

Research article

The Madagascar palm genome provides new insights on the evolution of Apocynaceae specialized metabolism

Clément Cuello^a, Hans J. Jansen^b, Cécile Abdallah^a,
 Duchesse-Lacours Zamar Mbadinga^a, Caroline Birer Williams^a, Mickael Durand^a,
 Audrey Oudin^a, Nicolas Papon^c, Nathalie Giglioli-Guivarc'h^a, Ron P. Dirks^b,
 Michael Krogh Jensen^d, Sarah Ellen O'Connor^e, Sébastien Besseau^{a, **},
 Vincent Courdavault^{a, *}

^a *Biomolécules et Biotechnologies Végétales, EA2106, Université de Tours, 37200, Tours, France*

^b *Future Genomics Technologies, 2333 BE, Leiden, the Netherlands*

^c *Univ Angers, Univ Brest, IRF, SFR ICAT, F-49000, Angers, France*

^d *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs, Lyngby, Denmark*

^e *Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany*

ARTICLE INFO

Keywords:

Apocynaceae

Evolution

Biosynthetic gene clusters

Specialized metabolites

Alkaloids

ABSTRACT

Specialized metabolites possess diverse interesting biological activities and some cardenolides and monoterpene indole alkaloids- (MIAs) derived pharmaceuticals are currently used to treat human diseases such as cancers or hypertension. While these two families of biocompounds are produced by specific subfamilies of *Apocynaceae*, one member of this medicinal plant family, the succulent tree *Pachypodium lamerei* Drake (also known as Madagascar palm), does not produce such specialized metabolites. To explore the evolutionary paths that have led to the emergence and loss of cardenolide and MIA biosynthesis in *Apocynaceae*, we sequenced and assembled the *P. lamerei* genome by combining Oxford Nanopore Technologies long-reads and Illumina short-reads. Phylogenomics revealed that, among the *Apocynaceae* whose genomes have been sequenced, the Madagascar palm is so far the species closest to the common ancestor between MIA producers/non-MIA producers. Transposable elements, constituting 72.48% of the genome, emerge as potential key players in shaping genomic architecture and influencing specialized metabolic pathways. The absence of crucial MIA biosynthetic genes such as strictosidine synthase in *P. lamerei* and non-*Rauvolfioideae* species hints at a transposon-mediated mechanism behind gene loss. Phylogenetic analysis not only showcases the evolutionary divergence of specialized metabolite biosynthesis within *Apocynaceae* but also underscores the role of transposable elements in this intricate process. Moreover, we shed light on the low conservation of enzymes involved in the final stages of MIA biosynthesis in the distinct MIA-producing plant families, inferring independent gains of these specialized enzymes along the evolution of these medicinal plant clades. Overall, this study marks a leap forward in understanding the genomic dynamics underpinning the evolution of specialized metabolites biosynthesis in the *Apocynaceae* family, with transposons emerging as potential architects of genomics restructuring and gene loss.

* Corresponding author. E.A. 2106 Biomolécules et Biotechnologies Végétales, University of Tours, France.

** Corresponding author.

E-mail addresses: sebastien.besseau@univ-tours.fr (S. Besseau), vincent.courdavault@univ-tours.fr (V. Courdavault).

Abbreviations

ADH	alcohol dehydrogenase
BUSCO	Benchmarking Universal Single-Copy Orthologs
GO	geissoschizine oxidase
Ks	synonymous substitutions per synonymous sites
LAMT	loganic acid O-methyltransferase
LTR	long terminal repeat
MEP	methylerythritol phosphate
MIA	monoterpene indole alkaloid
P450	cytochrome P450
PAS	precondylocarpine acetate synthase
SAT	stemmadenine-O-acetyltransferase
SGD	strictosidine β -D-glucosidase
SLS	secologanin synthase
STR	strictosidine synthase
TDC	tryptophan decarboxylase
TE	transposable element
WGD	whole genome duplication

1. Introduction

Comprising around 350 genera and 5000 species, the *Apocynaceae* family, commonly known as the dogbane family, exhibits a global distribution, thriving in varied ecosystems from tropical rainforests to arid deserts [1]. Plants from this family showcase a plethora of morphological structures, reproductive strategies, physiological adaptations and phytochemical diversity that have evolved over millennia [2–5]. In addition to its ecological significance, *Apocynaceae* has been a source of invaluable contributions to human health and well-being through its rich specialized metabolites repertoire including monoterpene indole alkaloids (MIAs) and cardenolides [6,7].

Cardenolides are a group of biocompounds produced by plant species mainly belonging to the *Apocynaceae*, *Plantaginaceae*, and *Brassicaceae* families [7,8]. They exhibit a broad range of biological activities including cardiotoxic and anti-tumor activities [9,10]. They are characterized by a steroid core structure and a lactone ring [11]. The biosynthesis of cardenolides starts with the condensation of two farnesyl-diphosphates by squalene synthase [12]. From squalene, several enzymatic routes are proposed to generate the three putative sterol precursors of cardenolides: cholesterol, β -sitosterol and campesterol [13–15]. These three precursors can be used as substrates by DpCYP87A106 and CpCYP87A103 to produce pregnenolone [16]. In *Digitalis purpurea*, pregnenolone is further modified by 3 β -hydroxysteroid dehydrogenase and progesterone 5 β -reductase to produce pregnanolone [17–20]. Even though extensive research has been performed on sterol-derived cardenolides biosynthesis, enzymatic steps from pregnanolone towards the different types of cardenolides remains unidentified in any plant species.

MIAs are a group of specialized metabolites produced by plant species mainly belonging to the *Apocynaceae*, *Loganiaceae* and *Rubiaceae* families (*Gentianales* order) and a few *Nyssaceae* (*Cornales* order) [6]. In the *Apocynaceae* family, MIAs are specifically produced by the *Rauvolfioideae* subfamily [2]. Non-MIA producing *Apocynaceae* species are thus reported as non-*Rauvolfioideae* species. A broad range of MIAs exhibit a wide range of biological activities, including anti-tumor, anti-inflammatory, and anti-microbial properties [6,21]. MIA backbones are characterized by an indole ring and a monoterpene unit, which are biosynthetically derived from the shikimate and monoterpene iridoid pathways, respectively [22]. The shikimate pathway starts with the condensation of phosphoenolpyruvate and D-erythrose-4-phosphate by the 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase followed by nine steps resulting in the formation of tryptophan (Fig. 1) [23]. This essential amino acid is further decarboxylated by tryptophan decarboxylase (*TDC*) forming tryptamine, the indole precursor of MIA [24]. The monoterpene iridoid moiety is synthesized from geranyl diphosphate produced by geranyl-diphosphate synthase from isopentyl-diphosphate and from dimethylallyl-diphosphate synthesized from the MEP pathway (Fig. 1) [25]. A series of seven enzymes leads to the synthesis of loganic acid from geranyl-pyrophosphate [26]. Loganic acid is further O-methylated by loganic acid O-methyltransferase (*LAMT*) and then converted to secologanin, the monoterpene iridoid precursor of MIA, by secologanin synthase (*SLS*) (Irmeler et al., 2000; Murata et al., 2008; Dugé de Bernonville et al., 2015) (Fig. 1) [27–29]. Over the last three decades, the biosynthesis of MIAs has been extensively studied in the iconic plant *Catharanthus roseus* also known as the Madagascar periwinkle [30]. The pathway starts with the condensation of tryptamine and secologanin by strictosidine synthase (*STR*), the first identified Pictet Spengler enzyme [31,32]. The resulting intermediate and first MIA of the whole pathway, strictosidine, is then deglycosylated by strictosidine β -D-glucosidase (*SGD*) [33,34] and further modified by a series of enzymatic reactions to produce dihydroprecondylocarpine acetate, the last molecule that can be considered as part of a central MIA pathway (Fig. 1) [35]. This central MIA pathway contains several branching points resulting to the wide range of MIAs (Fig. 1). Strictosidine aglycones are the precursors of heteroyohimbane and yohimbane alkaloids such as the antihypertensive ajmalicine [36–40]. Geissoschizine is the precursor of sarpagane alkaloids such as the antiarrhythmic ajmaline [41,42]. Dihydroprecondylocarpine is the precursor of iboga alkaloids such as the antiaddictive ibogaine [43,44] and aspidosperma alkaloids such

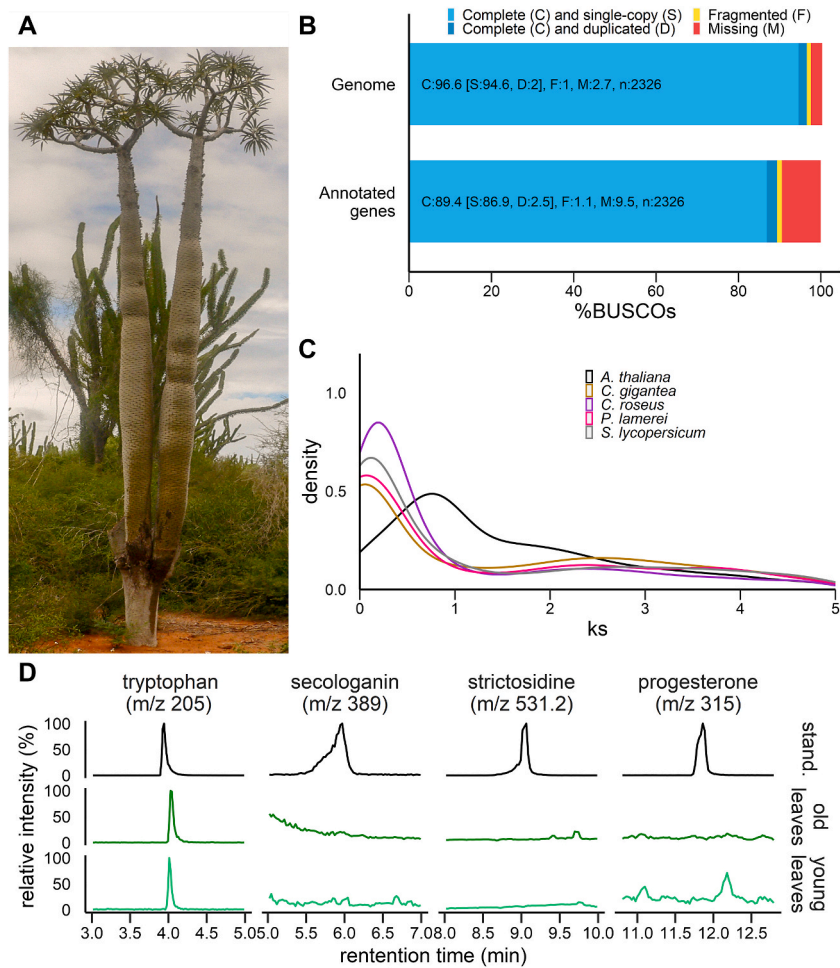


Fig. 2. General characteristics of the assembled genome and specialized metabolite profile from *P. lamerei*. (A) *P. lamerei*, commonly known as Madagascar palm tree, is a succulent spiky tree. (B) BUSCO scores of genome and annotated genes based on *Eudicotyledons* lineage. (C) Synonymous substitution rate (Ks) distribution plot for *P. lamerei* paralogs compared to other *Eudicotyledons*. (D) Tryptophan (m/z 205), secologanin (m/z 389), strictosidine (m/z 531.2) and progesterone (m/z 315) profiles of *P. lamerei* old and young leaves.

2. Results and discussion

2.1. Genome sequencing, assembly and annotation

Flye (v.2.8.3) was used to assemble the *P. lamerei* genome using Oxford Nanopore Technologies (ONT) long-reads, resulting in a 968.8 Mb assembly spread across 3029 scaffolds. A double polishing using Illumina short-reads and pilon (v.1.23) next resulted in a 968.6 Mb assembly distributed across 3029 scaffolds (Table 1). Interestingly, this genome is the second largest *Apocynaceae* genome

Table 1
Genome assembly features of *Pachypodium lamerei*.

