# Complex scaffold remodeling in plant triterpene biosynthesis

**Authors:** Ricardo De La Peña[1]†, Hannah Hodgson[2]†, Jack Chun-Ting Liu[3]†, Michael J. Stephenson[4], Azahara C. Martin[5], Charlotte Owen[2], Alex Harkess[6], Jim Leebens-Mack[7], Luis E. Jimenez[1], Anne Osbourn[2]* and Elizabeth S. Sattely[1,8]*

**Affiliations:**

[1]Department of Chemical Engineering, Stanford University; Stanford, CA 94305, US.

[2]Department of Biochemistry and Metabolism, John Innes Centre; Norwich Research Park, Norwich NR4 7UH, UK.

[3]Department of Chemistry, Stanford University; Stanford, CA 94305, US.

[4]School of Chemistry, University of East Anglia; Norwich Research Park, Norwich NR4 7TJ, UK.

[5]Department of Crop Genetics, John Innes Centre; Norwich Research Park, Norwich NR4 7UH, UK.

[6]HudsonAlpha Institute for Biotechnology; Huntsville, AL 35806, US.

[7]Department of Plant Biology, 4505 Miller Plant Sciences, University of Georgia; Athens, GA 30602, US.

[8]Howard Hughes Medical Institute, Stanford University; Stanford, CA 94305, US.

† These authors contributed equally to this work

* Corresponding author. Email: Anne Osbourn anne.osbourn@jic.ac.uk, Elizabeth S. Sattely sattely@stanford.edu

**Abstract:** Triterpenes with complex scaffold modifications are widespread in the plant kingdom. Limonoids are an exemplary family that are responsible for the bitter taste in citrus (e.g., limonin) and the active constituents of neem oil, a widely used bioinsecticide (e.g., azadirachtin). Despite the commercial value of limonoids, a complete biosynthetic route has not been described. Here, we report the discovery of 22 enzymes, including a pair of neofunctionalized sterol isomerases, that catalyze 12 unique reactions in the total biosynthesis of kihadalactone A and azadiraone, products that bear the signature limonoid furan. These results enable access to valuable limonoids and provide a template for discovery and reconstitution of triterpene biosynthetic pathways in plants that require multiple skeletal rearrangements and oxidations.

**One-Sentence Summary:** Discovery of 22 enzymes responsible for the production of bioactive limonoids with complex scaffold rearrangements from Citrus and Meliaceae species.

## Main Text

Among numerous complex triterpenes that are found in the plant kingdom, limonoids are particularly notable given their wide range of biological activities and structural diversity that stems from extensive scaffold modifications (*1*, *2*). Produced by mainly two families in the Sapindales, Rutaceae (citrus) and Meliaceae (mahogany) (*3*), these molecules bear a signature furan and include over 2,800 known structures (*4*, *5*). Azadirachtin, a well-studied limonoid, exemplifies the substantial synthetic challenge for this group of molecules, with 16 stereocenters and 7 quaternary carbons. Few synthetic routes to limonoids have been reported (*6*), (*7*), (*8*), and, more generally, complete biosynthetic pathways to triterpenes with extensive scaffold modifications have remained elusive. This lack of production routes limits the utility and biological investigation of clinical candidates from this diverse compound class (*9*).

Around 90 limonoids have also been reported to have anti-insect activity (*2*), and several have also been found to target mammalian receptors and pathways (*4*). For example, azadirachtin (Fig. 1), the main component of biopesticides derived from the neem tree (*Azadirachta indica*), is a potent antifeedant, active against >600 insect species (*9*). Perhaps related to antifeedant activity, Rutaceae limonoids such as nomilin, obacunone, and limonin (Fig. 1) that accumulate in *Citrus* species at high levels (*3*) are partially responsible for the "delayed bitterness" of citrus fruit juice, which causes serious economic losses for the citrus juice industry worldwide (*10*). In mammalian systems, several limonoids have shown inhibition of HIV-1 replication (*11*) and anti-inflammatory activity (*12*). Some limonoids of pharmaceutical interest have also been associated with specific mechanisms of action: gedunin (Fig. 1) and nimbolide (fig. S1) exert potent anti-cancer activity through Hsp90 inhibition (*13*) and RNF114 blockade (*14*, *15*), respectively.

Limonoids are unusual within the triterpene class due to their extensive biosynthetic scaffold rearrangements. They are referred to as tetranortriterpenoids because their signature tetracyclic, triterpene scaffold (protolimonoid) loses four carbons during the formation of a signature furan ring to give rise to the basic C26 limonoid structure (Fig. 1). A range of modifications can then occur to the basic limonoid scaffold through the cleavage of one or more of the four main rings (*16*, *17*) (fig. S1). Radioactive isotope labeling studies suggest that most Rutaceae limonoids are derived from a nomilin-type intermediate (*seco*-A,D ring scaffolds) whereas Meliaceae limonoids are derived from an azadirone-type intermediate (intact A ring) (Fig. 1) (*4*, *5*,*18*, *19*). It is proposed that at least two main scaffold modifications are conserved in both plant families: a C-30 methyl shift of the protolimonoid scaffold (*apo*-rearrangement) and the conversion of the hemiacetal ring of melianol (1) to a mature furan ring with a concomitant loss of the C-25~C-28 carbon side chain (Fig. 1) (*20*). Additional modifications specific to Rutaceae and Meliaceae would then yield the nomilin- and azadirone-type intermediates. The diversity and array of protolimonoid structures isolated beyond melianol (**1**) (fig. S1) hint at a series of possible conserved biosynthetic transformations, including hydroxylation and/or acetoxylation on C-1,C-

7 and C-21, which suggests involvement of cytochrome P450s (CYPs), 2-oxoglutarate-dependent dioxygenases (2-ODDs) and acetyltransferases.

Despite extensive interest in the biology and chemistry of complex plant triterpenes over the last half century, few complete biosynthetic pathways have been described. A notable exception is the disease resistance saponin from oat, avenacin A-1, whose pathway consists of 4 CYP-mediated scaffold modifications and 6 side-chain tailoring steps (*21*). Barriers to pathway reconstitution of complex triterpenes include a lack of knowledge of the structures of key intermediates, order of scaffold modification steps, instability of pathway precursors, and the challenge of identifying candidate genes for the anticipated >10 enzymatic transformations required to generate advanced intermediates. Limonoids are no exception; to date, only the first three enzymatic steps to the protolimonoid melianol (**1**) from the primary metabolite 2,3-oxidosqualene have been elucidated (Fig. 1) (*20*). In this work, we used systematic transcriptome and genome mining, phylogenetic and homologous analysis, coupled with *N. benthamiana* as a heterologous expression platform, to identify suites of candidate genes from *Citrus sinensis* and *Melia azedarach* that can be used to reconstitute limonoid biosynthesis.

***Identification of candidate limonoid biosynthetic genes***

One genome of Rutaceae plants (*C. sinensis* var. Valencia) and several transcriptome resources, including from Citrus and Meliaceae plants (two from *A. indica* and one from *M. azedarach*) were previously used to identify the first three enzymes in the limonoid pathway (*20*). These included an oxidosqualene cyclase (*Cs*OSC1 from *C. sinensis*, *Ai*OSC1 from *A. indica*, and *Ma*OSC1 from *M. azedarach*), and two CYPs (*Cs*CYP71CD1/*Ma*CYP71CD2 and *Cs*CYP71BQ4/*Ma*CYP71BQ5) that complete the pathway to melianol (*20*). To identify enzymes that further tailor melianol (**1**), we expanded our search to include additional sources. For Rutaceae enzyme identification, we included publicly available microarray data compiled by the Network inference for Citrus Co-Expression (NICCE) (*22*). For Meliaceae enzyme identification, we generated additional RNA-seq data and a reference-quality genome assembly and annotation.

Of publicly available microarray data for Citrus, fruit datasets were selected for in depth analysis as *CsOSC1* expression levels were highest in the fruit and it has been implicated as the site of limonin biosynthesis and accumulation (*19*). Gene co-expression analysis was first performed on the Citrus fruit dataset using only *CsOSC1* as the bait gene. This revealed promising candidate genes exhibiting highly correlated expression with *CsOSC1* (fig. S2). As we characterized more limonoid biosynthetic genes (as described below) we also included these as bait genes to enhance the stringency of co-expression analysis and further refine the candidate list. The top-ranking candidate list is rich in genes typically associated with secondary metabolism (Fig. 2A). The list specifically included multiple predicted CYPs, 2-ODDs and acetyltransferases, consistent with the proposed biosynthetic transformations.

Efforts to identify and clone candidate genes from *M. azedarach* have previously been limited by the lack of a reference genome with high-quality gene annotations and by the lack of suitable transcriptomic data for co-expression analysis (i.e. multiple tissues, with replicates). Therefore, in parallel to our search in *Citrus*, we generated genomic and transcriptomic resources for *M. azedarach.* A pseudochromosome level reference-quality *M. azedarach* genome assembly was generated using PacBio long-read and Hi-C sequencing technologies (table S1, fig. S3). Although the assembled genome size (230 Mbp) is smaller than available literature predictions for this species of 421 Mbp (*23*), the chromosome number (1n=14) matches literature reports (*23*) and was confirmed by karyotyping (fig. S4). The genome assembly annotation predicted 22,785 high-confidence protein coding genes (Fig. 2B, table S1). BUSCO assessment (*24*) of this annotation confirmed the completeness of the genome, as 93% of expected orthologs are present as complete single copy genes (comparable to 98% in the gold standard *Arabidopsis thaliana*) (Fig. 2B, table S1).

Illumina paired-end RNA-seq reads were generated for three different *M. azedarach* tissues (7 different tissues in total, with four replicates of each tissue, table S2), previously shown to differentially accumulate and express limonoids and their biosynthetic genes (*20*). Read-counts were generated by aligning RNA-Seq reads to the genome annotation, and EdgeR (*25*) was used to identify a subset of 18,151 differentially expressed genes (P-value < 0.05). The known melianol biosynthetic genes *MaOSC1*, *MaCYP71CD2* and *MaCYP71BQ5* (*20*) were used as bait genes for co-expression analysis across the sequenced tissues and the resulting ranked list was filtered by their Interpro domain annotations to enrich for relevant biosynthetic enzyme-coding genes. This informed the selection of 17 candidate genes for further investigation for functional analysis along with Citrus candidates (Fig. 2C).

***Citrus CYP88A51 and Melia CYP88A108 act with different melianol oxide isomerases (MOIs) to form distinct proto-limonoid scaffolds***

Top-ranking genes from both the *Citrus* and *Melia* candidate lists (Fig. 2A, 2C) were tested for function by *Agrobacterium*-mediated transient expression in *N. benthamiana* with the previously reported melianol (**1**) biosynthetic enzymes *Cs*OSC1, *Cs*CYP71CD1, and *Cs*CYP71BQ4 or *Ai*OSC1, *Ma*CYP71CD2, and *Ma*CYP71BQ4. LC/MS analysis of crude methanolic extracts from *N. benthamiana* leaves revealed that the expression of either *Cs*CYP88A51 or *Ma*CYP88A108, in combination with their respective melianol biosynthesis genes, led to the disappearance of melianol (**1**) and the accumulation of multiple mono-oxidized products (Fig. 3A, fig. S5 to S6). This result suggested that, while these CYP88A enzymes accept melianol as a substrate, the resulting products could be unstable or undergo further modification by endogenous *N. benthamiana* enzymes.

Despite the accumulation of multiple related metabolites, we continued to screen additional co-expressed candidate genes for further activity. This screen included homologs of *A. thaliana HYDRA1*, an ER membrane protein known as a sterol isomerase (SI) (two from the *Citrus*

candidate list, and one from the *Melia* list). SIs are exclusively associated with phytosterol and cholesterol biosynthesis, where they catalyze double bond isomerization from the C-8 to the C-7 position. They are present in all domains of life and are required for normal development of mammals (*26*), plants (*27*) and yeast (*28*). Testing of these putative *SI*s through transient *Agrobacterium*-mediated gene expression in *N. benthamiana* resulted in a marked change of the metabolite profile with the accumulation of a single mono-oxidized product with no mass change (Fig. 3A, fig. S7). We suspected that these enzymes were able to capture unstable intermediates and promote isomerization of the C30 methyl group required to generate mature limonoids. These sterol isomerases are therefore re-named melianol oxide isomerases, *CsMOI1-3* and *MaMOI2,* because of their ability to generate isomers of mono-oxidized melianol products.

*SI*s are typically found as single copy genes in given plant species. Surprisingly, we found additional putative *SI* genes in the *C. sinensis* and *M. azedarach* genomes, four and three, respectively (fig. S8). Phylogenetic analysis of *SIs* across a set of diverse plant species revealed that *SIs* from *C. sinensis* and *M. azedarach* fall into two distinct sub-clades (Fig. 3B). The more conserved of these clades contained one sequence from each species (*CsSI* and *MaSI*), whilst the more divergent clade contained the remaining *SIs* (*CsMOI1-3* and *MaMOI1*,*2*). This suggested that *CsSI* and *MaSI* are the conserved genes involved in phytosterol biosynthesis. Comparison of all *C. sinensis* and *M. azedarach* SI/MOI protein sequences showed that *Cs*MOI2 is ~93% identical at the protein level to *Cs*MOI3 and ~83% to *Ma*MOI2, but only ~54% and ~60% similar to *Cs*MOI1 and *Cs*SI, respectively (Fig. 3C). Although *CsMOI1*, *CsMOI2*, and *MaMOI2* ranked among the top 100 genes in our co-expression analysis lists (Fig. 3D), *CsSI*, *MaMOI1* and *MaSI* do not co-express with limonoid biosynthetic genes. The absence of *CsMOI3* from this list is attributed to the lack of specific microarray probes required for expression monitoring. Notably, screening of *Cs*SI in the *N. benthamiana* expression system did not change the product profile of *Cs*CYP88A51, consistent with its predicted involvement in primary metabolism based on the phylogenetic analysis (Fig. 3A).

To determine the chemical structures of the isomeric products formed through the action of these MOIs, we carried out large-scale expression experiments in *N. benthamiana* and isolated 13.1 mg of pure product. NMR analysis revealed the product of *Ma*MOI2 to be the epimeric mixture *apo*-melianol (**3**) bearing the characteristic limonoid scaffold with a migrated C-30 methyl group on C-8, a C-14/15 double bond, and C-7 hydroxylation (Fig. 3E, table S3) (*29*). Although the structure of the direct product of *Cs*MOI2 was not determined until after the discovery of two additional downstream tailoring enzymes, NMR analysis also confirmed C-8 methyl migration (table S4). These data indicate that, as predicted by sequence analysis, *Cs*MOI2 and *Ma*MOI2 indeed are functional homologs and catalyze a key step in limonoid biosynthesis by promoting an unprecedented methyl shift. Analysis of the product formed with expression of *Cs*MOI1, indicated the presence of a metabolite with a different retention time relative to *apo*-melianol (**3**) (Fig. 3A). Isolation and NMR analysis of (**4'**), a metabolite derived from (**4**) after inclusion of

two additional tailoring enzymes (table S5), indicated C-30 methyl group migration to C-8 and cyclopropane ring formation via bridging of the C18 methyl group to C-14.

Based on the characterized structures, we proposed that in the absence of MOIs, the CYP88A homologs form the unstable C-7/8 epoxide (**2**), which may either spontaneously undergo a Wagner-Meerwein rearrangement via C-30 methyl group migration and subsequent epoxide-ring-opening or degrade through other routes to yield multiple rearranged products (**2a**), (**2b**), (**2c**) and (**3**) (Fig. 3E). MOIs appear to stabilize the unstable carbocation intermediate and isomerize it to two types of limonoids: *Cs*MOI2, *Cs*MOI3 and *Ma*MOI2 form the C-14/15 double bond scaffold (classic limonoids) while *Cs*MOI1 forms the cyclopropane ring scaffold (glabretal limonoids). Glabretal limonoids have been isolated from certain Meliaceae and Rutaceae species before but are less common (*30*, *31*). *Cs*CYP88A51, *Ma*CYP88A108 and two different types of MOIs are thus responsible for rearrangement from melianol (**1**) to either (**3**) or (**4**) through an epoxide intermediate (**2**). These MOIs represent neofunctionalization of sterol isomerases from primary metabolism in plants.

***Characterization of conserved tailoring enzymes L21AT and SDR***

Having enzymes identified for the methyl shift present in the limonoids, we continued screening other candidate genes (Fig. 2A, 2C) for activity on (**3**) towards downstream products. BAHD-type acetyltransferases (named *Cs*L21AT or *Ma*L21AT, limonoid 21-*O*-acetyltransferse) and short-chain dehydrogenase reductases (*Cs*SDR and its homolog *Ma*SDR) result in the loss of compound (**3**), and the accumulation of acetylated and a dehydrogenated products, respectively (fig. S9 to S12). While the sequence of events can be important for some enzymatic transformations in plant biosynthesis, L21AT and SDR homologs appear to have broad substrate specificity. Our data suggests that L21AT can act on (**1**) or (**3**), and SDR is active on all intermediates after the OSC1 product (fig. S13 to S14), suggesting a flexible reaction order in the early biosynthetic pathway.

Furthermore, the products formed from the modification of (**3**) by both Citrus and Melia L21AT and SDR homologs were purified by large-scale *N. benthamiana* expression and structurally determined by NMR to be 21*(S)*-acetoxyl-apo-melianone (**6**) (Fig. 4A**,** table S4, table S6 to S7, fig. S15). (**6**) is a protolimonoid previously purified from the Meliaceae species *Chisocheton paniculatus (32)* and is also detectable in *M. azedarach* tissues (fig. S16). L21AT likely stereoselectively acetylates the 21-(*S*) isomer; a possible role for this transformation is stabilization of the hemiacetal ring observed as an epimeric mixture in melianol (**1**) (*20*) and *apo*-melianol (**3**) (table S3). Overall, our results indicated that L21AT acetylates the C21 hydroxyl and SDR oxidizes the C3 hydroxyl to the ketone on early protolimonoid scaffolds.

***Citrus and Melia cytochrome P450s catalyze distinct limonoid A-ring modifications***

Further *Citrus* and *Melia* candidate screens (Fig. 2A, 2C) supports activity of two *Citrus* CYPs, *Cs*CYP716AC1 and *Cs*CYP88A37, that are each capable of oxidizing (**6**) directly to (**7**) and (**8**) or consecutively to (**9**) (Fig. 4A, fig. S17 to S19), and that one CYP from *Melia* (*Ma*CYP88A164, a homolog of *Cs*CYP88A37) is also capable of oxidizing (**6**) to (**8**) (Fig. 4A, fig. S20). Purification and NMR analysis of the downstream product (**9**) revealed it to be 1-hydroxy-luvungin A, which bears an A-ring lactone (table S8). Additional NMR product characterization suggests that *Cs*CYP716AC1 is responsible for A-ring lactone formation and *Cs*CYP88A37 is responsible for C1 hydroxylation (table S9). Although the exact order of oxidation steps to (**9**) appeared to be interchangeable for *Cs*CYP716AC1 and *Cs*CYP88A37, incomplete disappearance of (**6**) by *Cs*CYP88A37 suggested that oxidation by *Cs*CYP716AC1 takes precedence (fig. S19).

In the absence of *Cs*SDR, neither *Cs*CYP716AC1 nor *Cs*CYP88A37 result in an oxidized protolimonoid scaffold, suggesting the necessary involvement of the C-3 ketone for further processing (fig. S21). These results, in combination with NMR characterization, indicated that *Cs*CYP716AC1 is likely responsible for Baeyer-Villiger oxidation to the A-ring lactone structure signature of Rutaceae limonoids. Comparative transcriptomics in *M. azedarach* revealed the lack of an obvious *Cs*CYP716AC1 homolog. The closest Melia enzyme to *Cs*CYP716AC1 is truncated, not co-expressed with melianol biosynthetic genes, and only shares 63% protein identity (table S10). These results highlight a branch point between biosynthetic routes in the Rutaceae and Meliaceae families.

***Acetylations complete tailoring in both Citrus and Melia protolimonoid scaffolds and set the stage for furan ring biosynthesis***

Subsequent *Citrus* and *Melia* gene candidate screens (Fig. 2A, 2C) revealed further activity of BAHD acetyltransferases. *Cs*L1AT and its homolog *Ma*L1AT (named limonoid 1-*O*-acetyltransferase) appear to be active on (**9**) and (**8**), respectively (fig. S22 to S23). When *Cs*L1AT was co-expressed with the biosynthetic genes for (**9**), a new molecule (**11**) with mass corresponding to acetylation of (**9**) was observed. When *Cs*CYP88A37 was omitted, acetylation of (**7**) was not observed (fig. S24), suggesting that *Cs*L1AT acetylates the C-1 hydroxyl of (**9**) to yield (**11**). However, when *Cs*CYP716AC1 was omitted from the Citrus candidates or when MaL1AT was tested, the dehydration scaffold (**10**) accumulated (fig. S23 to S24). Large-scale transient plant expression, purification, and NMR analysis of the dehydration product showed that the structure (**10**) (table S11 to S12) contains a C-1/2 double bond and is an epimer of a previously reported molecule from *A. indica* (*33*). (**10**) also accumulates in *M. azedarach* extracts (fig. S16). Two more co-expressed *Citrus* and *Melia* acetyltransferase homologs, *Cs*L7AT and *Ma*L7AT, (named limonoid 7-*O*-acetyltransferase) were found to result in acetylated scaffolds (**12**) and (**13**); modification at the C-7 hydroxyl was confirmed by the

purification and NMR analysis of (**13**) and its degradation product (**13'**) (Fig. 2A, 2C, fig S25 to S26, table S13 to table S14).

