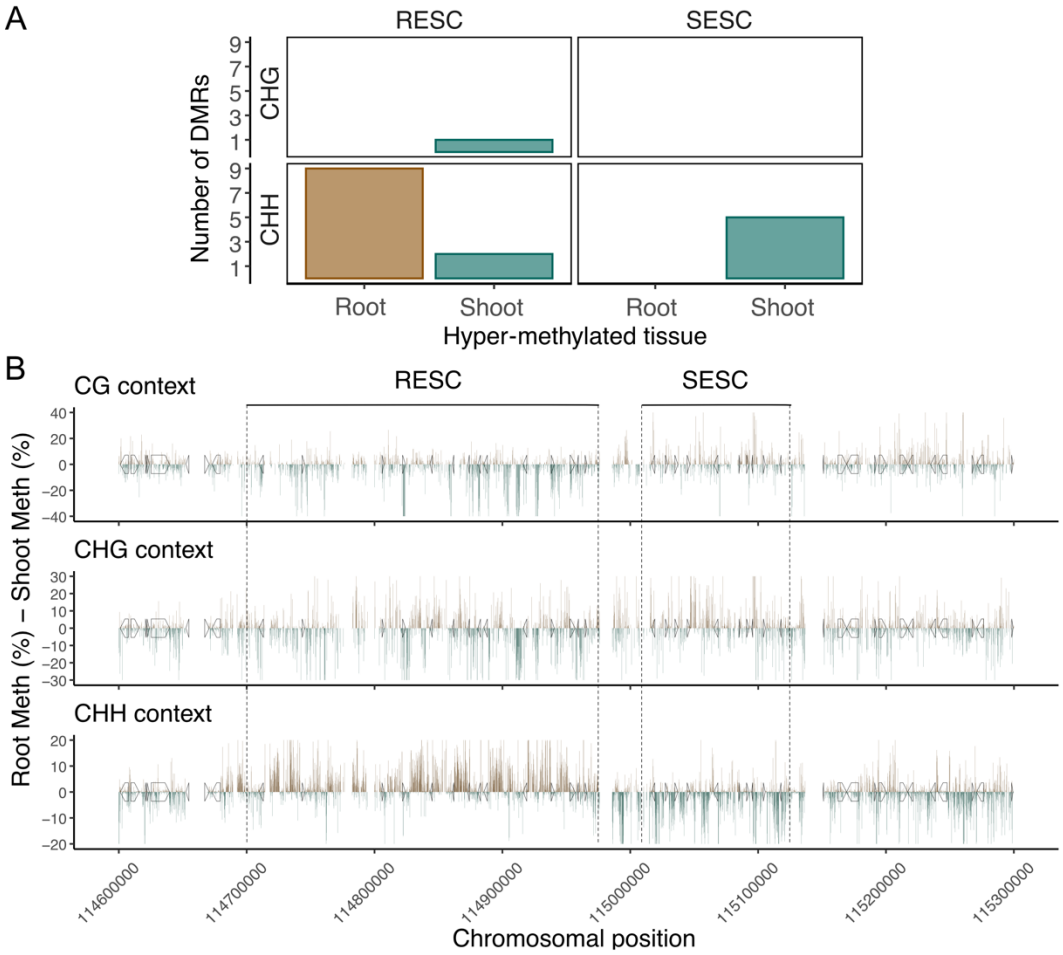


Figure 4. **DNA methylation at the withanolide BGC locus in *P. grisea*. A)** Number of differentially methylated regions (DMRs) identified within the RESC and SESC, respectively, in the comparison of root vs. shoot tissue. **B)** Differential DNA methylation between root and shoot tissue along the BGC and its flanking regions. Average methylation was calculated in non-overlapping 100 bp windows in root and shoot for each sequence context separately; windows that did not have above-minimum coverage information for both tissues were removed. The data show the per-context difference between root and shoot methylation for each window; positive values (in brown) represent higher methylation in root; negative values (in green) represent higher methylation in shoot tissue.

At a global level, the RESC also exhibited considerably higher CHH methylation in root tissue, where that sub-cluster is expressed, compared to shoot, in line with the DMR hypermethylation (**Figure 4B**). The SESC displayed the opposite trend, i.e., CHH methylation was highest in shoot tissue. CHH methylation differences primarily occurred in intergenic

345 regions. While we did not observe a noticeable pattern for CHG methylation, CG methylation
346 followed the opposite, albeit less pronounced trend than CHH methylation: in the RESC, CG
347 methylation was negatively correlated with expression in the respective tissue. In sum, our
348 data showed a marked shift in tissue-specific DNA methylation occurring along the genomic
349 region that separates the two sub-clusters.

350 *The withanolide BGC in Solanaceae underwent lineage-specific gene duplications*

351 Withanolides are widespread across *Solanaceae* (Misico et al., 2011). While they are absent
352 in most *Solanum* species, they have been isolated, among others, from the late-diverging
353 *Physalideae* tribe, as well as from early-diverging *Petunia* (in the form of petuniasteroids), and
354 are present in e.g. species from the genera *Datura*, *Nicandra* or *Lycium* (Hansel et al. 1975;
355 Elliger et al. 1989; Misico et al. 2011; Huang et al. 2020; Pan et al. 2024).

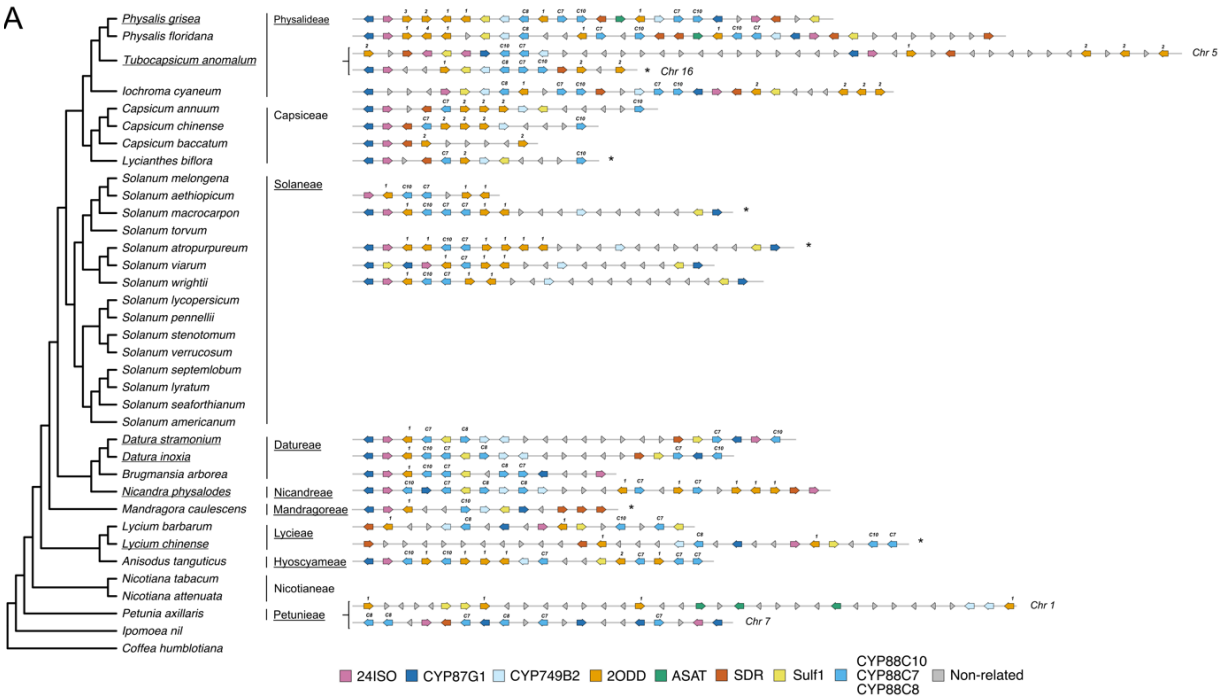
356 Given the ubiquitous presence of withanolide-related metabolites in several *Solanaceae*
357 genera, we asked how conserved the withanolide BGC was across the *Solanaceae*, and
358 whether the tandem duplication and sub-functionalization was a feature commonly associated
359 with this cluster. We mined the available *Solanaceae* genome assemblies, including species
360 and genera known to produce withanolides (**Supplementary Table 2**). Using Orthofinder
361 (Emms and Kelly 2019), we defined orthogroups; within the same orthogroup we searched for
362 the orthologs most similar to the genes in the *P. grisea* withanolide BGC. Among the list of
363 candidate orthologs in each surveyed species, we further confirmed which of these were
364 physically clustered in the genome.

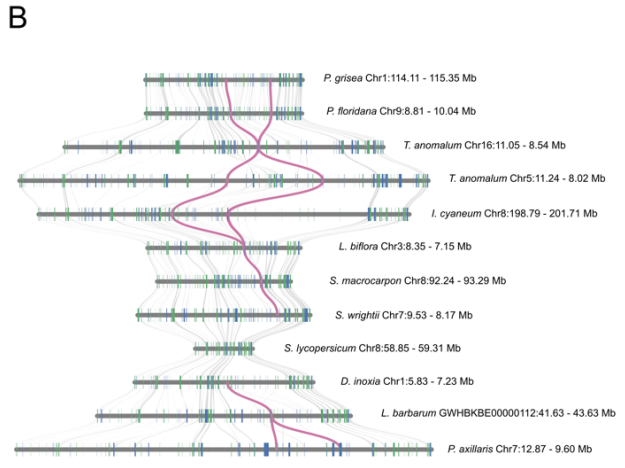
365 We identified orthologs of *24ISO* and of several of the *P. grisea* withanolide BGC genes
366 located in what appeared to be BGCs across most of the tribes included in our study. Genes
367 encoding for 24ISO orthologs and associated gene clusters were markedly absent in many
368 *Solanum* species, particularly in tomato- and potato-related species (*S. lycopersicum*, *S.*
369 *pennelli*, *S. stenotomum*, *S. verrucosum*, *S. septemlobum*, *S. lyratum*, *S. septemlobum*, *S.*
370 *seaforthianum* and *S. americanum*). In turn, we found a cluster in eggplant-related and several
371 other *Solanum* species (*S. aethiopicum*, *S. macrocarpon*, *S. torvum*, *S. atropurpureum*, *S.*
372 *viarum*, *S. wrightii*), but withanolides have not yet been reported in any of these (**Figure 5A**;
373 **Supplementary Figure 10**). Conversely, the genomes of all species with reported withanolide
374 production (or of their close relatives) contained at least one *24ISO* copy located in a putative
375 withanolide BGC. Synteny analysis of the gene cluster itself as well as of the flanking
376 chromosomal environment revealed that both were largely syntenic, indicating that the
377 withanolide BGC in *P. grisea* originated from an early common ancestral cluster that existed
378 already before the branching off of *Petunia* (**Figure 5B**). The cluster appears to have been
379 lost in several tribes and species, but conserved in many others.

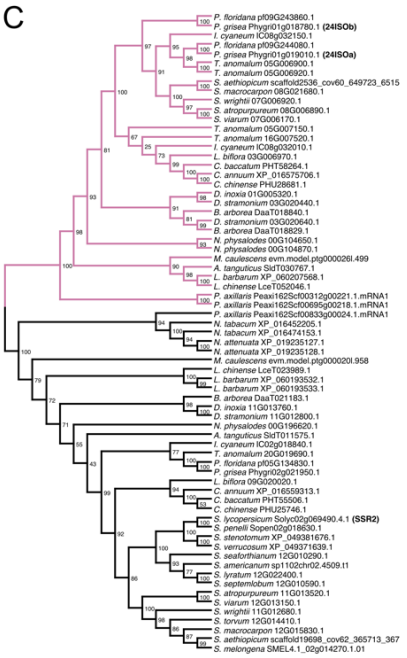
380 The tandem duplication that resulted in two paralogs of *24ISO* and seems to have been at the
381 origin of cluster sub-functionalization was apparent throughout the *Physalideae* tribe (including
382 the genera *Physalis*, *Tubocapsicu*m, and *Iochroma*). To test whether this conserved BGC
383 architecture was accompanied with conserved tissue-specific expression, we generated RNA-
384 seq data from two *Physalideae* species that we could sample in the botanical garden in
385 Munich, Germany. From the tree-like *Iochroma arborescens*, we could sample only leaf tissue;
386 from the shrub *Eriolarynx australis* (also frequently listed as *I. australe*), we sampled leaf and
387 root tissue (**Supplementary Dataset 1**). Our data showed that in leaves, the equivalent of the
388 *P. grisea* SESC, i.e., the sub-cluster located towards the proximal end of the whole BGC,
389 higher expression than its RESC equivalent (**Figure 5D**). Expression of the latter, on the other
390 hand, was weaker in general but appeared to be slightly stronger in root tissue (only
391 measurable for *E. australis*) (**Figure 5D**). We observed the same pattern in the SESC from
392 *Tubocapsicum anomalum* in publicly available RNA-seq from young leaves
393 (**Supplementary Figure 11**). Altogether, this indicates that the tandem duplication and sub-
394 functionalization of the cluster is common to species of the *Physalidea* clade.

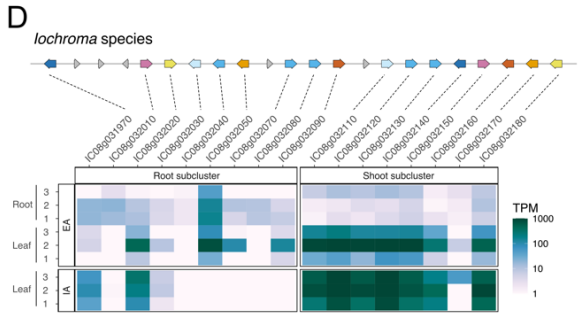
395 Our phylogenomic analysis also revealed several tribes with apparent cluster duplication,
396 namely *Datureae*, *Nicandreae*, and *Petunieae*, usually accompanied by the core duplication
397 of the *24ISO*, *CYP87G1* and the *CYP88C7* (**Figure 5A**). The amino-acid-sequence-based
398 phylogenetic analysis of all the genes within the *P. grisea* withanolide BGC suggested that, in
399 these tribes, these duplications were tribe-specific and unrelated to the duplication of the sub-
400 cluster architecture in the *Physalideae* (**Figure 5C; Supplementary Figures 12-20**). To
401 further test this hypothesis, we retrieved publicly available transcriptomic data for stem, root,
402 and leaf tissue of *Datura stramonium*. In contrast to the *Physalideae* species, the cluster in *D.*
403 *stramonium* did not show tissue-specific expression of different sub-clusters (**Figure 5E**). In
404 line with the different cluster architecture in comparison to the BGC in *Physalideae*, this
405 suggested an evolutionarily independent duplication event that did not result in sub-
406 functionalization.

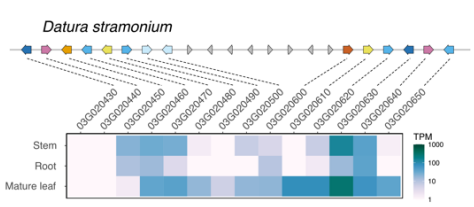
407 **Figure 5 (next page). An ancestral withanolides biosynthetic gene cluster in *Solanaceae* underwent several**
408 **lineage-specific gene duplications. A)** Phylogenetic relationship between the different *Solanaceae* species
409 included in the analysis (outgroups *Ipomoea nil* and *Coffea humblotiana*) and their respective withanolide BGC-
410 like loci. Underlined species and/or tribes are those in which withanolide-related metabolites have been detected
411 (**Supplementary Table 2**). An asterisk at the end of cluster indicates that this cluster has been inverted for
412 representation purposes. Numbers above the *CYP88C* genes indicate the respective subfamily; numbers above
413 2ODD genes indicate the respective orthogroup that they belong to. **B)** Microsynteny between the genomic region
414 containing the withanloide BGC in *P. grisea* and different representatives of the *Solanaceae* tribes included in this
415 study. Coloured lines indicate *24ISO* orthologs. Grey lines connect other orthologs. **C)** Amino-acid-sequence
416 maximum likelihood tree with the best-fit model selected by ModelFinder represented as a cladogram of the
417 orthogroup containing the Pg24ISO. Numbers indicate ultra-fast bootstrap values. **D)** and **E)** Heatmap showing the
418 expression pattern of each gene within the cluster across different tissues in the *Iochroma* species *Eriolarynx*
419 *australis* (EA) and *Iochroma arborescens* (IA) **(D)** and in *Datura stramonium* **(E)**. Expression levels are presented
420 as transcripts per million (TPM).


A
Physalis grisea
Physalis floridana
Tubocapsicum anomalum
Iochroma cyaneum
Capsicum annuum
Capsicum chinense
Capsicum baccatum
Lycianthes biflora
Solanum melongena
Solanum aethiopicum
Solanum macrocarpon
Solanum torvum
Solanum atropurpureum
Solanum viarum
Solanum wrightii
Solanum lycopersicum
Solanum pennellii
Solanum stenotomum
Solanum verrucosum
Solanum septemlobum
Solanum lyratum
Solanum seaforthianum
Solanum americanum
Datura stramonium
Datura inoxia
Brugmansia arborea
Nicandra physalodes
Mandragora caulescens
Lycium barbarum
Lycium chinense
Anisodus tanguticus
Nicotiana tabacum
Nicotiana attenuata
Petunia axillaris
Ipomoea nil
Coffea humblotiana
Physalideae
Capsiceae
Solaneae
Datureae
Nicandreae
Mandragoreae
Lycieae
Hyoscyameae
Nicotianeae
Petunieae
Chr 5
Chr 16
Chr 1
Chr 7
24ISO
CYP87G1
CYP749B2
2ODD
ASAT
SDR
Sulf1
CYP88C10
CYP88C7
CYP88C8
Non-related


B
P. grisea Chr1:114.11 - 115.35 Mb
P. floridana Chr9:8.81 - 10.04 Mb
T. anomalum Chr16:11.05 - 8.54 Mb
T. anomalum Chr5:11.24 - 8.02 Mb
I. cyaneum Chr8:198.79 - 201.71 Mb
L. biflora Chr3:8.35 - 7.15 Mb
S. macrocarpon Chr8:92.24 - 93.29 Mb
S. wrightii Chr7:9.53 - 8.17 Mb
S. lycopersicum Chr8:58.85 - 59.31 Mb
D. inoxia Chr1:5.83 - 7.23 Mb
L. barbarum GWHBKBE00000112:41.63 - 43.63 Mb
P. axillaris Chr7:12.87 - 9.60 Mb


C
24ISO
SSR2


D
Iochroma species
Root subcluster
Shoot subcluster
Root
Leaf
EA
IA
TPM
1000
100
10
1


E
Datura stramonium
Stem
Root
Mature leaf
TPM
1000
100
10
1

421

422

423 # Discussion

424 Next to frequent biosynthetic operons in bacteria, the organisation of collinear genes in BGCs
425 has long been considered to be exclusive to fungi and plants; recent evidence suggests that
426 they also exist in the animal kingdom (Scesa et al., 2022). BGCs likely ensure the co-
427 inheritance of genes encoding for the subsequent steps in a biosynthetic cascade, which
428 would safeguard against the accumulation of potentially toxic intermediates (Nützmann and
429 Osbourn 2015). BGCs have also been proposed to facilitate the formation of metabolons, i.e.,
430 temporary complexes composed of the sequential enzymes of a biosynthetic process
431 (Nützmann and Osbourn 2015). While these two aspects of BGCs remain hypothetical, it is
432 certain that BGCs coordinate gene expression to ensure co-expression of the biosynthetic
433 genes across tissues and treatments and hence the correct spatio-temporal production of the
434 metabolites (Nützmann and Osbourn 2015). However, the mechanisms and temporal events
435 that lead to the formation and diversification of BGCs remain unresolved (Polturak et al. 2022;
436 Smit and Lichman 2022). Similarly, our understanding of BGC transcriptional regulation in
437 plants is limited.

438 Here, we have identified a BGC underlying withanolide production in *P. grisea*. We show how
439 a partial duplication resulted in the formation of two tandemly arranged sub-clusters, which
440 are expressed in different tissues and appear to contribute to the production of tissue-specific
441 withanolide compounds. Multi-layer epigenomic analysis revealed that the two sub-clusters
442 are isolated from each other at the chromatin level, and that the tissue-specific expression is
443 regulated epigenetically.