Length of genome assembly (bp)	968,622,212
Number of scaffolds	3029
N50 of scaffolds (Mb)	1.67
L50 of scaffolds	145
GC content (%)	35.34
Longest scaffold (Mb)	10.422
scaffolds >50 kb (%)	96.75
QV	31.1
Number of protein-coding genes	16,176
Average gene length (bp)	5434
Average transcripts number per gene	1.6

available after *V. thoursii* (1351.2 Mb, [62]). A clear sign of the excellent quality of the assembled genome is the base-level QV of 31.1, which corresponds to more than 99.999% correctness, and the k-mer completeness of 92.5% (Table 1). Overall, the assembled genome is 96.6% complete (Fig. 2B) with a low duplication rate (2.0%), according to the identification of core Eudicotyledons Benchmarking Universal Single-Copy Orthologs (BUSCO). This BUSCO score is similar to the other Apocynaceae genomes (Supplementary Figure S1A) with the exception of *Apocynum venetum* and *V. minor*, both presenting a slightly higher duplication (14.4% and 36.6%, respectively), and *Calotropis procera*, being moderately less complete (85.3%).

On the basis of available resources, genomes of MIA-producing Apocynaceae (i.e. from the Rauvolfioideae subfamily) generally contain around 30,000 genes (*C. roseus*: 37,298 genes [55]; *V. thoursii*: 33,300 genes [62]; *V. minor*: 29,624 genes [57]). Despite being the second largest Apocynaceae genome characterized to date, *P. lamerei* displays the lowest number of genes (16,176 protein encoding genes) (Fig. 3A). Interestingly, *Gelsemium sempervirens*, a MIA-producing Gentianales close to Apocynaceae (Fig. 3A), features relatively similar gene contents than non-Rauvolfioideae species genomes (i.e. *C. gigantea*, *C. procera*, *Asclepias syriaca*, *A. venetum* and *P. lamerei*) (Fig. 3A). This suggests that the higher gene content in Rauvolfioideae species likely results from gene duplication occurring after Rauvolfioideae/non-Rauvolfioideae divergence. We also noted that all these predicted gene sets achieved good completeness scores according to Eudicotyledons BUSCO, ranging from 81% (*C. procera*) to 92.6% (*A. venetum*) (Supplementary Figure S1B). In addition, the combination of BLASTp and BLASTx against UniProt database and hmmscan against the Pfam database led to the functional annotation of 85.2% of the predicted genes (13,777 of the 16,176 genes, Supplementary Table S1).

Upon the identification of paralogous gene pairs in each species and estimation of the synonymous substitutions per synonymous site (Ks) for each gene pair, whole genome duplication (WGD) events were investigated. Paralogs were thus arranged according to age (Ks). In this case, genes being contemporarily duplicated showed a high-density peak at low Ks. Duplicated genes progressively disappear ending in an L-shape pattern. Important increases in paralogs at a particular Ks are caused by large-scale duplication such as WGD events. These events usually appear as secondary peaks in the Ks plot. *Arabidopsis thaliana*, *C. roseus*, *P. lamerei* and *Solanum*

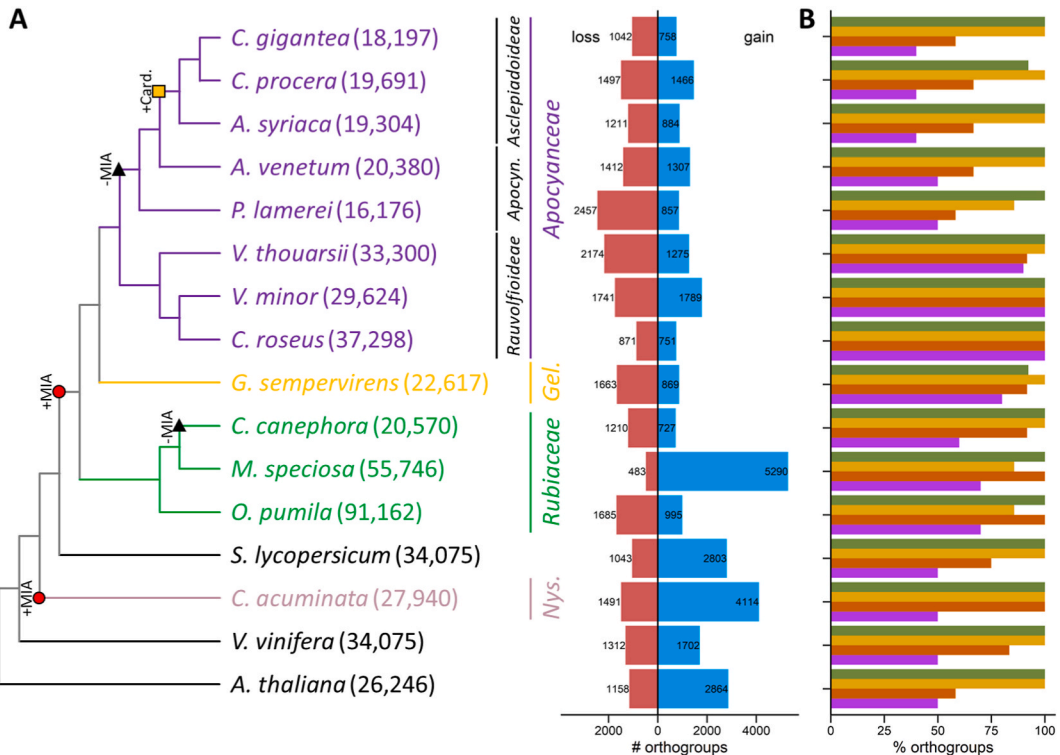


Fig. 3. Comparative genomic analysis of *P. lamerei* and 15 other plant species. (A) Phylogenetic tree. Colored branches represent MIA-producing plant families including eight Apocynaceae (purple), one Gelsemiaceae (Gel., yellow), three Rubiaceae (green) and one Nyssaceae (Nys., pink). Apoc.: Apocynoideae subfamily. Species: *C. gigantea*: *Calotropis gigantea*, *C. procera*: *Calotropis procera*, *A. syriaca*: *Asclepias syriaca*, *A. venetum*: *Apocynum venetum*, *P. lamerei*: *P. lamerei*, *V. thoursii*: *Voacanga thoursii*, *V. minor*: *Vinca minor*, *C. roseus*: *Catharanthus roseus*, *G. sempervirens*: *Gelsemium sempervirens*, *C. canephora*: *Coffea canephora*, *M. speciosa*: *Myrtagyna speciosa*, *O. pumila*: *Ophiorrhiza pumila*; *S. lycopersicum*: *Solanum lycopersicum*, *C. acuminata*: *Camptotheca acuminata*, *V. vinifera*: *Vitis vinifera*, *A. thaliana*: *Arabidopsis thaliana*. Numbers in brackets represent the number of annotated genes in each genome. Red circles correspond to MIA biosynthesis appearance events (+MIA). Black triangles correspond to MIA biosynthesis disappearance events (-MIA). Yellow square corresponds to cardenolides biosynthesis appearance events (+card.). For each species, orthogroups gain (blue, right) and loss (red, left) are shown on the barplot. (B) Percent of MIA biosynthesis related orthogroups found in each species. Green: indole, orange: MEP, Brown: Iridoid, Violet: Central MIA pathway. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

lycopersicum all showed a secondary peak at around $Ks = 2.5$ (Fig. 2C), which corresponds to the well-described and conserved whole-genome triplication event common to *Eudicotyledons* [63]. No additional peak could be identified in *P. lamerei* Ks plot, thus indicating the absence of any other recent WGD event.

Since TE are known to play important roles in genetic instability and genome evolution [64], the TE composition of the *P. lamerei* genome was also inferred. Based on this analysis, it appears that 72.48% of *P. lamerei* genome is composed of TE, most of them being long terminal repeat (LTR, Supplementary Figure S2). The majority of *P. lamerei* LTR are Copia-type (37.41%), Gypsy-type LTR accounting for 17.58% (Supplementary Figure S2). Such Copia-Gypsy ratio is mainly found in *Rauvolfioideae* species while non-*Rauvolfioideae* tend to have a Copia-Gypsy ratio of 1 or less (Supplementary Figure S2). In accordance with their genome size, *P. lamerei* displays the second highest proportion of TE after *V. thourarsii* [62]. The similar genome size and TE proportion in these two species suggest that *P. lamerei* and *V. thourarsii* are likely among the closest species to a common ancestor that led to the differentiation of *Rauvolfioideae* and non-*Rauvolfioideae* *Apocynaceae* species.

2.2. Conservation of specialized metabolism biosynthetic genes

To ensure completeness of the predicted proteome, we first looked for orthologs of genes involved in the biosynthesis of flavonoids, a prominent group of specialized metabolites broadly conserved in flowering plants. We first performed a targeted metabolomics analysis on key intermediates of the flavonoid pathway to ensure their occurrence in the studied species *P. lamerei* and the MIA-biosynthesis model plant *C. roseus*. As previously reported in *C. roseus* [65,66], we mainly identified final products of the flavonoids pathway such as glycosylated flavonols (*i.e.* kaempferol-3-glucoside and quercetin-3-glucoside) and both anthocyanidins and anthocyanins (*i.e.* cyanidin and pelargonidin) in both species (Table 2). We also identified the flavonoid precursor phenylalanine. Interestingly, no flavan-3-ols (*i.e.* catechin and epicatechin) could be detected in the analyzed samples. As flavonoids could be detected in *P. lamerei*, we thus performed a BLASTp analysis of functionally characterized proteins from *A. thaliana*, *Vitis vinifera*, *C. roseus* and *Desmodium uncinatum* against the *P. lamerei* and *C. roseus* predicted proteomes (Supplementary Tables S2-S3). We were able to identify high confidence orthologs with similar identity and coverage between the two species for the eight genes involved in the biosynthesis of kaempferol, quercetin, pelargonidin, and cyanidin from the condensation of 4-coumaroyl-CoA and 3-malonyl-CoA (Supplementary Table S4). This is in accordance with the detection of the glycone form of all these end products in *P. lamerei* and *C. roseus*. Conversely, no catechin nor epicatechin was detected in *P. lamerei* and *C. roseus* samples corroborating with our incapacity to identify high confidence orthologs for leucoanthocyanidin reductase and anthocyanidin reductase in both species. These results confirmed the gene annotation quality and that no orthologs would potentially escape our identification procedure.

On this basis, we next performed a similar analysis with MIA and cardenolides biosynthetic genes. We created a database containing functionally characterized proteins involved in shikimate, MEP, iridoid, MIA and cardenolides pathways [67]. We then performed a BLASTp analysis of this database against predicted proteomes of all studied species and conserved hits of at least 90% coverage and 50% identity (Supplementary Tables S5-S20). We deliberately used low identity and coverage thresholds to maximize ortholog identification probability. As expected, this resulted in a higher number of putative orthologs for multigenic families such as alcohol dehydrogenases (ADHs) and cytochromes P450 (P450s). For instance, these low thresholds led to the identification of 10 SLS-like genes in *C. roseus* while only four *bona fide* SLS exist in this species (Supplementary Table S21). Based on this approach, we were able to retrieve orthologs from genes of shikimate, MEP, steroids and cardenolide pathways for all species with only few exceptions probably due to phylogenetic distance (Supplementary Table S21-S22). We also identified orthologs for most of the genes of iridoid pathway for all *Apocynaceae* with the exception of 8-hydroxygeraniol oxidoreductase and loganic acid O-methyltransferase for all non-*Rauvolfioideae* species (50–90% identity, 94–100% coverage).