Taken together, these data suggest that three acetyltransferases (L1AT, L7AT, and L21AT) act in the biosynthesis of the tri-acetylated 1,7,21-*O*-acetyl protolimonid (**13**) (Fig. 4A). However, we also observed the accumulation of two di-acetylated intermediates, (**11**) (1,21-*O*-acetyl) and (**11a**) (1,7-*O*-acetyl) when testing gene sets that lead to accumulation of (**13**) (fig. S27). This observation hints at the possibility of multiple sequences for enzymatic steps that comprise a metabolic network, at least in the context of pathway reconstitution in the heterologous host *N. benthamiana*.

***Downstream enzymes complete the biosynthesis to the furan-containing products azadirone (18) and kihadalactone A (19)***

With acetylation established, the key enzymes involved in the C4 scission implicated in furan ring formation still remained elusive. It was unclear which enzyme classes could catalyze these modifications. We screened gene candidates via combinatorial transient expression in *N. benthamiana* as previously described and ultimately identified three active candidate pairs (one from each species): the aldo-keto reductases (*Cs*AKR/*Ma*AKR), the CYP716ADs (*Cs*CYP716AD2/*Ma*CYP716AD4), and the 2-ODDs (named limonoid furan synthase, *Cs*LFS/*Ma*LFS) (Fig. 2A, 2C). Systematic testing of these gene sets resulted in the accumulation of the furan-containing molecules azadirone (**18**) and kihadalactone A (**19**), two limonoids present in the respective native species. When *Cs*AKR/*Ma*AKR was tested alone in our screens, we identified the appearance of a new peak with mass corresponding to reductive deacetylation of (**12**) or (**13**) (fig. S28 to S29). The product generated by expression of the *Melia* gene set in *N. benthamiana* was purified and characterized via NMR analysis to be the 21,23-diol (**14**) (Fig. 4A, table S15). Thus, the corresponding *Cs*AKR product (**15**) was proposed to share the same diol motif.

Transient expression of *Ma*CYP716AD4 or *Cs*CYP716AD2 with the biosynthetic genes for (**14**) or (**15**) resulted in two new pairs of peaks, each with C4 loss. Proposed structures indicate a $C_4H_6O$ fragment loss (**16a and 17a**) and a $C_4H_{10}O$ fragment loss (**16b and 17b**) from their respective precursors (Fig. 4A**,** fig. S30 to S31). It is unclear whether these observed masses correspond to the true products of CYP716ADs or whether these are further modified by endogenous *N. benthamiana* enzymes. CYP716AD products are proposed to contain C-21 hydroxyl and C-23 aldehyde functionalities (**16c and 17c**) which could also spontaneously form the five-membered hemiacetal ring (**16d and 17d**) (Fig. 4A, fig. S32). A new peak with a mass equivalent to (**16c or 16d**) is identifiable alongside (**16a** and **16b**) when transiently expressing *Ma*CYP716AD4 with the biosynthetic genes required for accumulation of (**14**) (fig. S31). We found that additional co-expression of LFS with the characterized genes that result in (**16**) and (**17**) yields accumulation of products (**18**) and (**19**) (fig. S33 to S34). Based on the predicted

chemical formula, MS fragmentation pattern, and NMR analysis (fig. S33, table S16), we proposed the product of *Cs*LFS to be kihadalactone A (**19**), a known furan-containing limonoid (*34*) previously identified in extracts from the Rutaceae plant *Phellodendron amurense*. We detected the presence of (**19**) in *P. amurense* seed samples (fig. S35), confirming prior reports of accumulation. Similarly, when *Ma*LFS was included in the co-expression, a new product with a mass equivalent to the furan-containing limonoid azadirone (**18**) was observed (fig. S34). The production of azadirone (**18**) in *N. benthamiana* was confirmed by comparison to an analytical standard (fig. S36, table S17) (isolated from *A. indica* leaf powder and analyzed by NMR). In addition, we detected azadirone in extracts from three Meliaceae species (fig. S36).

Taken together, we have discovered the 10- and 11-step biosynthetic transformations that enable a reconstitution of the biosynthesis of two known limonoids, azadirone (**18**) and kihadalactone A (**19**), as well as an enzyme catalyzing the formation of the alternative glabretal scaffold (*Cs*MOI1). Sequential introduction of these enzymes into *N. benthamiana* transient co-expression experiments demonstrate step-wise transformations leading to (**18**) and (**19**) (Fig. 4B). All of the enzymes involved in the biosynthesis of (**18**) and (**19**), except *Cs*CYP716AC1, are homologous pairs, and show a gradual decreasing trend in protein identity from 86% for the first enzyme pair *Cs*OSC1/*Ma*OSC1 to 66% for *Cs*LFS/*Ma*LFS. Despite the varied protein identities (Fig. 4B), these homologous enzymes from Melia or Citrus can be used to create functional hybrid pathways comprising a mix of species genes, supporting a promiscuous evolutionary ancestor for each of the limonoid biosynthetic enzymes (fig. S37).

## Discussion

A major challenge in elucidating pathways that involve many (e.g. >10) enzymatic steps is to determine whether the observed enzymatic transformations in a heterologous host are "on-pathway" and, if so, in what order they occur. It is important to note that while all enzymes described in Fig. 4 play a role in the production of final limonoid products, the sequence of enzymatic steps shown by the arrows is proposed based on the accumulation of observed metabolites after addition of each enzyme in the *N. benthamiana* heterologous expression system, and other sequences of steps are possible. For example, we've shown that *Cs*AKR likely doesn't accept hemi-acetal (**13**) directly as a substrate (fig. S38) despite our observation that it accumulates as a major metabolite when all upstream enzymes are expressed. Although one expects a pathway without *Cs*L21AT to still be functional as the C-21 acetal product (**11a**) appears to undergo reduction by *Cs*AKR to yield (**15**), attempts to drop out *Cs*L21AT led to significantly reduced yield of (**19**) (fig. S39), suggesting that *Cs*L21AT might have other unexpected roles in the pathway. In addition, reconstitution of several partial pathways indicates that some pathway enzymes can accept multiple related substrates. For example, each step after *apo*-melianol can diverge into multiple pathways, likely due to the promiscuity of these enzymes. Taken together, these data indicate that enzymes in limonoid biosynthesis might collectively function as a metabolic network (fig. S40). Further study of each individual enzyme

*in vitro* with purified substrate will be required to quantify substrate preference. This metabolic network observed in *N. benthamiana* suggests one possible strategy for how Rutaceae species access such a diverse range of limonoids; we anticipate that additional enzymes will further expand the network, e.g. for the oxidative cleavage of ring C, ultimately resulting in the most extensively rearranged and modified limonoid scaffolds isolated to date, e.g. azadirachtin (Fig. 1).

Among the 12 chemical transformations catalyzed by the 22 enzymes characterized in this study, several are not previously known in plant specialized metabolism. For example, MOI1 and MOI2, which appear to have evolved from sterol isomerases, are capable of catalyzing two different scaffold rearrangements despite their conserved active site residues (Fig. S41). The co-localization of the limonoid biosynthetic gene *MaMOI2* with two other non-limonoid SI genes in the *M. azedarach* genome is consistent with the origin of *MaMOI2* by tandem duplication and neofunctionalization (fig. S42); this genomic arrangement is conserved in *Citrus* on chromosome 5 as well. Furthermore, recent findings demonstrate a similar role of these enzymes in quassinoid biosynthesis (*35*). Other noteworthy enzymatic reactions in the limonoid pathway include C-4 scission and furan ring installation that generate an important pharmacophore of the limonoids. Although furan-forming enzymes have been reported from other plants (*36*, *37*), (*38*), the AKR, CYP716AD and 2-ODD module described here represents a new mechanism of furan formation via the oxidative cleavage of a C-4 moiety. Along with the sterol isomerases (MOIs), the AKR and 2-ODDs add to the growing pool of enzyme families (*39*, *40*) associated with primary sterol metabolism that appear to have been recruited to plant secondary triterpene biosynthesis, likely due to the structural similarities between sterols and tetracyclic triterpenes.

Limonoids are only one of many families of triterpenes from plants with complex scaffold modifications. Other examples include the *Schisandra* nortriterpenes (*41*), quinonoids (*42*), quassinoids (*43*), and dichapetalins (*42*); each represent a large collection of structurally diverse terpenes that contain several members with potent demonstrated biological activity but no biosynthetic route. Despite the value of these complex plant triterpenes, individual molecular species are typically only available through multi-step chemical synthesis routes or isolation from producing plants, limiting drug development (*15*) and agricultural utility (*9*). Many are only easily accessible in unpurified extract form that contains multiple chemical constituents; for example, azadirachtin, one of the most potent limonoids, can only be obtained commercially as a component of neem oil. Our results demonstrate that pathways to triterpenes with complex scaffold modifications can be reconstituted in a plant host, and the gene sets we describe enable rapid production and isolation of naturally-occurring limonoids. We anticipate that bioproduction of limonoids will serve as an attractive method to generate clinical candidates for evaluation, and that stable engineering of the limonoid pathway could be a viable strategy for sustainable crop protection.

## Figures

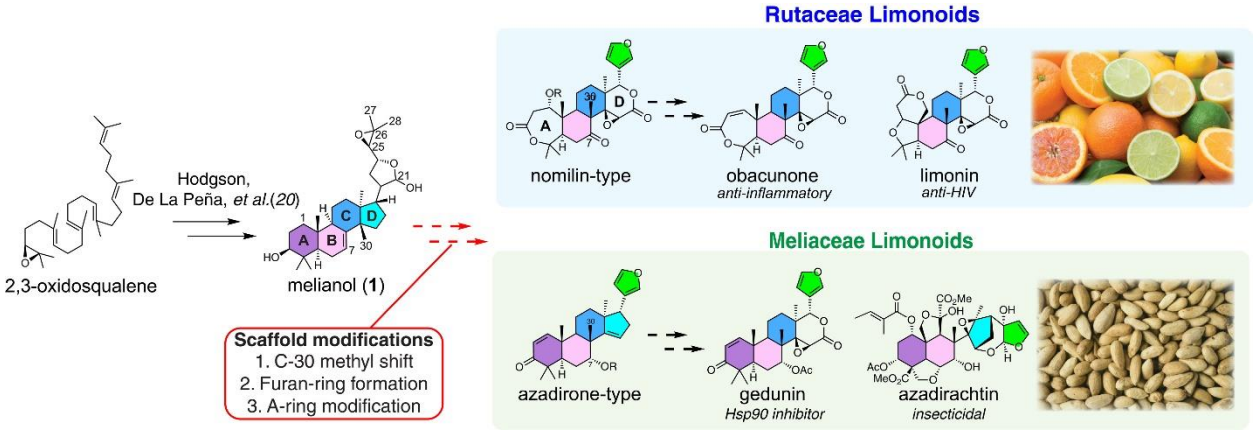



**Fig. 1. Structures of Rutaceae and Meliaceae limonoids and proposed biosynthetic pathway.** We previously characterized three conserved enzymes from both *Citrus* and *Melia* species that catalyze the formation of the protolimonoid melianol (**1**) from 2,3-oxidosqualene (*20*). Additionally, conserved scaffold modifications like C-30 methyl shift, furan-ring formation, and A-ring modification are proposed to convert protolimonoids to true limonoids. Beyond this, Rutaceae limonoids differ from Meliaceae limonoids in two key structural features: *seco*-A,D ring and C-7 modification, which are proposed to be the result of Rutaceae and Meliaceae specific modifications. Exceptions to this rule could potentially arise from late-stage species-specific tailoring (fig. S43). Rutaceae limonoids are derived from nomilin-type intermediates while Meliaceae limonoids are proposed to originate from azadirone-type intermediates. While the exact point of pathway divergence is unknown, comparative analysis of the various protolimonoid structures suggested that C-1, C-7, C-21 hydroxylation and/or acetoxylation are part of the conserved tailoring process. Obacunone and limonin are commonly found in various *Citrus* species (adapted photo by IgorDutina on iStock with standard license) and are responsible for the bitterness of their seeds. Azadirachtin (the most renowned Meliaceae limonoid) accumulates at high levels in the seeds of neem tree (photo by JIC photography), which are the source of commercial neem biopesticides.

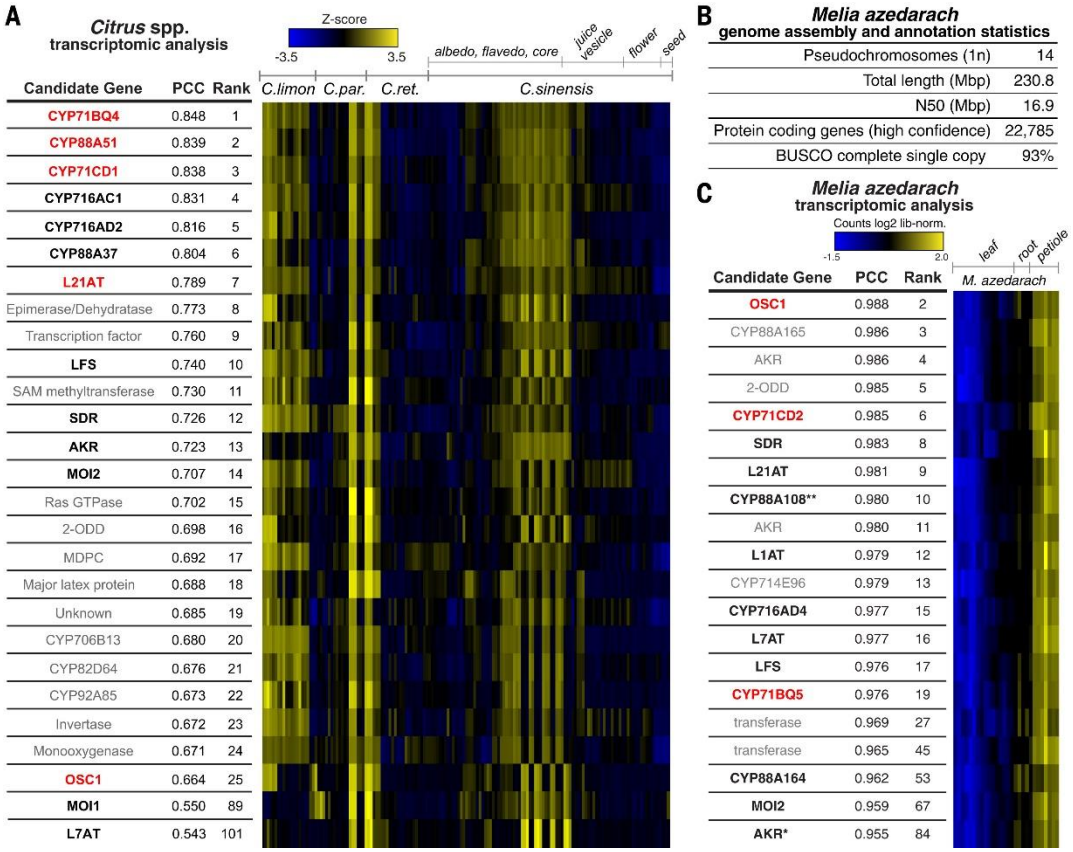


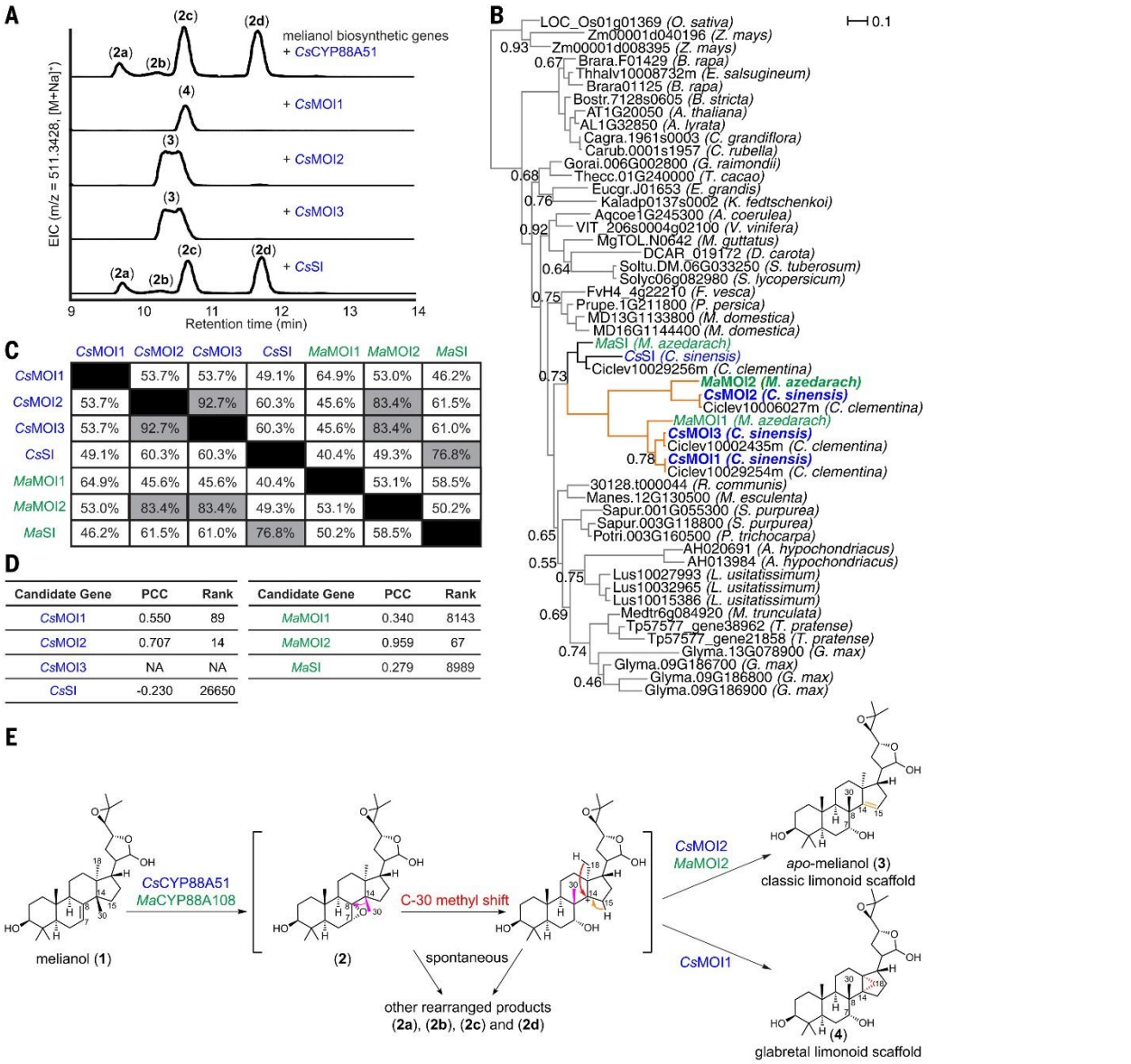

**Fig. 2. Genomic and transcriptomic analysis of *Citrus* and *Melia* resources.**

(**A**) Co-expression analysis of *C. sinensis* publicly available microarray expression data from NICCE (*22*) using *CsOSC1*, *CsCYP71CD1*, *CsCYP71BQ4*, *CsCYP88A51* and *CsL21AT* as bait genes. Linear regression analysis was used to rank the top 25 genes based on Pearson's correlation coefficient (PCC) to the bait genes of interest. Heat map displays Z-score calculated from $\log_2$ normalized expression across the fruit dataset. The reported PCC value corresponds to the average value calculated using each bait gene. Genes in red indicate bait genes used in analysis and genes in black are functional limonoid biosynthetic genes (table S18). Functional candidates outside of the top 25 genes are also included. For identification of individual bait genes used in this analysis see fig. S2. Enzymes have been abbreviated as follows: MOI = melianol oxide isomerase; CYP = cytochrome P450; L21AT = limonoid C-21-*O*-acetyltransferase; SDR = short-chain dehydrogenase; L1AT = limonoid C-1-*O*-acetyltransferase; L7AT = limonoid C-7-*O*-acetyltransferase; AKR = aldo-keto reductase; LFS = limonoid furan synthase; OSC = oxidosqualene cyclase.

(**B**) Summary of *Melia azedarach* pseudo-chromosome genome assembly and annotation statistics (fig. S3 to S4, table S1 to S2).