444 *An ancestral withanolide BGC with lineage-specific diversification in Solanaceae*

445 By anchoring our search for a withanolide BGC on the *24ISO* orthologs in *P. grisea*, we
446 identified a BGC-like arrangement of 22 genes, including several encoding CYP450s from
447 different families and subfamilies, as well as six *2-ODD* genes. In terms of size (close to 400
448 kb) and gene content, this makes the withanolide BGC one of the largest plant BGC identified
449 to date, comparable to the benzylisoquinoline alkaloid BGC in opium poppy (*Papaver*
450 *somniferum*; 500 kb, 15 genes) (Guo et al. 2018). VIGS-mediated suppression confirmed the
451 contribution of several genes to withanolide production (**Figure 1**). Our comparative
452 phylogenomic and transcriptomic analyses showed that the withanolide BGC architecture and
453 the division into two distinctly expressed sub-clusters are conserved across the *Physalideae*.
454 The identification of orthologous clusters and independent duplication events across the
455 *Solanaceae* family suggest the existence of an early ancestral cluster. We detected signals of
456 at least two independent duplication events, one in the *Physalideae* clade, and one in the

457 *Datureae*/*Nicandreae*/*Petunieae* clade, which resulted in different architectures and
458 functionalities.

459 In contrast, our analyses indicate the loss of the withanolide BGC in many species of the
460 genus *Solanum*, particularly in those more closely related to tomato and potato. This is
461 noteworthy, since it has been postulated that the biosynthesis of steroidal glycoalkaloids
462 (prominent in tomato and potato and presumably in many related species) and that of
463 withanolides, which branch off the same precursors, usually do not co-occur in the same
464 species. Few exceptions to this have been reported to date, including *D. stramonium*, which
465 produces both steroidal glycoalkaloids and diverse withanolides (Pan et al. 2024), and in
466 which we identified an orthologous withanolide BGC (**Figure 1**). Although we found putative
467 withanolide BGCs in several *Solanum* species, evidence of withanolide production in these
468 species remains to be confirmed in future studies. Conversely, withanolides have been
469 detected outside the *Solanaceae* family, and it will be interesting to investigate the genetic
470 architecture underlying withanolide production in those species.

471 *A sub-functionalisation that results in tissue-specific activity*

472 Our transcriptomic transect across different tissues, developmental stages and treatments
473 revealed that the proximal and distal parts of the withanolide BGC in *P. grisea* are expressed
474 tissue-specifically. In particular, the proximal set of 14 genes (the RESC) showed a strictly
475 root-specific expression that was also maintained upon stimulating elicitors such as MeJA.
476 The expression of the eight distal genes (the SESC) was most prominent in the aerial tissues,
477 albeit less tightly restricted than the RESC. In line with this finding, we detected tissue-specific
478 accumulation of different withanolide compounds. This suggests that the expression of the
479 respective paralogs in either root or shoot might lead to the production of slightly different
480 withanolide compounds and raises the question about the physiological and ecological
481 function of this compartmentalisation. Since withanolides play a role in organismic interactions
482 and act as antibiotics as well as in the repelling of herbivores (Huang et al. 2020), one can
483 speculate that root-specific and shoot-specific withanolides could play respective roles in the
484 defence against niche-specific microbes, nematodes, or insects. In future, the genetic and
485 biochemical characterization of the two-sub-clusters will help understand their role in the
486 chemical ecology of *P. grisea* and its relatives.

487 In line with the differential transcriptional activity in roots and shoots, our ATAC-seq data
488 revealed regions of increased local chromatin accessibility within the RESC and SESC,
489 respectively. Recent studies in the medicinal plants *Artemisia annua* and *Catharanthus roseus*
490 have shown that tissue- and even cell-type-specific local chromatin accessibility correlate with
491 BGC expression and the biosynthesis of artemisinin and monoterpene indole alkaloids,

492 respectively (Zhou et al. 2021; Li et al. 2023). Such tissue-specific local chromatin accessibility
493 allows TFs and other DNA-binding proteins to bind and to specifically activate gene
494 expression. The next step in elucidating withanolide BGC expression will therefore be to
495 identify the respective TFs and their binding sites. Our motif search of accessible regions
496 returned bHLH and MYB TF binding motifs in root, and C2H2 zinc-finger binding motifs in
497 shoot tissue (**Supplementary Table 3**), which will serve as an anchor for the future search of
498 regulating TFs. An alternative approach is to search for TFs that are co-expressed along with
499 either the RESC or SESC. Using our RNA-seq data across tissues and treatments, we
500 identified, among others, two basic helix-loop-helix (bHLH) TFs in the corresponding
501 expression modules. Members of this TF family are generally known to regulate specialised
502 pathways through jasmonate (JA) signalling (Goossens et al. 2017). *Phygri08g015540* was
503 co-expressed with the SESC in module M1. Interestingly, its closest homolog in tomato
504 (*Solyc01g096370*) has been shown to be involved in the regulation of steroidal glycoalkaloids
505 (Li et al. 2020). In module M2, which contained the RESC genes, we identified
506 *Phygri08g006590*, whose closest homolog in Arabidopsis (bHLH20/NAI1; AT2G22770) is
507 involved in glucosinolate accumulation and defence (Gao and Dubos 2024). Follow-up genetic
508 analysis of these and other candidate TFs, including e.g. ethylene response factors (ERFs),
509 which have been shown to regulate terpenoid biosynthesis across various plant lineages
510 (Shoji et al. 2023), is necessary to elucidate the regulatory mechanisms that control sub-
511 cluster expression.

512 *The sub-functionalisation of the withanolide BGC is maintained at the chromatin level*

513 The withanolide BGC in *P. grisea* combines two features that have been reported for other
514 plant BGCs individually, but never in combination: a defined organisation at the chromatin
515 level, and a bipartite architecture. Recent studies have shown that plant BGCs can be
516 organised in TADs or TAD-like structures, i.e., in regions with extensive 3D chromatin-
517 chromatin interactions, and that the configuration of the TAD influences the (co-)expression
518 of the genes located within the BGC. For example, in leaves of *A. thaliana*, the thalianol cluster
519 is contained in a condensed B compartment and transcriptionally inactive, while in roots it is
520 located in a relaxed-chromatin A compartment and transcriptionally active (Liu et al. 2020). In
521 opium poppy, the BGC encoding the biosynthesis of benzylisoquinoline alkaloids also forms
522 a TAD (Guo et al. 2018). Liu et al. recently reported on a *p*-menthane BGC in a *Lamiaceae*,
523 *Schizonepeta tenuifolia*, that underwent an inverted duplication to form a bipartite cluster (Liu
524 et al. 2023). Another *Lamiaceae* cluster, the miltiradiene BGC, exhibits gene duplications and
525 even displays tissue-specific expression differing along different parts of the cluster (Bryson
526 et al. 2023). The withanolide BGC described in our study exemplifies how a duplication of
527 parts of a BGC can result in the formation of two tandemly arranged sub-clusters that end up

being isolated from each other at the chromatin level: the two sub-clusters are arranged in two separate TAD-like domains. Their condensation levels differed between tissues and negatively correlated with expression; especially the RESC showed more intense interactions in leaves than in roots. Since the RESC was not expressed in the shoot (and *vice versa* the SESC was only lowly expressed in roots), the Hi-C, along with the ATAC-seq data, suggest that both sub-clusters adopt a more compacted conformation in the tissue in which their expression is suppressed. Such a compaction likely restricts the accessibility for TFs and the primary transcriptional machinery. These observations are in line with the patterns observed for the thalianol cluster in *A. thaliana* (Liu et al. 2020; Conneely et al. 2022; Méteignier et al. 2023).

Unexpectedly, asymmetric CHH methylation was positively correlated with gene expression in both sub-clusters. While CHH methylation is generally associated with transcriptional suppression, it should be noted that this relies mostly on data from *A. thaliana* and several grass crops, and that our understanding of the epigenome in *Solanaceae* is rather limited. To our knowledge, this is the first observation of differential DNA methylation correlating with BGC activity. Further investigations into the chromatin landscape of *P. grisea* and the withanolide BGC are necessary, e.g. on the role of post-translational histone modifications and variants, which have been shown to be associated with structural organisation of the thalianol BGC (Liu et al. 2020), as well as on the potential impact of the gene-depleted 50-kb spacer sequence that separates the two sub-clusters and that features a prominent accessible chromatin region in its centre (**Figure 2B**).

Conclusion

In summary, our study highlights how, based on chromatin- and epigenome-level organisation, eukaryotic BGCs can diversify and gain additional functionality. It moreover showcases that *P. grisea*, an emerging genetic model species in the *Solanaceae* family (Van Eck 2022; Dale et al. 2024), is amenable to further investigations into the biosynthesis of withanolides, a group of diverse and multi-functional plant specialised metabolites that are of interest in the context of plant protection and human health (Huang et al. 2020). We are aware that the tissue-specific analysis of the withanolide BGC presented here might yet mask a more refined structuring of withanolide biosynthesis, one that might be resolved in future by analyses at the single-cell level.

559 # Materials and Methods

560 *Plant material*

561 Seeds for *P. grisea* were kindly provided by Sebastian Soyk (University of Lausanne), from
562 the line sequenced for the *P. grisea* genome (He et al. 2023).

563 *Virus-induced gene silencing*

564 Gene silencing was done using a tobacco rattle virus (TRV)-based VIGS system (Dinesh-
565 Kumar et al. 2003) as described in (Knoch et al. 2018). Inserts were cloned into pTRV2 using
566 XbaI and KpnI restriction sites and the resulting plasmids cloned into *Agrobacterium* GV3101.
567 Fragments for genes to be silenced, except for Pg24ISO, were amplified from plasmids
568 containing the corresponding *P. grisea* cDNA gene sequences. Primers used in this study are
569 listed in **Supplementary Table 4**. The construct to silence Pg24ISO was synthesized by
570 Azenta (Genewiz) based on the Ws24ISO construct described in Knoch et al (2018), with XbaI
571 and KpnI restriction sites added. pTRV2 containing a *P. grisea* phytoene desaturase (PDS)
572 fragment was used to verify VIGS and to determine time of harvest. pTRV2 containing a
573 fragment of GFP was used as negative control. True leaves of 3-week-old *P. grisea* seedlings
574 were infiltrated with *Agrobacterium* harboring pTRV1 and pTRV2 in a 1:1 ratio by syringe
575 infiltration. Six plants were infiltrated per construct. Bleaching of PDS control plants was
576 beginning about 3 weeks after infiltration and samples were harvested about 4 weeks after
577 infiltration. The five top leaves from each plant were harvested and pooled for one sample,
578 and aliquots of the same sample were used for qPCR and metabolomics analysis.

579 *Quantitative real-time PCR (qPCR)*

580 RNA was extracted using RNeasy Plant Mini Kit (Qiagen) including on-column Dnase
581 treatment. cDNA was synthesized using the RevertAid H Minus First Strand cDNA Synthesis
582 Kit (Thermo Scientific). For qPCR 2 µl 100x diluted cDNA was used in 10 µl reaction volume.
583 Real-time PCR was done with the QuantStudio5 (Applied Biosystems) system using FastStart
584 Universal SYBR Green Master (Roche). Primers used in the analysis are listed in
585 **Supplementary Table 4**. Expression levels in VIGS plants were calculated as $2^{(-\Delta\Delta CT)}$ for each
586 gene relative to Tubulin and GFP control plants. Five or six biological replicates were analyzed
587 for each construct.

588 *Elicitor treatment*

589 Seeds were sterilized for 15 min in 6% NaClO, washed 1x with 100% EtOH and 3x with water,
590 and germinated on ½ Murashige & Skoog medium. Seedlings were grown horizontally in three

591 rows on square plates in a Polyklima chamber at 24˚C, 16 hours light (160 µmol), 8 hours dark
592 cycle and 50% relative humidity for 2 weeks. For elicitor treatment, seedlings were transferred
593 to Petri dishes (15 seedlings per dish, 4 replicates per condition) containing 30 mL water and
594 left overnight on the bench. Water was removed and replaced with 30 mL elicitor solution (50
595 µM MeJa, 10 µM ABA, 1 µM Flg22, or mock); seedlings were incubated for 4 h before
596 harvesting. Seedlings were dried with tissue paper; shoots and roots were separated with a
597 razor blade. The tissue was flash-frozen in liquid nitrogen and ground to a fine powder.

598 *RNA-sequencing and transcriptomic analysis*

599 Plants were grown in standard potting soil in the greenhouse. Different tissues were sampled
600 as follows from 4 individuals: 1) young leaves, not fully expanded; 2) mature leaves, first fully
601 expanded leaf; 3), petioles from mature leaves; 4) stems, top 10 cm; 5) calyx; 6) fruit, unripe,
602 green stage. Seedlings for harvesting of two-week old roots and shoots were grown in sand
603 at 24˚C, 60 % relative humidity, and a 16 h light / 8 h dark cycle. Six to ten seedlings were
604 pooled per sample, for a total of four biological replicates. Harvested tissue was flash-frozen
605 in liquid nitrogen and ground to a fine powder. Samples were stored at -80˚C until analysis
606 and aliquots of the same sample were used for metabolomic and transcriptomic analysis.

607 RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). In each set, each
608 tissue/condition was replicated four times. After quality control, we discarded one replicate
609 each for “fruit” and one for “young leaf” from the subsequent analyses (**Supplementary**
610 **Figure 2**). We prepared polyA-RNA sequencing libraries from 400 ng of total RNA input using
611 the NEBNext Ultra II RNA library preparation kit (New England Biolabs). Libraries were
612 sequenced at IMGM (Martinsried, Germany) on a Novaseq instrument (Illumina) single-end
613 100 bp reads. Sequencing and mapping are summarised in **Supplementary Dataset 1**;
614 quality control is shown in **Supplementary Figure 2**.

615 *Iochroma* samples were collected in the Botanical Garden in Munich, Germany. A pool of 2-3
616 young leaves from each plant were used for each replicate. Root samples could be harvested
617 from only one plant; we sampled tissue from three different positions. *Iochroma* samples, RNA
618 was extracted with Quick-RNA plant Miniprep Kit (Zymo Research).

619 For all the RNA-seq datasets, RNA-seq mapping and quantification was done using the nf-
620 core pipeline rnaseq v3.9 (https://nf-co.re/rnaseq/3.9) implemented in Nextflow 21.10.3 (Di
621 Tommaso et al. 2017; Ewels et al. 2020). Briefly, adapter and quality trimming was performed
622 using TrimGalore v0.6.7 (Krueger et al., 2021), a wrapper tool around Cutadapt (v3.4) (Martin
623 2011) and FastQC (v0.12.1) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
624 Reads were mapped to the *P. grisea* (He et al. 2023), *I. cyaneum* (Powell et al. 2022), *T.*
625 *anomalum, L. biflora* and *D. stramonium* (Wu et al. 2023) genome assemblies using STAR

626 v2.7.10a (Dobin et al. 2013) and quantified using salmon v1.5.2 (Patro et al. 2017). Feature
627 counts were normalised by gene length and sequencing depth using transcripts per million
628 (TPM), that were used to inspect the expression of the withanolide BGC locus. Differential
629 gene expression was performed with the DESeq2 (v1.44.0) R package (Love et al. 2014)
630 using raw feature counts. For the co-expression analysis, WGCNA implemented in the R
631 package CEMiTool (v1.28.0) was used (Zhang and Horvath 2005; Langfelder and Horvath
632 2008; Russo et al. 2018). As input, the raw counts of *P. grisea* tissue and elicitors RNA-seq
633 datasets were used after *vst* transformation (Love et al. 2014).

634 In the case of *T. anomalum, L. biflora* and *D. stramonium*, reads were obtained from the
635 repository described in the original manuscript (Wu et al., 2023). At the time we mapped the
636 first time the RNA-seq in *P. grisea* assembly, we noticed that several RESC genes
637 (*CYP87G1b*, *2ODD4*, *CYP749B2b* and *CYP88C10*) were mis-annotated, probably due to the
638 lack of expression of these genes in aerial tissue, used to annotate the genome in the original
639 manuscript (Lemmon et al. 2018; He et al. 2023). Accordingly, we manually re-annotated
640 those genes to a full CDS (re-annotation can be found in the manuscript repository).

641 *ATAC-seq data generation and analysis*

642 Plants were grown in standard potting soil in the greenhouse. For each replicate,
643 approximately ~ 300 mg of young leaf and ~ 1g of root of a pool of 2-3 3-week-old seedlings
644 were harvested to isolate nuclei. The nuclei isolation protocol was based on Sikorskaite et al.
645 (Sikorskaite et al. 2013) with slight modifications. Briefly, nuclei were extracted from the tissue
646 on a petri dish on ice by chopping with a razor blade in nuclei isolation buffer (NIB: 10 mM
647 MES-KOH pH 5.4, 10 mM NaCl, 200 mM sucrose, 0.1 mM spermine, 0.5 mM spermidine, 1
648 mM DTT and 1% BSA, 0.35% TritonX-100 - in the case of leaf tissue- or 0.75% for root tissue).
649 The supernatant was filtered through a 40 µm strainer and the nuclei were pellet by
650 centrifuging at 500 rcf for 5 min at 4°C. The pellet nuclei were resuspended in NIB, always
651 using bored-open pipette tips, and filtered with a 20 µm strainer. The filtered nuclei were added
652 on top of a NIB + 35% percoll solution and centrifuged at 500 rcf for 10 min at 4°C. The
653 supernatant was removed and the nuclei resuspended in TAPS solution (10mM TAPS-NaOH,
654 pH8.0; 5 mM $MgCl_2$). We then dyed the nuclei with 1 µg/mL DAPI of 1 replicate of leaf and
655 root. We estimated the number of nuclei in the root and leaf sample with a hemocytometer.
656 We treated ~50k nuclei with 2.5 µL of Tagment DNA enzyme (Illumina) in 25 µl of Tagment
657 DNA Buffer in a final volume of 50 µL at 37°C for 60 min. The DNA was then purified using
658 the Monarch® PCR & DNA Cleanup Kit (New England Biolabs #T1030). The tagmented DNA
659 was amplified with the NEB II Ultra Q5 with 11 cycles using primers previously described in
660 https://www.iaea.org/sites/default/files/warthmann-illumina-multiplexlibrary-protocol-
661 2021.pdf. The DNA was recovered using AMPure beads. As an input sample, we extracted

662 gDNA from a 3-week-old seedling that was subjected to the same tagmentation procedure
663 and amplification. Libraries were sequenced at Novogene Germany on a NovaSeqX Plus
664 platform (Illumina), using 2 x 150 bp paired-end reads.

665 ATAC-seq reads quality control and mapping was performed using the nf-core pipeline
666 atacseq (v2.1.2) (https://nf-co.re/atacseq/2.1.2) implemented in Nextflow v23.10.0 (Di
667 Tommaso et al. 2017; Ewels et al. 2020; Patel et al. 2023). Briefly, quality and adapter
668 trimming of raw reads was performed using TrimGalore! v0.6.7
669 (https://github.com/FelixKrueger/TrimGalore). Read mapping was performed with bowtie2
670 (v2.4.4) (Langmead and Salzberg 2012). After mapping, duplicates were marked with Picard
671 (v3.0.0) (https://broadinstitute.github.io/picard/). Subsequently, reads mapping to
672 mitochondrial and chloroplast genomes or to blacklisted regions were removed using
673 samtools (v1.17) (Danecek et al. 2021). Additionally, reads that were marked as duplicates,
674 not marked as primary alignments, unmapped or mapping to multiple locations were also
675 removed with samtools. Reads containing more than 4 mismatches, were soft-clipped or had
676 an insert size > 2kb were removed with bamtools (v2.5.2) (Barnett et al. 2011). The alignment-
677 level QC for the resulting filtered bam files were evaluated with picard. Normalised bigWig files
678 scaled to 1 million mapped reads were generated using bedtools genomecov (v2.30.0)
679 (Quinlan and Hall 2010) and bedGraphToBigWig (v445) (Kent et al. 2010).

680 The resulting filtered bam files from the pipeline were used to call narrow peaks with hmmratac
681 (Tarbell and Liu 2019) implemented in MACS3 (3.0.2) (Zhang et al. 2008). For each tissue,
682 we used all the replicates as input. Leaf and root accessible regions were filtered for a score
683 >100. Bedtools intersect was used to get consensus narrow peaks of leaf and root tissues.
684 The raw counts of consensus narrow peaks were used to compute the PCA with DESeq2.
685 The Homer (c4.11) annotatePeaks and findMotifsGenome algorithms (Heinz et al. 2010) were
686 used to annotate narrow peaks on each tissue and to find enriched motifs in the RESC (using
687 root narrow peaks) and in the SECS (using leaf narrow peaks). Normalised bigWig files were
688 used as input for deeptools (v3.2.0) (Ramírez et al. 2016) to compute metaplots over the
689 annotated genes scaled to 3kb. Visualisation of the chromatin accessibility of the withanolide
690 BGC locus was performed with igv (Robinson et al. 2011).