Interestingly, concerning the key MIA biosynthetic enzyme *STR*, only MIA-producing *Gentianales* present at least one CrSTR orthologous gene while no ortholog was identified in other species. As previously reported, *C. acuminata* (*Cornales*) uses an alternative seco-iridoid pathway producing strictosidinic acids via strictosidinic acid synthase that may explain why no CrSTR-like enzyme could be found in this species [68]. It also suggests that MIA biosynthesis in *Gentianales* and *Cornales* may have emerged independently (*i.e.*

Table 2

Detected flavonoids in *P. lamerei* and *C. roseus*.

Metabolite			<i>P. lamerei</i>		<i>C. roseus</i>		
	Molar mass	RT ^a	Young leaves	Old leaves	Flowers	Young leaves	Old leaves
phenylalanine	165.19	2.7	++	++	++	++	++
naringenin	272.25	12.5	–	–	–	–	–
kaempferol	286.23	13.0	–	–	–	–	–
kaempferol-3-glucoside	448.38	8.9	+	–	++	+	–
quercetin	302.25	11.3	–	–	–	–	–
quercetin-3-glucoside	464.38	7.8	++	–	++	–	–
pelargonidin	271.24	7.9	–	+	+	++	++
cyanidin	287.24	6.9	+	+	+	–	–
cyanidine-3-glucoside	484.83	5.9	+	+	+	+	+
catechin	290.26	4.8	–	–	–	–	–
epicatechin	290.26	5.9	–	–	–	–	–

^a Retention time.

by convergent evolution). Similarly, no CrSTR-like enzyme has been identified in *Apocynoideae* and *Asclepiadoideae* genomes. Thus, the absence of this enzyme, together with the lack of detection of secologanin nor strictosidine in *P. lamerei* leaf extracts (Fig. 2D) suggests that an evolutionary loss of MIA biosynthesis may have occurred in non-*Rauvolfioideae* species among *Apocynaceae*, similarly to what happened in *Rubiaceae* family [58,69]. This statement is reinforced by the lack of many other MIA biosynthetic genes such as SGD, geissoschizine oxidase (GO), stemmadenine-*O*-acetyltransferase (SAT) or precondylocarpine acetate synthase (PAS) which probably did not evolve in the absence of the precursor strictosidine (Supplementary Table S21). It is worth noting that no SAT-like genes could be found in *Gelsemiaceae* nor *Rubiaceae* species as well as no GO-like genes could be found in *Rubiaceae* species. Such orthologous pattern suggests that enzymes modifying the strictosidine aglycone derivatives may have appeared independently in these families leading to family-specific MIA skeletons (e.g. quinoline-type in *Rubiaceae* or corynanthe-type in *Apocynaceae*).

2.3. Evolutionary divergence of *P. lamerei*

To gain a more comprehensive understanding of the evolution of specialized metabolites biosynthesis in *Apocynaceae*, a maximum-likelihood phylogenetic tree was established (Fig. 3A). Therefore, protein coding genes from 8 *Apocynaceae* species (3 *Rauvolfioideae*, 2 *Apocynoideae* and 3 *Asclepiadoideae*), 4 non-*Apocynaceae* MIA-producing species (2 *Rubiaceae*, 1 *Gelsemiaceae* and 1 *Nyssaceae*) and 4 non-*Apocynaceae* non-MIA producing species (*A. thaliana*, *C. canephora*, *S. lycopersicum* and *V. vinifera*) were compared. 24,413 orthogroups were defined from these 16 protein datasets covering 94.1% of all proteins. 387 of them were single-copy orthogroups and used to set up the phylogenetic tree. The resulting phylogenetic tree is in agreement with *Apocynaceae* classification based on other approaches including morphological characterization and plastome analysis (Fig. 3A) [3,70]. In addition to their common genomic features, *V. thoursii* and *P. lamerei* are the first species to outgroup in *Rauvolfioideae* and non-*Rauvolfioideae* subfamilies.

The evolution of orthogroups was then analyzed along the phylogenetic tree (Fig. 3B–Supplementary Table S23). Orthogroups associated with shikimate and MEP pathways are present in all studied genomes suggesting that these pathways share a common evolutionary origin. Conversely, iridoid pathway-associated orthogroups are only identified in *Nyssaceae*, *Rubiaceae*, *Gelsemiaceae* and *Rauvolfioideae* species. Such distribution suggests that iridoid biosynthesis has emerged in the *Magnoliopsida* common ancestor and subsequently experienced independent losses in *Solanaceae* and non-*Rauvolfioideae* for instance. Obviously, orthogroups associated with MIA biosynthesis are mainly identified in *Rauvolfioideae* species. In addition, most of these MIA biosynthesis-associated orthogroups could also be found in *G. sempervirens*, *O. pumila* and *M. speciosa*. Interestingly, as the phylogenetic distance from *C. roseus* increases, fewer orthogroups associated with MIA biosynthesis are identified. This detection pattern therefore suggests an independent acquisition of post-strictosidine aglycone modifying enzymes, probably linked to the increase in gene contents in *Rauvolfioideae* species. This hypothesis could be strengthened by the future identification of genes coding for enzymes involved in MIAs biosynthesis in *Rubiaceae* and/or *Gelsemiaceae*, provided that the new orthogroups associated with these genes are not detected in *Rauvolfioideae* species.

2.4. Evolutionary scenario of biosynthetic gene clusters in *Apocynaceae*

To study the conservation of gene organization between chromosomes and species, a synteny analysis between *P. lamerei*, its close relative *V. thoursii* [62] and *C. roseus* chromosome-scale genome has been performed [55] (Supplementary Figure S3). The dot plot comparing those species indicated an apparent collinearity between scaffolds. There was a greater collinearity between *P. lamerei* and *V. thoursii* genomes than between *P. lamerei* and *C. roseus* genomes. While *P. lamerei* genome is only partially homologous with *C. roseus* genome, all *P. lamerei* scaffolds are homologous to *V. thoursii* scaffolds despite some major insertion/deletion events. Hence, given their greater collinearity, their phylogenomic linkage, their similar genome size, their proportion of TE and LTR family ratio, *P. lamerei* and *V. thoursii* are so far the most closely related species and therefore the closest species to the last common ancestor of *Rauvolfioideae* and non-*Rauvolfioideae*.

Given the low conservation of iridoid pathway genes and the absence of secologanin, we firstly examined the collinearity between *P. lamerei* and the genomic regions surrounding iridoid pathway genes in *C. roseus*. Although we were able to identify 3 highly conserved regions (Supplementary Figure S4), no syntenic region could be observed for 8HGO, IO, 7DLS, 7DLH, LAMT and SLS. It is worth noting that several tandem duplicates could be observed in both species for IRS and especially 7DLGT. These duplications could give rise to new functions via subfunctionalization throughout evolution [71–73]. It is noteworthy that despite a great conservation of G8H between *C. roseus* and *P. lamerei* (77.89%id, 100%cov), *P. lamerei* ortholog is greatly less expressed in all analyzed samples compared to *Rauvolfioideae* species (Supplementary Figure S5). Thus, the lack of collinearity in genomic regions surrounding most of iridoid pathway genes together with the low expression of G8H suggests a progressive degeneration of this pathway, the evolutionary processes of which remain to be discovered.

In fungi, bacteria and plants, genes involved in specific metabolic pathways frequently group together in the same genomic region [74]. These groups of genes, referred to as biosynthetic gene clusters, have been described in several MIA-producing plant species including the iconic *C. roseus* [52,53,55,56], *V. minor* [57], *R. stricta* [58] and *G. sempervirens* [53]. Microsyntenic study of these gene clusters enabled the identification of key enzymes in MIA biosynthesis as exemplified with the recent discovery of vincadifformine 16-hydroxylase in *V. minor* [57]. It is worth noting that genes encoding STR and TDC often cluster together as in *C. roseus*, *G. sempervirens*, *R. stricta* and *O. pumila* [52,53,55,56,58,69]. A recent single-cell approach in *C. roseus* unraveled a MATE transporter in this region that was found to be involved in secologanin transport from the cytosol to the vacuole [56]. By microsyntenic analysis between *C. roseus* and *P. lamerei*, we identified a similar [STR-TDC-MATE] region in *P. lamerei* contig 1749 (Fig. 4A). This chromosomal portion of *P. lamerei* is flanked by two highly conserved regions and, although it lacks any STR-like sequences, exhibits high TDC

(reverse strand) conservation. Interestingly, only one transporter-like gene can be found in *P. lamerei* whereas two MATE genes have been annotated in this region in *C. roseus*, similarly to *G. sempervirens* region [53]. Overall, these preliminary observations thus strengthen the hypothesis that the absence of MIA production in *P. lamerei* and non-*Rauvolfioideae* species results from STR deletion through evolution in *Apocynaceae*.

In such a scenario, transposable elements might have played a major role in the proposed evolutionary mechanism. Indeed, while we observe a significantly higher proportion of genes in the *C. roseus* cluster, the syntenic region in *P. lamerei* exhibits a specific TE enrichment in the corresponding chromosome portion (Fig. 4B). Similar observations can be made throughout the genome which can explain the huge differences in genome size between these two species. It has to be noted that no significant differences in gene proportion in the cluster, its associated contig or chromosome and the whole genome can be observed for each species. Conversely, *P. lamerei* genome is significantly richer in TE than the cluster and contig 1749 and *C. roseus* cluster is significantly enriched in TE compared to chromosome 5 and its whole genome (Fig. 4B). This observation thus reinforces the putative implication of TE in [STR-TDC-MATE] biosynthetic cluster evolution. In addition, STR in *C. roseus* is flanked by two S0:0000544 helitrons (Supplementary Figure S6). Interestingly, the proportion of this autonomous rolling-circle DNA transposon is significantly higher in the *P. lamerei* region than the *C. roseus* region (*P. lamerei*: 0.38/kb, *C. roseus*: 0.08/kb, *P*-value: 3.005e-06) while being smaller. Helitrons exhibit ‘copy-and-paste’ and ‘cut-and-paste’ modes of transposition [75] as well as an excisive mode [76]. Hence, S0:0000544 helitron might have been a key player in the evolutionary mechanism leading to STR deletion in non-*Rauvolfioideae* species.

3. Conclusion

The comprehensive genomic exploration of *P. lamerei* genome in comparison with other *Apocynaceae* species offers intriguing insights into the evolutionary dynamics within this plant family. The assembly of the *P. lamerei* genome reveals a remarkable genome size of 968.6 Mb, making it the second largest *Apocynaceae* genome known to date. Interestingly, *P. lamerei* and the other non-*Rauvolfioideae* genomes challenge the norm for MIA-producing *Apocynaceae* genomes, presenting a lower-than-expected gene count such