(**C**) Expression pattern of *M. azedarach* limonoid candidate genes selected based on PCC to melianol biosynthetic genes (*MaOSC1*, *MaCYP71CD2* and *MaCYP71BQ5* (*20*), shown in red) and biosynthetic annotation. Heatmap (constructed using Heatmap3 V1.1.1 (*44*), with scaling by row (gene)) includes genes that are ranked within the top 87 for co-expression and are annotated

with one of six interpro domains of biosynthetic interest (IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR020471 (Aldo/keto reductase), IPR002347 (Short-chain dehydrogenase/reductase SDR), IPR001128 (Cytochrome P450), IPR003480 (Transferase) and IPR007905 (Emopamil-binding protein)). Asterisks indicate the following: (*) full-length gene identified in transcriptomic rather than genomic data via sequence similarity to *Cs*AKR ((table S10, table S19), (**) gene previously identified as homolog of limonoid co-expressed gene from *A. indica (20)*). Genes shown in black are newly identified functional limonoid biosynthetic genes (this study) (table S10).




| | CsMOI1 | CsMOI2 | CsMOI3 | CsSI | MaMOI1 | MaMOI2 | MaSI |
|---|---|---|---|---|---|---|---|
| CsMOI1 | | 53.7% | 53.7% | 49.1% | 64.9% | 53.0% | 46.2% |
| CsMOI2 | 53.7% | | 92.7% | 60.3% | 45.6% | 83.4% | 61.5% |
| CsMOI3 | 53.7% | 92.7% | | 60.3% | 45.6% | 83.4% | 61.0% |
| CsSI | 49.1% | 60.3% | 60.3% | | 40.4% | 49.3% | 76.8% |
| MaMOI1 | 64.9% | 45.6% | 45.6% | 40.4% | | 53.1% | 58.5% |
| MaMOI2 | 53.0% | 83.4% | 83.4% | 49.3% | 53.1% | | 50.2% |
| MaSI | 46.2% | 61.5% | 61.0% | 76.8% | 50.2% | 58.5% | |

| Candidate Gene | PCC | Rank |
|---|---|---|
| CsMOI1 | 0.550 | 89 |
| CsMOI2 | 0.707 | 14 |
| CsMOI3 | NA | NA |
| CsSI | -0.230 | 26650 |

| Candidate Gene | PCC | Rank |
|---|---|---|
| MaMOI1 | 0.340 | 8143 |
| MaMOI2 | 0.959 | 67 |
| MaSI | 0.279 | 8989 |

**Fig. 3. Characterization of melianol oxide isomerases (MOIs).**

(**A**) Characterization of products generated via overexpression of MOIs and SI using transient gene expression in *N. benthamiana*. Liquid chromatography–mass spectrometry (LC-MS) extracted ion chromatograms (EICs) resulting from overexpression of *At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, and *Cs*MOIs and *Cs*SI in *N. benthamiana*. Representative EICs are shown (n=3).

(**B**) Phylogenetic tree (Bayesian) of sterol isomerase (SI) genes from high-quality plant genomes. SI sequences from 33 plant species were identified and downloaded from Phytozome via pFAM assignments (PF05241). Branch supports are provided (excluding those >0.95) and monocot SIs have been used as an outgroup. Enzymes that have melianol oxide isomerase activity when tested by *Agrobacterium*-mediated expression in *N. benthamiana* with melianol (**1**) biosynthetic genes and *Cs*CYP88A51 or *Ma*CYP88A108, have been renamed MOI, e.g. *Cs*MOI1-3 and *Ma*MOI2. Characterized MOIs from *C. sinensis* and *M. azedarach* selected for further analysis are bolded and their respective tree branches are indicated in orange. Genes from *Citrus* are shown in blue and those from *Melia* are shown in green.

(**C**) Percentage protein identity of MOIs and SIs from *C. sinensis* and *M. azedarach*, those with sequence similarity greater than 75% are highlighted in gray.

(**D**) Co-expression of MOIs and SIs from *C. sinensis* and *M. azedarach* displaying rank and PCC as outlined in Fig. 2A, 2C.

(**E**) Proposed mechanism of *Cs*CYP88A51/*Ma*CYP88A108, *Cs*MOI2/*Ma*MOI2 and *Cs*MOI1. *Cs*CYP88A51/*Ma*CYP88A108 first oxidizes the C7,C8 position of melianol (**1**) to yield an unstable epoxide intermediate (**2**), which can undergo spontaneous C-30 methyl shift from C-14 to C-8 (highlighted in red). Either (**2**) or the methyl shifted product spontaneously form a series of oxidized products (**2a - 2d**). In the presence of MOIs, the rearrangement of (**2**) is guided to form either (**3**) or (**4**) and no (**2a**), (**2b**), (**2c**), and (**2d**) are observed. Structures of (**2a**), (**2b**), (**2c**) and (**2d**) are not determined but their MS fragmentation patterns suggest they are isomeric molecules resulting from a single oxidation of melianol (**1**), which doesn't exclude the possibility them of being (**2**), (**3**), or (**4**) (as shown for *Ailanthus altissima* CYP71BQ17 (*35*)).

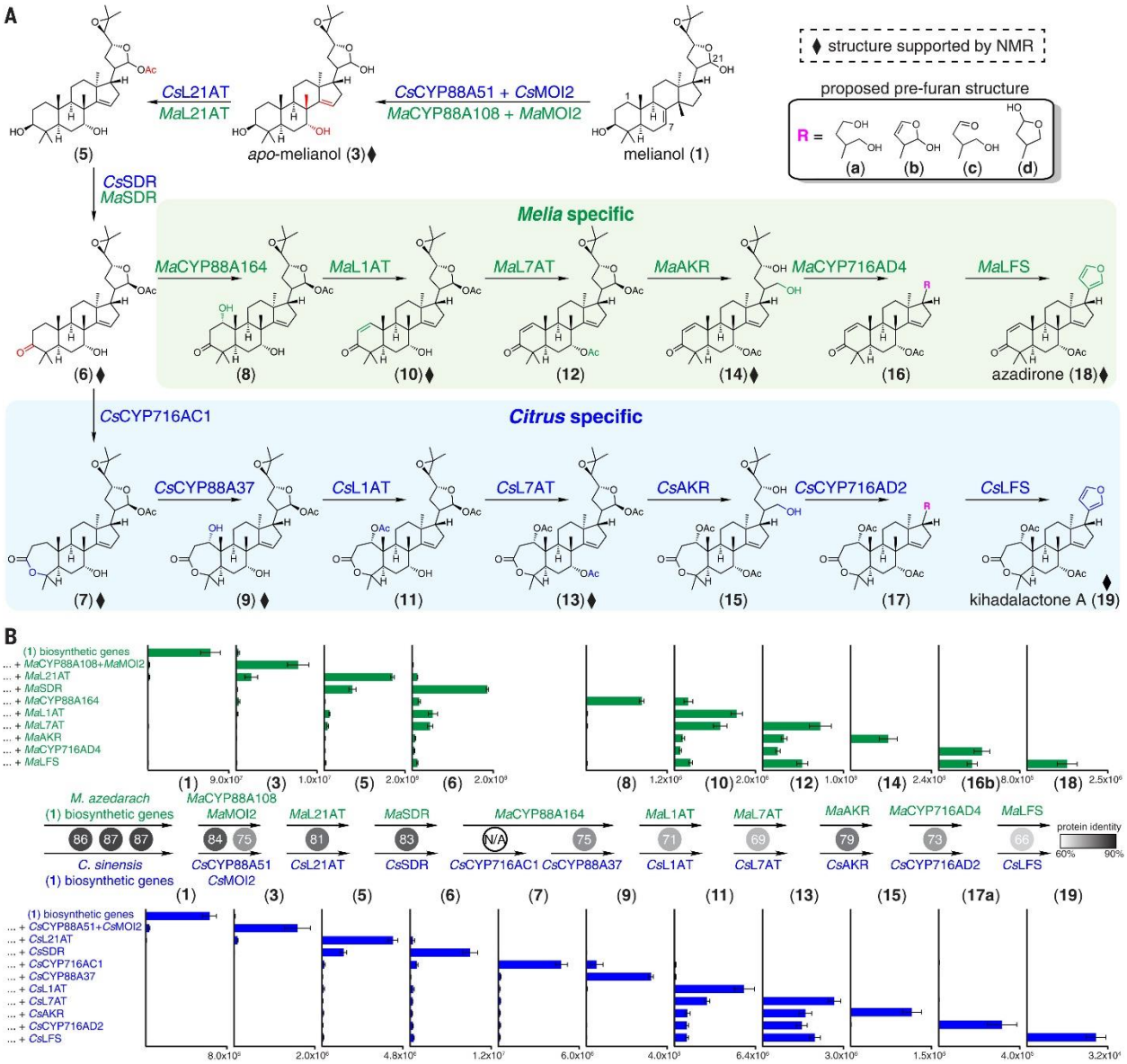



**Fig. 4. Complete biosynthetic pathway to azadirone (18) and kihadalactone A (19).** (**A**) Gene sets that lead to the production of azadirone (**18**) and kihadalactone A (**19**) in *N. benthamiana* leaves. Genes from *Citrus* are shown in blue and those from *Melia* are shown in green. The arrow reflects accumulation of the metabolites after addition of the associated enzyme as shown in Panel B rather than true enzymatic substrate-product relationship. In addition, limonoids biosynthesis likely proceeds as a network; other possible reaction sequences are shown in fig S40. Diamonds represent intermediates whose structures were supported either by NMR analysis of the purified product or comparison with an authentic standard (**18**). (**3**), (**6**), (**9**), (**10**), (**13**) and (**14**) were purified from *N. benthamiana* leaf extracts expressing the respective biosynthetic gene sets and analyzed by NMR; the structures of (**7**) and (**19**) are supported by partial NMR. Additionally, a side product (**20**), formed in experiments with all pathway enzymes up to and including *Ma*CYP716AD4 but without *Ma*L7AT (fig. S44) was purified and confirmed by NMR (table S20); similar activity was observed for *Cs*CYP716AD2 (fig. S45, supplementary

text). Enzymes have been abbreviated as follows: MOI = melianol oxide isomerase; CYP = cytochrome P450; L21AT = limonoid C-21-*O*-acetyltransferase; SDR = short-chain dehydrogenase; L1AT = limonoid C-1-*O*-acetyltransferase; L7AT = limonoid C-7-*O*-acetyltransferase; AKR = aldo-keto reductase; LFS = limonoid furan synthase.
(**B**) Integrated peak area of extracted ion chromatogram (EIC) for each pathway intermediates produced in *N. benthamiana* after sequential co-expression of individual enzymes. Values and error bars represent the mean and the standard error of the mean; n=6 biological replicates. Percentage identity between homologous proteins are shown in numbers in the circles and colored in gray scale. (**1**) biosynthetic genes comprise *Ma*OSC1/*Cs*OSC1, *Ma*CYP71CD2/*Cs*CYP71CD1, and *Ma*CYP71BQ5/*Cs*CYP71BQ4. *Cs*CYP88A37 is a homolog to *Ma*CYP88A164 while *Cs*CYP716AC1 has no *Melia* homolog.

**Acknowledgments**

We would like to acknowledge Jasmine Staples for performing the extractions of Meliaceae material sourced from Kew Gardens and Stephen Lynch (head of NMR facility, Stanford University, Department of Chemistry) for helpful discussion on NMR structural elucidation.

**Funding:** Work undertaken in E.S.S's laboratory was supported by the National Institutes of Health (NIH) grants (NIH U01 GM110699 and NIH R01 GM121527). The work undertaken in A.O's laboratory was partly funded through a Biotechnology and Biological Sciences Research Council Industrial Partnership (BBSRC) Award with Syngenta (BB/T015063/1) and A.O's lab is also supported by the John Innes Foundation and the BBSRC Institute Strategic Programme Grant 'Molecules from Nature - Products and Pathways' (BBS/E/J/00PR9790).

**Author contributions:** R.D.L.P., H.H., A.O. and E.S.S. conceived of the project and, with assistance from J.C.T.L., designed the research. H.H. generated and analyzed *M. azedarach* genome and transcriptome data. R.D.L.P. analyzed Citrus gene expression data and selected candidate genes from Citrus and, with J.C.T.L., expressed and characterized biosynthetic genes and metabolic products. L.E.J. assisted with isolation of Citrus intermediates. H.H. analyzed the Melia sequence resources and selected, expressed and characterized Melia biosynthetic genes and metabolic products. J.C.T.L and M.S. performed NMR analysis on the Citrus and Melia

products, respectively. J.L.M advised on *M. azedarach* genomics. A.H. performed chromatin cross-linking and DNA extraction on *M. azedarach* tissues for Hi-C analysis by Phase Genomics. A.C.M. performed karyotyping on *M. azedarach* roots. C.O. combined the pseudo-chromosome level genome assembly with the *M. azedarach* annotation and constructed the phylogenetic tree. R.D.L.P, J.C.T.L., H.H., A.O. and E.S.S. analyzed the data and wrote the manuscript.

**Competing interests:** The authors declare they have no competing interests.

**Data and materials availability:** All *Citrus* genes in this study have been deposited on (XXXX) with the accession numbers XXXXX. The *Melia azedarach* genome has been deposited on NCBI (PRJNA906622), along with the accompanying RNA-seq data (PRJNA906055). Coding sequences for the functional *M. azedarach* genes described in this study have also been deposited on Genbank with the accession numbers OP947595-OP947604.

## Supplementary Materials

Materials and Methods

Figs. S1 to S45

Tables S1 to S24

Data S1 (Full NMR spectral data for isolated compounds)

References 45-82

## References


1. Y. Y. Zhang, H. Xu, Recent progress in the chemistry and biology of limonoids. *RSC Adv.* **7**, 35191–35220 (2017).

2. Q.-G. Tan, X.-D. Luo, Meliaceous Limonoids: Chemistry and Biological Activities. *Chem. Rev.* **111**, 7437–7522 (2011).

3. A. Roy, S. Saraf, Limonoids: overview of significant bioactive triterpenes distributed in plants kingdom. *Biol. Pharm. Bull.* **29**, 191–201 (2006).

4. F. Mulani, S. Nandikol, H. Thulasiram, Chemistry and Biology of Novel Meliaceae Limonoids, , doi:10.26434/chemrxiv-2022-2bpb9.

5. J. Luo, Y. Sun, Q. Li, L. Kong, Research progress of meliaceous limonoids from 2011 to 2021. *Nat. Prod. Rep.* **39**, 1325–1365 (2022).

6. S. Yamashita, A. Naruko, Y. Nakazawa, L. Zhao, Y. Hayashi, M. Hirama, Total Synthesis of Limonin. *Angew. Chem. Int. Ed Engl.* **54**, 8538–8541 (2015).

7. G. E. Veitch, E. Beckmann, B. J. Burke, A. Boyer, S. L. Maslen, S. V. Ley, Synthesis of azadirachtin: a long but successful journey. *Angew. Chem. Int. Ed Engl.* **46**, 7629–7632 (2007).

8. J. Li, F. Chen, H. Renata, Thirteen-Step Chemoenzymatic Synthesis of Gedunin, , doi:10.26434/chemrxiv-2022-tqvbw.

9. E. D. Morgan, E. David Morgan, Azadirachtin, a scientific gold mine. *Bioorganic & Medicinal Chemistry*. **17** (2009), pp. 4096–4105.

10. M. Puri, S. S. Marwaha, R. M. Kothari, J. F. Kennedy, Biochemical Basis of Bitterness in Citrus Fruit Juices and Biotech Approaches for Debittering. *Critical Reviews in Biotechnology*. **16**

(1996), pp. 145–155.

11. L. Battinelli, F. Mengoni, M. Lichtner, G. Mazzanti, A. Saija, C. M. Mastroianni, V. Vullo, Effect of limonin and nomilin on HIV-1 replication on infected human mononuclear cells. *Planta Med.* **69**, 910–913 (2003).

12. X. Luo, Z. Yu, B. Yue, J. Ren, J. Zhang, S. Mani, Z. Wang, W. Dou, Obacunone reduces inflammatory signalling and tumour occurrence in mice with chronic inflammation-induced colorectal cancer. *Pharm. Biol.* **58**, 886–897 (2020).

13. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. **313**, 1929–1935 (2006).

14. J. N. Spradlin, X. Hu, C. C. Ward, S. M. Brittain, M. D. Jones, L. Ou, M. To, A. Proudfoot, E. Ornelas, M. Woldegiorgis, J. A. Olzmann, D. E. Bussiere, J. R. Thomas, J. A. Tallarico, J. M. McKenna, M. Schirle, T. J. Maimone, D. K. Nomura, Harnessing the anti-cancer natural product nimbolide for targeted protein degradation. *Nat. Chem. Biol.* **15**, 747–755 (2019).

15. B. Tong, J. N. Spradlin, L. F. T. Novaes, E. Zhang, X. Hu, M. Moeller, S. M. Brittain, L. M. McGregor, J. M. McKenna, J. A. Tallarico, M. Schirle, T. J. Maimone, D. K. Nomura, A Nimbolide-Based Kinase Degrader Preferentially Degrades Oncogenic BCR-ABL. *ACS Chem. Biol.* **15**, 1788–1794 (2020).

16. V. Domingo, J. F. Arteaga, J. F. Quílez del Moral, A. F. Barrero, Unusually cyclized triterpenes: occurrence, biosynthesis and chemical synthesis. *Nat. Prod. Rep.* **26**, 115–134 (2009).

17. W. J. Baas, Naturally occurring seco-ring-A-triterpenoids and their possible biological significance. *Phytochemistry*. **24**, 1875–1889 (1985).

18. P. Ou, S. Hasegawa, Z. Herman, C. H. Fong, Limonoid biosynthesis in the stem of Citrus limon. *Phytochemistry*. **27** (1988), pp. 115–118.

19. S. Hasegawa, Biochemistry of Limonoids in *Citrus*. *ACS Symposium Series* (2000), pp. 9–30.

20. H. Hodgson, R. De La Peña, M. J. Stephenson, R. Thimmappa, J. L. Vincent, E. S. Sattely, A. Osbourn, Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *Proceedings of the National Academy of Sciences*. **116**, 17096–17104 (2019).

21. Y. Li, A. Leveau, Q. Zhao, Q. Feng, H. Lu, J. Miao, Z. Xue, A. C. Martin, E. Wegel, J. Wang, A. Orme, M.-D. Rey, M. Karafiátová, J. Vrána, B. Steuernagel, R. Joynson, C. Owen, J. Reed, T. Louveau, M. J. Stephenson, L. Zhang, X. Huang, T. Huang, D. Fan, C. Zhou, Q. Tian, W. Li, Y. Lu, J. Chen, Y. Zhao, Y. Lu, C. Zhu, Z. Liu, G. Polturak, R. Casson, L. Hill, G. Moore, R. Melton, N. Hall, B. B. H. Wulff, J. Doležel, T. Langdon, B. Han, A. Osbourn, Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* **12**, 1–13 (2021).

22. D. C. J. Wong, C. Sweetman, C. M. Ford, Annotation of gene function in citrus using

gene expression information and co-expression networks. *BMC Plant Biology*. **14** (2014), , doi:10.1186/1471-2229-14-186.

23. D. Ohri, A. Bhargava, A. Chatterjee, Nuclear DNA Amounts in 112 Species of Tropical Hardwoods-New Estimates. *Plant Biol.* **6**, 555–561 (2004).

24. 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**, 3210–3212 (2015).

25. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).

26. N. Braverman, P. Lin, F. F. Moebius, C. Obie, A. Moser, H. Glossmann, W. R. Wilcox, D. L. Rimoin, M. Smith, L. Kratz, R. I. Kelley, D. Valle, Mutations in the gene encoding 3β-hydroxysteroid-Δ8,Δ7- isomerase cause X-linked dominant Conradi-Hünermann syndrome. *Nature Genetics*. **22** (1999), pp. 291–294.

27. M. Souter, J. Topping, M. Pullen, J. Friml, K. Palme, R. Hackett, D. Grierson, K. Lindsey, hydra Mutants of Arabidopsis are defective in sterol profiles and auxin and ethylene signaling. *Plant Cell*. **14**, 1017–1031 (2002).

28. S. Silve, P. H. Dupuy, C. Labit-Lebouteiller, M. Kaghad, P. Chalon, A. Rahier, M. Taton, J. Lupker, D. Shire, G. Loison, Emopamil-binding protein, a mammalian protein that binds a series of structurally diverse neuroprotective agents, exhibits delta8-delta7 sterol isomerase activity in yeast. *J. Biol. Chem.* **271**, 22434–22440 (1996).

29. W. Herz, H. Grisebach, G. W. Kirby, *Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products* (Springer Vienna).

30. G. Ferguson, P. Alistair Gunn, W. C. Marsh, R. McCrindle, R. Restivo, J. D. Connolly, J. W. B. Fulke, M. S. Henderson, Triterpenoids from Guarea glabra(meliaceae): a new skeletal class identified by chemical, spectroscopic, and X-ray evidence. *Journal of the Chemical Society, Chemical Communications* (1973), p. 159.