691 *Hi-C data generation and analysis*

692 Leaf samples and root samples consisted of a pool of two and four seedlings, respectively.
693 Approximately 0.5 g of fresh tissue was collected and fixed with 1% formaldehyde in MC buffer
694 (10 mM potassium phosphate, pH 7.0; 50 mM NaCl; 0.1M sucrose; 0.01% Triton X-100) by
695 applying a vacuum for 30 mins at room temperature. Formaldehyde was neutralised by
696 washing the samples with MC buffer supplemented with 0.15 M glycine and applying a vacuum

697 for 10 min at room temperature. Subsequently, the tissue was rinsed with fresh MC buffer
698 twice for 5 min and further washed with water. Next, the tissue was snap-frozen and stored at
699 -80 until library preparation.

700 Library preparation was started with tissue homogenization with ice-cold nuclei isolation buffer
701 (20 mM HEPES, pH 8.0, 250 mM sucrose, 1 mM MgCl2, 5 mM KCl, 40% glycerol, 0.25% Triton
702 X-100, 0.1 mM PMSF, 0.1% 2-mercaptoethanol); subsequently, the homogenate was filtered
703 with double-layered miracloth (Millipore). After washing twice with nuclei isolation buffer and
704 one round with 1× NEB Dpn II buffer (NEB), nuclei pellet was resuspended with 250 µl 0.5%
705 SDS and split into five tubes. The subsequent steps followed essentially a published protocol
706 (Wang et al. 2023). Nuclei penetration was then performed at 62 °C for 5min, followed by
707 adding 145 µl water and 25 µl 10% Triton X-100 each tube to quench SDS. Subsequently,
708 chromatin digestion was carried out at 37°C overnight with 50 U DpnII (NEB) in each tube,
709 and stopped by incubating at 62°C for 20 min the next day. For sticky end filling and chromatin
710 labelling, 1 µl of 10 mM dTTP, 1 µl of 10 mM dATP, 1 µl of 10 mM dGTP, 10 µl of 1 mM biotin-
711 14-dCTP, 29 µl water and 40 U Klenow fragment (Thermo Fisher) were added, and further
712 incubated at 37°C for 2 h. After adding 663 µl water, 120 µl 10× blunt-end ligation buffer
713 (300 mM Tris–HCl, 100 mM MgCl2, 100 mM DTT, 1 mM ATP, pH 7.8) and 40 U T4 DNA ligase
714 (Thermo Fisher), proximity ligation was performed at room temperature for 4 h. Next, five tubes
715 of nuclei pellet were combined by resuspending with 650 µl SDS buffer (50 mM Tris–HCl, 1%
716 SDS, 10 mM EDTA, pH 8.0). After proteinase K (Thermo Fisher) treatment at 55 °C for 30 min,
717 de-cross-linking was carried out by adding 30 µl 5 M NaCl and incubating at 65 °C overnight.
718 The next day, following DNA recovery and RNA digestion with RNase A, 3 µg purified Hi-C
719 DNA was sonicated in 130 µl TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) with a Q800R3
720 sonicator (QSONICA) by using the parameters: 25% amplitude, 15 s ON, 15 s OFF, pulse-on
721 time for 4.5 min, to achieve fragment size shorter than 500 bp. Sonicated products were further
722 purified with MinElute PCR purification kit (QIAGEN) and fragments larger than 300 bp were
723 recovered with Ampure beads. Subsequently, the DNA was mixed with 0.5 µl 10 mM dTTP,
724 0.5 µl 10 mM dATP and 5 U T4 DNA polymerase (Thermo Fisher) in a 50 µl reaction volume
725 and incubated at 20 °C for 30 min to remove biotin from unligated DNA. After DNA recovery
726 with Ampure beads, end repair and adaptor ligation were performed following the manual of
727 NEBNext Ultra II DNA Library Prep Kit (NEB). The ligated DNA was then affinity pulled-down
728 with Dynabeads MyOne Streptavidin C1 beads (Invitrogen), and amplified with 13 PCR cycles.

729 Hi-C libraries were sequenced at the LAFUGA Gene Center Munich on a NextGen 500
730 instrument (Illumina) using 2 x 60 bp paired-end reads. Sequencing and mapping statistics
731 are summarised in **Supplementary Dataset 3**; quality control is shown in **Supplementary**
732 **Figure 7**.

733 Hi-C mapping was performed with the nf-core pipeline nf-core/hic (v2.1.0) (Servant et al. 2023)
734 implemented in nextflow v24.04.2. Briefly, read quality control was performed with FastQC
735 (v0.11.9). Hi-C data processing was done using HiC-Pro (Servant et al. 2015). HiC-pro
736 employs bowtie2 (v2.4.4) to perform the mapping using a two-step strategy that rescues reads
737 spanning the ligation site, following with the detection of valid interaction products and
738 duplicates removal to finally generate the raw contact maps.

739 For all the samples and replicates, we randomly selected the same number of valid pairs (
740 ~41 mio), in reference to the sample with the lowest number of valid pairs (leaf replicate 2).
741 We used pre from juicer tools (1.9.9) (Durand et al. 2016b) to transform the raw contact maps
742 from HiC-pro (.AllValidPairs) into .hic format. Downstream analyses were performed in the R
743 environment (v4.4.1). Strawr (v0.0.92) (Durand et al. 2016a) was used to parse the .hic
744 matrices into R with a resolution of 50,000 bp and normalised with the Knight and Ruiz (KR)
745 method (Knight and Ruiz 2013). Quality control of replicates was performed applying Pearson
746 correlation and representing the relationship of the samples with hierarchical clustering
747 implemented in the R package pheatmap (v1.0.12).

748 After confirming the clustering of the replicates by tissue, we combined all the replicates of
749 one tissue into one sample and re-run the nf-core/hic pipeline, to increase the number of valid
750 pairs per sample. Pre was used to convert the raw contact maps into .hic format. Juicebox
751 (Durand et al., 2016) was used to visually explore the genome-wide contact matrix. Hi-C
752 matrices were parsed into R using strawr. The withanolide BGC locus was explored with KR
753 normalised Hi-C matrices at 20-kb resolution. To compute the PCA for the A/B and sub-A/B
754 compartment annotation, we calculated Pearson's correlation of the KR normalised
755 observed/expected matrices at 500-kb and 20-kb resolution.

756 *Whole-genome bisulfite sequencing and DNA methylation analysis*

757 The identical samples as transcriptomic and metabolomic analyses were used. Genomic DNA
758 (gDNA) was extracted using the DNeasy Plant Mini Kit (Qiagen). Libraries were prepared from
759 100 ng input gDNA as described in (10.1371/journal.pgen.1010452) using the NEBNext Ultra
760 II DNA library preparation kit (New England Biolabs), in combination with the EZ-96 DNA
761 Methylation-Gold MagPrep kit (Zymo Research) for bisulfite conversion. We used 0.01% λ-
762 phage DNA for bisulfite conversion control. Libraries were sequenced at Novogene (UK) on a
763 NovaSeqX platform (Illumina) with 2 x 150 bp reads. Sequencing and mapping statistics are
764 summarised in **Supplementary Dataset 4**; quality control is shown in **Supplementary**
765 **Figures 8 and 9**.

766 WGBS-seq reads quality control and mapping was performed using the nf-core pipeline
767 methylseq (2.6.0) implemented in Nextflow v23.10.0 (Di Tommaso et al. 2017; Ewels et al.

768 2020, 2024), following the bwa-meth (v0.2.2) (Pedersen et al. 2014) and MethylDackel
769 (v0.6.0) (https://github.com/dpryan79/MethylDackel) workflow. Briefly, the first three
770 nucleotides from the 5' end of read 1 and 2 were trimmed. Raw data quality control was
771 performed with FastQC (v0.12.1) and adapter sequence trimming with cutadapt (3.4)
772 implemented in Trimgalore (v0.6.7). After adapter trimming, the first three bases from the 3'
773 end of read 1 and 2 were also trimmed. Reads were aligned with bwa-meth (0.2.2). Duplicates
774 were then marked with picard (v3.0.0). The methylation calls were extracted with
775 MethylDackel (v0.6.0) with a minimum depth of 5 independent reads for MethylDackel to
776 report a methylation call. Alignment quality control (estimation of genome coverage, insert
777 sizes of mapped reads and GC content distribution) was performed with qualimap (v2.2.2-
778 dev) (Okonechnikov et al. 2016). Alignment statistics were evaluated with samtools (v1.17).

779 MethylScore (v0.2) (Hüther et al. 2022) implemented in Netflow version v21.10.6 was used to
780 call differentially methylated regions between all samples with default parameters. Non-
781 conversion rate was calculated for the λ-phage genome as the total of reads supporting
782 methylation for a methylated cytosine divided by the total number of reads. The per-site
783 methylation frequency was calculated based on (Schultz et al. 2012) as the number of reads
784 reporting a methylated base divided by the total reads mapping to that base. To calculate the
785 relative frequency of methylation rates, sites with a coverage ≥10 were considered. Bedtools
786 subcommands map and makewindows (v2.30.0) were used to calculate the mean site
787 methylation in windows along the genome.

788 *Metabolite analysis*

789 Metabolite extraction was performed as described in (Salem et al. 2020) with some minor
790 modifications. Briefly, 1 ml of pre-cooled (−20°C) methyl tert-butyl ether:methanol (3:1 v/v) was
791 added to homogenised tissues. After, tubes were thoroughly vortexed for 1 min and then
792 incubated on an orbital shaker (1000 rpm) for 15 min at 4°C followed by sonication for 15 min.
793 For phase separation, a volume of 500 µl of water:methanol (3:1 v/v) was added to each tube
794 and the samples were again thoroughly vortexed for 1 min. After that, the samples were
795 centrifuged at 13,300 g for 5 min. The dried polar aliquots were resuspended in
796 water:methanol (1:1 v/v) and used for secondary metabolites analysis.

797 Secondary metabolites were profiled by the Waters Acquity ultra performance liquid
798 chromatography (UPLC) system coupled to the Q Exactive Orbitrap mass detector according
799 to the previously published protocol (Giavalisco et al. 2009). The UPLC system was equipped
800 with a HSS T3 C18 reversed-phase column (100 × 2.1 mm i.d., 1.8-µm particle size; Waters,
801 Milford, MA, USA) that was operated at a temperature of 40°C. The mobile phases consisted
802 of 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B). The

803 flow rate of the mobile phase was 400 µl/min, and 3 µl of sample was loaded per injection. The
804 UPLC was connected to a Q Exactive Orbitrap focus (Thermo Fisher Scientific, Waltham, MA,
805 USA) via a heated electrospray source (Thermo Fisher Scientific). The spectra were recorded
806 using full scan and MS2 in both positive and negative ionisation detection, covering a mass
807 range from m/z 100 to 1500. The resolution was set to 25 000, and the maximum scan time
808 was set to 250 msec. The sheath gas was set to a value of 60, while the auxiliary gas was set
809 to 35. The transfer capillary temperature was set to 150°C, while the heater temperature was
810 adjusted to 300°C. The spray voltage was fixed at 3 kV, with a capillary voltage and a skimmer
811 voltage of 25 and 15 V, respectively. Mass spectrometry spectra were recorded from minute
812 0 to 19 of the UPLC gradient. Molecular masses, retention time, and associated peak
813 intensities were extracted from the raw files using refinerms (version 5.3; GeneData), and
814 xcalibur software (Thermo Fisher Scientific). Metabolite identification and annotation were
815 performed using standard compounds, data-dependent method (ddMS2) fragmentation, the
816 literature, and metabolomics databases (Alseekh et al. 2021). To identify additional
817 withanolide features, we searched the LOTUS database (Rutz et al. 2022) for withanolides
818 and downloaded structure information for these. Using the exact mass, we identified
819 withanolide structures in our dataset as those with a match in mass with less than 5 ppm
820 difference to the exact mass of reported structures. Common leaf-specific withanolide features
821 in the different tissues and in the VIGS untargeted metabolomic experiments were defined as
822 features differing ≤0.02 in mass/charge ratio (m/z) and retention time.

823 *Withanolide BGC orthologs detection*

824 OrthoFinder (v2.5.4) (Emms and Kelly, 2019, 2015) was used to infer orthogroups containing
825 *P. grisea* withanolide BGC genes. Each orthogroup was aligned using MAFFT (v7.490) (Katoh
826 et al. 2002; Katoh and Standley 2013). IQTREE (v2.1.4-beta) (Minh et al. 2020) was used to
827 infer the maximum likelihood (ML) tree with the best-fit model selected by ModelFinder
828 (Kalyaanamoorthy et al. 2017) with 1000 replicates of ultrafast bootstrapping (Hoang et al.
829 2018). Trees were represented with Dendroscope (v3.8.2) (**Supplementary Figures 12-20**).
830 Genes within the same clade as *P. grisea* withanolide BGC genes were considered as
831 potentially clustered homologs. The genomic location upstream and downstream of these
832 homologs was further analyzed to confirm clustering.

833 Microsynteny analysis and figures were done with MCScan implemented in Python (v3.9) with
834 JCVI utility libraries (v1.1.11) following the package workflow:
835 https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version). Briefly, protein FASTA files
836 and GFF3 files of each species were used as input. When needed, the FASTA files were
837 processed using Biopython's SeqIO package (Cock et al. 2009). Pairwise ortholog and
838 synteny blocks search was performed with default settings using a lift-over distance of 5, a c-

839 score = 0.99 and a maximum of 3 iterations (6 in the case of *T. anomalum*). Multi-synteny
840 blocks were constructed combining the syntenic blocks detected on every species used with
841 respect to *P. grisea*.

842 *Data availability*

843 The code used for all analyses will be made available upon final publication of the manuscript
844 (https://github.com/spriego/Priego-Cubero_and_Knoch_et_al_Physalis). Sequencing data
845 will be deposited in the ENA Short Read Archive (www.ebi.ac.uk/ena/) upon final publication
846 under project no. PRJEB80835, submission no. ERA30854395.

847 Author contributions

848 S.P.C., E.K., and C.B. conceived the study. S.P.C. carried out the genomic, epigenomic,
849 transcriptomic, and phylogenomic analyses. E.K. carried out metabolomic analysis, and
850 together with K.H.B and P.C. carried out molecular cloning and VIGS. Z.W. prepared Hi-C
851 samples and analysed the data. S.A. prepared analytical samples and analysed metabolic
852 profiles. A.R.F., C.L. and C.B. supervised the work. S.P.C. and C.B. wrote the manuscript with
853 contributions from all authors.

854 Acknowledgements

855 We would like to thank Niklas Schandry, Liza Rouyer and Peter Kerpan for critical reading of
856 the manuscript and constructive feedback, and Zhihui Bao and Arlen Slaymaker for support
857 in the laboratory work. We thank Simon Pfanzelt (Staatliche Naturwissenschaftliche
858 Sammlungen Bayerns) for plant material of *E. australis* and *I. arborescens*, Sebastian Soyk
859 (University of Lausanne) for providing *P. grisea* seeds, David Nelson (University of
860 Tennessee) for naming the P450s, Norman Warthman (IAEA Vienna) for sharing the
861 tagmentation protocol, Robert J. Schmitz (University of Georgia) for sharing the ATAC-seq
862 protocol, and the LMU Biocenter Genome Centre for technical support. We acknowledge the
863 Graduate School Life Science Munich (LSM) for their valuable support to S.P.C. This work
864 was funded by the European Union's Horizon 2020 research and innovation programme by
865 the European Research Council (ERC), Grant Agreement No. 716823 'FEAR-SAP' (C.B.); by
866 the Deutsche Forschungsgemeinschaft (DFG) in the frame of TRR356 'Plant-Microbe' (grant
867 no. 491090170) and the project 'Momidiverse' (grant no. 457739273) to C.B.; and by the
868 LMUexcellent programme by the Ludwig Maximilians Universität München (C.B., E.K.). For
869 computational analyses, we used the BioHPC-Genomics compute cluster at the Leibniz
870 Rechenzentrum (DFG; grant no. 450674345). S.A. and A.R.F. acknowledge the European

Union's Horizon 2020 research and innovation programme, project PlantaSYST (SGA-CSA No: 739582 under FPA No: 664620) and the BG05M2OP001-1.003-001-C01 project, financed by the European Regional Development Fund through the Bulgarian ‘Science and Education for Smart Growth’ Operational Programme,

Declaration of conflict of interest

The authors declare no conflict of interest.

References

Akiyama R, Umemoto N, and Mizutani M. Recent advances in steroidal glycoalkaloid biosynthesis in the genus *Solanum*. Plant Biotechnol Tokyo. 2023:**40**(3):185–191. https://doi.org/10.5511/plantbiotechnology.23.0717b

Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, C. Ewald J, Fraser PD, Giavalisco P, Hall RD, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods. 2021:**18**(7):747–756. https://doi.org/10.1038/s41592-021-01197-1

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, and Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011:**27**(12):1691–1692. https://doi.org/10.1093/bioinformatics/btr174

de Bernonville TD, Maury S, Delaunay A, Daviaud C, Chaparro C, Tost J, O'Connor SE, and Courdavault V. Developmental methylome of the medicinal plant *Catharanthus roseus* unravels the tissue-specific control of the monoterpene indole alkaloid pathway by DNA methylation. Int J Mol Sci. 2020:**21**(17):6028. https://doi.org/10.3390/ijms21176028

Bryson AE, Lanier ER, Lau KH, Hamilton JP, Vaillancourt B, Mathieu D, Yocca AE, Miller GP, Edger PP, Buell CR, et al. Uncovering a miltiradiene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory. Nat Commun. 2023:**14**(1):343. https://doi.org/10.1038/s41467-023-35845-1

Cárdenas PD, Almeida A, and Bak S. Evolution of Structural Diversity of Triterpenoids. Front Plant Sci. 2019:**10**:1523. https://doi.org/10.3389/fpls.2019.01523

Cigler M, Imrichova H, Frommelt F, Caramelle L, Depta L, Rukavina A, Kagiou C, Hannich JT, Mayor-Ruiz C, Superti-Furga G, et al. Orpinolide disrupts a leukemic dependency on cholesterol transport by inhibiting OSBP. Nat Chem Biol. 2024:1–10. https://doi.org/10.1038/s41589-024-01614-4

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinforma Oxf Engl. 2009:**25**(11):1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Conneely LJ, Berkowitz O, and Lewsey MG. Emerging trends in genomic and epigenomic regulation of plant specialised metabolism. Phytochemistry. 2022:**203**:113427. https://doi.org/10.1016/j.phytochem.2022.113427

Dale SM, Tomaszewski E, Lippman Z, and Van Eck J. Engineering the future of *Physalis grisea*: A focus on agricultural challenges, model species status, and applied improvements. PLANTS PEOPLE PLANET. 2024:**n/a**(n/a). https://doi.org/10.1002/ppp3.10536

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:**10**(2):giab008. https://doi.org/10.1093/gigascience/giab008

Dhar N, Razdan S, Rana S, Bhat WW, Vishwakarma R, and Lattoo SK. A Decade of Molecular Understanding of Withanolide Biosynthesis and In vitro Studies in Withania somnifera (L.) Dunal: Prospects and Perspectives for Pathway Engineering. Front Plant Sci. 2015:**6**.