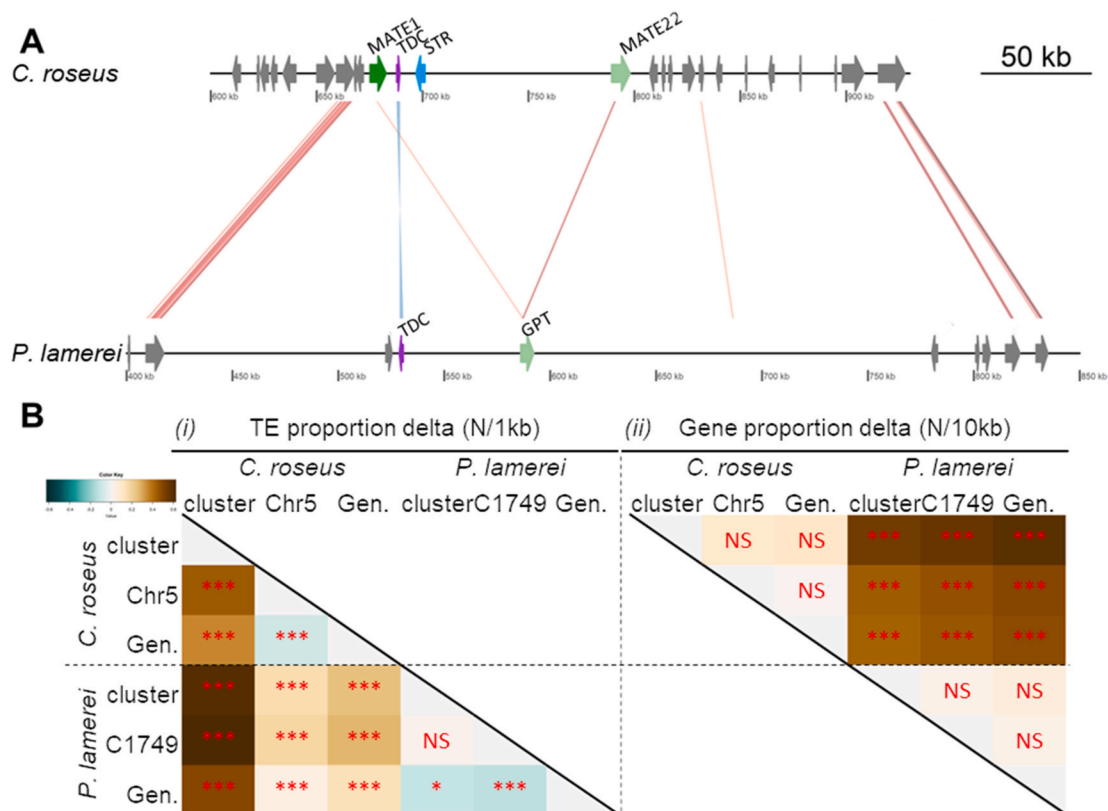


Fig. 4. Biosynthetic gene cluster evolution. (A) Synteny between *C. roseus* and *P. lamerei* scaffolds that are involved in strictosidine biosynthesis. Blue: strictosidine synthase (STR), purple: tryptophan decarboxylase (TDC), green: MATE transporter, light green: transporter, grey: other function, GPT: glucose-6-phosphate transporter. (B) Transposable element (i, N/1 kb) and gene (ii, N/10 kb) proportion delta (Proportion in row subtracted to proportion in col) in the biosynthetic cluster, its associated scaffold/chromosome (Chr5: *C. roseus* chromosome 5, C1749: *P. lamerei* contig 1749) and the whole genome (Gen.). The proportion delta was calculated as follows: $\Delta = \text{Proportion in the row condition} - \text{Proportion in the column condition}$. *P*-value: 0 “*****” 0.001 “****” 0.01 “***” 0.05 “NS” 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

as the 16,176 protein-encoding genes for *P. lamerei*. This deviation from the typical gene count observed in MIA-producing *Apocynaceae* genomes raises questions about the genomic landscape of *P. lamerei* and its implications for specialized metabolism. The phylogenetic analysis places *P. lamerei* as an important species regarding the study of *Apocynaceae* evolution. Comparative genomics further reveals a similarity between *P. lamerei* and *V. thouarsii*, indicating a close relatedness among *Apocynaceae* species. The notable presence of TEs, especially LTR elements, in the *P. lamerei* (72.48%) and *V. thouarsii* (75.16% [62]) genomes raises questions about their role in shaping genomic architecture. It is likely that the common ancestor of these two species may have undergone horizontal transfer of TE, known to contribute to genome diversification [77]. This phenomenon could then have led to a sharp increase in genome size, which over time contracted through the removal of DNA in other *Apocynaceae* species such as *C. roseus*. As the insertion and removal of TE is often imprecise, it can indirectly affect the surrounding sequences which can lead to high duplication and reshuffling. These genomic rearrangements could explain the absence of certain MIA biosynthetic genes, including STR, in non-*Rauvolfioideae* species. Furthermore, TEs have been described to represent an important mechanism for gene evolution in rice and *Poaceae* [78,79]. Thus, they could also be a determinant factor in the emergence post-strictosidine aglycone modifying enzymes, as exemplified in rice [79]. This reinforces the potential interest of this genomic data set for the study of genetics and epigenetics mechanisms associated with the evolution, diversification and emergence of specialized metabolites biosynthetic pathways. Moreover, beyond widening our knowledge on MIA biosynthesis evolution, this new genome could improve our understanding of cardenolides emergence in *Asclepiadoideae* when this biosynthesis pathway is more elucidated.

4. Materials and methods

4.1. Sample collection, DNA extraction and sequencing

Pachypodium lamerei plants were purchased from *A l'ombre des figuiers* (achat-vente-palmiers.com). Nuclei were extracted from young leaves using a nuclei isolation protocol [80]. The nuclear pellet was resuspended in Qiagen buffer G2 with RNaseA and proteinaseK, and DNA extraction was further continued according to the instructions of Qiagen's Genomic Tip/100G protocol (Qiagen, Venlo, The Netherlands). DNaseq library was performed by Future Genomics Technologies (Leiden, The Netherlands) using Nextera Flex kit (Illumina, San Diego, USA) for Illumina sequencing and ONT 1D ligation sequencing kit (Oxford Nanopore Technologies Ltd, Oxford, United-Kingdom) for Nanopore sequencing. Illumina libraries were sequenced in paired-end mode (2x150 bp) using Illumina NovaSeq 6000 technology. ONT libraries were sequenced on Nanopore PromethION flowcell (Oxford Nanopore Technologies Ltd, Oxford, United-Kingdom) with the guppy version 3.0.3 high-accuracy basecaller.

4.2. De novo genome assembly

The *P. lamerei* genome assembly was performed by Future Genomics Technologies (Leiden, The Netherlands). After adapter removal using porechop [81], ONT reads were assembled using Flye assembler (v.2.8.3, [82]). Contigs were twice polished with Illumina reads using pilon (v.1.23, [83]).

4.3. RNA extraction and sequencing

Pachypodium lamerei plants were obtained from *ikhebeencactus* (ikhebeencactus.nl). RNA was extracted from liquid nitrogen flash-frozen leaves, spikes and roots using PureLink RNA mini kit (Thermo Fisher Scientific, Illkirch-Graffenstaden, France). RNA library construction and sequencing was performed using Illumina NovaSeq 6000 technology. Raw RNA-seq data have been deposited under the SRA accession numbers SRR25397418 (leaves), SRR25397417 (spikes), SRR25397419 (roots).

4.4. Gene model prediction and functional annotation

RNaseq reads were aligned to the *P. lamerei* reference genome using HiSAT2 (2.2.1 [84]). Subsequently, transcriptome was assembled for each alignment using StringTie (v2.1.7 [85]) and merged using stringtie merge option. A combination of BLASTX on predicted transcripts and BLASTp on TransDecoder (v5.5.0 [86]) predicted ORFs against the Uniprot database as well as hmmscan (v3.1b2 [87]) against the PFAM database (<https://pfam.xfam.org/>) was used to assess putative function of each gene model.

A similar strategy was used for gene modeling and gene functional annotation of *A. venetum*, *A. syriaca* and *C. procera*. Genomic sequence for each species was retrieved from GenBank accession numbers GCA_019593545.1, GCA_027405835.1 and GCA_004801955.1, respectively. RNaseq reads were retrieved for each species from SRA accession numbers: SRR17163778 (leaves), SRR17163779 (stems) and SRR17163780 (roots) for *A. venetum*; SRR5117431 (buds) for *A. syriaca*; and SRR8281638, SRR8281639, SRR8281640, SRR8281641, SRR8281642 and SRR8281643 (all corresponding to leaves under different salinity conditions) for *C. procera*.

4.5. Assembly completeness assessment

Merqury (v. 1.3 [88]) and the stat program from BBMap tool (v.38.94 [89]) were used to assess *P. lamerei* genome assembly quality. To evaluate assembly and gene model completeness, Benchmarking Universal Single-Copy Orthologs (BUSCO v.5.2.2 [90]) with default parameters was used using a plant-specific database of 2326 single copy orthologs (eudicots_odb10). Gene model statistics

were retrieved using `agat_sp_statistics` from the AGAT package (v.0.8.0 [91]).

4.6. Transposable elements prediction and annotation

Transposable elements (TE) were identified and annotated using the sensitive mode of Extensive *de novo* TE annotator (EDTA v.1.9.5 [92]). This pipeline combines long-terminal repeat (LTR) annotation using `LTR_finder` (v. 1.07 [93]) and `LTRharvest` included in `GenomeTools` (v.1.5.10 [94]); terminal inverted repeat annotation (TIR) using `Generic repeat finder` (v.1.0 [95]) and `TIR-learner` (v.2.5 [96]); and `Helitrons` annotation using `HelitronScanner` (v.1.1 [97]). Apart from that, TE size thresholds are employed to avoid erroneous findings. Tandem repeats and short sequences are thereby defined as TIR shorter than 80 bp, LTR shorter than 100 bp and helitrons shorter than 100 bp. LTR are further filtered using `LTR_retriever` (v.2.9.0 [98]) to prevent erroneous LTR findings. If TIR candidates do not exceed 600 bp, they are categorized as MITEs. EDTA advanced filters are used for additional filtering TIR and Helitrons (for further information, see Ref. [92]). The TE library that was obtained is then used to mask the genome. `RepeatModeler` (v.2.0.1, default settings [99]) is then used on the unmasked portion of the genome to find non-LTR retrotransposons and unclassified TE that have eluded structure-based TE identification methods. Finally, gene-related sequences are removed by EDTA using the supplied CDS sequences.

4.7. Whole-genome duplication analysis

Transcript sequences of *P. lamerei*, *A. venetum*, *A. syriaca*, *C. procera*, *V. thouarsii* [62], *V. minor* [57], *C. roseus* [55], *Arabidopsis thaliana* [100], *Mytragyna speciosa* [101], *Solanum lycopersicum* [102], *C. acuminata* [103], *C. gigantea* [104], *G. sempervirens* [53], *Vitis vinifera* [105] and *Ophiorrhiza pumila* [69] were used to assume whole genome duplication (WGD) events using the `DupPipe` pipeline [106]. Each dataset was subject to discontinuous `MegaBLAST` [107,108] to find duplicated gene pairs (40% sequence similarity over 300bp). Using `BLASTx` (v.2.6.0–1 [109]), the open reading frame of each gene pair was determined from the NCBI's plant RefSeq protein database (as of May 21, 2021), preserving solely the best hit sequence (sequence similarity threshold: 30% across 150 aa). Following that, DNA sequence alignment was done using `GeneWise` [110] against the best hit homologous protein and its translation. For each gene pair, `MUSCLE` (v.3.6 [111]) aligned the amino acid sequences, which then served as a guide for `RevTrans` (v.1.4 [112]) to align the nucleic acids. In order to assess the divergence times between gene pairs, substitutions per synonymous site (Ks) using the `Codeml`'s F3x4 model from the `PAML` software (v.4.9 [113]) were calculated.

4.8. Orthology analysis and phylogenetic tree reconstruction

Protein sequences of at least 30 amino acids from *P. lamerei*, *A. venetum*, *A. syriaca*, *C. procera*, *V. thouarsii* [62], *V. minor* [57], *C. roseus* [55], *A. thaliana* [100], *M. speciosa* [101], *S. lycopersicum* [102], *C. acuminata* [103], *C. gigantea* [104], *G. sempervirens* [53], *V. vinifera* [105] and *O. pumila* [69] were compared in order to build gene families. `CD-HIT` (v.4.7 [114]) was used for each species. In each `CD-HIT` cluster, the longest representative protein was selected. The subsequent protein datasets were used as input for `OrthoFinder` (v.2.5.4 [115]) using the following parameters: `S diamond -M msa -A muscle -T raxml-ng`. From the 387 single-copy orthogroups, a maximum-likelihood phylogenetic tree was built. In order to identify orthogroup loss and growth across the phylogenetic tree, `Cafe5` (v.4.2.1 [116]) was employed.