31. A.-R. Choi, I.-K. Lee, E.-E. Woo, J.-W. Kwon, B.-S. Yun, H.-R. Park, New Glabretal Triterpenes from the Immature Fruits of Poncirus trifoliata and Their Selective Cytotoxicity. *Chem. Pharm. Bull.* . **63**, 1065–1069 (2015).

32. J. Connolly, C. Labbé, D. S. Rycroft, D. A. H. Taylor, Tetranortriterpenoids and related compounds. Part 22. New apotirucailol derivatives and tetranortriterpenoids from the wood and seeds of chisocheton paniculatus (meliaceae). *J. Chem. Soc. Perkin 1*, 2959–2964 (1979).

33. G. Chianese, S. R. Yerbanga, L. Lucantoni, A. Habluetzel, N. Basilico, D. Taramelli, E. Fattorusso, O. Taglialatela-Scafati, Antiplasmodial Triterpenoids from the Fruits of Neem, Azadirachta indica. *J. Nat. Prod.* **73**, 1448–1452 (2010).

34. K. Kishi, K. Yoshikawa, S. Arihara, Limonoids and protolimonoids from the fruits of Phellodendron amurense. *Phytochemistry*. **31** (1992), pp. 1335–1338.

35. L. Chuang, S. Liu, J. Franke, Post-Cyclase Skeletal Rearrangements in Plant Triterpenoid Biosynthesis by a Pair of Branchpoint Isomerases. *bioRxiv* (2022), p. 2022.09.23.508984.

36. A. Muchlinski, M. Jia, K. Tiedge, J. S. Fell, K. A. Pelot, L. Chew, D. Davisson, Y. Chen, J. Siegel, J. T. Lovell, P. Zerbe, Cytochrome P450-catalyzed biosynthesis of furanoditerpenoids in the bioenergy crop switchgrass ( *Panicum virgatum* L.). *The Plant Journal*. **108** (2021), pp. 1053–1068.

37. C. M. Bertea, M. Schalk, F. Karp, M. Maffei, R. Croteau, Demonstration that menthofuran synthase of mint (Mentha) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Arch. Biochem. Biophys.* **390**, 279–286 (2001).

38. J.-J. Song, X. Fang, C.-Y. Li, Y. Jiang, J.-X. Li, S. Wu, J. Guo, Y. Liu, H. Fan, Y.-B. Huang, Y.-K. Wei, Y. Kong, Q. Zhao, J.-J. Xu, Y.-H. Hu, X.-Y. Chen, L. Yang, A 2-oxoglutarate-dependent dioxygenase converts dihydrofuran to furan in Salvia diterpenoids. *Plant Physiol.* **188**, 1496–1506 (2021).

39. G. Polturak, M. Dippe, M. J. Stephenson, R. Chandra Misra, C. Owen, R. H. Ramirez-Gonzalez, J. F. Haidoulis, H.-J. Schoonbeek, L. Chartrain, P. Borrill, D. R. Nelson, J. K. M. Brown, P. Nicholson, C. Uauy, A. Osbourn, Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2123299119 (2022).

40. X. Qi, S. Bakht, B. Qin, M. Leggett, A. Hemmings, F. Mellon, J. Eagles, D. Werck-Reichhart, H. Schaller, A. Lesot, R. Melton, A. Osbourn, A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proceedings of the National Academy of Sciences*. **103**, 18848–18853 (2006).

41. Y.-M. Shi, W.-L. Xiao, J.-X. Pu, H.-D. Sun, Triterpenoids from the Schisandraceae family: an update. *Nat. Prod. Rep.* **32**, 367–410 (2015).

42. R.-Y. Kuo, K. Qian, S. L. Morris-Natschke, K.-H. Lee, Plant-derived triterpenoids and analogues as antitumor and anti-HIV agents. *Nat. Prod. Rep.* **26**, 1321–1344 (2009).

43. Z. Guo, S. Vangapandu, R. W. Sindelar, L. A. Walker, R. D. Sindelar, Biologically active quassinoids and their chemistry: potential leads for drug design. *Curr. Med. Chem.* **12**, 173–190 (2005).

44. S. Zhao, Y. Guo, Q. Sheng, Y. Shyr, Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinformatics*. **15**, 1–2 (2014).

45. M. Giolai, P. Paajanen, W. Verweij, L. Percival-Alwyn, D. Baker, K. Witek, F. Jupe, G. Bryan, I. Hein, J. D. G. Jones, Targeted capture and sequencing of gene-sized DNA molecules. *Biotechniques*. **61**, 315–322 (2016).

46. J.-M. Belton, R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan, J. Dekker, Hi–C: a comprehensive technique to capture the conformation of genomes. *Methods*. **58**, 268–276 (2012).

47. M.-D. Rey, G. Moore, A. C. Martín, Identification and comparison of individual chromosomes of three accessions of Hordeum chilense, Hordeum vulgare, and Triticum aestivum by FISH. *Genome*. **61**, 387–396 (2018).

48. D. J. MacKenzie, M. A. McLean, S. Mukerji, M. Green, Improved RNA extraction from woody plants for the detection of viral pathogens by reverse transcription-polymerase chain reaction. *Plant Dis.* **81**, 222–226 (1997).

49. L. Venturini, S. Caim, G. G. Kaithakottil, D. L. Mapleson, D. Swarbreck, Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience*. **7**, giy093 (2018).

50. D. Mapleson, L. Venturini, G. Kaithakottil, D. Swarbreck, Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *Gigascience*. **7**, giy131 (2018).

51. E. S. A. Sollars, A. L. Harper, L. J. Kelly, C. M. Sambles, R. H. Ramirez-Gonzalez, D. Swarbreck, G. Kaithakottil, E. D. Cooper, C. Uauy, L. Havlickova, Genome sequence and genetic diversity of European ash trees. *Nature*. **541**, 212–216 (2017).

52. B. J. Clavijo, L. Venturini, C. Schudoma, G. G. Accinelli, G. Kaithakottil, J. Wright, P. Borrill, G. Kettleborough, D. Heavens, H. Chapman, An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* **27**, 885–896 (2017).

53. A. Hallab, thesis, Universitäts-und Landesbibliothek Bonn (2015).

54. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).

55. T. Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, E. Huala, The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*. **53**, 474–485 (2015).

56. UniProt Consortium, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

57. B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).

58. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**, 1236–1240 (2014).

59. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

60. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

61. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108–e108 (2013).

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

63. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 1–9 (2010).

64. C. Evans, J. Hardin, D. M. Stoebel, Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* **19**, 776–792 (2018).

65. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. D. Zeng, Z. H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–U130 (2011).

66. 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. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

67. F. Sainsbury, E. C. Thuenemann, G. P. Lomonossoff, pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).

68. R. D. L. Peña, R. De La Peña, E. S. Sattely, Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nature Chemical Biology*. **17** (2021), pp. 205–212.

69. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell, A. Osbourn, A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**, 185–193 (2017).

70. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

71. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. **17**, 754–755 (2001).

72. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell, A. Osbourn, A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**, 185–193 (2017).

73. M. J. Stephenson, J. Reed, B. Brouwer, A. Osbourn, Transient expression in nicotiana benthamiana leaves for triterpene production at a preparative scale. *J. Vis. Exp.* (2018).

74. S. Haldar, P. B. Phapale, S. P. Kolet, H. V. Thulasiram, Expedient preparative isolation, quantification and characterization of limonoids from Neem fruits. *Anal. Methods*. **5**, 5386–5391 (2013).

75. D. M. Goodstein, S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D. S. Rokhsar, Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–86 (2012).

76. R. Tautenhahn, G. J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).

77. J. Connolly, W. R. Phillips, D. A. Mulholland, D. A. H. Taylor, Spicatin, a protolimonoid from Entandrophragma spicatum. *Phytochemistry*. **20**, 2596–2597 (1981).

78. C. Arenas, L. Rodriguez-Hahn, Limonoids from Trichilia havanensis. *Phytochemistry*. **29**, 2953–2956 (1990).

79. T. Long, A. Hassan, B. M. Thompson, J. G. McDonald, J. Wang, X. Li, Structural basis for human sterol isomerase in cholesterol biosynthesis and multidrug recognition. *Nat. Commun.* **10**, 2452 (2019).

80. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. **29**, 1072–1075 (2013).

81. B. Rodríguez, Complete assignments of the1H and13C NMR spectra of 15 limonoids. *Magn. Reson. Chem.* **41**, 206–212 (2003).

82. M. G. Bausher, N. D. Singh, S.-B. Lee, R. K. Jansen, H. Daniell, The complete chloroplast genome sequence of Citrus sinensis (L.) Osbeck var “Ridge Pineapple”: organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol.* **6**, 21 (2006).

Science
AAAS

Supplementary Materials for

# Complex scaffold remodeling in plant triterpene biosynthesis

Ricardo De La Peña[1]†, Hannah Hodgson[2]†, Jack Chun-Ting Liu[3]†, Michael J. Stephenson[4], Azahara C. Martin[5], Charlotte Owen[2], Alex Harkess[6], Jim Leebens-Mack[7], Luis E. Jimenez[1], Anne Osbourn[2]* and Elizabeth S. Sattely[1,8]*

[1]Department of Chemical Engineering, Stanford University; Stanford, CA 94305, US.

[2]Department of Biochemistry and Metabolism, John Innes Centre; Norwich Research Park, Norwich NR4 7UH, UK.

[3]Department of Chemistry, Stanford University; Stanford, CA 94305, US.

[4]School of Chemistry, University of East Anglia; Norwich Research Park, Norwich NR4 7TJ, UK.

[5]Department of Crop Genetics, John Innes Centre; Norwich Research Park, Norwich NR4 7UH, UK.

[6]HudsonAlpha Institute for Biotechnology; Huntsville, AL 35806, US.

[7]Department of Plant Biology, 4505 Miller Plant Sciences, University of Georgia; Athens, GA 30602, US.

[8]Howard Hughes Medical Institute, Stanford University; Stanford, CA 94305, US.

† These authors contributed equally to this work

* Corresponding author. Email: Anne Osbourn anne.osbourn@jic.ac.uk, Elizabeth S. Sattely sattely@stanford.edu

**This PDF file includes:**

Materials and Methods
Figs. S1 to S45
Tables S1 to S24
Captions for Data S1

**Other Supplementary Materials for this manuscript include the following:**

Data S1 - Full NMR spectral data for isolated compounds

## Materials and Methods

Generation of *Melia azedarach* genome assembly, annotation and RNA-seq dataset

Two *Melia azedarach* plants (individuals '02' and '11), purchased in 2016 (Crûg Farm Plants) and maintained (as described (*20*)) in a John Innes Centre greenhouse, were utilized for all sequencing experiments described. Raw RNA-seq reads and genome assembly (with annotation for assembled pseudo-chromosomes) have been submitted to NCBI under the BioProject numbers PRJNA906055 and PRJNA906622 respectively.

High molecular weight (HMW) genomic DNA (average 58 Kbp in length) was extracted from *M. azedarach* leaves (individual '11') using the modified CTAB protocol which includes the addition of proteinase K and RNase A (Qiagen) (*45*). From this, the Earlham Institute constructed a 20-30 Kbp PacBio shotgun library which was sequenced over 10 SMRT cells on a Sequel instrument. The resultant filtered subreads (over two million with an average length of 13 Kbp) were *de novo* assembled, utilizing the hierarchical genome assembly process 4 (HGAP-4, PacBio) tool to create a draft genome with a total length of 230 Mbp (550 contigs). The proximo Hi-C Plant Kit (Phase Genomics) was used for chromatin cross-linking and subsequent extraction of DNA from *M. azedarach* leaves (individual '11'), following this, Hi-C (*46*) was performed by Phase Genomics. The proximal tool was then used to generate a pseudo-chromosome level assembly based on chromatin interactions from the Hi-C analysis and the draft *M. azedarach* genome. A mis-assembly within the draft genome (contig 000011F) was identified during this process and subsequently split, which resulted in the generation of 14 pseudo-chromosomes in the final assembly. Karyotyping was performed on young *M. azedarach* root tips (individual '11'). The preparation of mitotic metaphase spreads was carried out as described previously (*47*). Chromosomes were counterstained with DAPI (1 µg/ml). Images were acquired using a Leica DM5500B microscope equipped with a Hamamatsu ORCA-FLASH4.0 camera and controlled by Leica LAS X software V2.0.

Seven different tissues (four replicates of each) were harvested for RNA extraction from *Melia azedarach* plants. These included: upper leaves, lower leaves, petiole (including rachis) and roots of a high salannin-producing individual '11' and upper leaves, lower leaves and petiole (including rachis) of a low salannin individual '02'. Tissues were immediately flash frozen in liquid nitrogen before being ground to a fine powder using a pre-cooled pestle and mortar. All tissues were harvested on the same day and extractions were performed in technical replicates. RNA extraction was performed using the MacKenzie-modified RNeasy Plant Mini Kit (Qiagen) protocol (*48*), with DNAase (Promega) treatment, performed on column. The Earlham Institute generated high-throughput Illumina stranded RNA libraries (150bp, paired end) of each of the 28 samples, which were multiplexed and sequenced over two lanes of a HiSeq 4000 instrument (Illumina). This generated over 635 million paired end reads (an average of 91 million per tissue (table S2)).

This RNA-seq dataset was utilized to by the Earlham Institute to generate a high quality structural genome annotation for *M. azedarach*, using their specialist plant genome annotation pipeline (including both Mikado (*49*) and Portcullis (*50*) tools), shown to be capable of annotating a diverse range of plant species (*51*, *52*). Functional annotation was generated using the Assignment of Human Readable Descriptions (*53*) (AHRD) V.3.3.3 tool. AHRD was provided with results of BLAST V2.6.0 (*54*) searches (e-value = 1e-5) against reference proteins from TAIR (*55*), UniProt (*56*), Swiss-Prot and TREMBL (*57*) datasets, along with interproscan (*58*) results.

Transcriptome data mining and analysis of *Citrus* dataset

Publicly available gene expression data from a collection of 297 Citrus datasets were downloaded from the Network Inference for Citrus Co-Expression (NICCE) (*22*). The dataset consisted of normalized expression data collected from multiple sources, tissues, and treatments (multiple Citrus spp., fruit, leaf, biotic stress, abiotic stress and age). Linear regression analysis to calculate Pearson's R coefficient on normalized expression levels was performed using *CsOSC1* as the bait gene (fig. S2). As additional genes were characterized, these were then used as bait genes along with previously characterized genes (*20*). These included using *CsCYP71CD1*, *CsCYP71BQ4*, *CsCYP88A51*, and *CsL21AT* as bait genes. The obtained list was ranked by decreasing Pearson's R coefficient (PCC). The top microarray probes (per bait gene list) were then mapped to the respective *Citrus sinensis* genes. Candidate genes were then annotated both via Pfam assignment and via the best blastx hit using the *Arabidopsis thaliana* proteome as a reference. The final list of candidates was further refined as needed to only include candidates with Pfam assignments belonging to desired biosynthetic genes.

Mining of *M. azedarach* resources for gene expression analysis

To process raw RNA-seq reads generated for *M. azedarach* and generate read counts, STAR V2.5 (*59*) was used to align all reads to the *M. azedarach* genome annotation (pooling all reads per replicate (directional and lane)) and Samtools V1.7 (*60*) was used to index the subsequent alignment. The featureCounts tool of subread V1.6.0 (*61*) was used to generate raw read counts by counting the number of reads overlapping with genes in each alignment.

Raw read counts were analyzed in R using DEseq2 V1.22.1 (*62*). Genes with zero counts were removed from the analysis, normalization was performed based on library size (to account for differences in number of reads sequenced for each replicate (*63*, *64*)) and subsequent counts were $\log_2$ transformed with a pseudo count of one. The resultant library-normalized $\log_2$ read counts were used for downstream analyses. Separately, differential expression analysis (to identify a subset of genes considered differentially rather than constitutively expressed) was performed by importing the raw read counts into an EdgeR (*25*) object and removing genes with low coverage (less than one count per million in more than four samples). Normalization (by library size) was performed using the 'trimmed mean of M-values' method. Finally to identify

differentially expressed genes, a genewise negative binomial generalized linear model (glmQLFit) was used with pairwise comparisons between all sample types. Using these differentially expressed genes as a subset, $\log_2$ library-normalized counts (generated by DEseq2 V1.22.1 (*62*)) for the 28 replicates were used to calculate Pearson's correlation coefficients (PCCs) for each gene to each of the known melianol biosynthetic genes *MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*. Genes were ranked based on their average PCC value against these three genes and then filtered to select only the genes with one of the following interpro annotations of biosynthetic interest; IPR005123 (Oxoglutarate/iron-dependent dioxygenase), IPR020471 (Aldo/keto reductase), IPR002347 (Short-chain dehydrogenase/reductase SDR), IPR001128 (Cytochrome P450), IPR003480 (Transferase) or IPR007905 (Emopamil-binding protein).

Although at rank 84 in this analysis (Fig. 2C), *MaAKR* can be considered co-expressed, it is not as strongly co-expressed as other functional genes, and was in fact first identified due to its sequence similarity to the functional Citrus gene *CsAKR*. The gene prediction for *MaAKR* in the *M. azedarach* genome is truncated (lacking 38 terminal amino acids due to two point mutations). To identify a full-length version*, de novo* transcriptome assembly was performed using Trinity V2.4.0 (*65*) following a standard protocol (*66*) and incorporating all petiole replicates from *M. azedarach* (individual '11' (pooled)). Transdecoder X5.5.0 (*66*) was used to generate structural annotations for this transcriptome. Subsequently the truncated *MaAKR* (table S10) sequence identified in the genome was used as a BLASTp query to identify the full length *MaAKR* sequence (table S20).

Cloning of candidate genes from *C. sinensis* and *M. azedarach*

mRNA from *Citrus sinensis* var. Valencia (Sweet orange) fruit buds (green immature fruit 1~3 cm in diameter) from one-year old plants were isolated using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) following the manufacturer's instructions. Tissues were flash-frozen in liquid nitrogen and ground using a pestle and mortar. cDNA was generated using Super Script IV First Strand Synthesis System (Invitrogen). Candidate genes from *C. sinensis* were cloned (via Gibson assembly) into pEAQ-HT vectors (*67*), and transferred into *Agrobacterium tumerificans* (strain *GV3101*) following methods which have been previously described (*68*). Candidate genes from *M. azedarach* were amplified from leaf and petiole cDNA, cloned (via gateway cloning) into pEAQ-HT-DEST1 vectors (*67*) and transferred into *Agrobacterium tumerificans* (strain *LBA4404*) following methods which have been previously described (*20*). Primers used for cloning of functional genes from *C. sinensis* and *M. azedarach* are listed (table S21 and table S22, respectively).

Characterization of *C. sinensis* and *M. azedarach* candidate genes through transient co-expression in N*icotiana benthamiana*

To understand the function of enzymes of interest, candidate genes from *C. sinensis* and *M. azedarach* were tested via co-expressing various combinations of candidate genes with the

previously characterized melianol biosynthetic genes (*AiOSC1/CsOSC1, MaCYP71CD2/CsCYP71CD1* and *MaCYP71BQ5/CsCYP71BQ4* (*20*)). This was performed by agroinfiltration of *A. tumefaciens* strains harboring the genes of interest in pEAQ vectors, following methods previously described (*20*, *69*). In addition to the limonoid biosynthetic genes, *Avena strigosa tHMGR* (encoding a truncated feedback insensitive HMG CoA-reductase that boosts triterpene yield (*69*)) was infiltrated in combination with *M. azedarach* candidate genes, while *A. thaliana* HMG CoA-reductase was used in combination with *C. sinensis* candidate genes.

Construction of sterol isomerase phylogenetic tree

Sterol isomerase sequences from high-quality plant genomes (33 species) were obtained from Phytozome (https://phytozome-next.jgi.doe.gov/) using a PFAM based search with PF05241 (EXPanded EBP superfamily). Full names of the species for which SI sequences were downloaded are as follows: *Amaranthus hypochondriacus, Aquilegia coerulea, Arabidopsis lyrata, Arabidopsis thaliana, Boechera stricta, Brassica rapa, Capsella grandiflora, Capsella rubella, Citrus clementina, Citrus sinensis, Daucus carota, Eucalyptus grandis, Eutrema salsugineuma, Fragaria vesca, Glycine max, Gossypium raimondii, Kalanchoe fedtschenkoi, Linum usitassimum, Malus domestica, Manihot esculenta, Medicago trunculata, Mimulus guttatus, Oryza sativa, Populus trichocarpa, Prunus persica, Ricinus communis, Salix pupurea, Solanum lycopersicum, Solanum tuberosum, Theobroma cacao, Trifolium pratense, Vitis vinifera* and *Zea mays.* Sequences with length of 150-400 amino acids were selected for analysis. Sterol isomerase sequences (Interpro: IPR007905 (Emopamil-binding protein)) from the newly generated *M. azedarach* genome were also included in this analysis.