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, and Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017:**35**(4):316–319. https://doi.org/10.1038/nbt.3820

Dinesh-Kumar SP, Anandalakshmi R, Marathe R, Schiff M, and Liu Y. Virus-induced gene silencing. . In. Plant Functional Genomics, E Grotewold, ed. (Humana Press: Totowa, NJ), pp. 287–293. https://doi.org/10.1385/1-59259-413-1:287

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013:**29**(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Doğan ES and Liu C. Three-dimensional chromatin packing and positioning of plant genomes. Nat Plants. 2018:**4**(8):521–529. https://doi.org/10.1038/s41477-018-0199-5

Dong P, Tu X, Chu P-Y, Lü P, Zhu N, Grierson D, Du B, Li P, and Zhong S. 3d chromatin architecture of large plant genomes determined by local A/B compartments. Mol Plant. 2017:**10**(12):1497–1509. https://doi.org/10.1016/j.molp.2017.11.005

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, and Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016a:**3**(1):99–101. https://doi.org/10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, and Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b:**3**(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002

Elliger CA, Waiss AC, Wong RY, and Benson M. Petuniasterones from *Petunia parodii* and *P. integrifolia*; unusual ergostane-type steroids. Phytochemistry. 1989:**28**(12):3443–3452. https://doi.org/10.1016/0031-9422(89)80362-0

Emms DM and Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:**20**(1):238. https://doi.org/10.1186/s13059-019-1832-y

Ewels P, Hüther P, Miller E, Sateesh_Peri, Spix N, bot nf-core, Peltzer A, F S, Alneberg J, Garcia MU, et al. nf-core/methylseq: Huggy mollusc. 2024. https://doi.org/10.5281/zenodo.10463781

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, and Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020:**38**(3):276–278. https://doi.org/10.1038/s41587-020-0439-x

Gao F and Dubos C. The Arabidopsis bHLH transcription factor family. Trends Plant Sci. 2024:**29**(6):668–680. https://doi.org/10.1016/j.tplants.2023.11.022

Giavalisco P, Köhl K, Hummel J, Seiwert B, and Willmitzer L. 13c isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. Anal Chem. 2009:**81**(15):6546–6551. https://doi.org/10.1021/ac900979e

Goossens J, Mertens J, and Goossens A. Role and functioning of bHLH transcription factors in jasmonate signalling. J Exp Bot. 2017:**68**(6):1333–1347. https://doi.org/10.1093/jxb/erw440

Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, Teodor R, Lu Y, Bowser TA, Graham IA, et al. The opium poppy genome and morphinan production. Science. 2018:**362**(6412):343–347. https://doi.org/10.1126/science.aat4096

Hansel R, Huang JT, and Rosenberg D. Zwei Withanolide aus *Lycium chinense*. Arch Pharm (Weinheim). 1975:**308**(8):653–654. https://doi.org/10.1002/ardp.19753080814

He J, Alonge M, Ramakrishnan S, Benoit M, Soyk S, Reem NT, Hendelman A, Van Eck J, Schatz MC, and Lippman ZB. Establishing *Physalis* as a *Solanaceae* model system enables genetic reevaluation of the inflated calyx syndrome. Plant Cell. 2023:**35**(1):351–368. https://doi.org/10.1093/plcell/koac305

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010:**38**(4):576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, and Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018:**35**(2):518–522. https://doi.org/10.1093/molbev/msx281

Huang J, Xu W, Zhai J, Hu Y, Guo J, Zhang C, Zhao Y, Zhang L, Martine C, Ma H, et al. Nuclear phylogeny and insights into whole-genome duplications and reproductive development of *Solanaceae* plants. Plant Commun. 2023:**4**(4):100595. https://doi.org/10.1016/j.xplc.2023.100595

Huang M, He J-X, Hu H-X, Zhang K, Wang X-N, Zhao B-B, Lou H-X, Ren D-M, and Shen T. Withanolides from the genus *Physalis*: a review on their phytochemical and pharmacological aspects. J Pharm Pharmacol. 2020:**72**(5):649–669. https://doi.org/10.1111/jphp.13209

Hüther P, Hagmann J, Nunn A, Kakoulidou I, Pisupati R, Langenberger D, Weigel D, Johannes F, Schultheiss SJ, and Becker C. MethylScore, a pipeline for accurate and context-aware identification of differentially methylated regions from population-scale plant whole-genome bisulfite sequencing data. Quant Plant Biol. 2022:**3**:e19. https://doi.org/10.1017/qpb.2022.14

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, and Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:**14**(6):587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Misawa K, Kuma K, and Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. NAR. 2002:**30**(14):3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:**30**(4):772–780. https://doi.org/10.1093/molbev/mst010

Kent WJ, Zweig AS, Barber G, Hinrichs AS, and Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics. 2010:**26**(17):2204–2207. https://doi.org/10.1093/bioinformatics/btq351

Kim J, Kang S-H, Park S-G, Yang T-J, Lee Y, Kim OT, Chung O, Lee J, Choi J-P, Kwon S-J, et al. Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in *Platycodon grandiflorus*, a medicinal plant. Hort Res. 2020:**7**:112. https://doi.org/10.1038/s41438-020-0329-x

Knight PA and Ruiz D. A fast algorithm for matrix balancing. IMA J Numer Anal. 2013:**33**(3):1029–1047. https://doi.org/10.1093/imanum/drs019

Knoch E, Sugawara S, Mori T, Poulsen C, Fukushima A, Harholt J, Fujimoto Y, Umemoto N, and Saito K. Third DWF1 paralog in *Solanaceae*, sterol Δ24-isomerase, branches withanolide biosynthesis from the general phytosterol pathway. Proc Natl Acad Sci USA. 2018:**115**(34):E8096–E8103. https://doi.org/10.1073/pnas.1807482115

Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008:**9**. https://doi.org/10.1186/1471-2105-9-559

Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012:**9**(4):357–359. https://doi.org/10.1038/nmeth.1923

Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D, Van Eck J, and Lippman ZB. Rapid improvement of domestication traits in an orphan crop by genome editing. Nat Plants. 2018:**4**(10):766–770. https://doi.org/10.1038/s41477-018-0259-x

Li C, Wood JC, Vu AH, Hamilton JP, Rodriguez Lopez CE, Payne RME, Serna Guerrero DA, Gase K, Yamamoto K, Vaillancourt B, et al. Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. Nat Chem Biol. 2023:**19**(8):1031–1041. https://doi.org/10.1038/s41589-023-01327-0

Li W-F, Ning G-X, Mao J, Guo Z-G, Zhou Q, and Chen B-H. Whole-genome DNA methylation patterns and complex associations with gene expression associated with anthocyanin biosynthesis in apple fruit skin. Planta. 2019:**250**(6):1833–1847. https://doi.org/10.1007/s00425-019-03266-4

Li Y, Chen Y, Zhou L, You S, Deng H, Chen Y, Alseekh S, Yuan Y, Fu R, Zhang Z, et al. Microtom metabolic network: rewiring tomato metabolic regulatory network throughout the growth cycle. Mol Plant. 2020:**13**(8):1203–1218. https://doi.org/10.1016/j.molp.2020.06.005

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009:**326**(5950):289–293. https://doi.org/10.1126/science.1181369

Liu C, Smit SJ, Dang J, Zhou P, Godden GT, Jiang Z, Liu W, Liu L, Lin W, Duan J, et al. A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in <i>Schizonepeta tenuifolia<i>. Mol Plant. 2023:**16**(3):533–548. https://doi.org/10.1016/j.molp.2023.01.004

Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, Paajanen P, Yant L, and Osbourn A. Formation and diversification of a paradigm biosynthetic gene cluster in plants. Nat Commun. 2020:**11**(1):5354. https://doi.org/10.1038/s41467-020-19153-6

Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014:**15**(12):550. https://doi.org/10.1186/s13059-014-0550-8

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011:**17**(1):10–12. https://doi.org/10.14806/ej.17.1.200

Méteignier L-V, Nützmann H-W, Papon N, Osbourn A, and Courdavault V. Emerging mechanistic insights into the regulation of specialized metabolism in plants. Nat Plants. 2022:1–9. https://doi.org/10.1038/s41477-022-01288-7

Méteignier L-V, Nützmann H-W, Papon N, Osbourn A, and Courdavault V. Emerging mechanistic insights into the regulation of specialized metabolism in plants. Nat Plants. 2023:**9**(1):22–30. https://doi.org/10.1038/s41477-022-01288-7

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, and Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:**37**(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Misico RI, Nicotra VE, Oberti JC, Barboza G, Gil RR, and Burton G. Withanolides and Related Steroids. . In. Progress in the Chemistry of Organic Natural Products Vol. 94, AD Kinghorn, H Falk, and J Kobayashi, eds, Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products. (Springer Vienna: Vienna), pp. 127–229. https://doi.org/10.1007/978-3-7091-0748-5_3

Nützmann H-W, Doerr D, Ramírez-Colmenero A, Sotelo-Fonseca JE, Wegel E, Di Stefano M, Wingett SW, Fraser P, Hurst L, Fernandez-Valverde SL, et al. Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. Proc Natl Acad Sci USA. 2020:**117**(24):13800–13809. https://doi.org/10.1073/pnas.1920474117

Nützmann H-W and Osbourn A. Regulation of metabolic gene clusters in *Arabidopsis thaliana*. New Phytol. 2015:**205**(2):503–510. https://doi.org/10.1111/nph.13189

Okonechnikov K, Conesa A, and García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016:**32**(2):292–294. https://doi.org/10.1093/bioinformatics/btv566

Pan J, Liu Y, Li X-Y, Wang S-Y, Wu J-T, Li M-M, Guan W, Mohammed Algradi A, Kuang H-X, and Yang B-Y. Cytotoxic withanolides from the leaves of *Datura stramonium* L. Chem Biodiv. 2024:**21**(2):e202301655. https://doi.org/10.1002/cbdv.202301655

Patel H, Espinosa-Carrasco J, Langer B, Ewels P, bot nf-core, Garcia MU, Syme R, Peltzer A, Talbot A, Behrens D, et al. nf-core/atacseq: [2.1.2] - 2022-08-07. 2023. https://doi.org/10.5281/zenodo.8222875

Patro R, Duggal G, Love MI, Irizarry RA, and Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017:**14**(4):417–419. https://doi.org/10.1038/nmeth.4197

Pedersen BS, Eyring K, De S, Yang IV, and Schwartz DA. Fast and accurate alignment of long bisulfite-seq reads. ArXiv14011129 Q-Bio. 2014.

Polturak G, Liu Z, and Osbourn A. New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. Curr Opin Green Sustain Chem. 2022:**33**:100568. https://doi.org/10.1016/j.cogsc.2021.100568

Polturak G and Osbourn A. The emerging role of biosynthetic gene clusters in plant defense and plant interactions. PLOS Pathog. 2021:**17**(7):e1009698. https://doi.org/10.1371/journal.ppat.1009698

Powell AF, Zhang J, Hauser D, Vilela JA, Hu A, Gates DJ, Mueller LA, Li F-W, Strickler SR, and Smith SD. Genome sequence for the blue-flowered Andean shrub *Iochroma cyaneum* reveals extensive discordance across the berry clade of *Solanaceae*. Plant Genome. 2022:**15**(3):e20223. https://doi.org/10.1002/tpg2.20223

Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:**26**(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. NAR. 2016:**44**(W1):W160–W165. https://doi.org/10.1093/nar/gkw257

Rana S, Bhat WW, Dhar N, Pandith SA, Razdan S, Vishwakarma R, and Lattoo SK. Molecular characterization of two A-type P450s, WsCYP98A and WsCYP76A from *Withania somnifera* (L.) Dunal: expression analysis and withanolide accumulation in response to exogenous elicitations. BMC Biotechnol. 2014:**14**(1):89. https://doi.org/10.1186/s12896-014-0089-5

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011:**29**(1):24–26. https://doi.org/10.1038/nbt.1754

Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, Hirata TDC, Lima DS, Passos FM, Fukutani KF, et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics. 2018:**19**(1):56. https://doi.org/10.1186/s12859-018-2053-1

Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, Graham JG, Stephan R, Page R, Vondrášek J, et al. The LOTUS initiative for open knowledge management in natural products research. eLife. 2022:**11**:e70780. https://doi.org/10.7554/eLife.70780

Salem MA, Yoshida T, Perez de Souza L, Alseekh S, Bajdzienko K, Fernie AR, and Giavalisco P. An improved extraction method enables the comprehensive analysis of lipids, proteins, metabolites and phytohormones from a single sample of leaf tissue under water-deficit stress. Plant J. 2020:**103**(4):1614–1632. https://doi.org/10.1111/tpj.14800

Sasse J, Schlegel M, Borghi L, Ullrich F, Lee M, Liu G-W, Giner J-L, Kayser O, Bigler L, Martinoia E, et al. Petunia hybrida PDR2 is involved in herbivore defense by controlling steroidal contents in trichomes. Plant Cell Env. 2016:**39**(12):2725–2739. https://doi.org/10.1111/pce.12828

Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T, Yamamoto T, Takebayashi Y, Kojima M, Sakakibara H, Aoki T, et al. Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. Plant Cell. 2014:**26**(9):3763–3774. https://doi.org/10.1105/tpc.114.130096

Schultz MD, Schmitz RJ, and Ecker JR. "Leveling" the playing field for analyses of single-base resolution DNA methylomes. Trends Genet. 2012:**28**(12):583–585. https://doi.org/10.1016/j.tig.2012.10.012

Servant N, bot nf-core, Ewels P, Garcia MU, Talbot A, Peltzer A, Miller E, Rossini R, Zohren J, Menden K, et al. nf-core/hic: nf-core/hic v2.1.0. 2023. https://doi.org/10.5281/zenodo.7994878

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, and Barillot E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015:**16**(1):259. https://doi.org/10.1186/s13059-015-0831-x

Shoji T, Sugawara S, Mori T, Kobayashi M, Kusano M, and Saito K. Induced production of specialized steroids by transcriptional reprogramming in *Petunia hybrida*. PNAS Nexus. 2023:**2**(10):pgad326. https://doi.org/10.1093/pnasnexus/pgad326

Sikorskaite S, Rajamäki M-L, Baniulis D, Stanys V, and Valkonen JP. Protocol: Optimised methodology for isolation of nuclei from leaves of species in the *Solanaceae* and *Rosaceae* families. Plant Methods. 2013:**9**(1):31. https://doi.org/10.1186/1746-4811-9-31

Smit SJ and Lichman BR. Plant biosynthetic gene clusters in the context of metabolic evolution. Nat Prod Rep. 2022:**39**(7):1465–1482. https://doi.org/10.1039/D2NP00005A

Tarbell ED and Liu T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. NAR. 2019:**47**(16):e91. https://doi.org/10.1093/nar/gkz533

Van Eck J. The Physalis Improvement Project: blending research with community science. EMBO Rep. 2022:**23**(1):e53918. https://doi.org/10.15252/embr.202153918

Wang N, Wang Z, Tzourtzou S, Wang X, Bi X, Leimeister J, Xu L, Sakamoto T, Matsunaga S, Schaller A, et al. The plant nuclear lamina disassembles to regulate genome folding in stress conditions. Nat Plants. 2023:**9**(7):1081–1093. https://doi.org/10.1038/s41477-023-01457-2

Wu Y, Li D, Hu Y, Li H, Ramstein GP, Zhou S, Zhang X, Bao Z, Zhang Y, Song B, et al. Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. Cell. 2023:**186**(11):2313-2328.e15. https://doi.org/10.1016/j.cell.2023.04.008

Xia G, Cao S, Chen L, and Qiu F. Natural withanolides, an update. Nat Prod Rep. 2022:**39**(4):784–813. https://doi.org/10.1039/D1NP00055A

Yang J-Y, Chen C-X, Zhao R-H, Hao X-J, and Liu H-Y. Plantagiolide F, a minor withanolide from Tacca plantaginea. Nat Prod Res. 2011:**25**(1):40–44. https://doi.org/10.1080/14786411003774304

Yang X, Gao S, Guo L, Wang B, Jia Y, Zhou J, Che Y, Jia P, Lin J, Xu T, et al. Three chromosome-scale Papaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nat Commun. 2021:**12**(1):6030. https://doi.org/10.1038/s41467-021-26330-8

Zhang B and Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005:**4**(1). https://doi.org/10.2202/1544-6115.1128

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008:**9**(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang Y-Y, Latzel V, Fischer M, and Bossdorf O. Understanding the evolutionary potential of epigenetic variation: a comparison of heritable phenotypic variation in epiRILs, RILs, and natural ecotypes of *Arabidopsis thaliana*. Heredity. 2018:**121**(3):257–265. https://doi.org/10.1038/s41437-018-0095-9

Zhang Z, Yang Y, Xu Y, Liu Y, Li H, and Chen L. Molecular targets and mechanisms of anti-cancer effects of withanolides. Chem-Biol Interact. 2023:**384**:110698. https://doi.org/10.1016/j.cbi.2023.110698

Zhao H, Yang M, Bishop J, Teng Y, Cao Y, Beall BD, Li S, Liu T, Fang Q, Fang C, et al. Identification and functional validation of super-enhancers in Arabidopsis thaliana. Proc Natl Acad Sci. 2022:**119**(48):e2215328119. https://doi.org/10.1073/pnas.2215328119

Zhou L, Huang Y, Wang Q, and Guo D. Chromatin accessibility is associated with artemisinin biosynthesis regulation in *Artemisia annua*. Molecules. 2021:**26**(4):1194. https://doi.org/10.3390/molecules26041194

1132 # Supplementary Material

1133 **Supplementary Figure 1. Overview of tissue selection for sequencing and metabolic**
1134 **analysis**

1135 **Supplementary Figure 2. RNA-seq quality control and weighted gene correlation**
1136 **network analysis.**

1137 **Supplementary Figure 3. VIGS-induced silencing of biosynthetic cluster genes.**

1138 **Supplementary Figure 4. ATAC-seq quality control.**

1139 **Supplementary Figure 5. ATAC-seq metaplots of leaf samples.**

1140 **Supplementary Figure 6. ATAC-seq metaplots of root samples.**

1141 **Supplementary Figure 7. Chromosomal conformation capture by Hi-C in *P. grisea*.**

1142 **Supplementary Figure 8. DNA methylation in *P. grisea*.**

1143 **Supplementary Figure 9. WGBS quality control.**

1144 **Supplementary Figure 10. Phylogenomic comparison of the withanolide BGC across**
1145 ***Solanaceae*, drawn to scale.**

1146 **Supplementary Figure 11. Expression of an orthologous BGC in *Physalideae* species.**

1147 **Supplementary Figure 12. Phylogenetic analysis based on CYP87G1 amino acid**
1148 **sequence.**

1149 **Supplementary Figure 13. Phylogenetic analysis based on the CYP88C subfamily**
1150 **members (CYP88C7, CYP88C8 and CYP88C10) amino acid sequence.**

1151 **Supplementary Figure 14. Phylogenetic analysis based on the CYP749B2 amino acid**
1152 **sequence.**

1153 **Supplementary Figure 15. Phylogenetic analysis based on the sulfotransferases (Sulfa**
1154 **and Sulfb) amino acid sequence.**

1155 **Supplementary Figure 16. Phylogenetic analysis based on the 2ODD amino acid**
1156 **sequences found in the first larger 2ODD orthogroup (2ODD1).**

1157 **Supplementary Figure 17. Phylogenetic analysis based on the 2ODD amino acid**
1158 **sequences found in the second larger 2ODD orthogroup (2ODD2).**

1159 **Supplementary Figure 18. Phylogenetic analysis based on the 2ODD amino acid**
1160 **sequences found in the smaller 2ODD orthogroup (2ODD3).**

1161 **Supplementary Figure 19. Phylogenetic analysis based on the SDRs (SDRa and SDRb)**
1162 **amino acid sequences.**

1163 **Supplementary Figure 20. Phylogenetic analysis based on the ASAT amino acid**
1164 **sequences.**

1165 **Supplementary Figure 21. Quality control analysis of the metabolomic datasets.**

1166 **Supplementary Table 1. Sequence identities and similarities between paralogs within**
1167 **the withanolide biosynthetic gene cluster in *P. grisea*.**

1168 **Supplementary Table 2. Genome assemblies used for genome mining of withanolide**
1169 **biosynthesis genes.**

1170 **Supplementary Table 3. Top 10 Homer Known Motif Enrichment Results for RESC and**
1171 **SESC peaks in root and leaf, respectively.**