4.9. Synteny analysis

P. lamerei genome was aligned with *V. thouarsii* and *C. roseus* genome using `minimap2` (v.2.24 [117]) and the following options: `cx asm20 -cs`. `D-Genies` (<https://dgenies.toulouse.inra.fr/> [118]) was used to visualize the generated `paf` file, selecting hits with at least 80% identity and ranking contigs by size. Syntenic regions between `STR/TDC/MATE1` gene cluster for *C. roseus* and `TDC/MATE1` gene cluster for *P. lamerei* were compared using `BLASTN` (v.2.6.0–1 [109]). The resulting hits between the clusters were filtered to only include alignments with an e-value of $1e-6$ and alignment length of 1500 bp and alignments were visualized using the `R` `genoPlotR` library (v.0.8.11 [119]). Gene and TE in microsyntenic regions was performed by comparing their proportion in the region of interest to their proportion in the corresponding contig by an exact Poisson test with the `poisson.test` function implemented in `stat` package (v. 4.1.1) in `R` (v. 4.1.1 [120]).

4.10. Transcript Abundance estimation

`Salmon` (v.0.6.0 [121]) with `-biasCorrect` and `-vbo` options was used to count RNAseq reads after they were pseudo-aligned onto the predicted transcripts. Abundance estimates were established as transcripts per million (TPM) and are presented in [Supplementary Table S24](#).

4.11. Targeted metabolomics analysis

To attest the presence of MIA and cardenolides, a targeted metabolomic analysis was carried out. Metabolites were extracted by vortexing 50 mg of lyophilized leaf powder with 1 mL of 0.1% formic acid methanol solution. Extracts were injected into the ultra-performance liquid chromatography system `ACQUITY` (Waters, Milford, MA, USA) coupled to the single quadrupole mass

spectrometer SQD2 (Waters, Milford, MA, USA) using conditions described in Ref. [122] over a 18 min gradient. The presence of tryptophan, secologanin and strictosidine, three key metabolites in MIA biosynthesis, as well as progesterone for cardenolides biosynthesis, was then checked by injecting the corresponding standards and a selected ion monitoring mode was used to collect data for tryptophan (m/z 205), secologanin (m/z 389), strictosidine (m/z 531.2) and progesterone (m/z 315).

To attest the presence of flavonoids, similar analysis was carried out on *C. roseus* and *P. lamerei*. Young and old leaves of *P. lamerei* and *C. roseus* and *C. roseus* flowers were collected and immediately flash-frozen in liquid nitrogen. Samples were freeze-dried for 48 h and subsequently ground using MM400 ball grinder (Retsch GmbH, Haan, Germany). 20 mg of powder for each sample were then added to 1 mL of 80% methanol and sonicated for 30 min. After 10 min centrifugation at $18,000\times g$ and $4\text{ }^{\circ}\text{C}$, 600 μL of supernatant were retrieved and five-fold diluted with the mobile phase (milliQ water – acetonitrile 5% - formic acid 0.1%). Extracts were then injected into the ultra-performance liquid chromatography system ACQUITY (Waters, Milford, MA, USA) coupled to the single quadrupole mass spectrometer XEVO (Waters, Milford, MA, USA) using conditions described in Ref. [123]. A selected ion monitoring mode was used to collect data for phenylalanine (m/z 166), naringenin (m/z 273), kaempferol (m/z 285), kaempferol-3-glucoside (m/z 449), quercetin (m/z 301), quercetin-3-glucoside (m/z 465), pelargonidin (m/z 272), cyanidin (m/z 288), cyanidin-3-glucoside (m/z 485), catechin (m/z 291) and epicatechin (m/z 291). To confirm metabolite identification, the corresponding standards were injected for all targets.

Grant information

This work was supported by EU Horizon 2020 research and innovation program [MIAMi project-Grant agreement N°814645]; ARD CVL Biopharmaceutical program of the Région Centre-Val de Loire [ETOPOCentre project]; ARP-IR from the Région Centre-Val de Loire [ScaleBio project] and ANR [project MIACYC – ANR-20-CE43-0010].

Data availability statement

Raw DNA-seq data, raw RNA-seq data and the genome assembly have been deposited in the NCBI database under the BioProject accession number: PRJNA997810 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997810>). The genome annotation, its functional annotation, transcripts sequences, predicted CDS and protein sequences as well as database containing functionally characterized proteins involved in shikimate, methylerythritol phosphate (MEP), iridoid, MIA, steroids and cardenolide pathways are available on the figshare (Cuello et al., 2023, <https://doi.org/10.6084/m9.figshare.23126816>).

CRedit authorship contribution statement

Clément Cuello: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Hans J. Jansen:** Methodology, Investigation, Formal analysis, Data curation. **Cécile Abdallah:** Writing – review & editing, Methodology, Investigation. **Duchesse-Lacours Zamar Mbadinga:** Writing – review & editing, Investigation. **Caroline Birer-Williams:** Writing – review & editing, Investigation. **Mickael Durand:** Writing – review & editing, Investigation. **Audrey Oudin:** Investigation. **Nicolas Papon:** Writing – review & editing, Investigation. **Nathalie Giglioli-Guivarc'H:** Writing – review & editing, Investigation. **Ron P. Dirks:** Validation, Formal analysis, Data curation. **Michael Krogh Jensen:** Project administration, Funding acquisition, Conceptualization. **Sarah Ellen O'Connor:** Project administration, Funding acquisition, Conceptualization. **Sébastien Besseau:** Writing – review & editing, Validation, Supervision, Investigation, Conceptualization. **Vincent Courdavault:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vincent Courdavault reports financial support was provided by Horizon Europe. Michael Krogh Jensen reports a relationship with BioMIA that includes: board membership and employment. Ron Dirks reports a relationship with Future Genomics Technologies that includes: board membership and employment. Hans Jensen reports a relationship with Future Genomics Technologies that includes: board membership and employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors benefited from the use of the cluster at the Centre de Calcul Scientifique en région Centre-Val de Loire.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28078>.

References

- [1] R.H. Patil, M.P. Patil, V.L. Maheshwari, Morphology, ecology, taxonomy, diversity, habitat and geographical distribution of the Apocynaceae family, in: *Apocynaceae Plants: Ethnobotany, Phytochemistry, Bioactivity and Biotechnological Advances*, Springer Nature Singapore, Singapore, 2023, pp. 1–11, https://doi.org/10.1007/978-981-99-5406-3_1.
- [2] A. Dey, A. Mukherjee, M. Chaudhury, Alkaloids from Apocynaceae: origin, pharmacotherapeutic properties, and structure-activity studies, *Stud. Nat. Prod. Chem.* 52 (2017) 373–488, <https://doi.org/10.1016/B978-0-444-63931-8.00010-2>.
- [3] M. Fishbein, T. Livshultz, S.C. Straub, A.O. Simões, J. Boutte, A. McDonnell, A. Foote, Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits, *Am. J. Bot.* 105 (3) (2018) 495–513, <https://doi.org/10.1002/ajb2.1067>.
- [4] J. Ollerton, S. Liede-Schumann, M.E. Endress, U. Meve, A.R. Rech, A. Shuttlesworth, H.A. Keller, M. Fishbein, L.O. Alvarado-Cardenas, F.W. Amorim, P. Bernhardt, F. Celep, Y. Chirango, F. Chiriboga-Arroyo, L. Civeyrel, A. Cocucci, L. Cranmer, I.C. da Silva-Batista, de Jager Linde, M.S. Depira, A. Domingos-Melo, C. Dvorsky, K. Agostini, L. Freitas, M.C. Gaglianone, L. Galetto, M. Gilbert, I. Gonzalez-Ramirez, P. Gorostiague, D. Goyder, L. Hachuy-Filho, A. Heiduk, A. howrd, G. Ionta, S.V. Isls-Hernandez, S.D. Johnson, L. Joubert, C.N. Kaiser-Bunbury, S. Kephart, A. Kidyoo, S. Koptur, C. Koschnitzke, E. Lamborn, T. Livshultz, I.C. Machado, S. Marino, L. Mema, K. Mochizuki, L.P. Cerdeira Morellato, C.K. Mrisha, E.W. Muiruri, N. Nakahama, V.T. Nascimento, C. Nuttman, P.O. Oliveira, C.I. Peter, S. Punekar, N. Rafferty, A. Rapini, Z.X. Ren, C.I. Rodriguez-Flores, L. Rosero, S. Sakai, M. Sazima, S.L. Steenhuisen, C. W. Tan, C. Torres, K. Trojelsgaard, A. Ushimaru, M.F. Vieira, A.P. Wiemer, T. Yamashiro, T. Nadia, J. Queiroz, Z. Quirino, The diversity and evolution of pollination systems in large plant clades: Apocynaceae as a case study, *Ann. Bot.* 123 (2) (2019) 311–325, <https://doi.org/10.1093/aob/mcy127>.
- [5] A.E. Mohammed, Z.H. Abdul-Hameed, M.O. Alotaibi, N.O. Bawakid, T.R. Sobahi, A. Abdel-Lateff, W.M. Alarif, Chemical diversity and bioactivities of monoterpene indole alkaloids (MIAs) from six Apocynaceae genera, *Molecules* 26 (2) (2021) 488, <https://doi.org/10.3390/molecules26020488>.
- [6] S.E. O'Connor, J.J. Maresh, Chemistry and biology of monoterpene indole alkaloid biosynthesis, *Nat. Prod. Rep.* 23 (2006) 532–547, <https://doi.org/10.1039/B512615K>.
- [7] S. Wen, Y. Chen, Y. Lu, Y. Wang, L. Ding, M. Jiang, Cardenolides from the Apocynaceae family and their anticancer activity, *Fitoterapia* 112 (2016) 74–84, <https://doi.org/10.1016/j.fitote.2016.04.023>.
- [8] T. Züst, S.R. Strickler, A.F. Powell, M.E. Mabry, H. An, M. Mirzaei, T. York, C.K. Holland, P. Kumar, M. Erb, G. Petschenka, J.M. Gomez, F. Perfectti, C. Müller, J.C. Pires, L.A. Mueller, G. Jander, Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae), *Elife* 9 (2020) e51712, <https://doi.org/10.7554/eLife.51712>.
- [9] J.M. Calderón-Montaño, E. Burgos-Morón, M.L. Orta, D. Maldonado-Navas, I. García-Domínguez, M. López-Lázaro, Evaluating the cancer therapeutic potential of cardiac glycosides, *BioMed Res. Int.* 2014 (2014), <https://doi.org/10.1155/2014/794930>.
- [10] S. Hafner, M. Schmiech, S.J. Lang, The cardenolide glycoside acovenoside a interferes with epidermal growth factor receptor trafficking in non-small cell lung cancer cells, *Front. Pharmacol.* 12 (2021) 611657, <https://doi.org/10.3389/fphar.2021.611657>.
- [11] A.A. Agrawal, G. Petschenka, R.A. Bingham, M.G. Weber, S. Rasmann, Toxic cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions, *New Phytol.* 194 (1) (2012) 28–45, <https://doi.org/10.1111/j.1469-8137.2011.04049.x>.
- [12] A. Pandey, V. Swarnkar, T. Pandey, P. Srivastava, S. Kanajiya, D.K. Mishra, V. Tripathi, Transcriptome and metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway from *Calotropis procera*, *Sci. Rep.* 6 (1) (2016) 34464, <https://doi.org/10.1038/srep34464>.
- [13] E. Caspi, D.O. Lewis, D.M. Piatak, K.V. Thimann, A. Winter, Biosynthesis of plant sterols. Conversion of cholesterol to pregnenolone in *Digitalis purpurea*, *Experientia* 22 (8) (1966) 506–507, <https://doi.org/10.1007/BF01898654>.
- [14] E. Caspi, D.O. Lewis, Progesterone: its possible role in the biosynthesis of cardenolides in *Digitalis lanata*, *Science* 156 (3774) (1967) 519–520, <https://doi.org/10.1126/science.156.3774.519>.
- [15] F. Milek, E. Reinhard, W.J.P.P. Kreis, Influence of precursors and inhibitors of the sterol pathway on sterol and cardenolide metabolism in *Digitalis lanata* Ehrh, *Plant Physiol. Biochem.* 35 (2) (1997) 111–121.
- [16] M. Kunert, C. Langley, R. Lucier, K. Ploss, C.E. Rodríguez López, D.A. Serna Guerrero, E. Rothe, S.E. O'Connor, P.D. Sonawane, Promiscuous CYP87A enzyme activity initiates cardenolide biosynthesis in plants, *Nat. Plants* (2023) 1–11, <https://doi.org/10.1038/s41477-023-01515-9>.
- [17] J. Munkert, M. Ernst, F. Müller-Urli, W. Kreis, Identification and stress-induced expression of three β -hydroxysteroid dehydrogenases from *Erysimum crepidifolium* Rchb. and their putative role in cardenolide biosynthesis, *Phytochemistry* 100 (2014) 26–33, <https://doi.org/10.1016/j.phytochem.2014.01.006>.
- [18] J. Munkert, C. Costa, O. Budeanu, J. Petersen, S. Bertolucci, G. Fischer, F. Müller-Urli, W. Kreis, Progesterone 5 β -reductase genes of the Brassicaceae family as function-associated molecular markers, *Plant Biol.* 17 (6) (2015) 1113–1122, <https://doi.org/10.1111/plb.12361>.
- [19] V. Herl, J. Frankenstein, N. Meitinger, F. Müller-Urli, W. Kreis, $\Delta 5$ - β -Hydroxysteroid dehydrogenase (β HSD) from *Digitalis lanata*. Heterologous expression and characterisation of the recombinant enzyme, *Planta Med.* 73 (7) (2007) 704–710, <https://doi.org/10.1055/s-2007-981537>.
- [20] V. Herl, D.C. Albach, F. Müller-Urli, C. Bräuchler, G. Heubl, W. Kreis, Using progesterone 5 β -reductase, a gene encoding a key enzyme in the cardenolide biosynthesis, to infer the phylogeny of the genus *Digitalis*, *Plant Systemat. Evol.* 271 (2008) 65–78, <https://doi.org/10.1007/s00606-007-0616-0>.
- [21] A.P. Macabeo, G.J. Alejandro, A.V. Hallare, W.S. Vidar, O.B. Villaflores, Phytochemical survey and pharmacological activities of the indole alkaloids in the genus *Voacanga* Thouars (Apocynaceae) - an update, *Phcog. Rev.* 3 (2009) 132–142.
- [22] B. St-Pierre, S. Besseau, M. Clastre, V. Courdavault, M. Courtois, J. Creche, E. Ducos, T. Dugé de Bernonville, C. Dutilleul, G. Glévarec, N. Imbault, A. Lanoue, A. Oudin, N. Papon, O. Pichon, N. Giglioli-Guivarc'h, Deciphering the evolution, cell biology and regulation of monoterpene indole alkaloids, *Adv. Bot. Res.* 68 (2013) 73–109, <https://doi.org/10.1016/B978-0-12-408061-4.00003-1>. Academic Press.
- [23] H. Maeda, N. Dudareva, The shikimate pathway and aromatic amino acid biosynthesis in plants, *Annu. Rev. Plant Biol.* 63 (2012) 73–105, <https://doi.org/10.1146/annurev-arplant-042811-105439>.
- [24] V. De Luca, C. Marineau, N. Brisson, Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases, *Proc. Natl. Acad. Sci. USA* 86 (8) (1989) 2582–2586, <https://doi.org/10.1073/pnas.86.8.2582>.
- [25] A. Contin, R. van der Heijden, A.W. Lefebvre, R. Verpoorte, The iridoid glucoside secologanin is derived from the novel triose phosphate/pyruvate pathway in a *Catharanthus roseus* cell culture, *FEBS Lett.* 434 (3) (1998) 413–416, [https://doi.org/10.1016/S0014-5793\(98\)01022-9](https://doi.org/10.1016/S0014-5793(98)01022-9).
- [26] K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. van der Krol, R. Lugaan, T. Ilc, R. Verpoorte, K.M. Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink, D. Werck-Reichhart, The seco-iridoid pathway from *Catharanthus roseus*, *Nat. Commun.* 5 (1) (2014) 3606, <https://doi.org/10.1038/ncomms4606>.
- [27] S. Irmeler, G. Schröder, B. St-Pierre, N.P. Crouch, M. Hotze, J. Schmidt, D. Strack, U. Matern, J. Schröder, Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase, *Plant J.* 24 (6) (2000) 797–804, <https://doi.org/10.1046/j.1365-3113x.2000.00922.x>.
- [28] J. Murata, J. Roepke, H. Gordon, V. De Luca, The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization, *Plant Cell* 20 (3) (2008) 524–542, <https://doi.org/10.1105/tpc.107.056630>. Epub 2008 Mar 7.
- [29] T. Dugé de Bernonville, E. Foureau, C. Parage, A. Lanoue, M. Clastre, M.A. Londono, A. Oudin, B. Houillé, N. Papon, S. Besseau, G. Glévarec, L. Atehortúa, N. Giglioli-Guivarc'h, B. St-Pierre, V. De Luca, S.E. O'Connor, V. Courdavault, Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced de novo assembly of a *Catharanthus roseus* transcriptome, *BMC Genom.* 16 (1) (2015) 619, <https://doi.org/10.1186/s12864-015-1678-y>.
- [30] N. Kulagina, L.V. Méteignier, N. Papon, S.E. O'Connor, V. Courdavault, More than a *Catharanthus* plant: a multicellular and pluri-organelle alkaloid-producing factory, *Curr. Opin. Plant Biol.* 67 (2022) 102200, <https://doi.org/10.1016/j.pbi.2022.102200>.
- [31] J.F. Treimer, M.H. Zenk, Purification and properties of strictosidine synthase, the key enzyme in indole alkaloid formation, *Eur. J. Biochem.* 101 (1979) 225–233, <https://doi.org/10.1111/j.1432-1033.1979.tb04235.x>.