Protein alignments were performed on this set of sequences using mafft (*70*) (FFT-NS-I method) with a maximum of 1000 iterations. The phylogenetic tree was generated using MrBayes (*71*), with a mixed amino acid probability model and MCMC analysis was performed over 1 million generations using 4 chains, 2 independent runs and a temperature of 0.7.

Extraction and analysis of limonoids and protolimonoids from Rutaceae species and *N. benthamiana* expressing candidate *C. sinensis* biosynthetic genes

*N. benthamiana* leaf tissue was collected 5-days post *Agrobacterium* infiltration using a 1 cm DIA leaf disc cutter. Each biological replicate consisted of 4 leaf discs from the same leaf (approx. 0.04 g FW leaves). Leaf discs were lyophilized overnight and placed inside a 2 mL safe-lock microcentrifuge tube (Eppendorf). 500 µL of methanol (Fisher Scientific, ACS & HPLC grade) was added to each sample, and these were then homogenized in a ball mill (Retsch MM 400) using 5 mm stainless steel beads and milled at 25 Hz for 2 min. After homogenization, the samples were centrifuged at 13,200 rpm for 10 min. Supernatants were filtered using either 0.20 or 0.45 µm PTFE filters (GE) before being subjected to LC-MS analysis.

LC-MS was carried using electrospray ionization (ESI) on positive mode on an Agilent 1260 HPLC coupled to an Agilent 6520 Q-TOF mass spectrometer. Separation was carried out using a 5 µm, 2 × 100 mm Gemini NX-C18 column (Phenomenex) using 0.1% formic acid in water (A) versus 0.1% formic acid in acetonitrile (B) run at 400 µL/min, room temperature. The following gradient of solvent B was used: 3% 0-1 min, 3%-30% 1-3 min, 30%-97% 3-18 min, 97% 18-22 min, 97%-3% 22-23 min and 3% 23-29 min. MS spectra was collected at m/z 50 - 1400. The ESI source was set as follows: 350 °C gas temperature, 10 L/min drying gas, 35 psi nebulizer, 3500 V VCap, 150 V fragmentor 65 V skimmer and 750 V octupole 1 RF Vpp.

MS/MS data (100-1700 m/z, 1.5 spectra/sec) was collected using the same instrument, column and gradient under targeted MSMS acquisition mode, with a narrow isolation width (~1.3 m/z) and collision energies of 20, 40 and 50 eV.

In addition, seeds of *Phellodendron amurense* (amur cork tree) were purchased from eBay, lyophilized as described above, and 2~3 seeds were homogenized in a ball mill (Retsch MM 400) using 5 mm stainless steel beads and milled at 25 Hz for 2 min in 2 mL ethyl acetate solvent (Fisher Scientific, HPLC grade). The extracts were air dried, redissolved in equal volume of methanol, and filtered using 0.45 µm PTFE filters (GE) before subjecting to LC-MS analysis.

Extraction and analysis of limonoids and protolimonoids from Meliaceae species and *N. benthamiana* expressing candidate Meliaceae biosynthetic genes

For each sample, 10 mg of freeze-dried plant material was weighed and then homogenized using Tungsten Carbide Beads (3 mm, Qiagen) with a TissueLyser (1000 rpm, 2 min). Samples were agitated at 18 °C for 20 min in 500 µl methanol (100%). Samples were transferred to a 0.22 µM filter mini-column (Geneflow) and filtered by centrifugation before being transferred to a glass analysis vial.

Unless otherwise stated, all UHPLC-MS experiments described relating to Meliaceae material and genes were performed with positive mode electrospray ionization (Dual AJS ESI) on an LC/Q-TOF instrument (6546, Agilent), with separation by on an 1290 infinity LC system equipped with a DAD (Agilent). 1 ul of sample was injected for separation on a Kinetex 2.6 µm XB-C18 100 Å 2.1 x 50 mm column (Phenomenex) using 0.1% formic acid in water (A) versus acetonitrile (B) at 500 µl/min and 40 °C. Separation was performed using the following gradient of solvent B: 37% 0-1 min (first minute of flow diverted to waste), 37-67% 1-11 min, 67-100% 11-11.5 min, 100% 11.5-13.5 min, 100-37% 13.5-14 min and 37% 14-15 min. Full MS spectra were collected (m/z 100-1000, 1 spectra/sec). Spray chamber and source parameters were as follows; 325 °C gas temperature, 10 L/min drying gas, 20 psi nebulizer, 3500 V VCap, 120 V fragmentor 45 V skimmer and 750 V octupole 1 RF Vpp. Reference masses used for calibration were 121.05087300 and 922.00979800. In addition DAD spectra (200-400 nm, 2 nm step) were collected.

In addition to metabolite extraction from infiltrated *N. benthamiana* and the *Melia azedarach* trees maintained at JIC, extraction and analysis was also performed on dried leaf material from 13 Meliaceae species (*Carapa guianensis, Cipadessa fruticosa, Dysoxylum spectabile, Khaya nyasica, Malleastrum mandenense, Melia azedarach, Nymania capensis, Toona sinensis, Trichilia havanensis, Turraea floribunda, Turraea obtusifolia, Turraea sericea* and *Turraea vogelioides*) sourced from Kew Gardens in 2017 (Nagoya Protocol compliant) and stored at -70 °C.

General considerations for the purification and characterization of limonoid intermediates from *N. benthamiana* expressing Citrus biosynthetic genes

Approximately 500 g of leaves from 60-100 infiltrated plants were cut into small pieces of approximately 0.25 $cm^2$ in area. Leaves were immediately flash frozen and lyophilized to complete dryness. Dried leaves were then grinded to powder using a mortar and pestle. Leaf powder was then placed in a 4 L flask (1 g FW leaves per 12.5 mL) with a magnetic stir bar and extracted using EtOAc for 72 h at room temperature with constant stirring. Extracts were filtered using vacuum filtration and dried using rotary evaporation. Flash chromatography was performed using a 7 cm DIA column loaded with silica (SiliaFlash® P60). Hexane (Fisher Scientific, ACS & HPLC grade) and ethyl acetate were used as running solvents. 500 mL fractions were collected via isocratic elution (60% hexane, 40% ethyl acetate). Fractions were analyzed via LC-MS, and those containing the compound of interest were pooled and dried using rotary evaporation. The dried samples were as then resuspended in approximately 1 mL of DMSO. The samples were then further purified using an Isolera Prime Biotage using a Sfår C18 Duo 12g column. Fractions were collected using water (A) and acetonitrile (B) as solvents. The following gradient of solvent B was used: 30% for 3 column volumes (CV), 30-80% for 25 CV, 80-100% 2 CV. Active fractions, as verified by LC-MS, were then dried to completion using rotary evaporation or lyophilization. For Citrus intermediates $^{1}H$ NMR and $^{13}C$ NMR spectra were acquired using a Varian Inova 600 MHz spectrometer at room temperature. Shifts are referenced to the residual solvent peak ($CDCl_3$, Acros Organics) and reported downfield in ppm using $Me_4Si$ as the 0.0 ppm internal reference standard.

General considerations for the purification and characterization of limonoid intermediates from *N. benthamiana* expressing *M. azedarach* biosynthetic genes and *A. indica*

To enable the purification of heterologously produced intermediates from *N. benthamiana*, large-scale vacuum infiltration of the relevant *A. tumefaciens* strains was performed as previously described (*72*, *73*), using 100-130 large-sized *N. benthamiana* plants. Once harvested and freeze-dried, a preliminary triterpene extraction was performed on the leaf material using a previously described method (*73*). Briefly, a speed extractor (Bucchi) was used to perform high temperature (100 °C) and pressure (130 bar) extraction from leaf material with ethyl acetate. Unless otherwise specified, the ambersep 900 hydroxide form beads (Sigma-Aldrich) recommended to remove chlorophylls (*73*) were not used, due to the presence of acetate groups in the compounds being isolated.

All Preparative HPLC was performed on an Agilent Technologies infinity system equipped with a 1290 infinity II fraction collector, a 1290 infinity II preparative pump and column oven, a 1260 infinity II quaternary pump, a 1260 infinity II Diode Array Detector (DAD), a 1260 infinity II ELSD and an infinity lab LC/MSD XT. Separation for preparative HPLC was performed on a 250 x 21.2 mm Luna® 5 µM C18(2) 100 Å column (Phenomonex), at 25 ml/min, with a collection:detector split of 1000:1 and the quaternary pump providing a make-up flow at 1.2 ml/min for the detectors. All preparative runs included a minimum of 3 min post-time at starting solvent percentage. Unless otherwise stated, MS data was collected via MM-ES+APCI scan mode, collecting data after 1.5 min with a mass range 200-1200 and collection of [M] or $[M+H]^+$ masses.

General considerations for NMR characterizations

Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of $^1$H, $^{13}$C, DEPT-135, DEPT-edited HSQC, HMBC and 2D NOESY or ROESY experiments. Where signals overlap $^1$H δ is reported as the center of the respective HSQC crosspeak. Multiplicities are described as, s = singlet, d = doublet, dd = doublet of doublets, dt = doublet of triplets, t = triplet, q = quartet, quint = quintet, tquin = triplet of quintets, m = multiplet, br = broad, appt = apparent.

Purification of *apo*-melianol (**3**) (via expression of *M. azedarach* genes)

Using vacuum infiltration 115 large *N. benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains harboring pEAQ-HT-DEST1 expression constructs of the following genes: *AstHMGR*, *AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108* and *MaMOI2*. Leaves were harvested and freeze-dried six days after infiltration, yielding 159.9 g of dried leaf material. Following the preliminary triterpene extraction method described above, and for this compound utilizing the ambersep 900 hydroxide form beads to remove chlorophyll, successive rounds of fractionation were performed utilizing an Isolera Prime (Biotage) as described in (table S23). Fractions containing the target were pooled, and to achieve final purification, subject to semi-preparative UHPLC, performed on an Agilent Technologies 1290 Infinity II system equipped with an Agilent Technologies 1290 infinity II Diode Array Detector (DAD), Agilent 1260 Infinity Evaporative Light Scattering Detector (ELSD) and an Agilent 1260 infinity II fraction collector. The sample was dissolved in a minimal volume of acetonitrile and injected in 200 µl aliquots. Separation was performed on a 250 x 10 mm S-5 µM 12 nm Pack pro C18 column (YMC) using water (A) versus 95% acetonitrile (B) at 4 ml/min and 40 ºC with the following gradient of solvent B; 68% 0-30 min, 68-100% 30-32 min, 100% 32-37 min, 100-41% 37-39 min and 41% 39-44 min. The fraction collector was programmed to collect between 22-25 min (with a maximum peak duration of 2 min) and to be triggered (threshold and peak) by detection of a peak from either the DAD or ELSD detector. DAD was set to collect signals with a wavelength of 205 nm and bandwidth of 4 nm. Fractions collected within this region (over 11

runs) were pooled and dried down. This yielded 13.1 mg of (**3**) as a white powder on which NMR was performed in $CDCl_3$ (table S3).

Purification of (**6**) (via expression of *C. sinensis* genes)

62 *N. benthamina* plants (5-6 week old) were vacuum infiltrated using *A. tumefaciens* mediated transient expression using equal volume of infiltrated strains (OD per strain = 0.2) harboring *AtHMGR, CsOSC1, CsCYP71CD1, CsCYP71BQ4, CsCYP88A51, CsMOI2, CsL21AT*, and *CsSDR*. 871.09 g of leaves were harvested 6 days post-infiltration, dried (yielding 106.89 g) and extracted in ethyl acetate following the standard procedure outlined above. Isolation and NMR analysis of (**6**) (table S4) was subsequently performed following the standard methods outlined above.

Purification of (**4'**) (via expression of *C. sinensis* genes)

43 *N. benthamina* plants (6-7 week old) were vacuum infiltrated using *A. tumefaciens* mediated transient expression using equal volume of infiltrated strains (OD per strain = 0.2) harboring *AtHMGR, CsOSC1, CsCYP71CD1, CsCYP71BQ4, CsCYP88A51, CsMOI1, CsL21AT* and *CsSDR*. 865.1 g of leaves were harvested 6 days post-infiltration, dried (yielding 105.5 g) and extracted in ethyl acetate following the standard procedure outlined above. Isolation and NMR analysis of (**4'**) (table S5) was subsequently performed following the standard methods outlined above

Purification of 21(*S*)-acetoxyl-*apo*-melianone (**6**) (via expression of *M. azedarach* genes)

Using vacuum infiltration (*72*, *73*), 121 large *N. benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains harboring pEAQ-HT-DEST1 expression constructs of *AstHMGR, AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108, MaMOI2, MaL21AT* and *MaSDR*. One week after infiltration, leaves were harvested and freeze-dried yielding 150.1 g of dried material. Following the preliminary extraction of triterpenes described above, successive rounds of fractionation were performed utilizing an Isolera Prime (Biotage) (table S23). Fractions containing the target were pooled and final purification was achieved by re-crystallisation. Briefly, hot ethanol (70 ºC) was added dropwise to the sample (heated to 70 ºC) until all solids had dissolved. The sample was then covered and left at room temperature for crystals to form. Crystals were washed in cold ethanol under vacuum and then filtered by dissolving the samples in methanol to allow collection. Initial recrystallisation was performed in triplicate, yielding ~300 mg of pale yellow product. The recrystallisation was repeated using this pale yellow product, to yield 77.25 mg of white product (**6**). 5 mg of product redissolved in $CDCl_3$ for NMR (table S6).

Purification of (**9**) (via expression of *C. sinensis* genes)

63 *N. benthamina* plants (5-6 week old) were vacuum infiltrated using *A. tumefaciens* mediated transient expression using equal volume of infiltrated strains (OD per strain = 0.2) harboring *AtHMGR, CsOSC1, CsCYP71CD1, CsCYP71BQ4, CsCYP88A51, CsMOI2, CsL21AT,*

*CsSDR, CsCYP716AC1* and *CsCYP88A37*. 864.2 g of leaves were harvested 6 days post-infiltration, dried (yielding 89.38 g) and extracted in ethyl acetate following the standard procedure outlined above. This resulted in the isolation of 20.3 mg of (**9**). NMR analysis was performed following the standard methods outlined above (table S8).

Purification of epi-neemfruitin B (**10**) (via expression of *M. azedarach* genes)

Using vacuum infiltration (*72*, *73*), 143 large *N. benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains harboring pEAQ-HT-DEST1 expression constructs of the following genes: *AstHMGR, AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108, MaMOI2, MaL21AT, MaSDR, MaCYP88A164* and *MaL1AT*. Eight days after infiltration, leaves were harvested and freeze-dried, yielding 112.5g of dried material. Following the preliminary extraction of triterpenes described above, successive rounds of fractionation were performed utilizing an Isolera Prime (Biotage) (table S23). Fractions containing the target were pooled and dissolved in minimal volume of methanol (3 ml) for final purification via injection (500-1200 µl) onto a preparative HPLC instrument. Separation was achieved using water (A) versus 95% acetonitrile (B) with the following gradient of solvent B; 42% 0-1 min, 42-73% 1-1.5 min, 73-100% 1.5-11.5 min, 100% 11.5-16.5 min and 100-42% 16.5-17 min. Fractions were collected between 8-11 minutes triggered by detection of an MS peak with a *m/z* of 526 [M] (threshold 5,000) and DAD peak (threshold of 5, wavelength 205 nm). Fractions were pooled and dried to yield 4 mg of a pale yellow product (**10**), which was dissolved in minimal ethanol and treated with activated charcoal to remove coloured impurities. This yielded 2.25 mg of purified product, which was dissolved in $CDCl_3$ for NMR (table S11).

Purification of (**13**) (via expression of *C. sinensis* genes)

31 *N. benthamina* plants (5-6 week old) were vacuum infiltrated using *A. tumefaciens* mediated transient expression using equal volume of infiltrated strains (OD per strain = 0.2) harboring *AtHMGR, CsOSC1, CsCYP71CD1, CsCYP71BQ4, CsCYP88A51, CsMOI2, CsL21AT, CsSDR, CsCYP716AC1*, *CsCYP88A37, CsL1AT, CsL7AT, CsAKR, CsCYP716AD2,* and *CsLFS*. 397.96 g of leaves were harvested 6 days post-infiltration, dried (yielding 40.07 g) and extracted in ethyl acetate following the standard procedure outlined above. This resulted in the isolation of 0.3 mg of (**13**) and 19.4 mg of (**13'**). NMR analysis was performed on both products (table S13 to S14) following the standard methods outlined above.

Purification of (**14**) (via expression of *M. azedarach* genes)

Using vacuum infiltration (*72*, *73*), 110 medium/large *N. benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains harboring pEAQ-HT-DEST1 expression constructs of *AstHMGR, AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108, MaMOI2, MaL21AT, MaSDR, MaCYP88A164, MaL1AT, MaL7AT* and *MaAKR.* Six days after infiltration, leaves were harvested and freeze-dried yielding 140.4 g of dried material. Following the preliminary extraction of triterpenes described above, initial fractionation was then performed utilizing an Isolera Prime (Biotage) (table S23). Fractions containing the target compound were

then subject to liquid-liquid partitioning (80% methanol:hexane, in triplicate). The 80% methanol fractions were pooled and re-dissolved in a minimal volume of methanol (10 ml) for final purification via injection (250-1000 µl) onto a preparative HPLC instrument. Separation was performed using water (A) versus 95% acetonitrile (B) with the following gradient of solvent B: 42%-100%, 0-15 min, 100% 15-19 min and 100-42% 19-19.5 min. Fractions were collected between 9-11.5 minutes triggered by a peak of *m/z* 570 $[M+ACN+H]^+$ (threshold 5,000). The $[M+ACN+H]^+$ adduct mass was used as an inputted mass rather than [M] or $[M+H]^+$ due to the high accumulation of acetonitrile adducts for this intermediate. Ten fraction collecting runs were performed and the pooled fractions yielded ~20 mg of product. Initially, 4 mg of product was dissolved in $CDCl_3$ for NMR, however this appeared to be converted to the known protolimonoid, gradifoliolenone (36) (fig. S1), in solution. A further 4 mg of product was dissolved in pyridine-$d_5$ however a suspected rotamer effect was observed. Therefore NMR characterization was finally performed by dissolving 5 mg of product in benzene-$d_6$ (table S15).

Purification of kihadalactone A (**19**) (via expression of *C. sinensis* genes)

63 *N. benthamina* plants (5-6 week old) were vacuum infiltrated using *A. tumefaciens* mediated transient expression using equal volume of infiltrated strains (OD per strain = 0.2) harboring *AtHMGR, CsOSC1, CsCYP71CD1, CsCYP71BQ4, CsCYP88A51, CsMOI2, CsL21AT, CsSDR, CsCYP716AC1, CsCYP88A37, CsL1AT, CsL7AT, CsAKR, CsCYP716AD2* and *CsLFS*. 705.7 g of leaves were harvested 6 days post-infiltration, dried (yielding 79.15 g) and extracted in ethyl acetate following the standard procedure outlined above. Isolation and NMR analysis of (**19**) (table S16) was subsequently performed following the standard methods outlined above.

Purification of azadirone (**18**) (from *A. indica* leaf powder)

224.9 g of neem (*A. indica*) leaf powder (purchased from H&C Herbal Ingredients Expert) was extracted following the preliminary triterpene extraction method described above. Following this the extract was partitioned between ethyl acetate (800 ml) and water (800 ml) which yielded 21.2 g of crude extract. Initial fractionation was then performed utilizing an Isolera Prime (Biotage) (table S23) following a method adapted from previous reports of azadirone isolation from *A. indica* fruits (*74*). Fractions containing azadirone were then dissolved in a minimal volume of methanol (with dropwise addition of ethyl acetate), before being filtered, through both a Sep-Pak vac 3cc C18 cartridge (Waters) and a minisart highflow PES 0.22 µM syringe filter (Sartorius), before injection (7000 µl) onto a preparative HPLC system. MS was collected via MM-ES+APCI in SIM mode, detecting and collecting for a *m/z* of 437.2 $[M+H]^+$. Fractions were collected between 14-20 min (threshold 5,000). Separation was performed using water (A) versus acetonitrile (B) with the following gradient of solvent B; 65% 0-1.5 min, 60-100% 1.5-26.5 min, 100%, 26.5-30 min and 100-65% 30-30.5 min. After 3 runs, fractions of azadirone (**18**) with a reasonable level of purity were pooled, yielding ~1 mg of purified product, which was dissolved in $CDCl_3$ for NMR (table S17).