1172 **Supplementary Table 4. Primers used for VIGS and qPCR.**

1173 **Supplementary Dataset 1. Sequence statistics of the RNA-seq dataset.**

1174 **Supplementary Dataset 2. Sequence statistics of the ATAC-seq dataset.**

1175 **Supplementary Dataset 3. Sequence statistics of the Hi-C-seq dataset.**

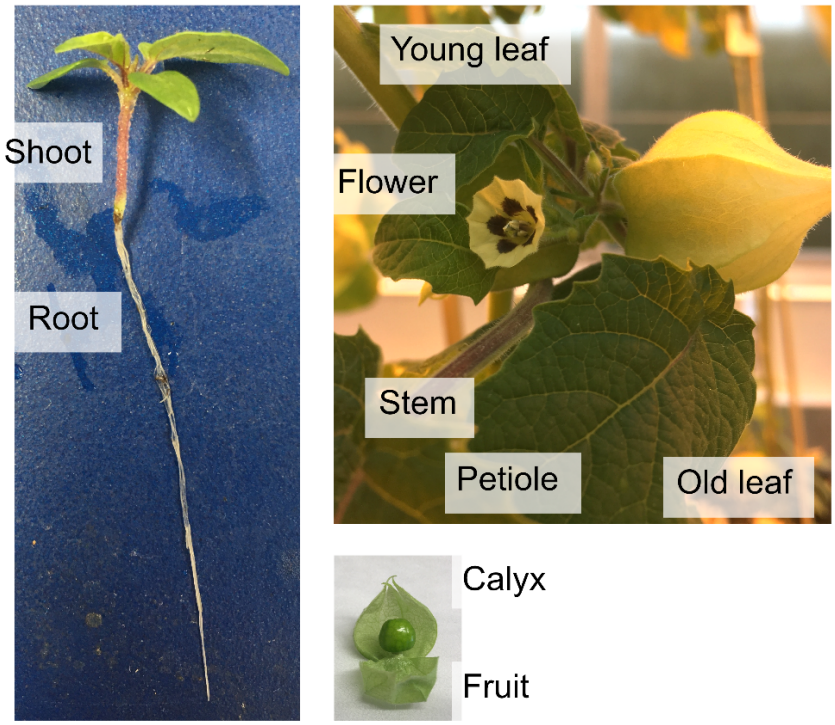
1176 **Supplementary Dataset 4. Sequence statistics of the WGBS dataset.**

1177

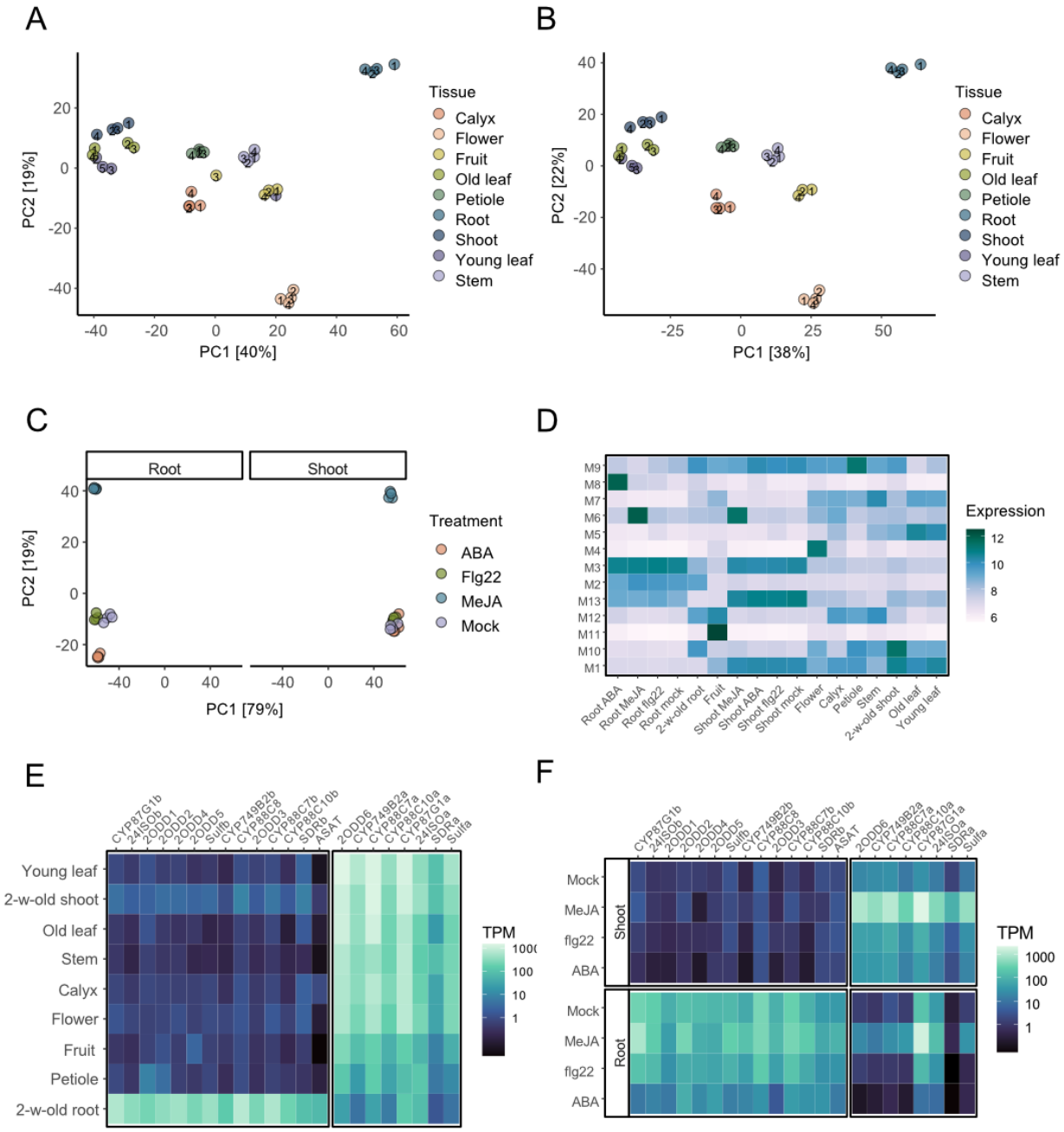
1178

1179

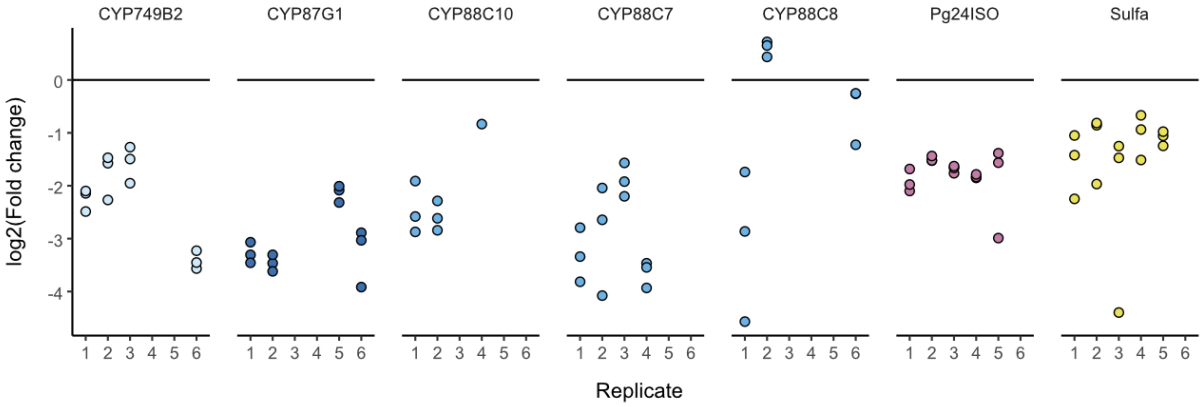
Supplementary Figures



Supplementary Figure 1. Overview of tissue selection for sequencing and metabolic analysis. Nine different tissues were selected for transcriptomic and metabolic analysis: root and shoot of 2-week-old seedlings (left); young leaf, mature leaf, petiole, stem, flower (open), calyx and fruit from adult plants (right). The same root and shoot samples were moreover used for WGBS library preparation.



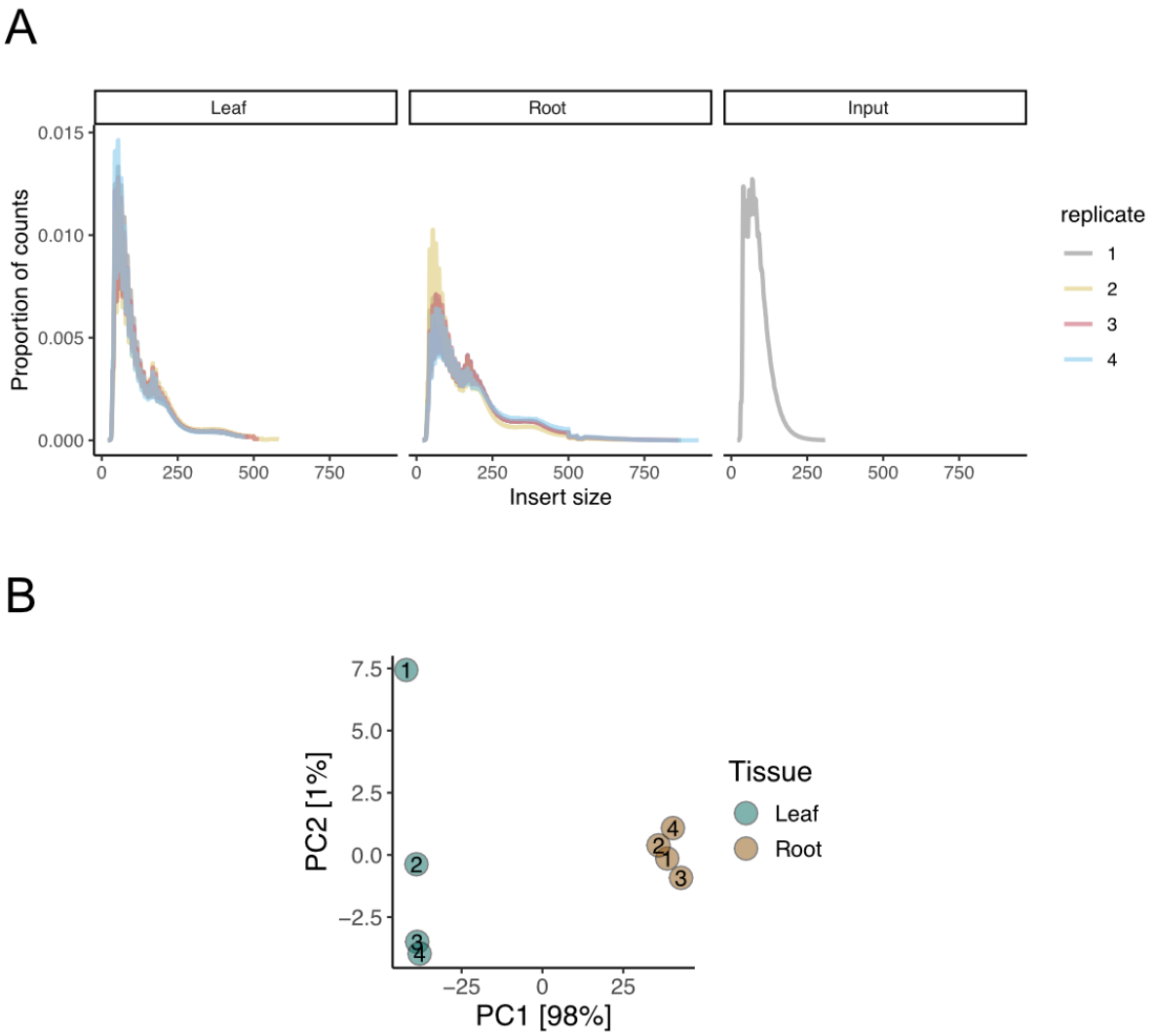
Supplementary Figure 2. RNA-seq quality control and weighted gene correlation network analysis. Principal components analysis of **A)** 'tissues' RNA-seq before removing outliers, **B)** 'tissues' RNA-seq after removing two outlier samples, and **C)** 'elicitors' RNA-seq experiments using the top 500 genes with the highest variance after variance-stabilising transformation using the DESeq2 package (Love et al. 2014). **D)** Mean expression of the genes in each co-expression module (M1-M13) detected in the different samples (x-axis). The expression of each sample is the mean of four replicates (three replicates in the case of 'Fruit' and 'Small leaf', after outlier removal). The expression was normalised using the variance stabilising transformation. **E)** Heatmap showing the expression pattern of each gene in the withanolide BGC across different tissues. Expression levels are presented as transcripts per million (TPM). **F)** Heatmap showing expression in TPM of the genes within the cluster upon mock treatment and in response to various elicitors: methyl jasmonate (MeJA), flagellin 22 (Flg22), and abscisic acid (ABA).



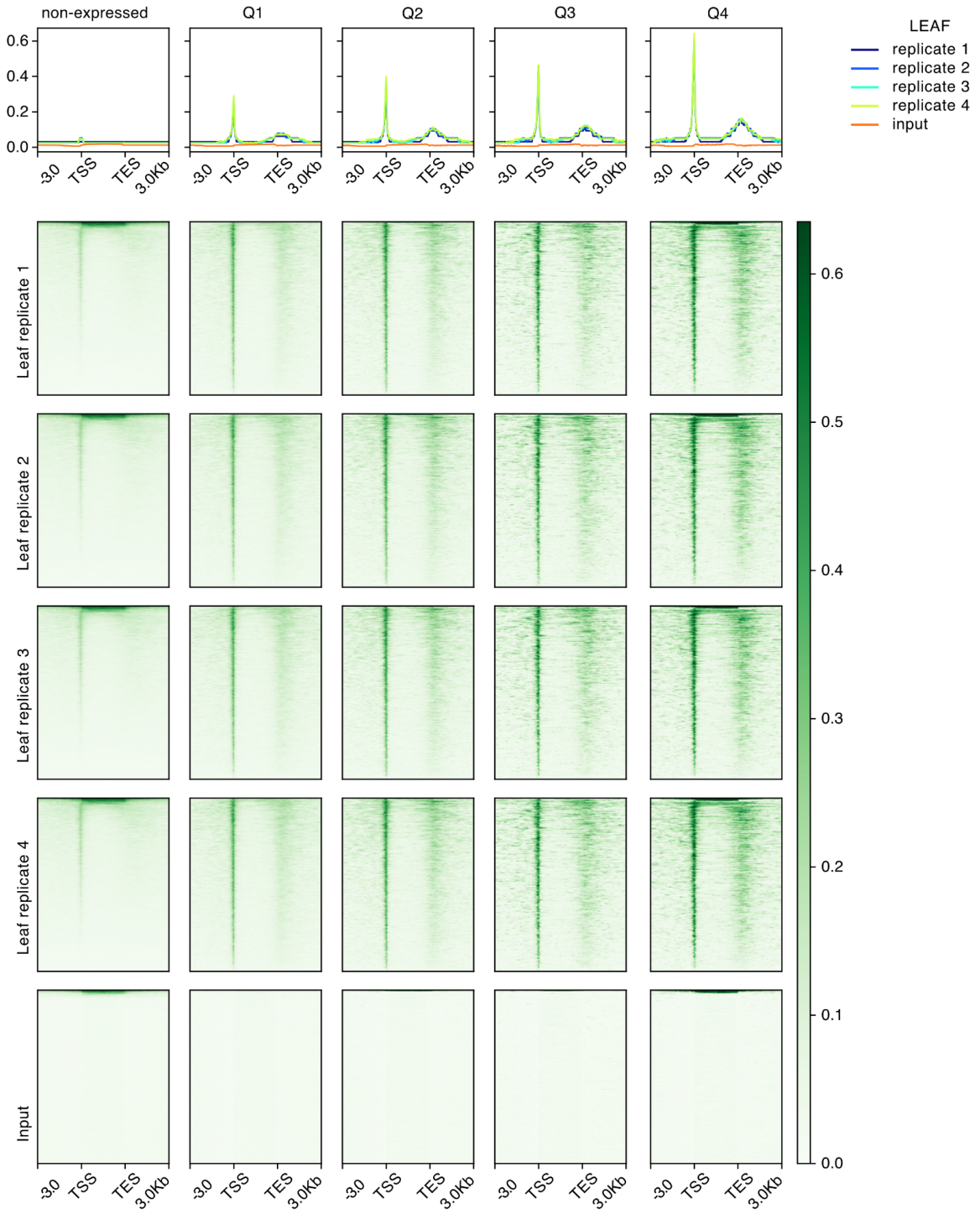
1200

1201 **Supplementary Figure 3. VIGS-induced silencing of biosynthetic cluster genes.** qPCR was performed on
1202 VIGS-infiltrated *N. benthamiana* leaves in six biological replicates (1 to 6), with three technical replicates each, for
1203 the genes indicated at the top. Fold-changes are expressed in relation to an anti-GFP control (mock); all data were
1204 normalised against expression of a *TUBULIN* reference gene. Data is shown for those replicates that were
1205 selected, based on qPCR results, for the metabolic analysis in **Figure 1E**.