- [32] T.D. McKnight, C.A. Roessner, R. Devagupta, A.I. Scott, C.L. Nessler, Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*, *Nucleic Acids Res.* 18 (16) (1990) 4939, <https://doi.org/10.1093/nar/18.16.4939>.
- [33] A. Geerlings, M.M.L. Ibanez, J. Memelink, R. van der Heijden, R. Verpoorte, Molecular cloning and analysis of strictosidine β -D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*, *J. Biol. Chem.* 275 (5) (2000) 3051–3056, <https://doi.org/10.1074/jbc.275.5.3051>.
- [34] I. Carqueijeiro, K. Koudounas, T. Dugé de Bernonville, L.J. Sepúlveda, A. Mosquera, D.P. Bomzan, A. Oudin, A. Lanoue, S. Besseau, P. Lemos Cruz, N. Kulagina, E.A. Stander, S. Eymieux, J. Burlaud-Gaillard, E. Blanchard, M. Clastre, L. Atehortúa, B. St-Pierre, N. Giglioli-Guivarc'h, N. Papon, D.A. Nagegowda, S. E. O'Connor, V. Courdavault, Alternative splicing creates a pseudo-strictosidine β -D-glucosidase modulating alkaloid synthesis in *Catharanthus roseus*, *Plant Physiol.* 185 (3) (2021) 836–856, <https://doi.org/10.1093/plphys/kiab075>.
- [35] L. Caputi, J. Franke, S.C. Farrow, K. Chung, R.M.E. Payne, T.-D. Nguyen, T.-T.T. Dang, I. Soares Teto Carqueijeiro, K. Koudounas, T. Dugé de Bernonville, B. Ameyaw, D.M. Jones, L.J.C. Vierira, V. Courdavault, S.E. O'Connor, Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle, *Science* 360 (6394) (2018) 1235–1239, <https://doi.org/10.1126/science.aat4100>.
- [36] A. Stavrinides, E.C. Tatsis, L. Caputi, E. Foureau, C.E. Stevenson, D.M. Lawson, V. Courdavault, S.E. O'Connor, Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity, *Nat. Commun.* 7 (1) (2016) 12116, <https://doi.org/10.1038/ncomms12116>.
- [37] A. Stavrinides, E.C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S.E. O'Connor, Unlocking the diversity of alkaloids in *Catharanthus roseus*: nuclear localization suggests metabolic channeling in secondary metabolism, *Chem. Biol.* 22 (3) (2015) 336–341, <https://doi.org/10.1016/j.chembiol.2015.02.006>.
- [38] F. Olawale, T.L. Adetunji, A.E. Adetunji, O. Iwaloye, I.M. Folurunso, The therapeutic value of alstonine: an updated review, *South Afr. J. Bot.* 152 (2023) 288–295, <https://doi.org/10.1016/j.sajb.2022.11.047>.
- [39] E.A. Stander, B. Lehka, I. Carqueijeiro, C. Cuello, F.G. Hansson, H.J. Jensen, T. Dugé de Bernonville, C. Birer Williams, V. Verges, E. Lezin, M.D.B.B. Lorensen, T.T. Dang, A. Oudin, A. Lanoue, M. Durand, N. Giglioli Guivarc'h, C. Janfelt, N. Papon, R.P. Dirks, S.E. O'Connor, M.K. Jensen, S. Besseau, V. Courdavault, The *Rauwolfia tetraphylla* genome suggests multiple distinct biosynthetic routes for yohimbane monoterpene indole alkaloids, *Commun. Biol.* 6 (2023) 1197, <https://doi.org/10.1038/s42003-023-05574-8>.
- [40] E. Lezin, I. Carqueijeiro, C. Cuello, M. Durand, H.J. Jansen, V. Verges, C. Birer Williams, A. Oudin, T. Dugé de Bernonville, J. Petrignet, N. Celton, B. St-Pierre, N. Papon, C. Sun, R.P. Dirks, S.E. O'Connor, M.K. Jensen, S. Besseau, V. Courdavault, A chromosome-scale genome assembly of *Rauwolfia tetraphylla* facilitates identification of the complete ajmaline biosynthetic pathway, *Plant Comm* (2023), <https://doi.org/10.1016/j.xplc.2023.100784>.
- [41] M. Ruppert, X. Ma, J. Stockigt, Alkaloid biosynthesis in *Rauwolfia*-cDNA cloning of major enzymes of the ajmaline pathway, *Curr. Org. Chem.* 9 (15) (2005) 1431–1444, <https://doi.org/10.2174/138527205774370540>.
- [42] J. Stockigt, S. Panjikar, M. Ruppert, L. Barleben, X. Ma, E. Loris, M. Hill, The molecular architecture of major enzymes from ajmaline biosynthetic pathway, *Phytochemistry Rev.* 6 (2007) 15–34, <https://doi.org/10.1007/s11101-006-9016-2>.
- [43] R.G. Dos Santos, J.C. Bouso, J.E. Hallak, The antiaddictive effects of ibogaine: a systematic literature review of human studies, *Journal of Psychedelic Studies* 1 (1) (2017) 20–28, <https://doi.org/10.1556/2054.01.2016.001>.
- [44] S.C. Farrow, M.O. Kamileen, L. Caputi, K. Bussey, J.E. Mundy, R.C. McAtee, C.R.J. Stephenson, S.E. O'Connor, Biosynthesis of an anti-addiction agent from the iboga plant, *J. Am. Chem. Soc.* 141 (33) (2019) 12979–12983, <https://doi.org/10.1021/jacs.9b05999>.
- [45] B. Ahmad, A. Banerjee, H. Tiwari, S. Jana, S. Bose, S. Chakrabarti, Structural and functional characterization of the Vindoline biosynthesis pathway enzymes of *Catharanthus roseus*, *J. Mol. Model.* 24 (2018) 1–14, <https://doi.org/10.1007/s00894-018-3590-2>.
- [46] M. Goboza, Y.G. Aboua, N. Chegou, O.O. Oguntibeju, Vindoline effectively ameliorated diabetes-induced hepatotoxicity by docking oxidative stress, inflammation and hypertriglyceridemia in type 2 diabetes-induced male Wistar rats, *Biomed. Pharmacother.* 112 (2019) 108638, <https://doi.org/10.1016/j.biopha.2019.108638>.
- [47] V. Courdavault, S.E. O'Connor, A. Oudin, S. Besseau, N. Papon, Towards the microbial production of plant-derived anticancer drugs, *Trends in Cancer* 6 (6) (2020) 444–448, <https://doi.org/10.1016/j.trecan.2020.02.004>.
- [48] N. Kulagina, G. Guirimand, C. Melin, P. Lemos-Cruz, I. Carqueijeiro, J.O. De Craene, A. Oudin, V. Heredia, K. Koudounas, M. Unlubayir, A. Lanoue, N. Imbault, B. St-Pierre, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, J. Marc, S. Besseau, V. Courdavault, Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories, *Microb. Biotechnol.* 14 (6) (2021) 2693–2699, <https://doi.org/10.1111/1751-7915.13898>.
- [49] J. Zhang, L.G. Hansen, O. Gudich, K. Viehrig, L.M. Lassen, L. Schrübbers, K.B. Adhikari, P. Rubaszka, E. Carrasquer-Alvarez, L. Chen, V. D'Ambrosio, B. Lehka, A.K. Haidar, S. Nallapareddy, K. Giannakou, M. Laloux, D. Arsovska, M.A.K. Jorgensen, L.J.G. Chan, M. Kristensen, H.B. Christensen, S. Sudarsan, E.A. Stander, E. Baidoo, C.J. Petzold, T. Wulff, S.E. O'Connor, V. Courdavault, M.K. Jensen, J.D. Keasling, A microbial supply chain for production of the anti-cancer drug vinblastine, *Nature* 609 (7926) (2022) 341–347, <https://doi.org/10.1038/s41586-022-05157-3>.
- [50] V. Salim, A.D. Jones, D. DellaPenna, *Camptotheca acuminata* 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism, *Plant J.* 95 (1) (2018) 112–125, <https://doi.org/10.1111/tpj.13936>.
- [51] P. Lemos Cruz, I. Carqueijeiro, K. Koudounas, D.P. Bomzan, E.A. Stander, C. Abdallah, N. Kulagina, A. Oudin, A. Lanoue, N. Giglioli-Guivarc'h, D. A. Nagegowda, N. Papon, S. Besseau, M. Clastre, V. Courdavault, Identification of a second 16-hydroxytabersonine-O-methyltransferase suggests an evolutionary relationship between alkaloid and flavonoid metabolisms in *Catharanthus roseus*, *Protoplasma* 260 (2) (2023) 607–624, <https://doi.org/10.1007/s00709-022-01801-x>.
- [52] F. Kellner, F. Geu-Flores, N.H. Sherden, S. Brown, E. Foureau, V. Courdavault, S.E. O'Connor, Discovery of a P450-catalyzed step in vindoline biosynthesis: a link between the aspidosperma and eburnamine alkaloids, *Chem. Commun.* 51 (2015) 7626–7628, <https://doi.org/10.1039/c5cc01309g>.
- [53] J. Franke, J. Kim, J.P. Hamilton, D. Zhao, G.M. Pham, K. Wiegert-Rininger, E. Crisovan, L. Newton, B. Vaillancourt, E. Tatsis, C.R. Buell, S.E. O'Connor, Gene discovery in *Gelsemium* highlights conserved gene clusters in monoterpene indole alkaloid biosynthesis, *ChemBiochem* 20 (2019) 83–87, <https://doi.org/10.1002/CBIC.201800592>.
- [54] C. Cuello, E.A. Stander, H.J. Jansen, T.D. Dugé De Bernonville, A. Oudin, C. Birer-Williams, A. Lanoue, N. Giglioli-Guivarc'h, N. Papon, M. Krogh Jensen, S. E. O'Connor, S. Besseau, V. Courdavault, An updated version of the Madagascar periwinkle genome, *F1000Research* 11 (2022), <https://doi.org/10.12688/f1000research.129212.1>.
- [55] S. Sun, X. Shen, Y. Li, Y. Li, S. Wang, R. Li, H. Zhang, G. Shen, B. Guo, J. Wei, J. Xu, B. St-Pierre, S. Chen, S. Chao, Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism, *Nat. Plants* 9 (2023) 179–190, <https://doi.org/10.1038/s41477-022-01291-y>.
- [56] C. Li, J.C. Wood, A.H. Vu, J.P. Hamilton, C.E. Rodriguez Lopez, R.M.E. Payne, D.A.S. Guerrero, K. Gase, K. Yomamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C.R. Buell, Single-cell multi-omics in the medicinal plant *Catharanthus roseus*, *Nat. Chem. Biol.* (2023), <https://doi.org/10.1038/s41589-023-01327-0>.
- [57] E.A. Stander, C. Cuello, C. Birer-Williams, N. Kulagina, H.J. Jansen, I. Carqueijeiro, L.V. Méteignier, V. Vergès, A. Oudin, N. Papon, R. Dirks, M.K. Jensen, S. E. O'Connor, T. Dugé de Bernonville, S. Besseau, V. Courdavault, The *Vinca* minor genome highlights conserved evolutionary traits in monoterpene indole alkaloid synthesis, *G3 Genes/Genomes/Genetics* (2022) jkac268, <https://doi.org/10.1093/g3journal/jkac268>.
- [58] J.S. Sabir, R.K. Jansen, D. Arasappan, V. Calderon, E. Noutahi, C. Zheng, S. Park, M.J. Sabir, M.N. Baeshen, N.H. Hajrah, M.A. Khyami, N.A. Baeshen, A. Y. Obaid, A.L. Al-Malki, D. Sankoff, N. El-Mabrouk, T.A. Ruhlman, The nuclear genome of *Rhazya stricta* and the evolution of alkaloid diversity in a medically relevant clade of Apocynaceae, *Sci. Rep.* 6 (1) (2016) 33782, <https://doi.org/10.1038/srep33782>.
- [59] F. Adabavazeh, S. Pourseyedi, N. Nadernejad, R. Razavizadeh, Hairy root induction in *Calotropis procera* and optimization of its phytochemical characteristics by elicitors, *Plant Cell Tissue Organ Cult.* (2023), <https://doi.org/10.1007/s11240-023-02481-y>.
- [60] C.M. Naidoo, Y. Naidoo, Y.H. Dewir, H.N. Murthy, S. El-Hendawy, N. Al-Suhaibani, Major bioactive alkaloids and biological activities of *Tabernaemontana* species (Apocynaceae), *Plants* 10 (2) (2021) 313, <https://doi.org/10.3390/plants10020313>.
- [61] J.K. Weng, J.H. Lynch, J.O. Matos, N. Dudareva, Adaptive mechanisms of plant specialized metabolism connecting chemistry to function, *Nat. Chem. Biol.* 17 (2021) 1037–1045, <https://doi.org/10.1038/s41589-021-00822-6>.