Purification of (**20**) (via expression of *M. azedarach* genes)

Using vacuum infiltration (*72*, *73*), 120 medium/large *N. benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains harboring pEAQ-HT-DEST1 expression constructs of *AstHMGR, AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108, MaMOI2, MaL21AT, MaSDR, MaCYP88A164, MaL1AT, MaAKR* and *MaCYP716AD4*. Eight days after infiltration, leaves were harvested and freeze-dried yielding 123.1 g of dried material. Following the preliminary extraction of triterpenes described above, initial fractionation was then performed utilizing an Isolera Prime (Biotage) (table S23). Fractions containing the target were then dissolved in a minimal volume of 80% acetonitrile (6 ml) before injection (500-1500 µl) onto a preparative HPLC system. For this product MS was collected via MM-ES+APCI in SIM mode, detecting and collecting for a mass of 503.4 $[M+H]^+$. Fractions were collected between 1.5-10 min (threshold 5,000). Initial separation was performed using water (A) versus acetonitrile (B) with the following gradient of solvent B: 60% 0-0.5 min, 60-75% 0.5-10 min, 75-100% 10-10.5 min, 100% 10.5-15 min and 100-60% 15-15.5 min. After 9 runs, fractions containing target were pooled and further purified by a second round of preparative HPLC, using the same instrument settings, but a different gradient consisting of water (A) versus methanol (B) with the following gradient of solvent B: 67% 0-0.5 min, 67-77% 0.5-20 min, 77-100% 20-20.5 min, 100% 20.5-24.5 min and 100-67% 24.5-25 min. After two injections, fractions containing the target were pooled, yielding ~0.6 mg of purified product (**20**), which was dissolved in benzene-$d_6$ for NMR (table S20).

## Supplementary text

### Off-target activity of *Ma*CYP716AD4/*Cs*CYP716AD2 activity on non C-7 *O*-acetylated substrates

The characterisation of *Cs*L7AT and *Ma*L7AT in the biosynthesis of early limonoids, e.g. azadirone (**18**) and kihadalactone A (**19**), was unexpected. The C-7 *O*-acetylation activity of *Cs*L7AT and *Ma*L7AT seems unnecessary for the biosynthesis of more elaborated limonoids like limonin and azadirachtin (Fig 1, fig. S1), most of which have C-7 ketone or hydroxyl instead of C-7 acetoxyl. However, when we omitted *Cs*L7AT in the full kihadalactone A (**19**) pathway, the expected (**19**) C-7 deacetylated product was not observed (fig. S45). Instead, an oxidized intermediate accumulates that still contains the full triterpene scaffold, indicating that C-7 *O*-acetylation is important for C-4 scission. Furthermore, C-7 *O*-acetylation also plays a key role in the Meliaceae pathway, as in the absence of *Ma*L7AT an analogous side-product (**20**) is made and structurally confirmed (fig. S44, fig. S32, table S20). These data suggest that *Ma*CYP716AD4/*Cs*CYP716AD2 activities require C-7 *O*-acetylation on the substrates, and downstream C-7 *O*-deacylation by a deacetylase would be required to reach more elaborated limonoids. However, we cannot exclude the possibility that *Cs*L7AT/*Ma*L7AT are not required for the pathway to more elaborated limonoids, but we are missing other key enzymes that would allow the proper functioning of *Ma*CYP716AD4/*Cs*CYP716AD2.

## Supplementary Figures

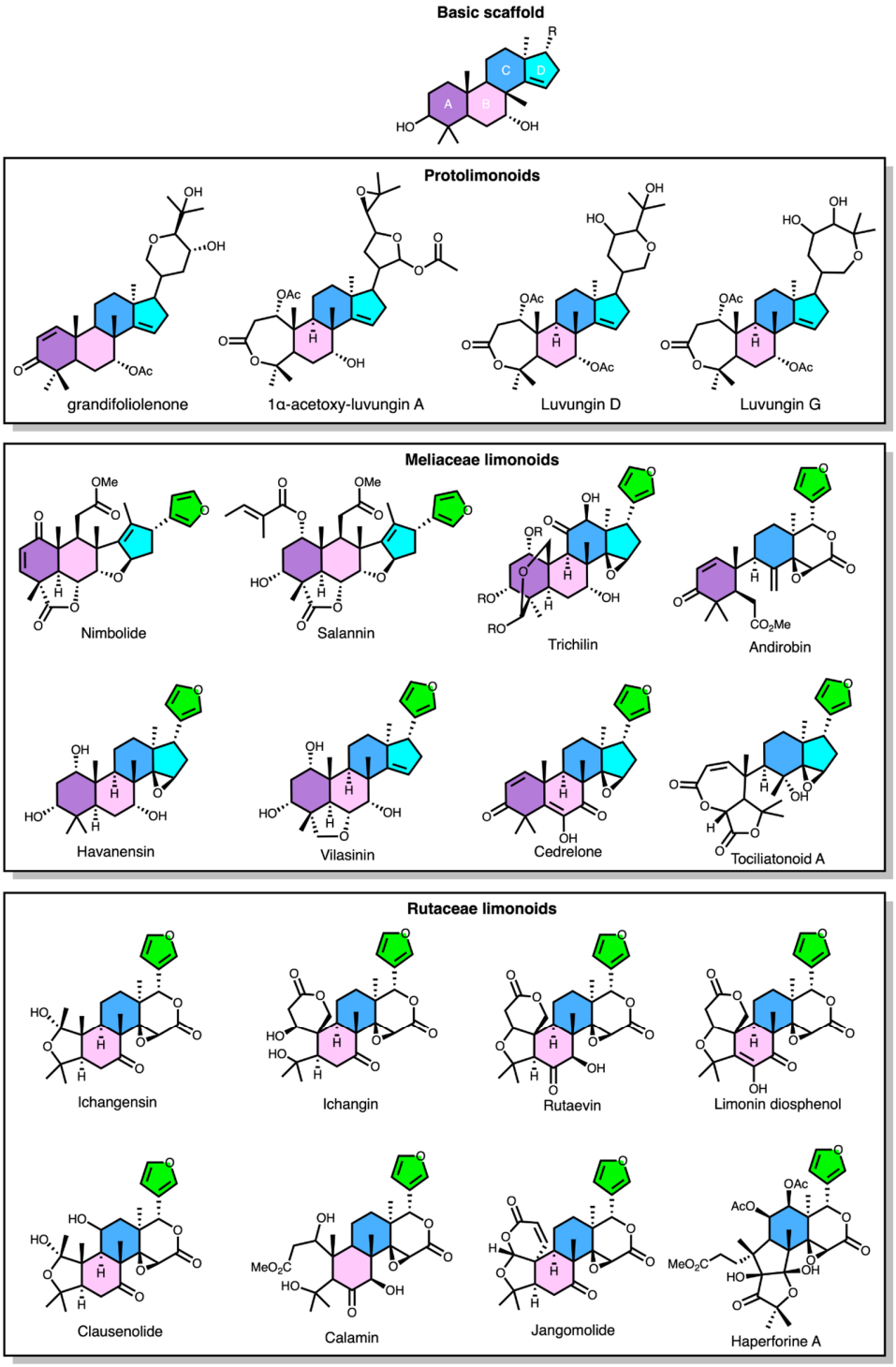



**Fig. S1. Supplementary limonoid and protolimonoid structures.**
Additional structures of protolimonoids, along with Meliaceae and Rutaceae limonoids relevant to the main text. Ring A-D are labeled on the basic scaffold on the top. The rings and the furan moiety are colored to show the cleavage and conservation of each ring.

| Candidate Gene | PCC | Rank |
|---|---|---|
| CsOSC1 | 1.000 | 0 |
| 2-ODD | 0.860 | 1 |
| CsL7AT | 0.780 | 2 |
| CsLFS | 0.770 | 3 |
| Epimerase/Dehydratase | 0.745 | 4 |
| CsCYP716AD2 | 0.731 | 5 |
| Glycosyl hydrolase | 0.713 | 6 |
| Unknown | 0.685 | 7 |
| CsCYP88A51 | 0.676 | 8 |
| CsAKR | 0.671 | 9 |
| CsCYP716AC1 | 0.656 | 10 |
| CsMOI2 | 0.652 | 11 |
| 2-ODD | 0.640 | 12 |
| CsSDR | 0.631 | 13 |
| Major latex protein | 0.622 | 14 |
| Phosphoglycerate mutase | 0.619 | 15 |
| Ras GTPase | 0.619 | 16 |
| SAM methyltransferase | 0.611 | 17 |
| Unknown | 0.606 | 18 |
| CsCYP71BQ4 | 0.600 | 19 |
| CsCYP88A37 | 0.577 | 39 |
| CsCYP7CD1 | 0.572 | 42 |
| CsL21AT | 0.514 | 79 |
| CsMOI1 | 0.368 | 1614 |

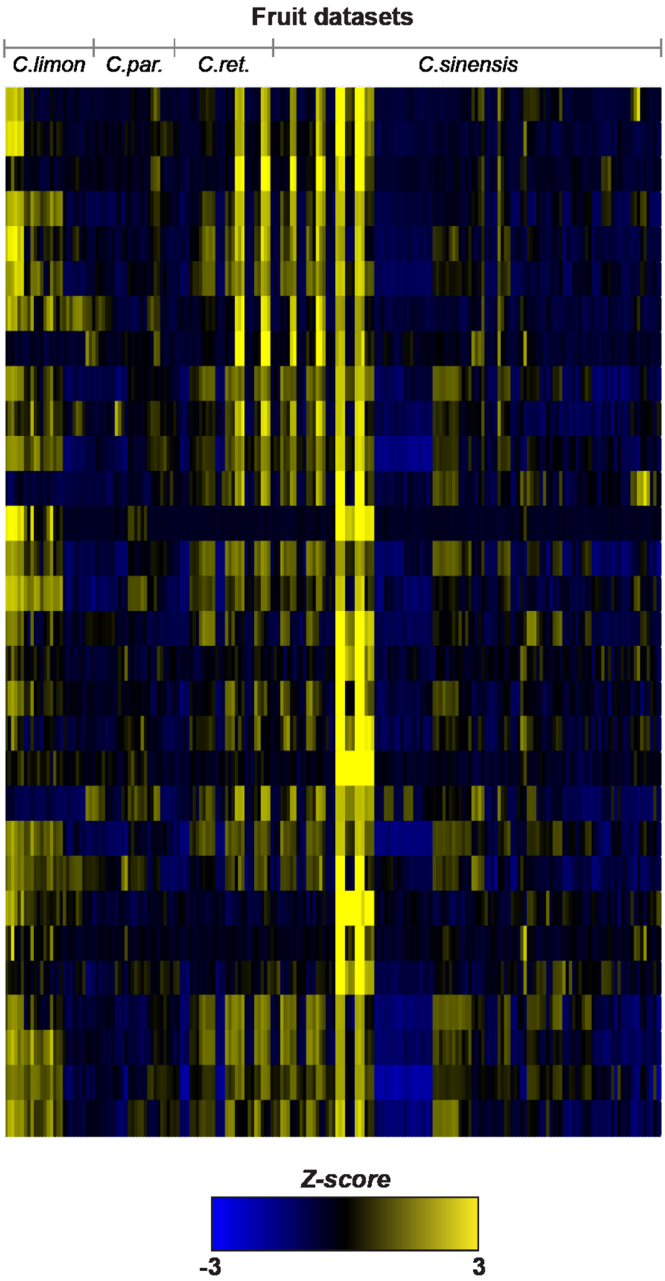



**Fig. S2. Co-expression analysis of the *C. sinensis* microarray expression data from Network inference for Citrus Co-Expression (NICCE) using *CsOSC1* as a bait gene.**
Linear regression analysis was used to rank the top 20 genes based on Pearson's correlation coefficient (PCC) to *CsOSC1*. Heat map displays Z-score calculated from $\log_2$ normalized expression across fruit datasets. Genes in red indicate candidates characterized in this study or our previous work (*20*). CYPs and acetyltransferases within the top 100 were selected for initial screening via *Agrobacterium*-mediated expression in *N. benthamiana* with *Cs*OSC1, *Cs*CYP71CD1 and *Cs*CYP71BQ4.

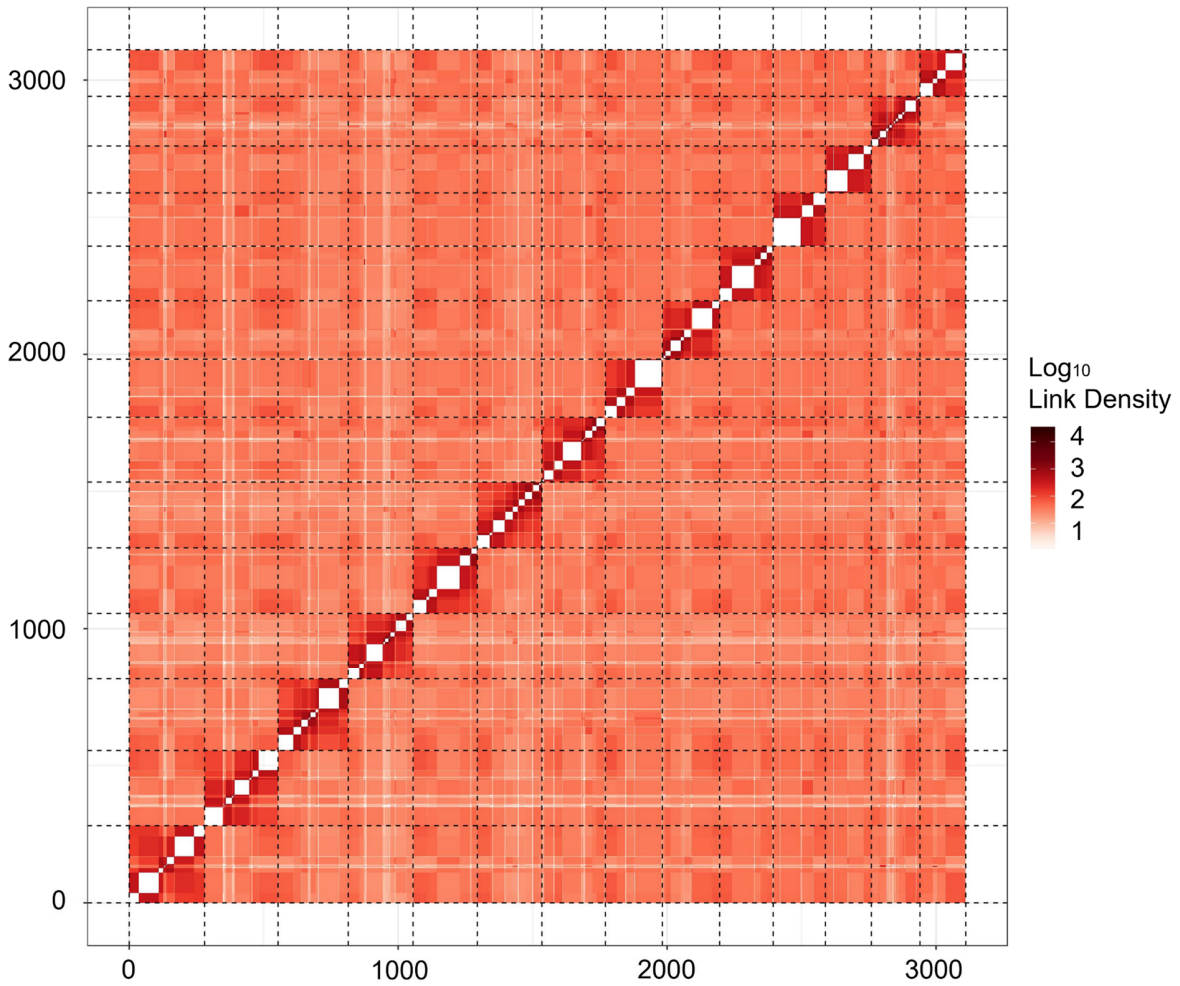



**Fig. S3. Hi-C post-scaffolding heatmap of *M. azedarach* genome.**
Analysis and generation of heatmap was performed by Phase Genomics. The genome was divided into 3,000 bins (length = 75,470 bp) for this analysis. The density of Hi-C links is plotted (red). Links between the same contig are not shown (white). White boxes therefore indicate draft assembly contigs.

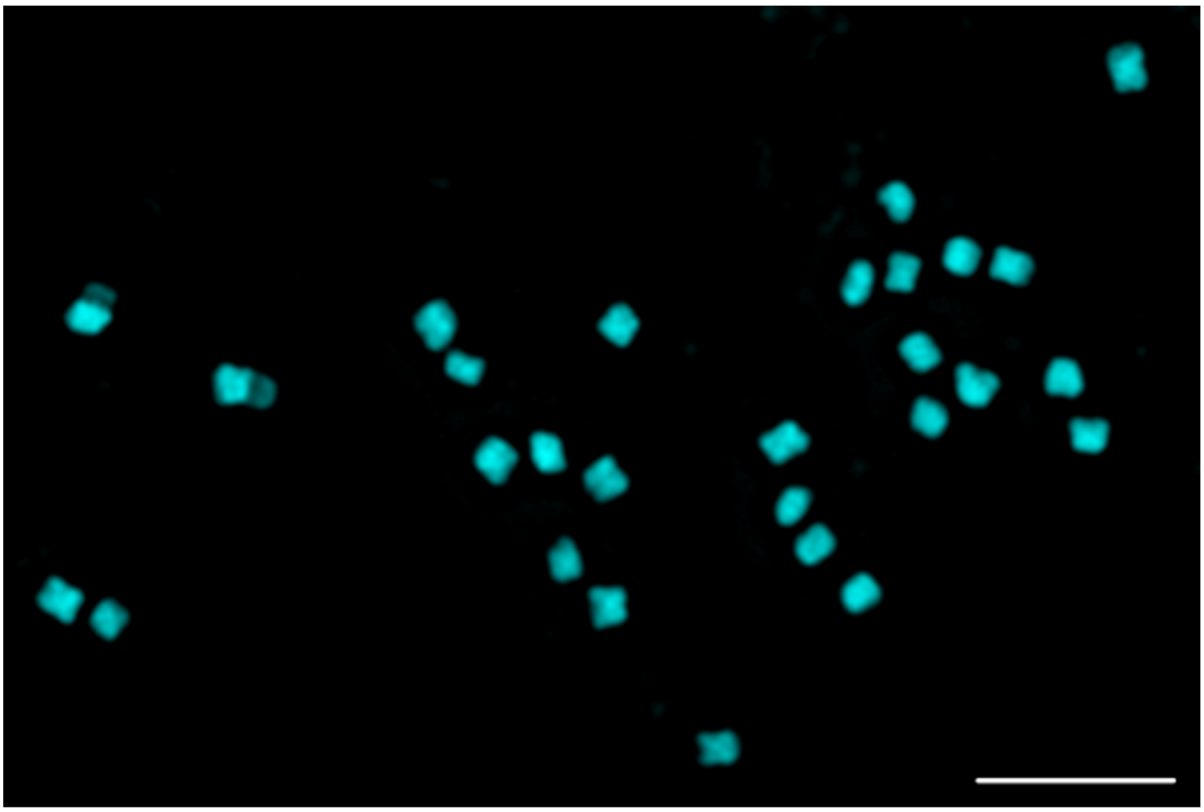


**Fig. S4. Karyotyping of *M. azedarach.***
Representative image of a mitotic metaphase spread of *M. azedarach* (individual '11') showing 28 chromosomes (2n=28). Chromosomes were counterstained with DAPI. Scale bar = 5 µm.