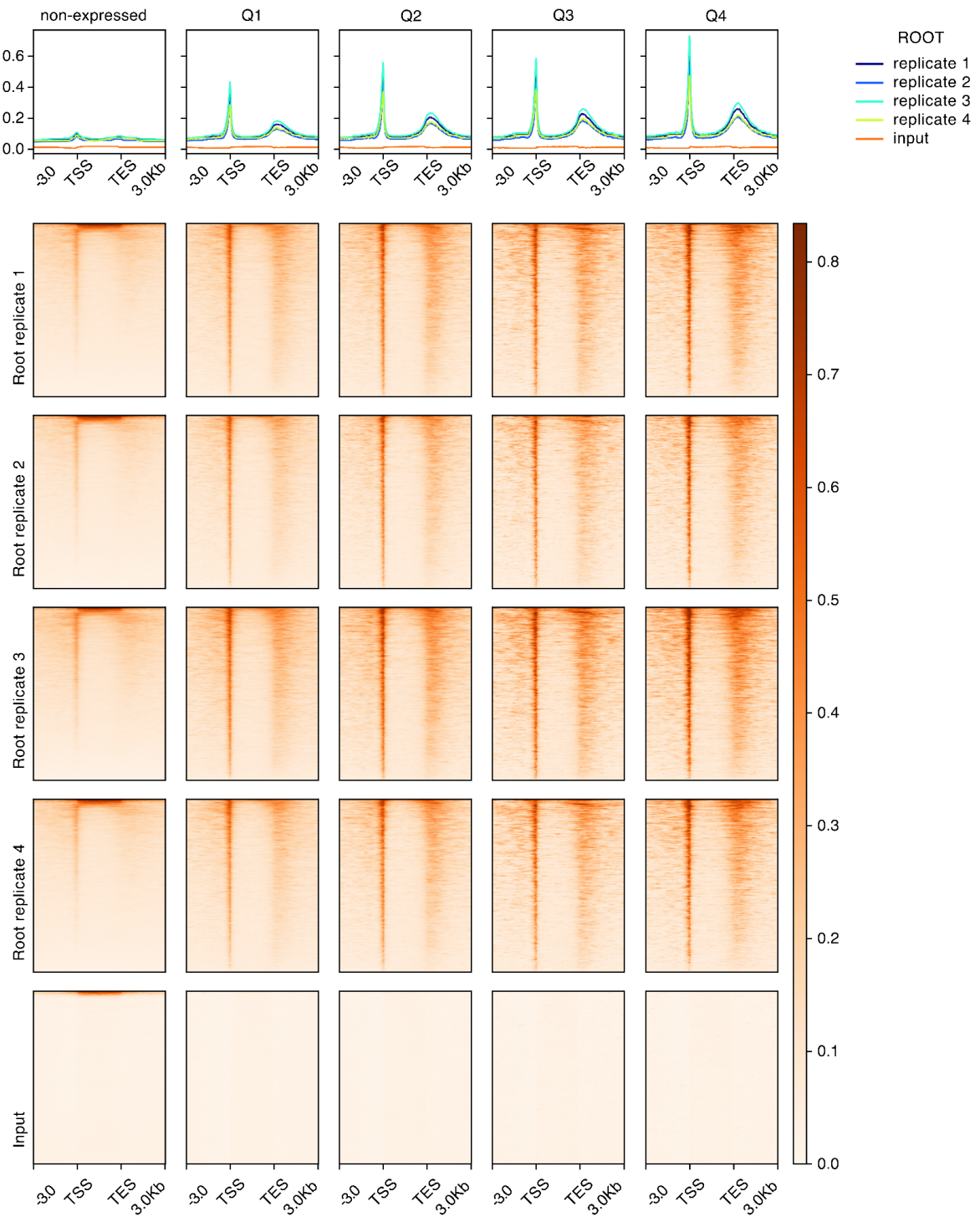
1206



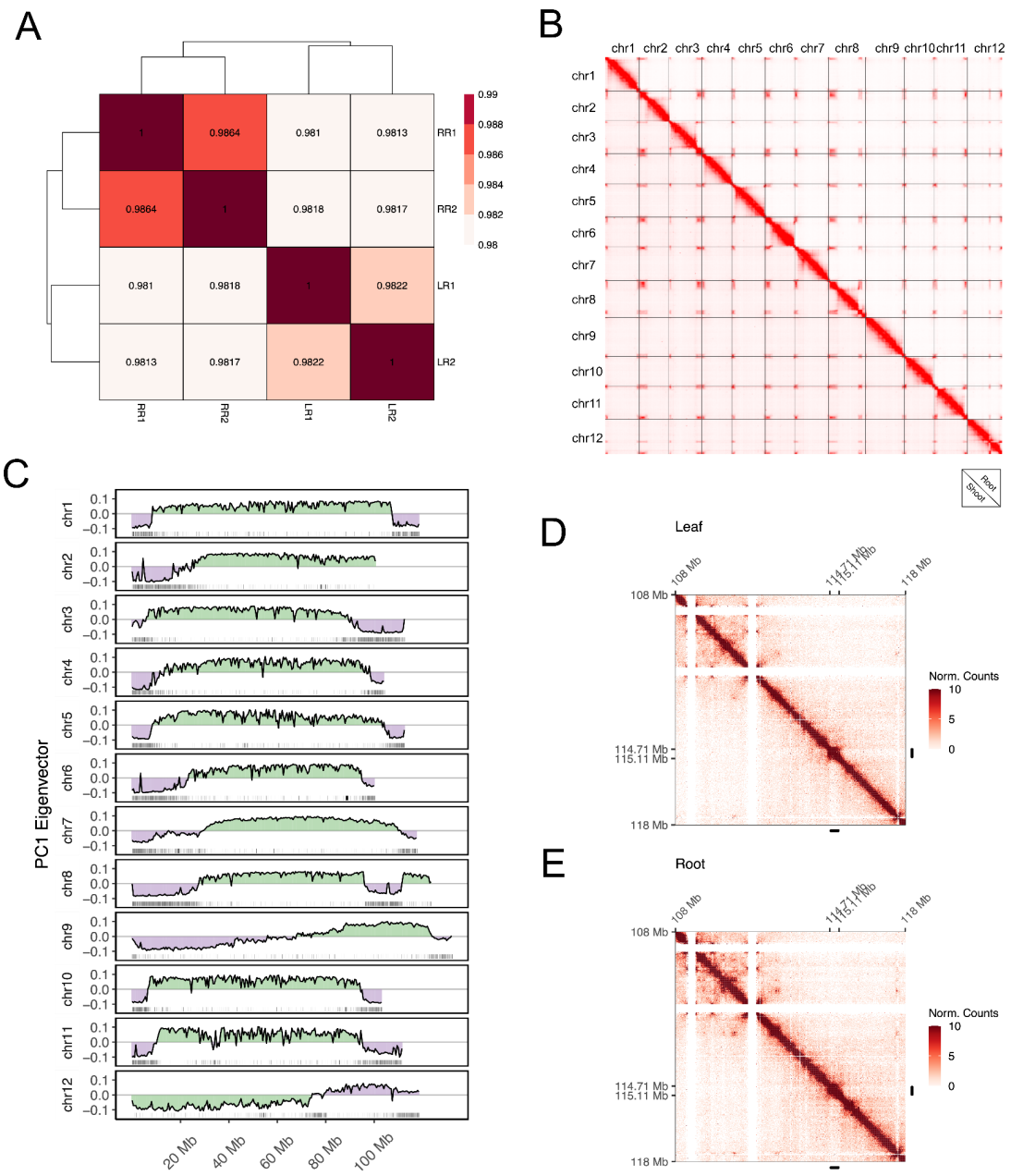
Supplementary Figure 4. ATAC-seq quality control. A) Insert size (bp) distribution of ATAC-seq paired-end reads in leaf, root and input samples. **B)** Principal components analysis of ATAC-seq samples using the top 500 consensus 'narrow peaks' with the highest variance after variance-stabilising transformation using the DESeq2 package (Love et al. 2014).



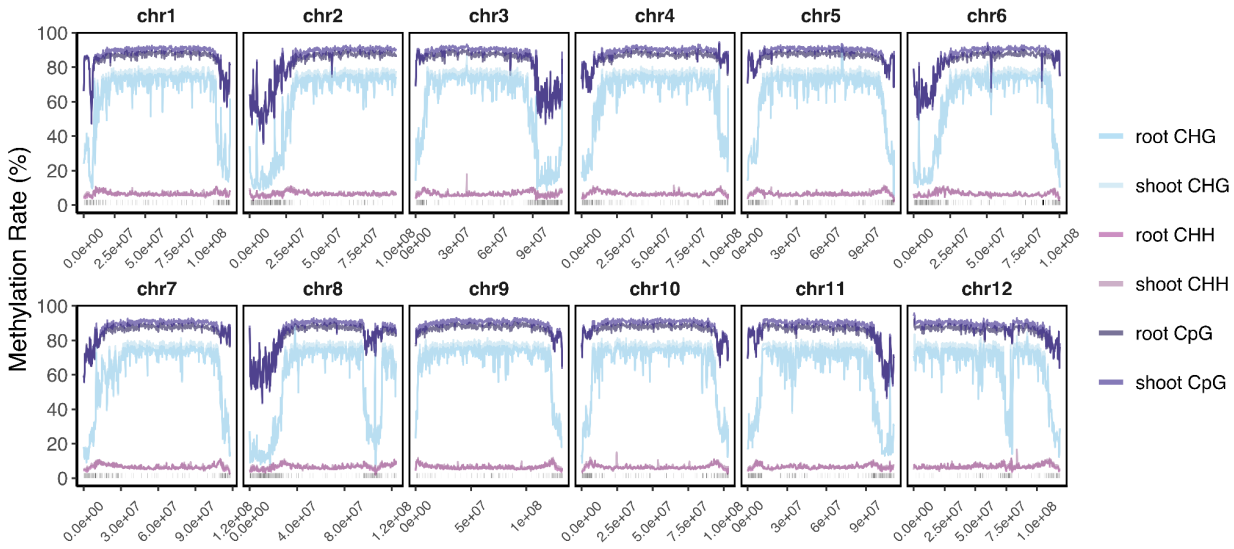
Supplementary Figure 5. ATAC-seq metaplots of leaf samples. Metaplots and corresponding heatmaps of the genome-wide ATAC-seq signal (normalised by reads per million) of each leaf and input sample. Data was computed for non-expressed genes and for expressed genes divided into four quantiles of expression level (Q1=lowest to Q4=highest) (based on normalised gene expression in 2-week-old shoots). The plot represents 3 kb upstream (-3.0), transcription starting site (TSS), transcription end site (TES), and 3 kb downstream of each gene (3.0 kb). The gene body region was scaled to 3,000 bp. The profile plot represents the median of all the values for each group of genes.



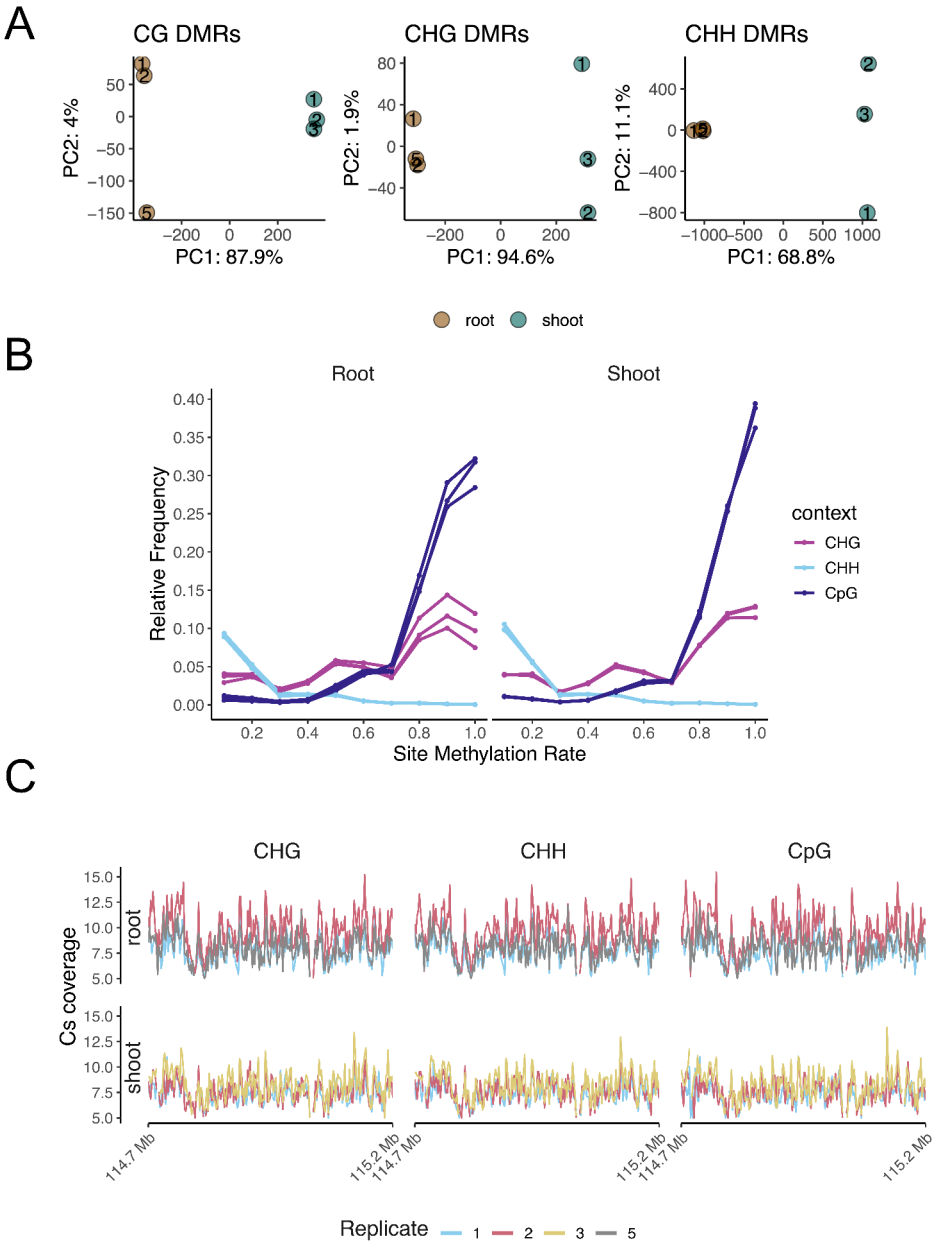
Supplementary Figure 6. ATAC-seq metaplots of root samples. Metaplots and corresponding heatmaps of the genome-wide ATAC-seq signal (normalised by reads per million) of each leaf and input sample. Data was computed for non-expressed genes and for expressed genes divided into four quantiles of expression level (Q1=lowest to Q4=highest) (based on normalised gene expression in 2-week-old roots). The plot represents 3 kb upstream (-3.0), transcription starting site (TSS), transcription end site (TES), and 3 kb downstream of each gene (3.0 kb). The gene body region was scaled to 3,000 bp. The profile plot represents the median of all the values for each group of genes.



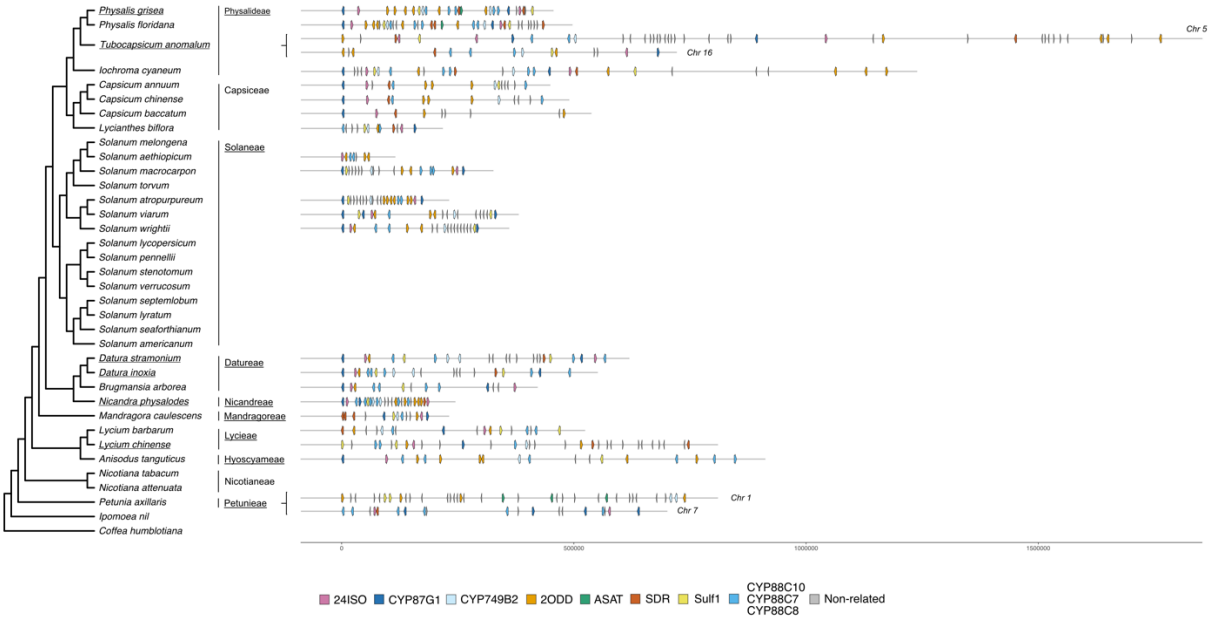
Supplementary Figure 7. **Chromosomal conformation capture by Hi-C in *P. grisea*. A)** Hierarchical clustering calculated from the Pearson's correlation of normalised Hi-C contact matrixes at 20-kb resolution from each sample: leaf replicates 1 and 2 (LR1, LR2); root replicates 1 and 2 (RR1, RR2). **B)** Genome-wide contact matrix of *P. grisea* at 1 Mb resolution, with shoot interactions displayed on the lower diagonal and root interactions on the upper diagonal. The colour intensity indicates frequency of interaction between two 1-Mb loci. **C)** A/B compartment annotation of *P. grisea* chromosomes. Negative (A compartment) and positive (B compartment) values represent the eigenvalues of the first principal component (PC) derived from the correlation coefficient matrix of the Observed/Expected Hi-C map for each chromosome at a 500-kb resolution. Gene density is represented at the bottom of each graph (black line). **D)** and **E)** Normalised Hi-C contact matrixes at 20-kb resolution covering the end of chromosome 1 (10 Mb) of leaf (**D)** and root (**E)** samples. Vertical and horizontal black lines indicate the withanolides BGC and denote reduced interaction with neighbouring chromatin.



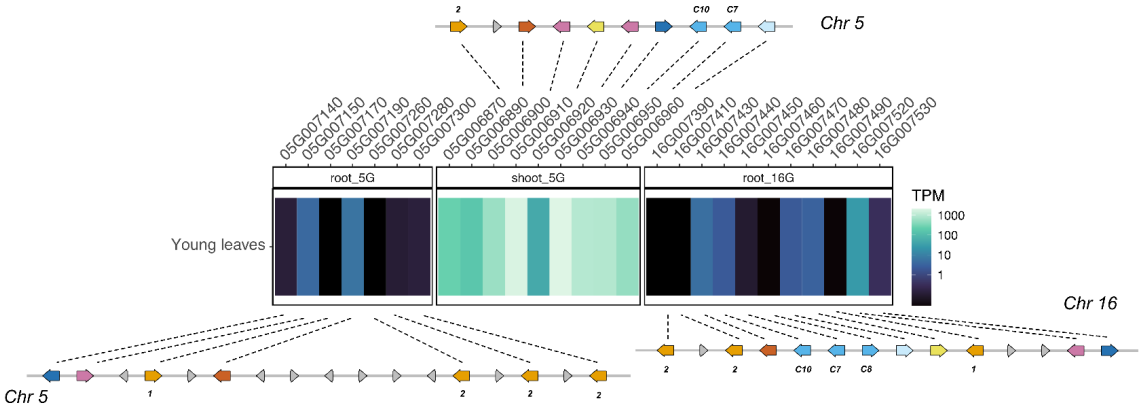
Supplementary Figure 8. DNA methylation in *P. grisea*. Genome-wide methylation distribution in the *P. grisea* genome, calculated as the average per-site methylation rate in 300 kb windows and averaged for the three replicates in each tissue. Dark lines above the x axis represent gene models.



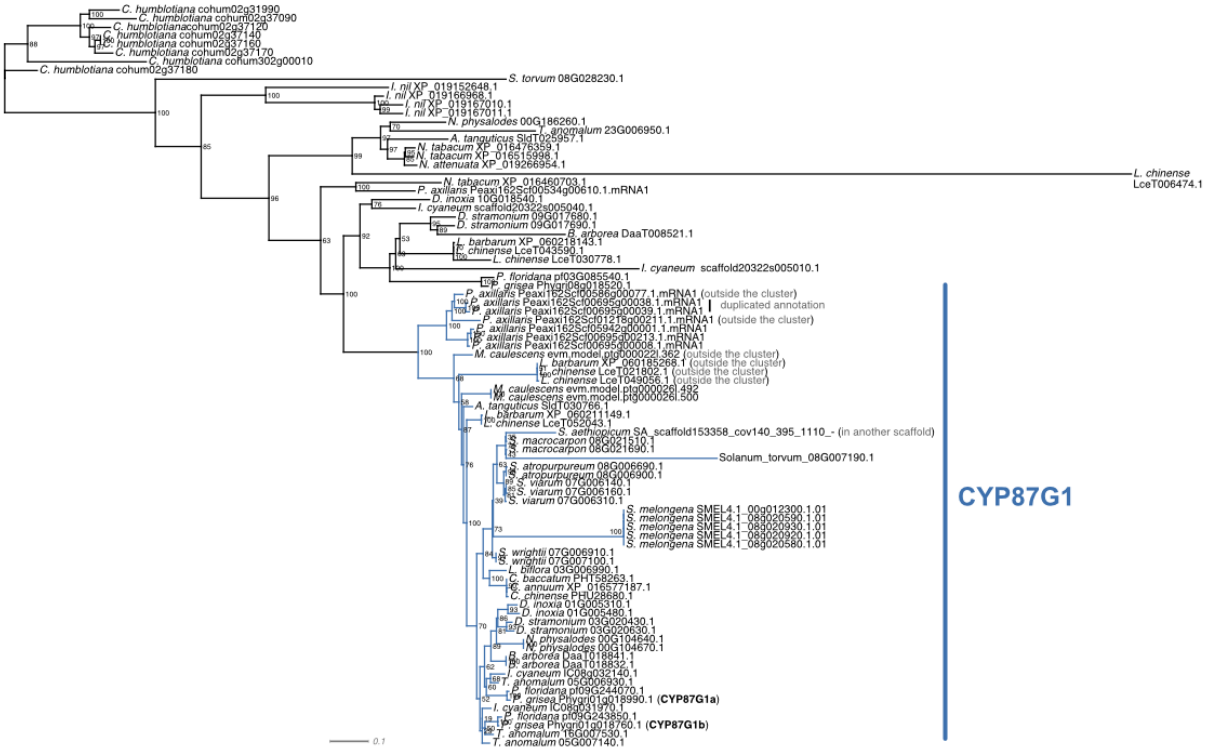
Supplementary Figure 9. WGBS quality control. A) Principal components analysis of the mean methylation rate in differentially methylated regions (DMRs) detected with MethylSore (Hüther et al. 2022). Number of DMRs: CG-DMRs = 322, CHG-DMRs = 183, CHH-DMRs = 9,522. **B)** Relative frequency of the site methylation rates in *P. grisea*. Each line per context represents one replicate. **C)** Mean cytosine coverage per context in 1 kb non-overlapping windows in the withanolide BGC locus.