- [62] C. Cuello, E.A. Stander, H.J. Jansen, T. Dugé de Bernonville, A. Lanoue, N. Giglioli-Guivarc'h, N. Papon, R.P. Dirks, M.K. Jensens, S.E. O'Connor, S. Besseau, V. Courdavault, Genome assembly of the medicinal plant *Voacanga thouarsii*, *Genome Biology and Evolution* 14 (11) (2022) evac158, <https://doi.org/10.1093/gbe/evac158>.
- [63] Y. Jiao, J. Leebens-Mack, S. Ayyampalayam, J.E. Bowers, M.R. McKain, J. McNeal, M. Rolf, D.R. Ruzicka, E. Wafula, N.J. Wickett, X. Wu, Y. Zhang, J. Wang, Y. Zhang, E.J. Carpenter, M.K. Ceyholos, T.M. Kutchan, A.S. Chanderbali, P.S. Soltis, D.W. Stevenson, R. McCombie, J.C. Pires, G.K.S. Wong, D.E. Soltis, C. W. Depamphilis, A genome triplication associated with early diversification of the core eudicots, *Genome Biol.* 13 (1) (2012) 1–14, <https://doi.org/10.1186/gb-2012-13-1-r3>.
- [64] M. Sahebi, M.M. Hanafi, A.J. van Wijnen, D. Rice, M.Y. Rafii, P. Azizi, M. Osma, S. Taheri, M.F. Abu Bakar, M.N. Mat Isa, Y.M. Noor, Contribution of transposable elements in the plant's genome, *Gene* 665 (2018) 155–166, <https://doi.org/10.1016/j.gene.2018.04.050>.
- [65] A. Piovan, R. Filippini, D. Favretto, Characterization of the anthocyanins of *Catharanthus roseus* (L.) G. Don in vivo and in vitro by electrospray ionization ion trap mass spectrometry, *Rapid Commun. Mass Spectrom.* 12 (7) (1998) 361–367, [https://doi.org/10.1002/\(SICI\)1097-0231\(19980415\)12:7<361::AID-RCM162>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0231(19980415)12:7<361::AID-RCM162>3.0.CO;2-U).
- [66] N.R. Mustafa, R. Verpoorte, Phenolic compounds in *Catharanthus roseus*, *Phytochemistry Rev.* 6 (2007) 243–258, <https://doi.org/10.1007/s11101-006-9039-8>.
- [67] C. Cuello, H.J. Jansen, C. Abdallah, D.L. Zamar, C. Birer-Williams, M. Durand, A. Oudin, N. Papon, N. Giglioli-Guivarc'h, R.P. Dirks, M.K. Jensen, S. E. O'Connor, S. Besseau, V. Courdavault, Datasets for *Pachypodium lamerei*, *Apocynum venetum*, *Asclepias syriaca* and *Calotropis procera* genomes, *FigShare* (2023), <https://doi.org/10.6084/m9.figshare.23126816>.
- [68] R. Sadre, M. Magallanes-Lundback, S. Pradhan, V. Salim, A. Mesberg, A.D. Jones, D. DellaPenna, Metabolite diversity in alkaloid biosynthesis: a multilane (Diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*, *Plant Cell* 28 (8) (2016) 1926–1944, <https://doi.org/10.1105/tpc.16.00193>.
- [69] A. Rai, H. Hirakawa, R. Nakabayashi, S. Kikuchi, K. Hayashi, M. Rai, H. Tsugawa, T. Nakaya, T. Mori, H. Nagasaki, R. Fukushi, Y. Kusuya, H. Takahashi, H. Uchiyama, A. Toyoda, S. Hikosaka, E. Goto, K. Saito, M. Yamazaki, Chromosome-level genome assembly of *Ophiorrhiza pumila* reveals the evolution of camptothecin biosynthesis, *Nat. Commun.* 12 (1) (2021) 1–19, <https://doi.org/10.1038/s41467-020-20508-2>, 12.
- [70] A.O. Simões, T. Livshultz, E. Conti, M.E. Endress, Phylogeny and systematics of the Rauvolfioideae (Apocynaceae) based on molecular and morphological evidence, *Ann. Mo. Bot. Gard.* 94 (2) (2007) 268–297, [https://doi.org/10.3417/0026-6493\(2007\)94\[268:PASOTR\]2.0.CO;2](https://doi.org/10.3417/0026-6493(2007)94[268:PASOTR]2.0.CO;2).
- [71] S.B. Cannon, A. Mitra, A. Baumgarten, N.D. Young, G. May, The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*, *BMC Plant Biol.* 4 (1) (2004) 1–21, <https://doi.org/10.1186/1471-2229-4-10>.
- [72] Y.L. Huang, L.K. Zhang, K. Zhang, S.M. Chen, J.B. Hu, F. Cheng, The impact of tandem duplication on gene evolution in Solanaceae species, *J. Integr. Agric.* 21 (4) (2022) 1004–1014, [https://doi.org/10.1016/S2095-3119\(21\)63698-5](https://doi.org/10.1016/S2095-3119(21)63698-5).
- [73] E. Rodgers-Melnick, S.P. Mane, P. Dharmawardhana, G.T. Slavov, O.R. Crasta, S.H. Strauss, A.M. Brunner, S.P. DiFazio, Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*, *Genome Res.* 22 (1) (2012) 95–105, <https://doi.org/10.1101/gr.125146.111>.
- [74] L.V. Méteignier, H.W. Nützmann, N. Papon, A. Osbourn, V. Courdavault, Emerging mechanistic insights into the regulation of specialized metabolism in plants, *Nat. Plants* 9 (1) (2023) 22–30, <https://doi.org/10.1038/s41477-022-01288-7>.
- [75] K. Hu, K. Xu, J. Wen, B. Yi, J. Shen, C. Ma, T. Fu, Y. Ouyang, J. Tu, Helitron distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species, *BMC Bioinf.* 20 (1) (2019) 1–20, <https://doi.org/10.1186/s12859-019-2945-8>.
- [76] Y. Li, H.K. Dooner, Excision of Helitron transposons in maize, *Genetics* 182 (1) (2009) 399–402, <https://doi.org/10.1534/genetics.109.101527>.
- [77] M. Park, A. Sarkhosh, V. Tsoлова, I. El-Sharkawy, Horizontal transfer of LTR retrotransposons contributes to the genome diversity of *Vitis*, *Int. J. Mol. Sci.* 22 (19) (2021) 10446, <https://doi.org/10.3390/ijms221910446>.
- [78] H. Chen, Y. Zhang, S. Feng, Whole-genome and dispersed duplication, including transposed duplication, jointly advance the evolution of TLP genes in seven representative Poaceae lineages, *BMC Genom.* 24 (2023) 290, <https://doi.org/10.1186/s12864-023-09389-z>.
- [79] N. Jiang, Z. Bao, X. Eddy SR, Zhang, S.R. Wessler, Pack-MULE transposable elements mediate gene evolution in plants, *Nature* 431 (2004) 569–573, <https://doi.org/10.1038/nature02953>.
- [80] G. Carrier, S. Santoni, M. Rodier-Goud, A. Canaguier, A. De Kochko, C. Dubreuil-Tranchant, P. This, J.M. Boursiquot, L. Le Cunff, An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods, *Am. J. Bot.* 98 (1) (2011) e13–e15, <https://doi.org/10.3732/ajb.1000371>.
- [81] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Completing bacterial genome assemblies with multiplex MinION sequencing, *Microb. Genom.* 14 (10) (2017) e000132, <https://doi.org/10.1099/mgen.0.000132>, 3.
- [82] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, *Nat. Biotechnol.* 37 (5) (2019) 540–546, <https://doi.org/10.1038/s41587-019-0072-8>, 2019 37.
- [83] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakhthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, *PLoS One* 9 (2014), <https://doi.org/10.1371/JOURNAL.PONE.0112963> e112963–944 e112963.
- [84] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, *Nat. Biotechnol.* 37 (8) (2019) 907–915, <https://doi.org/10.1038/s41587-019-0201-4>, 2019 37.
- [85] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, *Nat. Biotechnol.* 33 (3) (2015) 290–295, <https://doi.org/10.1038/nbt.3122>, 2015 33.
- [86] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. Williams, C.N. Dewey, R. Henschel, R.D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity, *Nat. Protoc.* 8 (2013), <https://doi.org/10.1038/nprot.2013.084>.
- [87] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, *Nucleic Acids Res.* 39 (2011) W29–W37, <https://doi.org/10.1093/NAR/GKR367>.
- [88] A. Rhie, B.P. Walenz, S. Koren, A.M. Phillippy, Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies, *Genome Biol.* 21 (2020) 245, <https://doi.org/10.1186/s13059-020-02134-9>.
- [89] B. Bushnell, *BBMap: A Fast, Accurate, Splice-Aware Aligner* (No. LBNL-7065E), Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), 2014.
- [90] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* 31 (2015) 3210–3212, <https://doi.org/10.1093/bioinformatics/btv351>.
- [91] J. Dainat, D. Hereñú, LucileSol pascal-git, NBISweden/AGAT: AGAT-v0.8.1, Zenodo, 2022, <https://doi.org/10.5281/zenodo.5834795>.
- [92] S. Ou, W. Su, Y. Liao, K. Chougule, J.R.A. Agda, A.J. Hellinga, C.S.B. Lugo, T.A. Elliott, D. Ware, T. Peterson, N. Jiang, C.N. Hirsch, M.B. Hufford, Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline, *Genome Biol.* 20 (2019) 275, <https://doi.org/10.1186/s13059-019-1905-y>.
- [93] Z. Xu, H. Wang, LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res.* 35 (2007) W265–W268. *Web Server issue*.
- [94] D. Ellinghaus, S. Kurtz, U. Willhoef, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons, *BMC Bioinf.* 9 (1) (2008) 18, <https://doi.org/10.1186/1471-2105-9-18>.
- [95] X. Shi, S. Cao, X. Wang, S. Huang, Y. Wang, Z. Liu, Y. Zhou, The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding, *Horticulture Research* 10 (5) (2023) uhad061, <https://doi.org/10.1093/hr/uhad061>.
- [96] W. Su, X. Gu, T. Peterson, TIR-learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome, *Mol. Plant* 12 (3) (2019) 447–460, <https://doi.org/10.1016/j.molp.2019.02.008>.
- [97] W. Xiong, L. He, J. Lai, H.K. Dooner, C. Du, HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes, *Proc. Natl. Acad. Sci. U.S.A.* 111 (28) (2014) 10263–10268, <https://doi.org/10.1073/pnas.1410068111>.