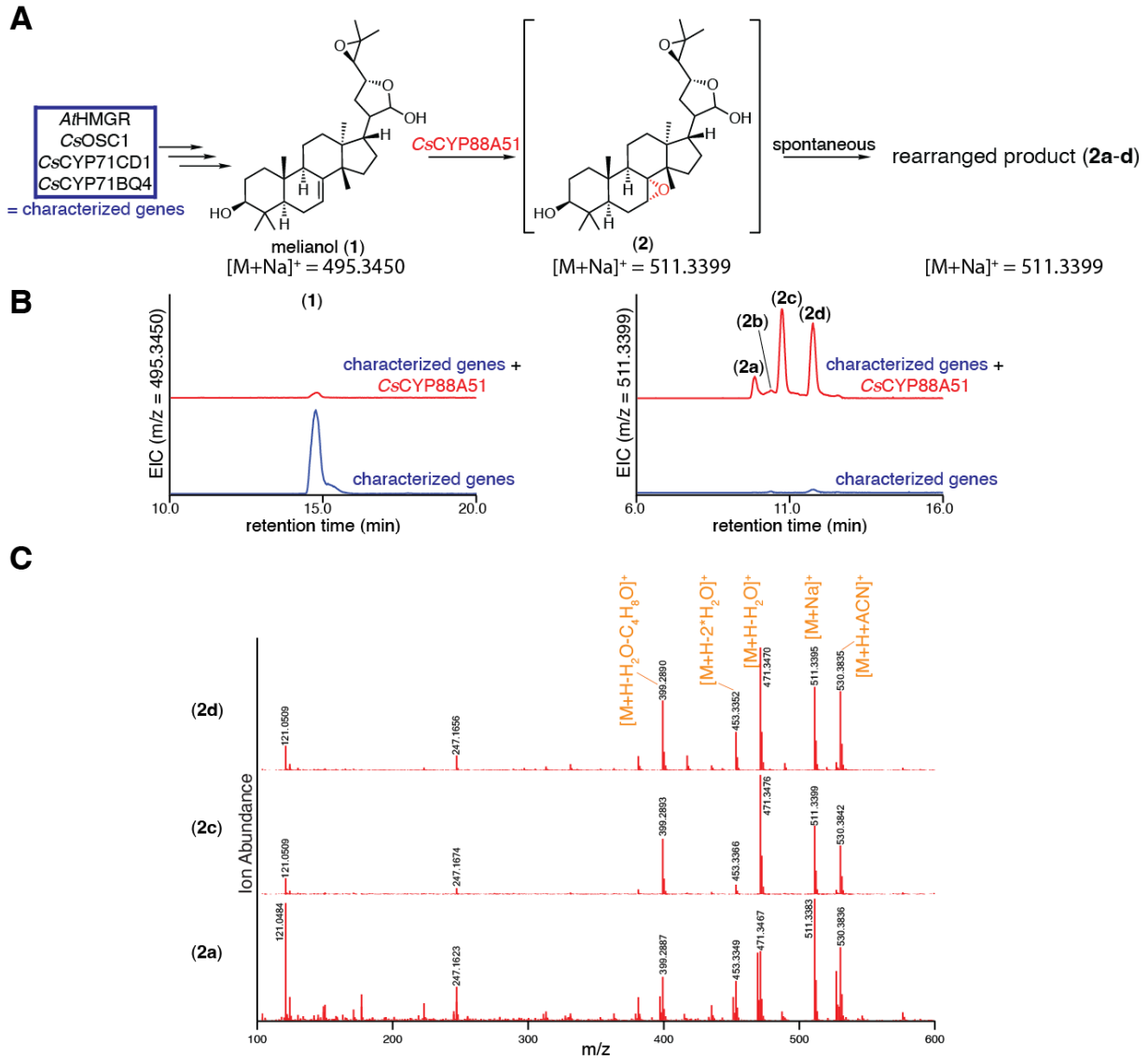



**Fig. S5. Characterization of *Cs*CYP88A51.**

(A) Predicted function of *Cs*CYP88A51 in converting (**1**) to an unstable epoxide intermediate (**2**), which spontaneously rearranges into uncharacterized products (**2a-d**), which all have the same mass as melianol (**1**) with a single oxidation. (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*CYP88A51 (red). EICs are displayed for *m/z* of 495.3450 (calculated mass for (**1**) $[M+Na]^+$) or 511.3399 (calculated mass for (**2**) $[M+Na]^+$). (C) Mass spectra of (**2a**), (**2c**) and (**2d**) in panel B with major adducts and fragments labeled. Note that $[M+Na]^+$ doesn't fragment well in MSMS and the parent peak $[M+H]^+$ is too low to be useful for MSMS analysis. Representative EICs and mass spectra are displayed for experiments of n=6.

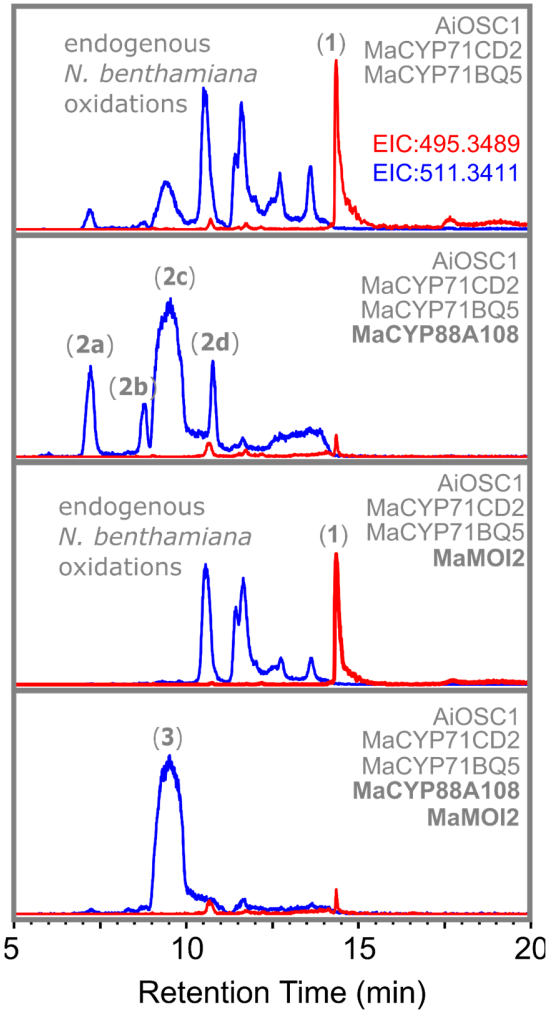



**Fig. S6. Individual activity of *Ma*CYP88A108 and *Ma*MOI2.**
Extracted ion chromatograms (EICs) for extracts of agro-infiltrated *N. benthamiana* leaves expressing melianol biosynthetic genes (*AiOSC1*, *MaCYP71CD2* and *MaCYP71BQ5*), with and without *MaCYP88A108* and *MaMOI2*. EICs displayed are for the *m/z* of [melianol (**1**)+Na]$^+$=495.3489 (red, calculated mass) and [*apo*-melianol (**3**)+Na]$^+$=511.3411 (blue, calculated mass). Alternate re-arrangment products (**2a-d**) with the same mass as *apo*-melianol are labeled in addition to melianol (**1**) and *apo*-melianol (**3**). For these LCMS traces, analysis was performed using an UHPLC-IT-TOF (Shimadzu) instrument following a method and methanol gradient previously described for the analysis of protolimonoids (*20*). Further characterization of *Ma*CYP88A108 and *Ma*MOI2 (being expressed together) using a Q-TOF instrument (Agilent) is available (fig. S7).

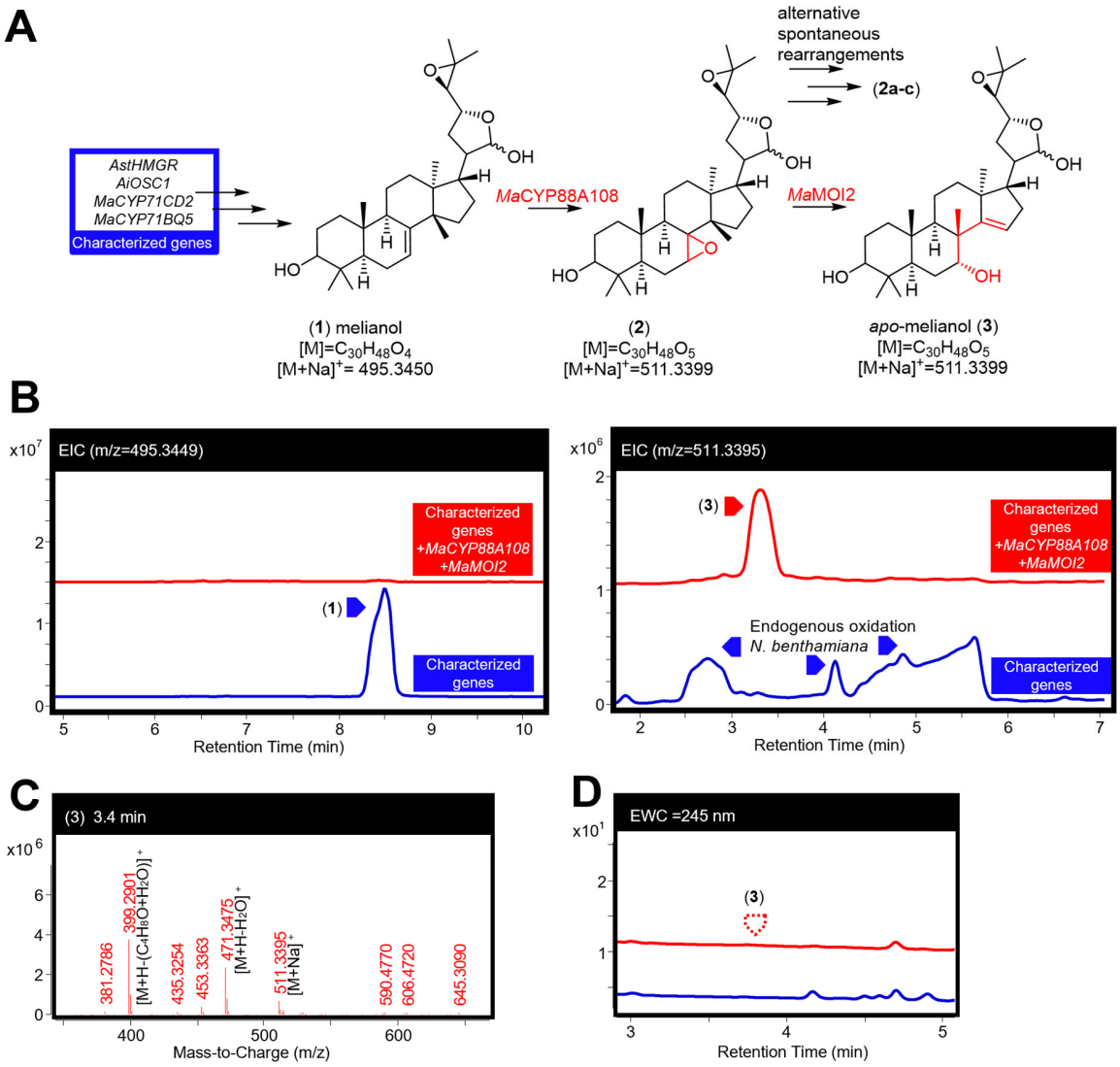



**Fig. S7. Characterization of *Ma*CYP88A108 and *Ma*MOI2.**
(A) Function of *Ma*CYP88A108 and *Ma*MOI2 in converting melianol (**1**) to the epimeric mixture *apo*-melianol (**3**), confirmed by NMR (table S3). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *MaCYP88A108* and *MaMOI2* (red). The EICs are displayed for *m/z*of 495.3449 (observed mass for [(**1**)+Na]$^+$) and 511.3395 (observed mass for [(**3**)+Na]$^+$). (C) Mass spectrum of (**3**) being heterologously produced in *N. benthamiana*. The main observed adduct ([M+Na]$^+$) and fragments (including loss of water [M+H-$H_2O$]$^+$, and loss of water and four-carbon epoxide containing fragment [M+H-($H_2O$+$C_4H_8O$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4nm) for the extracts displayed in panel B. Due to the lack of an enone system in (**3**) no UV peak is observed. Representative traces and spectra are displayed (n=6). Traces showing individual activity of *Ma*CYP88A108 and *Ma*MOI2 are available (fig. S6).

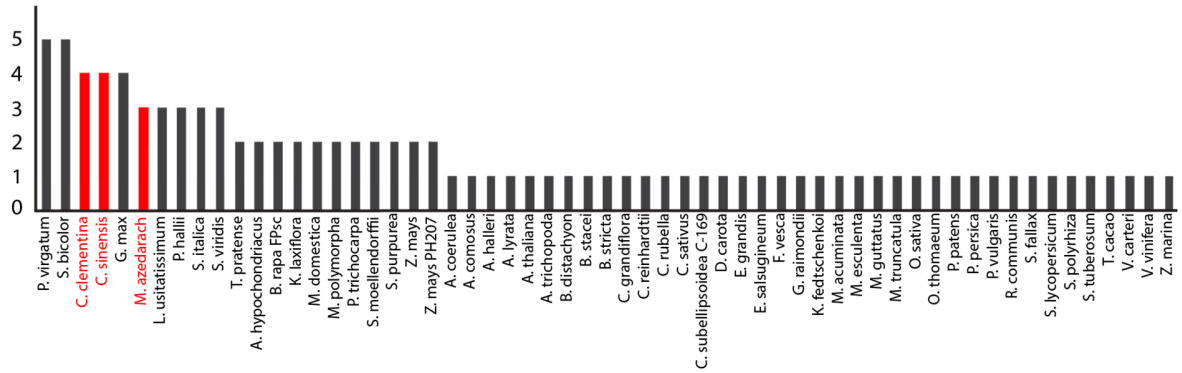



**Fig. S8. Histogram of the number of sterol isomerase genes present in high-quality plant genomes.**

Plant genomes from high-quality annotated genomes were downloaded from Phytozome (*75*). Sterol isomerases sequences were identified by pFAM assignment to EBP (PF05241).

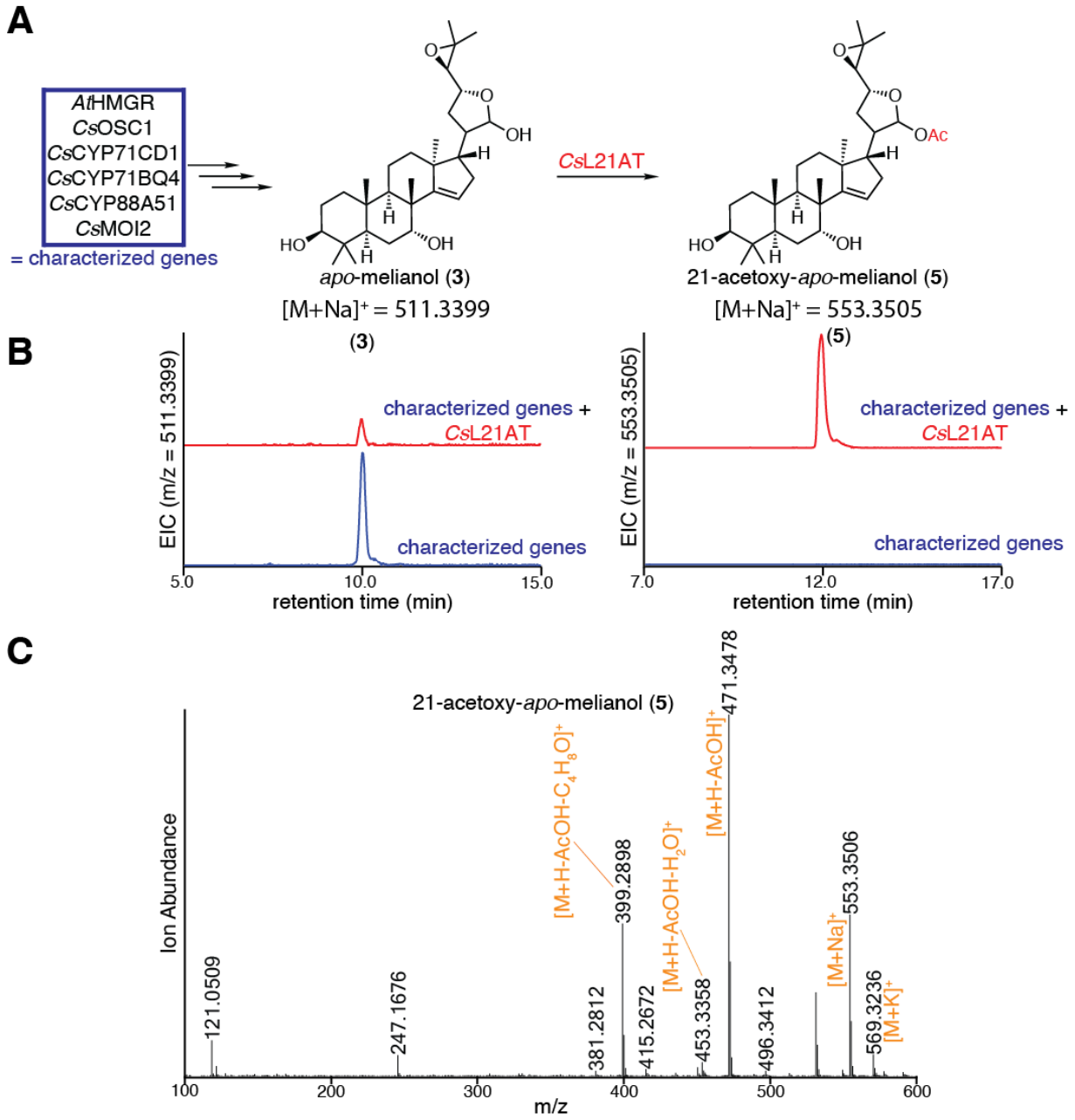



**Fig. S9. Characterization of *Cs*L21AT.**

(A) Predicted function of *Cs*L21AT in converting *apo*-melianol (**3**) to 21-acetoxyl-*apo*-melianol (**5**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*L21AT (red). EICs are displayed for *m/z* of 511.3399 (calculated mass for (**3**) $[M+Na]^+$) and 553.3505 (calculated mass for (**5**) $[M+Na]^+$). **(C)** Mass spectrum of (**5**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Note that $[M+Na]^+$ doesn't fragment well in MSMS and the parent peak $[M+H]^+$ is too low to be useful for MSMS analysis. Representative EICs and mass spectrum are displayed (n=6).

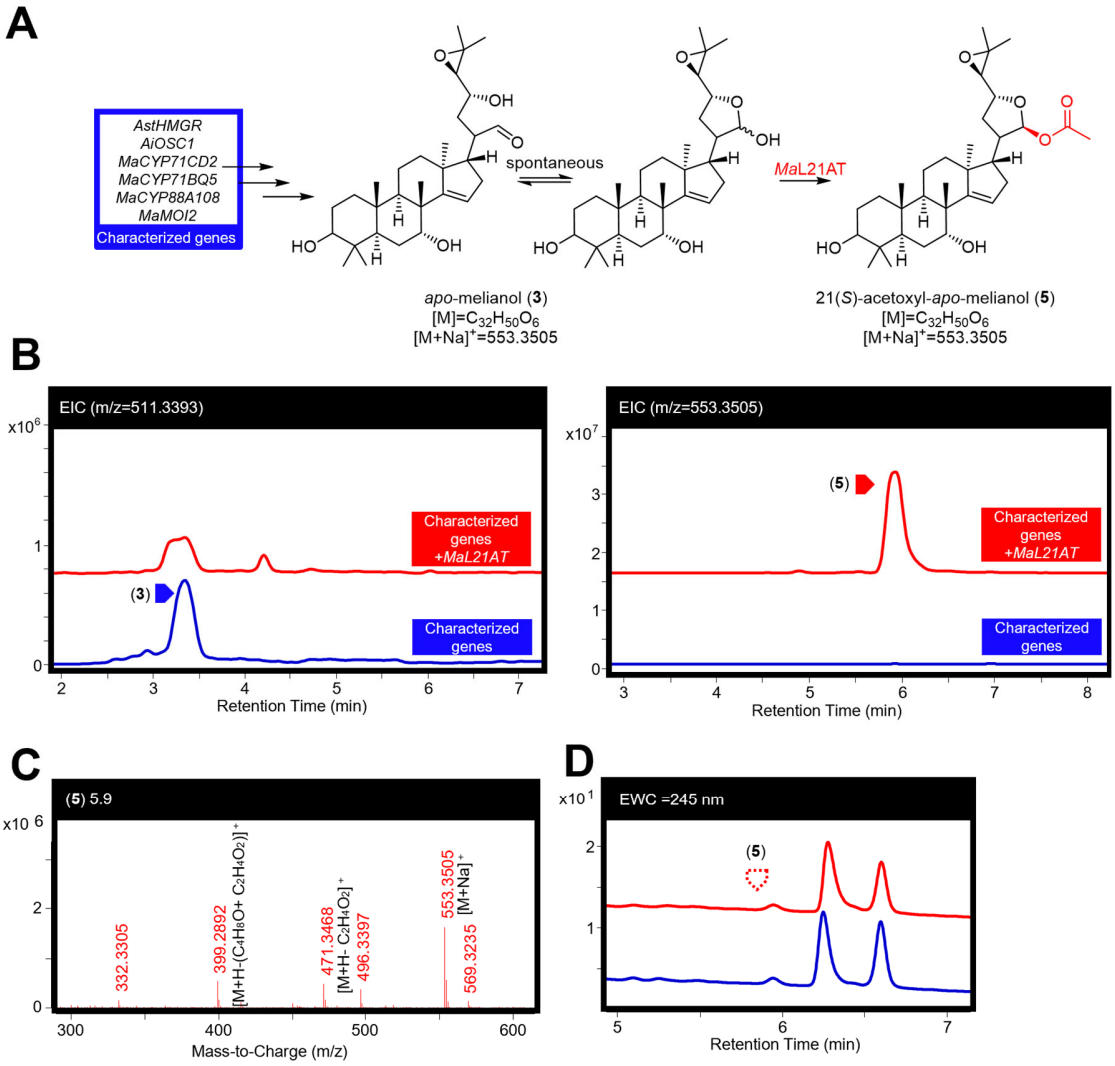



**Fig. S10. Characterization of *Ma*L21AT.**

(A) Function of *Ma*L21AT in producing 21-acetoxyl-*apo*-melianol (**5**) (confirmed by NMR of later product (**6**) (fig. S15, table S6 to S7)) from *apo*-melianol (**3**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Ma*L21AT (red). The EICs are displayed for *m/z* of 511.3393 (observed mass for [(**3**)+Na]$^+$) and 553.3505 (observed mass for [(**5**)+Na]$^+$). (C) Mass spectrum for (**5**) being heterologously produced in *N. benthamiana.* The main observed adduct ([M+Na]$^+$) and fragments (including loss of acetic acid [M+H-$C_2H_4O_2$]$^+$ and loss the four-carbon epoxide containing fragment and acetic acid [M+H-($C_4H_8O$+$C_2H_4O_2$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4nm) for the extracts displayed in panel B. Due to the lack of an enone system in (**5**) no UV peak is observed. Representative traces and spectrum are displayed (n=6).