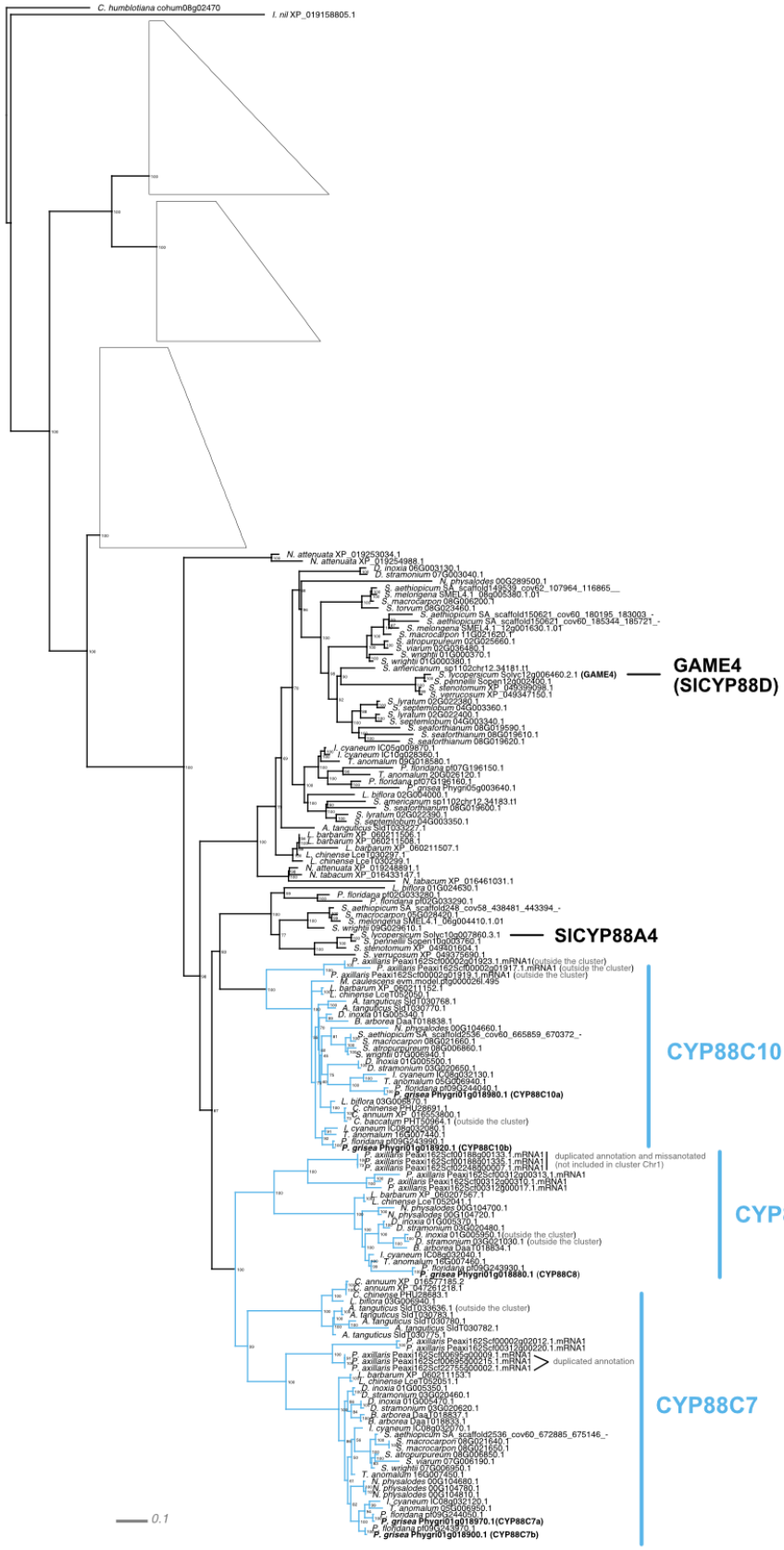
Supplementary Figure 10. Phylogenomic comparison of the withanolide BGC across *Solanaceae*, drawn to scale. The phylogenetic tree and cluster architectures are identical to the ones shown in Figure 5. Distances between genes are drawn to scale in this version.



Supplementary Figure 11. Expression of an orthologous BGC in *T. anomalum*. Expression is represented as transcript per million (TPM). Dashed lines connect each gene model with its respective expression data track.



Supplementary Figure 12. Phylogenetic analysis based on CYP87G1 amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea CYP87G1* paralogs (Phygri01g018760 and Phygri01g018990) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the CYP87G1 clade, which contains orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs. In the case of *P. axillaris*, duplicated annotations are indicated.

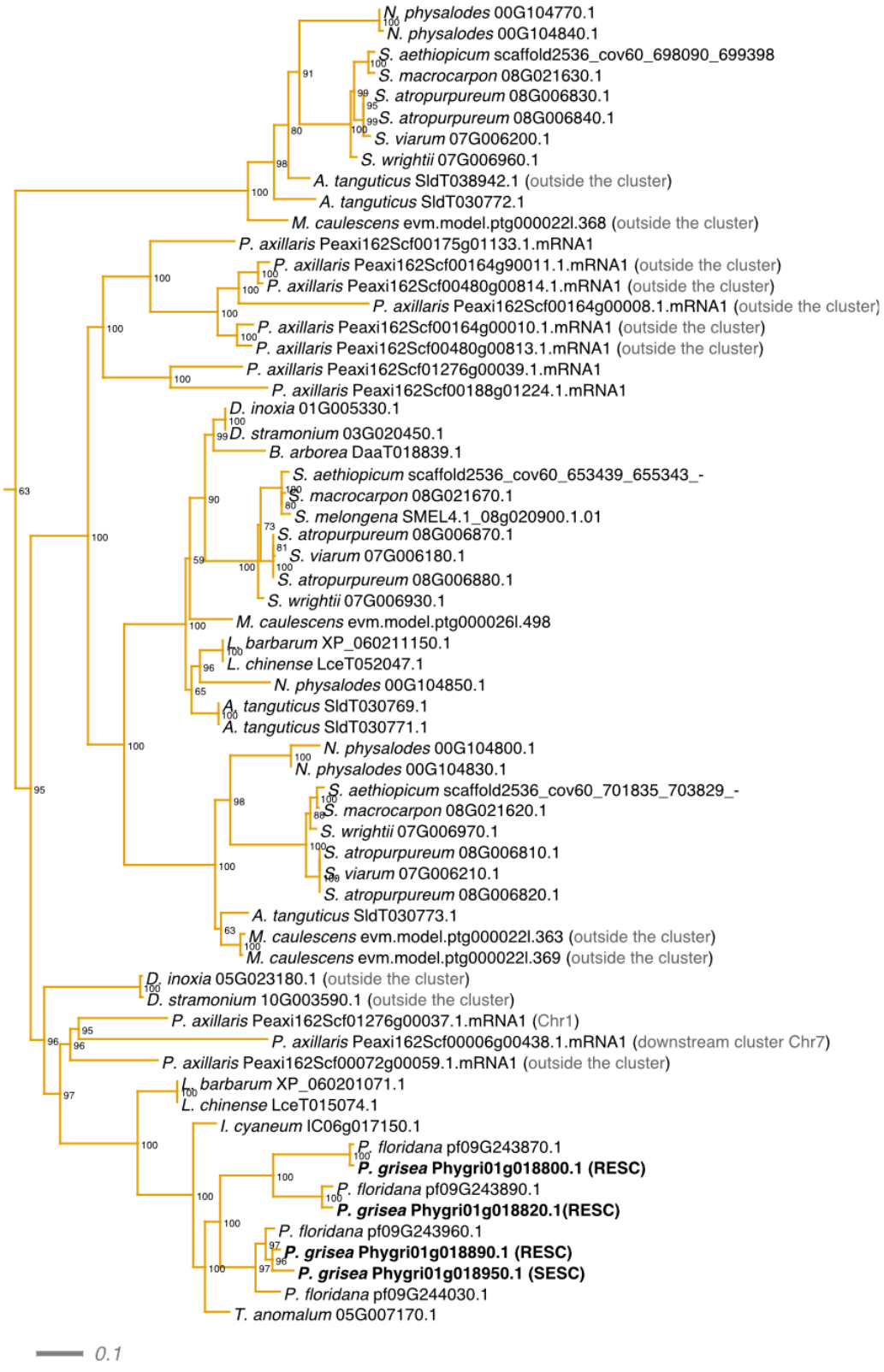


Supplementary Figure 13. Phylogenetic analysis based on the CYP88C subfamily members (CYP88C7, CYP88C8 and CYP88C10) amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea CYP88C7, CYP88C8* and *CYP88C10* orthologs and paralogs

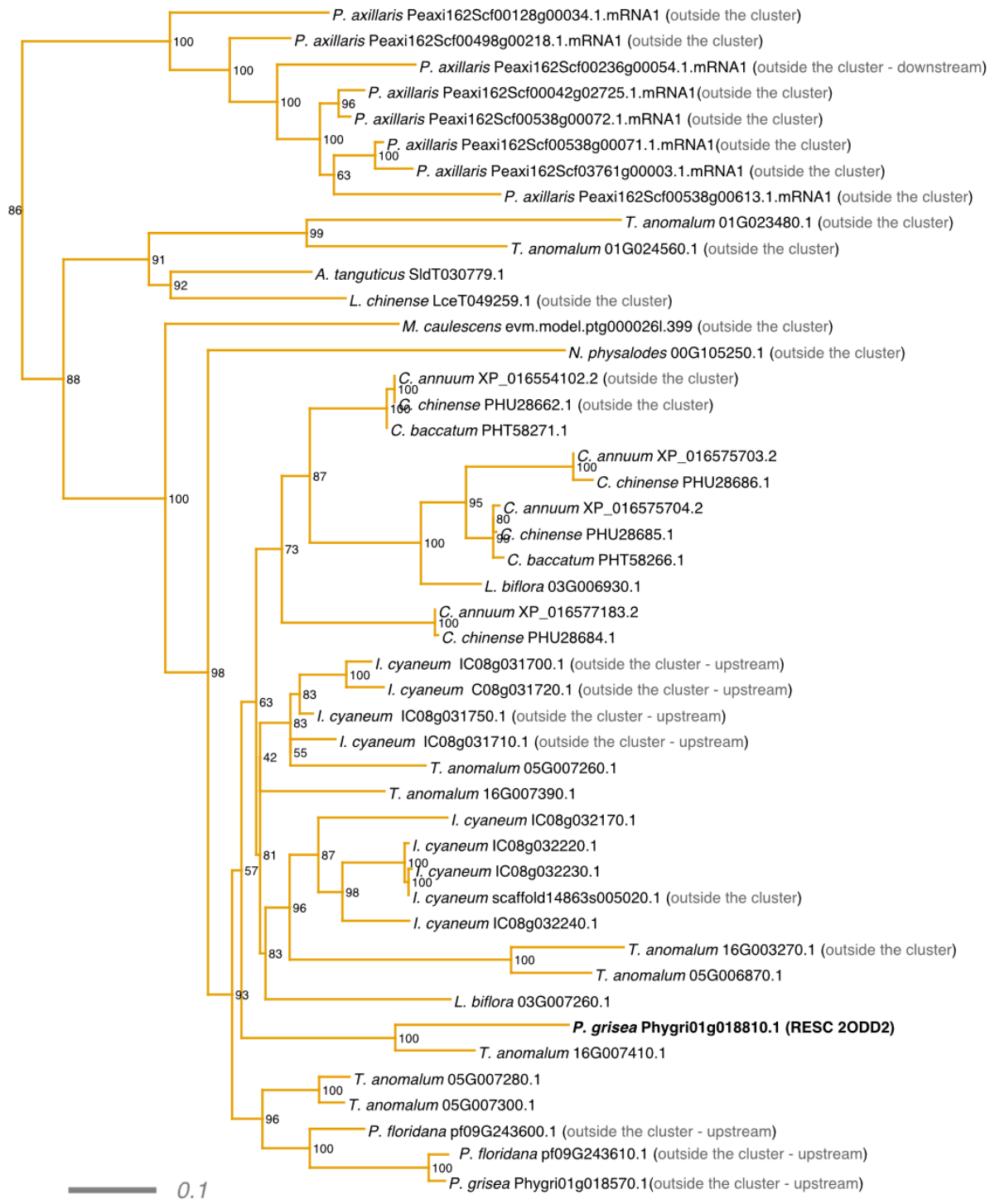
(Phygri01g018880, Phygri01g018900, Phygri01g018920, Phygri01g018970 and Phygri01g018980) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the CYP88C7, CYP88C8 and CYP88C10 clades, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs. In the case of *P. axillaris*, duplicated annotations are indicated.

Supplementary Figure 14. Phylogenetic analysis based on the CYP749B2 amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea CYP749B2* paralogs (Phygri01g018850 and Phygri01g018960) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the CYP749B2 clade, which contain orthologs potentially related (outside the cluster) or clustered in *Solanaceae* species orthologous withanolide BGCs. In the case of *P. axillaris*, duplicated annotations are indicated.

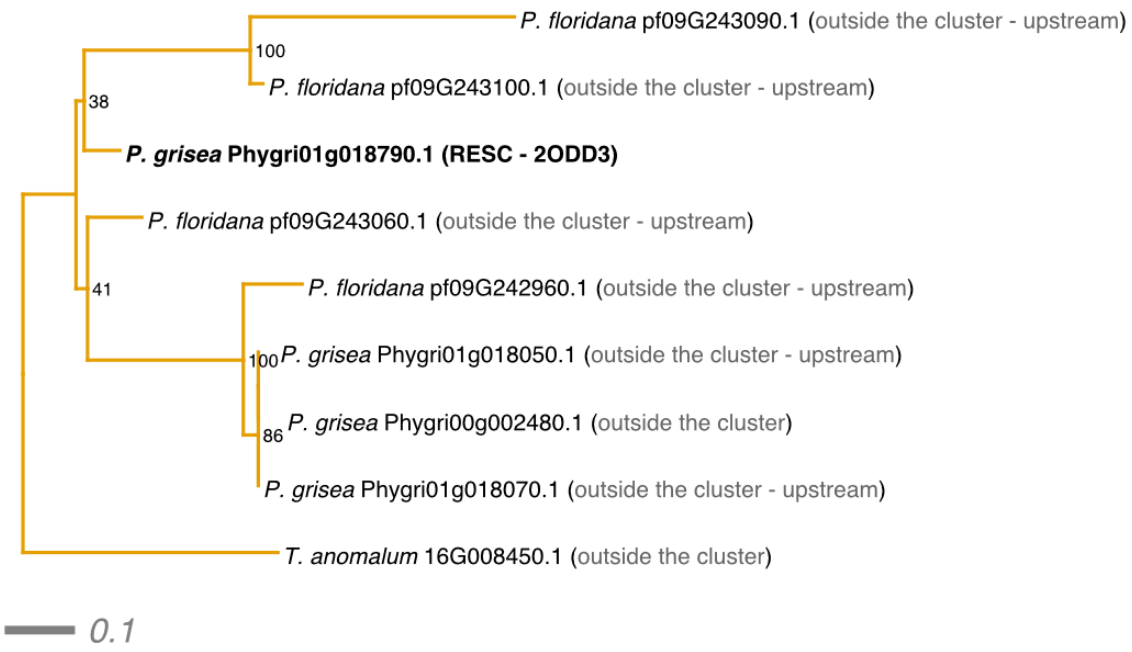
Supplementary Figure 15. Phylogenetic analysis based on the sulfotransferases (Sulfa and Sulfb) amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea Sulf* paralogs (Phygri01g018840 and Phygri01g019040) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the Sulf clade, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs. In the case of *P. axillaris*, duplicated annotations are indicated.



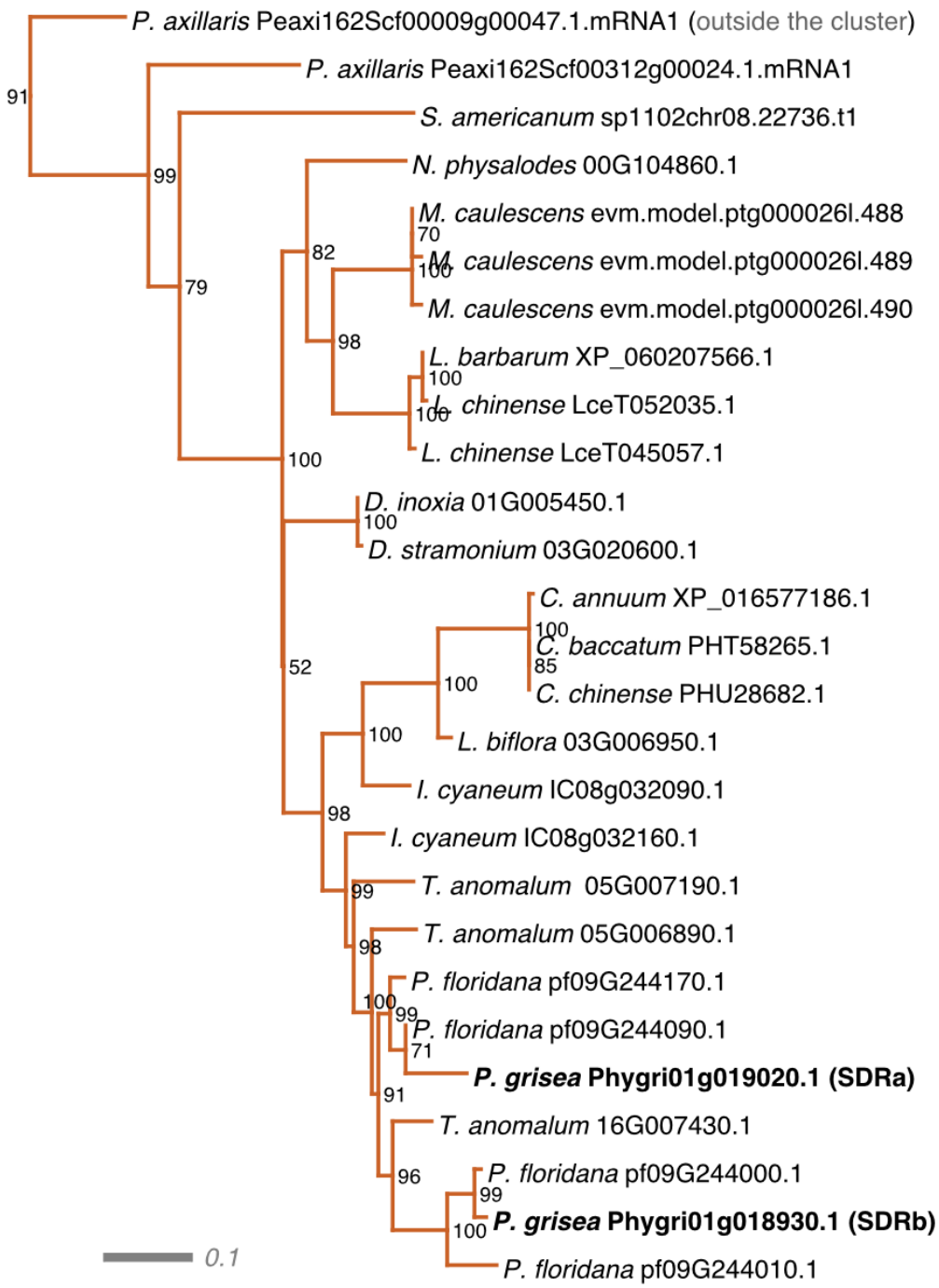
Supplementary Figure 16. Phylogenetic analysis based on the 2ODD amino acid sequences found in the larger 2ODD orthogroup (2ODD1). Subtree from the maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea* 2ODD1 paralogs (Phygri01g018800, Phygri01g018820, Phygri01g018890 and Phygri01g018950) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the 2ODD1 clade, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs.