- [98] S. Ou, N. Jiang, LTR retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons, *Plant Physiol.* 76 (2) (2018) 1410–1422.
- [99] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, A.F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families, *Proc. Natl. Acad. Sci. USA* 117 (17) (2020) 9451–9457, <https://doi.org/10.1073/pnas.1921046117>.
- [100] P. Lamesch, T.Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D.L. Alexander, M. Garcia-Hernandez, A.S. Karthikeyan, C.H. Lee, W.D. Nelson, L. Ploetz, S. Singh, A. Wensel, E. Huala, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools, *Nucleic Acids Res.* 40 (2012), <https://doi.org/10.1093/NAR/GKR1090>.
- [101] J. Brose, K.H. Lau, T.T.T. Dang, J.P. Hamilton, L.D.V. Martins, Britta Hamberger, Bjoern Hamberger, J. Jiang, S.E. O'Connor, C.R. Buell, The *Mitragyna speciosa* (Kratom) Genome: a resource for data-mining potent pharmaceuticals that impact human health, *G3 GenesGenomesGenetics* 11 (2021), <https://doi.org/10.1093/G3JOURNAL/JKAB058>.
- [102] P.S. Hosmani, M. Flores-Gonzalez, H. van de Geest, F. Maumus, L.V. Bakker, E. Schijlen, J. van Haarst, J. Cordewener, G. Sanchez-Perez, S. Peters, Z. Fei, J. J. Giovannoni, L.A. Mueller, S. Saha, An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps, 2019, <https://doi.org/10.1101/767764>.
- [103] M. Kang, R. Fu, P. Zhang, S. Lou, X. Yang, Y. Chen, T. Ma, Y. Zhang, Z. Xi, J. Liu, A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis, *Nat. Commun.* 12 (1) (2021) 1–12, <https://doi.org/10.1038/s41467-021-23872-9>, 2021 12.
- [104] G.M. Hoopes, J.P. Hamilton, J. Kim, D. Zhao, K. Wiegert-Rininger, E. Crisovan, C.R. Buell, Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides, *G3 GenesGenomesGenetics* 8 (2018) 385–391, <https://doi.org/10.1534/g3.117.300331>.
- [105] J. Shi, C. Liang, Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection, *Plant Physiol.* 180 (4) (2019) 1803–1815, <https://doi.org/10.1104/pp.19.00386>.
- [106] M.S. Barker, K.M. Dlugosch, L. Dinh, R.S. Challa, N.C. Kane, M.G. King, L.H. Rieseberg, EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics, *Evol. Bioinform. Online* 6 (2010) 143–149, <https://doi.org/10.4137/EBO.S5861>.
- [107] B. Ma, J. Tromp, M. Li, PatternHunter: faster and more sensitive homology search, *Bioinformatics* 18 (2002) 440–445, <https://doi.org/10.1093/BIOINFORMATICS/18.3.440>.
- [108] Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A Greedy Algorithm for Aligning DNA Sequences, 2004, pp. 203–214, <https://doi.org/10.1089/10665270050081478>. <https://home.liebertpub.com/cmb7>.
- [109] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, *BMC Bioinf.* 10 (2009) 421, <https://doi.org/10.1186/1471-2105-10-421>.
- [110] E. Birney, M. Clamp, R. Durbin, GeneWise and genomewise, *Genome Res.* 14 (2004) 988–995, <https://doi.org/10.1101/GR.1865504>.
- [111] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinf.* 5 (2004) 113, <https://doi.org/10.1186/1471-2105-5-113>.
- [112] R. Wernersson, A.G. Pedersen, RevTrans: multiple alignment of coding DNA from aligned amino acid sequences, *Nucleic Acids Res.* 31 (2003) 3537–3539, <https://doi.org/10.1093/NAR/GKG609>.
- [113] Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood, *Bioinformatics* 13 (1997) 555–556, <https://doi.org/10.1093/bioinformatics/13.5.555>.
- [114] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data, *Bioinformatics* 28 (2012) 3150–3152, <https://doi.org/10.1093/BIOINFORMATICS/BTS565>.
- [115] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.* 20 (2019), <https://doi.org/10.1186/S13059-019-1832-Y>.
- [116] F.K. Mendes, D. Vanderpool, B. Fulton, M.W. Hahn, CAFE 5 models variation in evolutionary rates among gene families, *Bioinformatics* 36 (2020) 5516–5518, <https://doi.org/10.1093/BIOINFORMATICS/BTAA1022>.
- [117] H. Li, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics* 34 (18) (2018) 3094–3100, <https://doi.org/10.1093/bioinformatics/bty191>.
- [118] F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an interactive, efficient and simple way, *PeerJ* 6 (2018) e4958, <https://doi.org/10.7717/peerj.4958>.
- [119] L. Guy, J. Roat Kultima, S.G.E. Andersson, genoPlotR: comparative gene and genome visualization in R, *Bioinformatics* 26 (2010) 2334–2335, <https://doi.org/10.1093/bioinformatics/btq413>.
- [120] R Core Team. R, A Language and Environment for Statistical Computing, 2020. Vienna: Austria.
- [121] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, *Nat. Methods* 14 (4) (2017) 417–419.
- [122] T. Dugé de Bernonville, I. Carqueijeiro, A. Lanoue, F. Lafontaine, P. Sánchez Bel, F. Liesecke, K. Musset, A. Oudin, G. Glévarec, O. Pichon, S. Besseau, M. Clastre, B. St-Pierre, V. Flors, S. Maury, E. Huguet, S.E. O'Connor, V. Courdavault, Folivory elicits a strong defense reaction in *Catharanthus roseus*: metabolomic and transcriptomic analyses reveal distinct local and systemic responses, *Sci. Rep.* 7 (2017) 40453, <https://doi.org/10.1038/srep40453>.
- [123] C. Elejalde-Palmett, K. Billet, A. Lanoue, J.O. De Craene, G. Glévarec, O. Pichon, M. Clastre, V. Courdavault, B. St Pierre, N. Giglioli-Guivarc'h, T. Dugé de Bernonville, S. Besseau, Genome-wide identification and biochemical characterization of the UGT88F subfamily in *Malus x domestica* Borkh, *Phytochemistry* 157 (2019) 135, <https://doi.org/10.1016/j.phytochem.2018.10.019>.