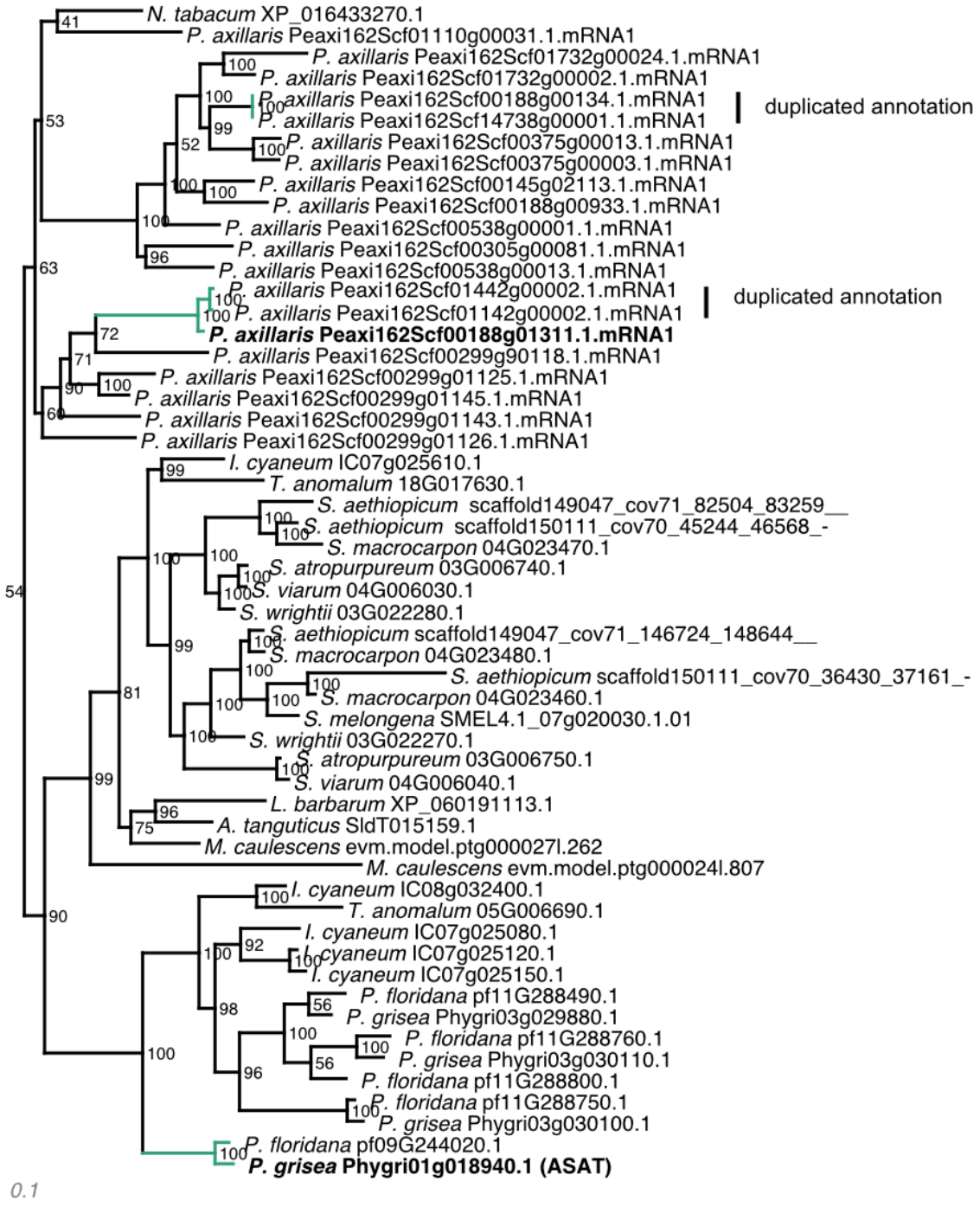
Supplementary Figure 17. Phylogenetic analysis based on the 2ODD amino acid sequences found in the second larger 2ODD orthogroup (2ODD2). Subtree from the maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea* 2ODD2 (Phygri01g018810) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the 2ODD2 clade, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs.



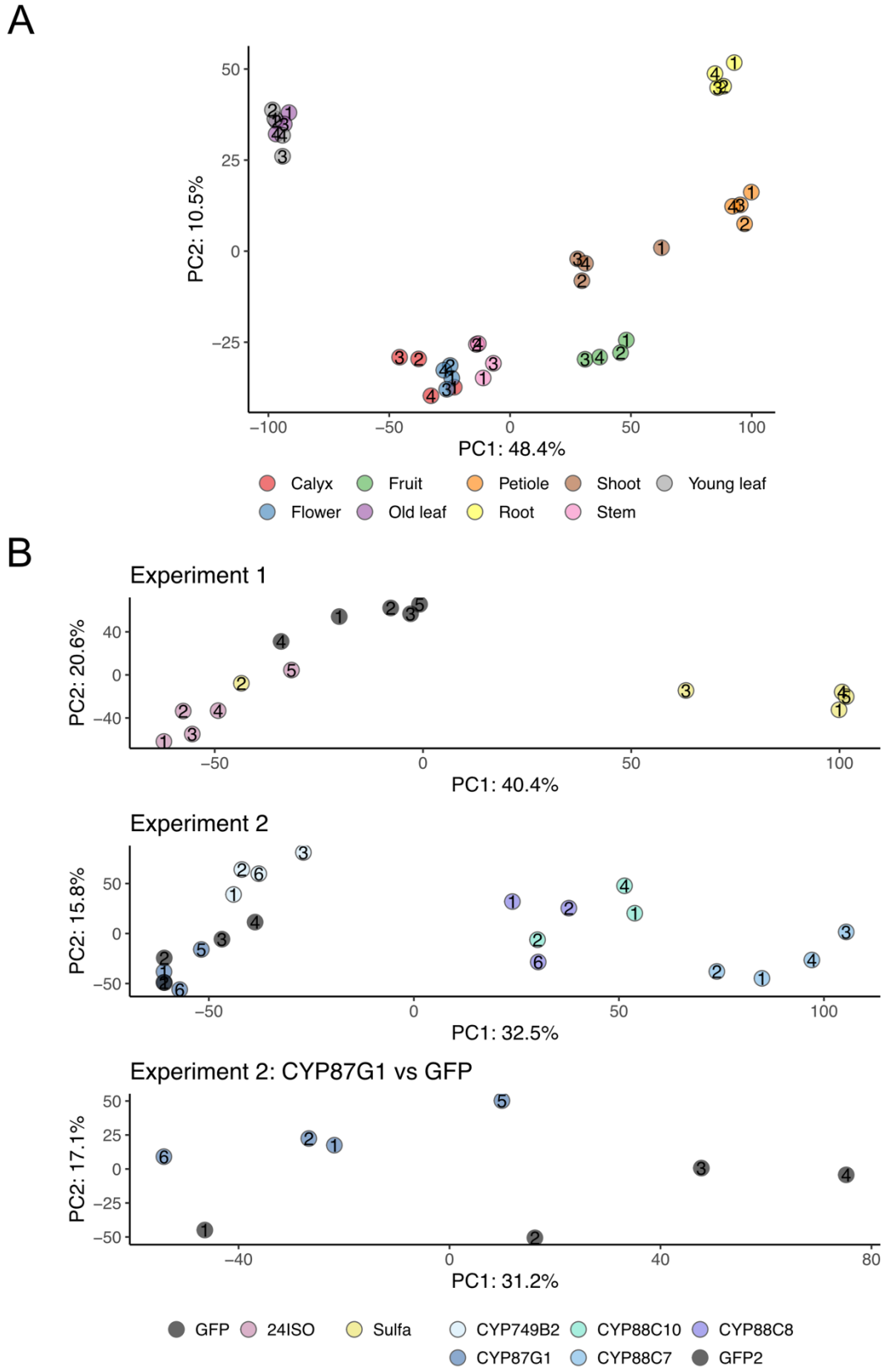
Supplementary Figure 18. Phylogenetic analysis based on the 2ODD amino acid sequences found in the smaller 2ODD orthogroup (2ODD3). Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea* 2ODD3 (Phygri01g018790) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the 2ODD3 clade, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs.



Supplementary Figure 19. Phylogenetic analysis based on the SDRs (SDRa and SDRb) amino acid sequences. Subtree from the maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea* SDRs (Phygri01g018930 and Phygri01g019020) in the Solanaceae species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the SDRs clade, which contain orthologs potentially related (but located outside the cluster) or clustered within orthologous *Solanaceae* withanolide BGCs.



Supplementary Figure 20. Phylogenetic analysis based on the ASAT amino acid sequences. Subtree from the maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *P. grisea* ASAT (Phygri01g018940) in the *Solanaceae* species included in **Supplementary Table 2**. Numbers represent 1,000 replicates of ultrafast bootstrapping. Coloured edges highlight the ASAT clustered in orthologous *Solanaceae* withanolide BGCs.



Supplementary Figure 21. Quality control analysis of the metabolomic datasets. A) and **B)** Principal component analysis of metabolite profiles from the different tissues **(A)** and the VIGS-transformed samples **(B)**. For each panel, the 4th quartile of the features with the highest variance (calculated from the log2 abundance) were considered. In **B),** 'Experiment 1' and 'Experiment 2' indicate independent experiments with their respective control (GFP and GFP2 for experiments 1 and 2, respectively).

1348 # Supplementary Tables

	CDS		Protein	
Pair	% identity	% similarity	% identity	% similarity
24ISO	91.9	91.9	94.9	97.9
CYP749B2	94.4	94.4	91.6	97.7
CYP87G1	93.5	93.5	92.8	97.1
CYP88C7	94.6	94.6	95	96.7
CYP88C8				
CYP88C10	84.4	84.4	80	89
Sulf	92.2	92.2	89.9	93.6

1349

1350 **Supplementary Table 1. Sequence identities and similarities between paralogs within the withanolide**
1351 **biosynthetic gene cluster in *P. grisea*.**

1352

Species	Version	Database	Reference	Reference withanolides detected (or related species)
Physalis grisea	v1.0	Original article repository	10.1093/plcell/koac305	https://doi.org/10.1016/j.fitote.202
Physalis floridana	GWHANUX00000000	Genome Warehouse	https://doi.org/10.1038/s41438-021-00705-w	2.105340
Tubocapsicum anomalum	GWHBKCL00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1016/j.bioorg.2021.104809; https://doi.org/10.1016/j.bioorg.2018.08.034
Iochroma cyaneum	v1.0	Solgenomics	https://doi.org/10.1002/tpg2.20223	https://doi.org/10.1016/0031-9422(93)80038-T; https://doi.org/10.1021/np50078a017
Capsicum annuum	GCF_002878395.1	NCBI	https://doi.org/10.1038/s41438-017-0011-0	
Capsicum chinense	GCA_002271885.2	NCBI	https://doi.org/10.1186/s13059-017-1341-9	
Capsicum baccatum	GCA_002271895.2	NCBI	https://doi.org/10.1186/s13059-017-1341-9	
Lycianthes biflora	GWHBKBD00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	
Solanum melongena	v4.1	Solgenomics	https://doi.org/10.1111/tpj.15313	DOIhttps://doi.org/10.1039/C5RA1
Solanum aethiopicum		GIGAnDB	https://doi.org/10.5524/100642; https://doi.org	2014D; doi:
Solanum macrocarpon	GWHBKBV00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	10.1080/14786410600888459;
Solanum torvum	GWHCAXF00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1246/bcsj.20180
Solanum atropurpureum	GWHBKBL00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	225;
Solanum viarum	GWHCAWV00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1248/cpb.49.161
Solanum wrightii	GWHBKCK00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	
Solanum lycopersicum	ITAG4.0	Solgenomics	https://doi.org/10.1038/nature11119; https://d	
Solanum pennelli	v2	Solgenomics	https://doi.org/10.1038/ng.3046	
Solanum stenotomum	GCF_019186545.1	NCBI	DOI: 10.1093/g3journal/jkab262	
Solanum verrucosum	GCF_900185275.1	NCBI		
Solanum septemlobum	GWHBKCE00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	
Solanum lyratum	GWHCAXE00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	
Solanum seaforthianum	GWHBKCD00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	
Solanum americanum	sp1102.v5	Figshare	https://doi.org/10.6084/m9.figshare.20454432.	
Datura stramonium	GWHBKBC00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1002/cbdv.202301655
Datura inoxia	GWHBJWV00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1016/j.steroids.2022.109010
Brugmansia arborea		Original article repository	https://doi.org/10.1038/s41467-023-44246-3	
Nicandra physalodes	GWHBKBF00000000	Genome Warehouse	https://doi.org/10.1016/j.cell.2023.04.008	https://doi.org/10.1016/j.phytochem.2017.02.009
Mandragora caulescens		Original article repository	https://doi.org/10.1038/s41467-023-44246-3	https://doi.org/10.1016/j.fitote.2010.05.013
Lycium barbarum	GCF_019175385.1	NCBI	https://doi.org/10.1038/s42003-021-02152-8	https://doi.org/10.1002/ardp.19753
Lycium chinense		Original article repository	https://doi.org/10.1038/s41467-023-44246-3	080814
Anisodus tanguticus		Original article repository	https://doi.org/10.1038/s41467-023-44246-3	https://doi.org/10.1021/np9902082
Nicotiana tabacum	GCF_000715135.1	NCBI	https://doi.org/10.1038/ncomms4833	
Nicotiana attenuata	GCF_001879085.1	NCBI	https://doi.org/10.1073/pnas.170007311	
Petunia axillaris	v.4.03; GCA_026929995.1	NCBI; CoGe id62433		https://doi.org/10.1016/0031-9422(89)80362-0
Ipomoea nil	GCF_001879475.1	NCBI	https://doi.org/10.1038/ncomms13295	
Coffea humblotiana		Solgenomics	https://doi.org/10.1038/s41598-021-87419-0	

Supplementary Table 2. Genome assemblies used for genome mining of withanolide biosynthesis genes.

Root tissue RESC peaks								
Motif Name	Consensus	P-value	Log P-value	q-value (Benjamini)	# of Target Sequences with	% of Target Sequences with Motif	# of Background Sequences	% of Background Sequences
IDD7(C2H2)/col-IDD7-DAP-Seq(GSE60143)/Homer	TTTGTCKTTTTN	1.00E-07	-1.64E+04	0.0001	22.0	36.07%	5004.2	10.30%
MYB17(MYB)/colamp-MYB17-DAP-Seq(GSE60143)/Homer	GGTAGGTGRG	1.00E-06	-1.52E+04	0.0001	18.0	29.51%	3596.1	7.40%
MGP(C2H2)/colamp-MGP-DAP-Seq(GSE60143)/Homer	TTTTGTCGTTTW	1.00E-06	-1.41E+04	0.0003	17.0	27.87%	3449.1	7.10%
MYB93(MYB)/colamp-MYB93-DAP-Seq(GSE60143)/Homer	GGTAGGTGRD	1.00E-05	-1.38E+04	0.0003	32.0	52.46%	11475.0	23.63%
ATY19(MYB)/col-ATY19-DAP-Seq(GSE60143)/Homer	YYCACCWACCAT	1.00E-05	-1.31E+04	0.0004	24.0	39.34%	7112.9	14.65%
At5g66730(C2H2)/colamp-At5g66730-DAP-Seq(GSE60143)/Homer	TTTGTCKTTTTK	1.00E-05	-1.25E+04	0.0006	16.0	26.23%	3445.5	7.09%
MYB63(MYB)/col-MYB63-DAP-Seq(GSE60143)/Homer	BHYACCWACCHH	1.00E-05	-1.24E+04	0.0006	18.0	29.51%	4350.9	8.96%
MNT(bHLH)/HepG2-MNT-ChIP-Seq(Encode)/Homer	DGCACACGTG	1.00E-05	-1.22E+04	0.0006	20.0	32.79%	5369.7	11.06%
MYB13(MYB)/col-MYB13-DAP-Seq(GSE60143)/Homer	HYCACCWACCHH	1.00E-05	-1.19E+04	0.0008	22.0	36.07%	6514.7	13.41%
AtIDD11(C2H2)/colamp-AtIDD11-DAP-Seq(GSE60143)/Homer	TTTGTCGTTT	1.00E-04	-1.14E+04	0.0011	17.0	27.87%	4199.3	8.65%
Leaf tissue SESC peaks								
At5g66730(C2H2)/colamp-At5g66730-DAP-Seq(GSE60143)/Homer	TTTGTCKTTTTK	1.00E-04	-1.05E+04	0.0278	11.0	24.44%	2690.0	5.57%
JKD(C2H2)/col-JKD-DAP-Seq(GSE60143)/Homer	TTTTGTCGTTTT	1.00E-04	-1.05E+04	0.0278	10.0	22.22%	2200.7	4.56%
LCL1(MYBrelated)/colamp-LCL1-DAP-Seq(GSE60143)/Homer	NAAAATATCTWHW\	1.00E-04	-9.42E+03	0.0278	9.0	20.00%	1999.1	4.14%
NPAS(bHLH)/Liver-NPAS-ChIP-Seq(GSE39860)/Homer	NVCACGTG	1.00E-03	-9.17E+03	0.0278	18.0	40.00%	7753.2	16.06%
NUC(C2H2)/col-NUC-DAP-Seq(GSE60143)/Homer	HAVAAAACGACAAA/	1.00E-03	-8.96E+03	0.0278	7.0	15.56%	1221.5	2.53%
IDD7(C2H2)/col-IDD7-DAP-Seq(GSE60143)/Homer	TTTGTCKTTTTN	1.00E-03	-8.66E+03	0.0291	12.0	26.67%	3895.4	8.07%
AtIDD11(C2H2)/colamp-AtIDD11-DAP-Seq(GSE60143)/Homer	TTTGTCGTTT	1.00E-03	-8.51E+03	0.0291	11.0	24.44%	3356.7	6.95%
IDD2(C2H2)/colamp-IDD2-DAP-Seq(GSE60143)/Homer	HAVAAAAMGACAAA	1.00E-03	-7.65E+03	0.0596	7.0	15.56%	1514.4	3.14%
SGR5(C2H2)/colamp-SGR5-DAP-Seq(GSE60143)/Homer	TTTGTCTTTTTT	1.00E-03	-7.27E+03	0.0780	14.0	31.11%	5922.6	12.26%
MGP(C2H2)/colamp-MGP-DAP-Seq(GSE60143)/Homer	TTTTGTCGTTTW	1.00E-02	-6.87E+03	0.1046	9.0	20.00%	2826.1	5.85%

Supplementary Table 3. Top 10 Homer Known Motif Enrichment Results for RESC and SESC peaks in root and leaf, respectively.

Primer	Sequence	Gene	Use
EK111	GGTGGTtctagaGATTGCAGTGGAAGGAGC	PDS	VIGS
EK112	TGGTGGgtaccGCCACCTTTTGACTCGAT	PDS	VIGS
EK139	GGTGGTtctagagactaagttagtctctga	CYP87G1	VIGS
EK140	TGGTGGgtaccccttgacaaaggctttg	CYP87G1	VIGS
EK141	GGTGGTtctagagagtgcttcaagaaagct	CYP88C7	VIGS
EK142	TGGTGGgtaccatcatcccatcttgaagg	CYP88C7	VIGS
EK143	GGTGGTtctagagctttccatagggagtc	CYP88C8	VIGS
EK144	TGGTGGgtaccCATCTGGAGATGATCTTC	CYP88C8	VIGS
EK145	GGTGGTtctagaCCTAATAGGTGGCTAT	CYP88C10	VIGS
EK146	TGGTGGgtaccAGCAAATGTATACTGAAG	CYP88C10	VIGS
EK147	GGTGGTtctagaTGATAATTGCCATTTTAG	CYP749B2_1	VIGS
EK148	TGGTGGgtaccTAATTTGGTGAAACAGTG	CYP749B2_1	VIGS
EK151	GGTGGTtctagaACTTTCAAGGCTCAAC	PgSulf1	VIGS
EK152	TGGTGGgtaccTTCCCTACATATGTAAA	PgSulf1	VIGS
EK157	CCAAGAGGTCTCTCGATGGC	TUB	qPCR
EK158	GTAAACTGCTCGCTCACCCT	TUB	qPCR
EK159	CGTCCTTCCTTTCTCGTGCT	24ISO_1	qPCR
EK160	TCCGCGCAGATCTAACATCC	24ISO_1	qPCR
EK161	TTCTTGACGTATGAGGACTTGA	Sulf1_1	qPCR
EK162	CCCATAAACTCAGCCAATTTC	Sulf1_1	qPCR
EK163	GGAACATCCTATCAAGAAGATAGGCTTT	CYP88C7	qPCR
EK164	AAATCGAAGAAGTAAGATACTAATGC	CYP88C7	qPCR
EK174	GTTTGATCAGCTCGCTCCTG	ASAT4l	qPCR
EK175	TGCTAGGCAAGTAGTGGAGC	ASAT4l	qPCR
EK192	CAATGTTCTTACACAAAGAGTGCTGTT	CYP87G1	qPCR
EK193	CTAAGTTCCTAAGGTATTTATGGGCAGAT	CYP87G1	qPCR
EK194	TCATGAAACACTGCGTTTGGGTA	CYP88C8	qPCR
EK195	GTTCATAGTGGTGTTTGCCATTCTAAA	CYP88C8	qPCR
EK196	ACCAACTAGAAAAGCAGCTGC	CYP88C10	qPCR
EK197	GCATCAAAGTAGACCGATACGGATCCT	CYP88C10	qPCR
EK198	GAGGGTCGCTTTAATGCTGATG	CYP749B2	qPCR
EK199	AGCAAAATAGAAAGATTTGCATTCTTC	YP749B2	qPCR
EK200	CGAGGCGAAGAGAGTACTGTAA	SDR1	qPCR
EK201	TTTTCTGCTTGACGCACTCA	SDR1	qPCR
EK202	AAACTTGCTCAAAGCGAGCAT	SDR2	qPCR
EK203	CCGGAGGCACCAAATTTCAAG	SDR2	qPCR

Supplementary Table 4. Primers used for VIGS and qPCR.