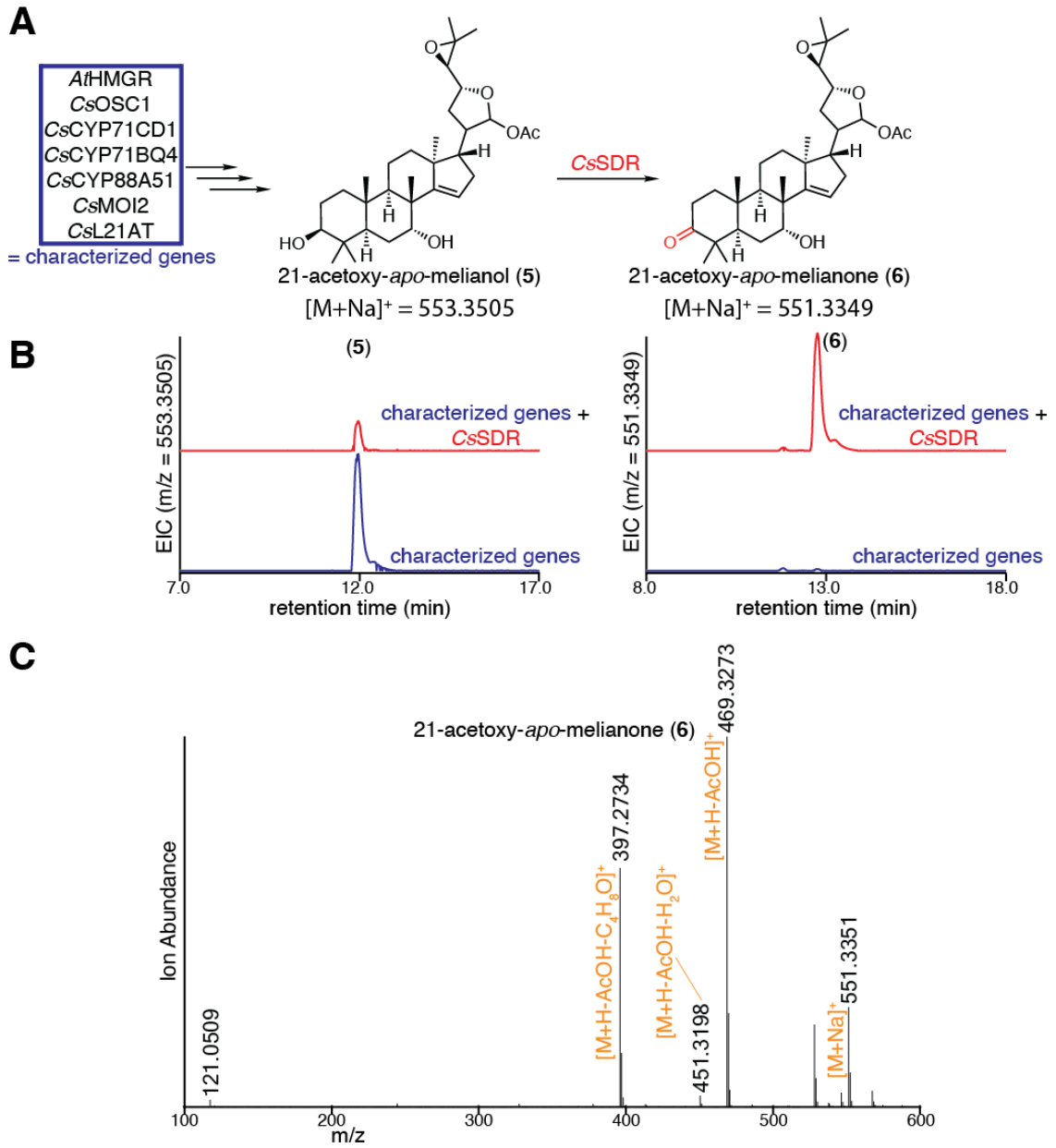



**Fig. S11. Characterization of *Cs*SDR.**

(A) Predicted function of *Cs*SDR in converting 21-acetoxyl-*apo*-melianol (**5**) to 21-acetoxyl-*apo*-melianone (**6**). (B) EICs for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*SDR (red). EICs are displayed for *m/z* of 553.3505 (calculated mass for (**5**) $[M+Na]^+$) or 551.3349 (calculated mass for (**6**) $[M+Na]^+$). (C) Mass spectrum of (**6**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Note that $[M+Na]^+$ doesn't fragment well in MSMS and the parent peak $[M+H]^+$ is too low to be useful for MSMS analysis. Representative EICs and mass spectrum are displayed (n=6).

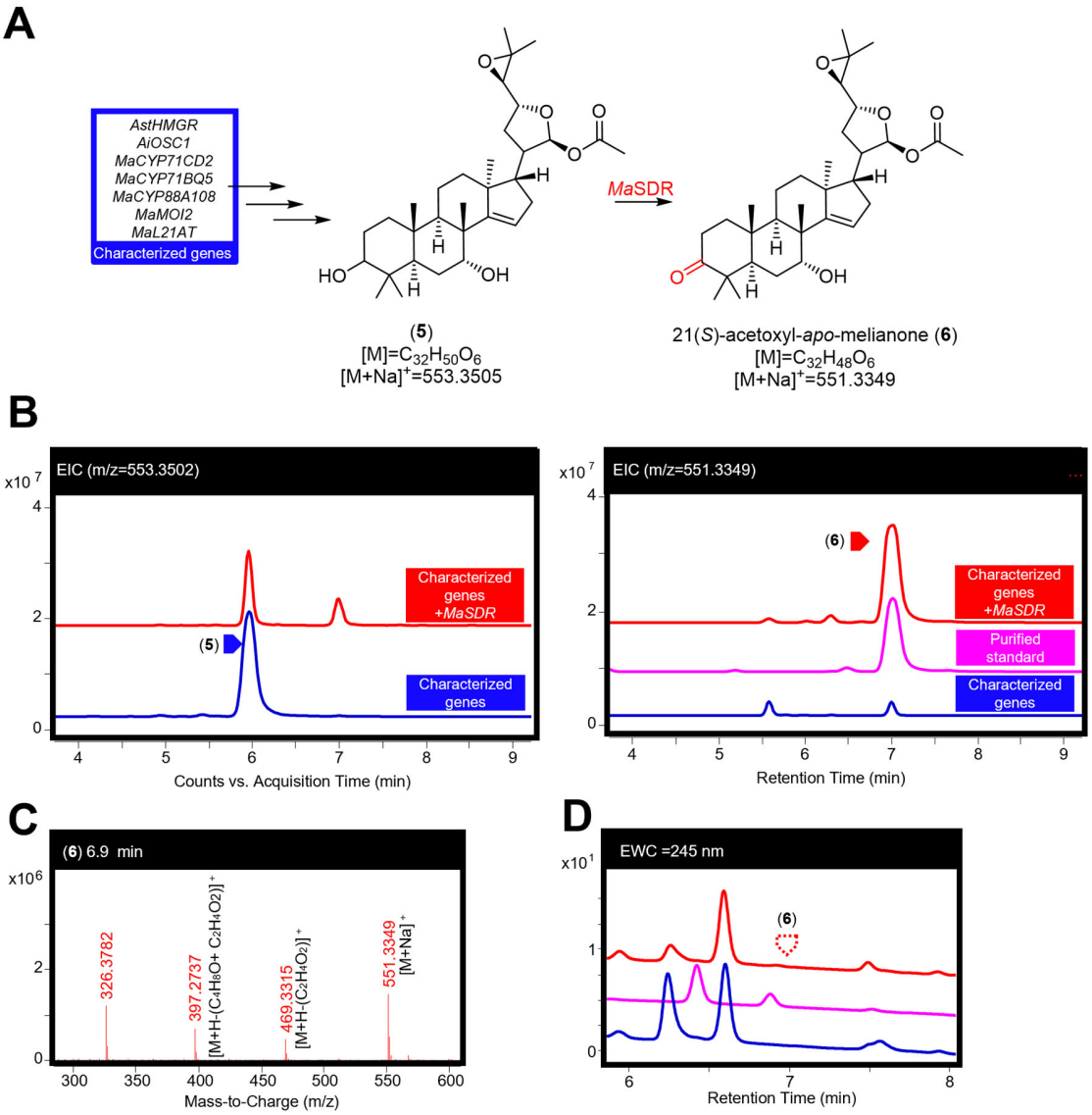



**Fig. S12. Characterization of *Ma*SDR.**

(A) Function of *Ma*SDR in producing 21(*S*)-acetoxyl-*apo*-melianone (**6**) (confirmed by NMR (fig. S15, table S6 to S7)) from 21-acetoxyl-*apo*-melianol (**5**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *MaSDR* (red), along with a purified standard (pink). The EICs are displayed for *m/z* of 553.3502 (observed mass for [(**5**)+Na]$^+$) and 551.3349 (observed mass [(**6**)+Na]$^+$). (C) Mass spectrum for (**6**) being heterologously produced in *N. benthamiana.* The main observed adduct ([M+Na]$^+$) and fragments (including loss of acetic acid [M+H-$C_2H_4O_2$]$^+$ and loss the four-carbon epoxide containing fragment and acetic acid [M+H-($C_4H_8O$+$C_2H_4O_2$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4nm) for extracts displayed in panel B. Due to the lack of an enone system in (**6**) no UV peak is observed. Traces of purified standards have been scaled for comparative purposes. Representative traces and spectrum are displayed (n=6).

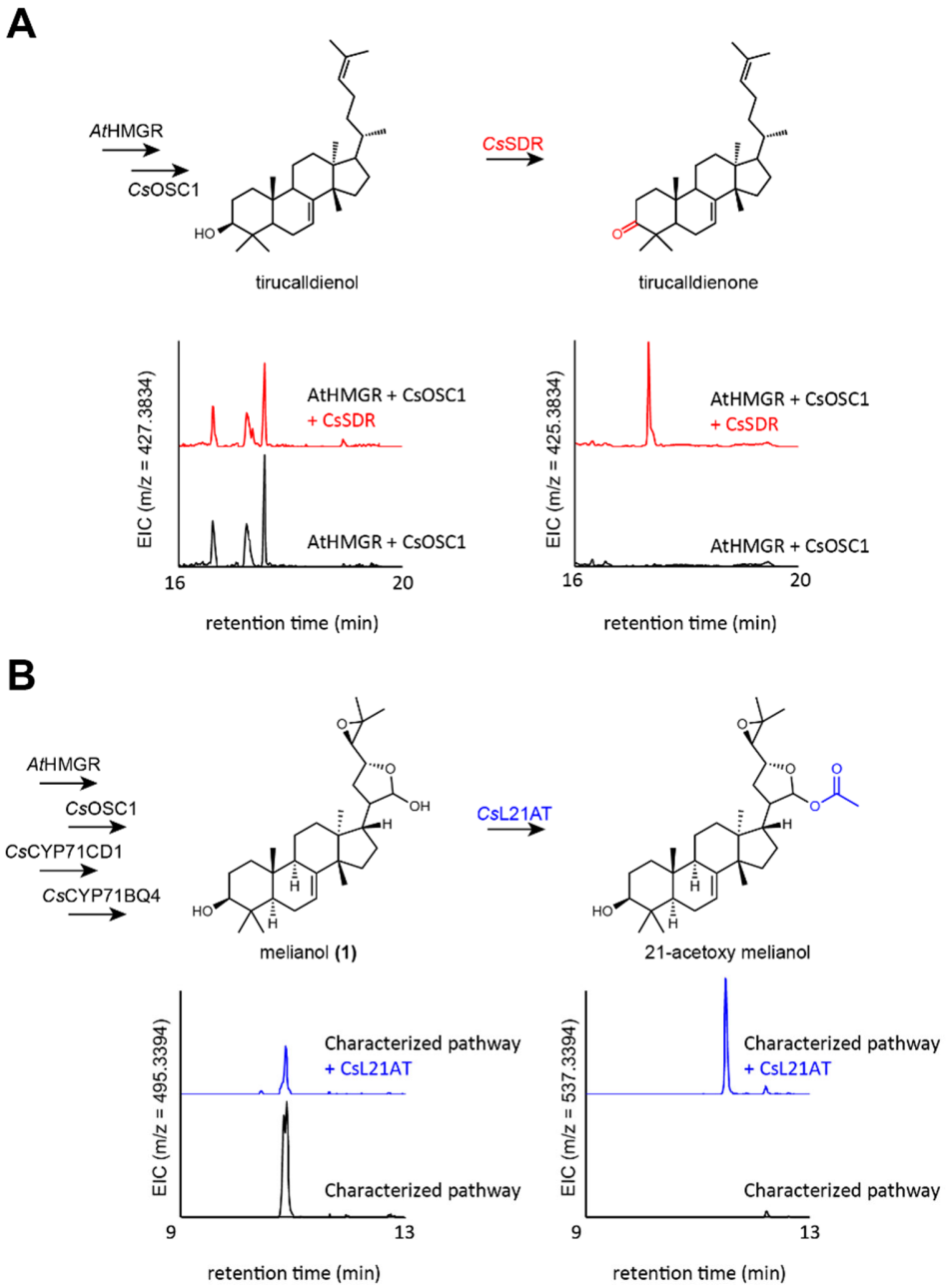



**Fig. S13. Substrate promiscuity of *Cs*L21AT and *Cs*SDR.**
Annotated extracted ion chromatograms (EICs) for extracts of agro-infiltrated *N. benthamiana* demonstrating the ability of *Cs*SDR and *Cs*L21AT to act on alternative scaffolds to *apo*-melianol (**3**), tirucalla-7,24-dien-3ß-ol and melianol (**1**) respectively. (A) EICs of *N. benthamiana* extracts infiltrated with *At*HMGR and *Cs*OSC1 alone (black) or with *Cs*SDR (red). EICs are displayed for observed *m/z* of tirucalla-7,24-dien-3ß-ol $[M+H]^+$ and tirucalla-7,24-dien-3-one $[M+H]^+$. (B) EICs of *N. benthamiana* extracts infiltrated with *At*HMGR, *Cs*OSC1, *Cs*CYP71CD1 and *Cs*CYP71BQ4 alone (black) or with *Cs*L21AT (blue). EICs are displayed for observed masses of melianol $[M+Na]^+$ and 21-acetoxyl-melianol $[M+Na]^+$. Representative EICs are displayed (n=3).

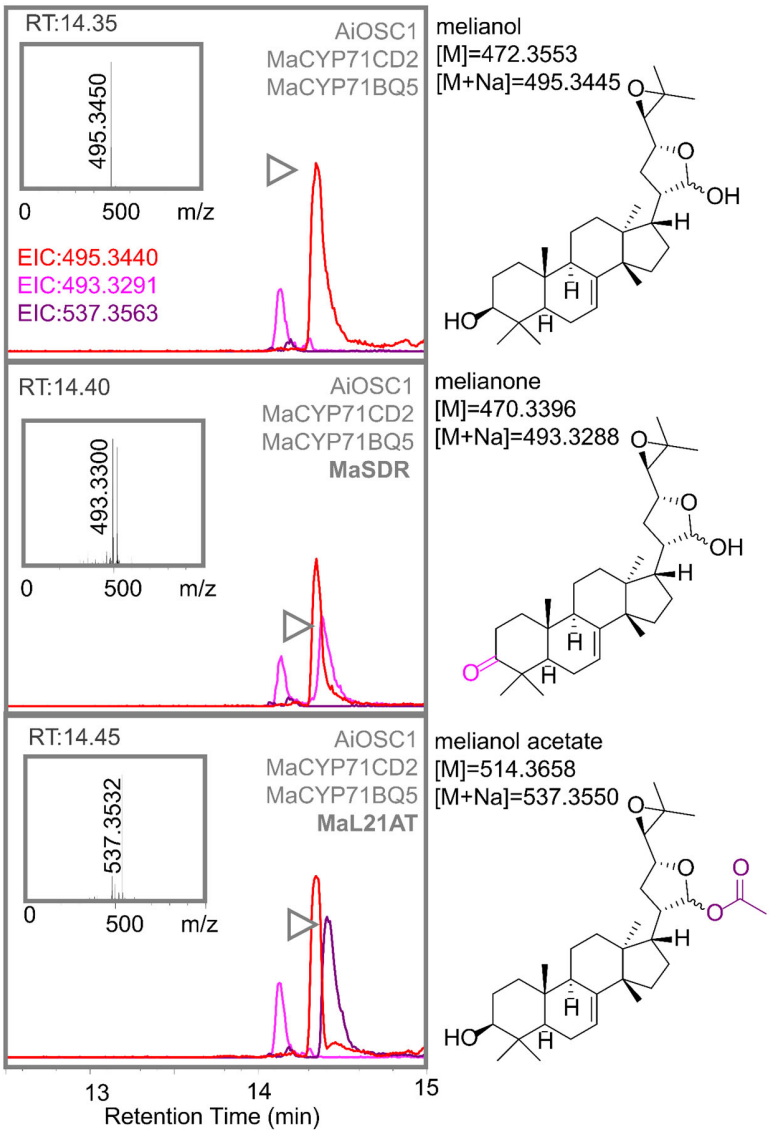



**Fig. S14. Substrate promiscuity of *Ma*SDR and *Ma*L21AT.**
Annotated UHPLC-IT-TOF generated extracted ion chromatograms (EICs) of methanol extracts of agro-infiltrated *N. benthamiana* leaves expressing *Ma*SDR and *Ma*L21AT in combination with melianol biosynthetic enzymes (*Ai*OSC1, *Ma*CYP71CD2 and *Ma*CYP71BQ5), demonstrating the ability of both *Ma*SDR and *Ma*L21AT to act on melianol (**1**) in addition to *apo*-melianol (**3**) (Fig. 4A). EICs displayed are for the following observed adducts: [melianol (**1**)+Na]$^+$=495.3440 (red), [melianone+Na]$^+$=493.3291 (pink) and [melianol acetate +Na]$^+$=537.3563 (purple). Mass spectra of new peaks (highlighted with gray arrows) are shown in the box. UHPLC-IT-TOF was performed using the methanol gradient previously described for the Shimadzu IT-TOF instrument (*20*). Predicted structures of highlighted peaks based on characterized enzymatic functions are shown on the right (with exact mass and calculated sodium adduct). Representative EICs and spectra are displayed (n=3).

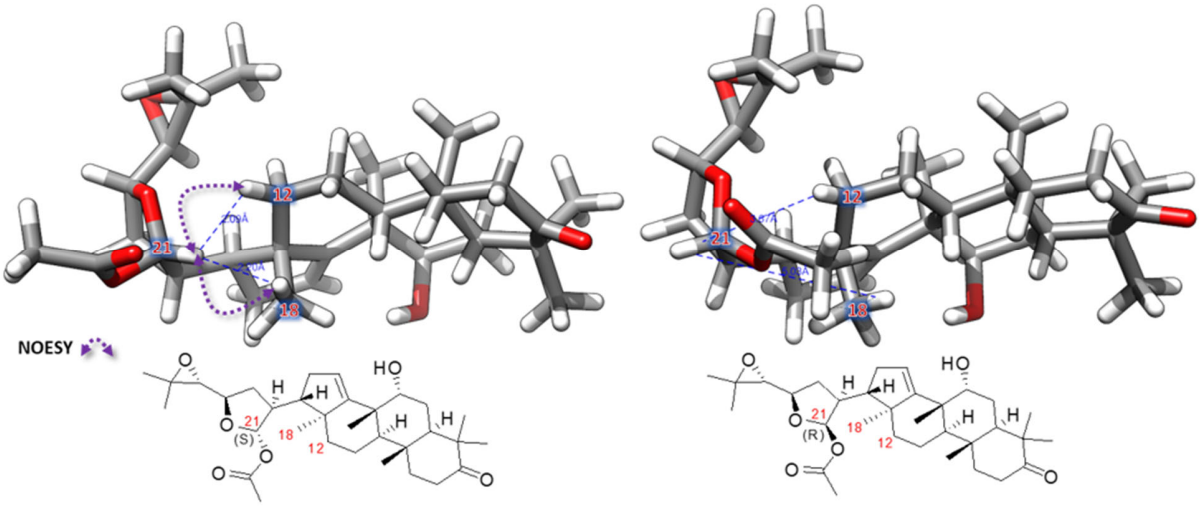



**Fig. S15. 3D models of 21(*S*)-acetoxyl-*apo*-melianone and 21(*R*)-acetoxyl-*apo*-melianone.** The 3D models of 21(*S*)-acetoxyl-*apo*-melianone (left) and 21(*R*)-acetoxyl-*apo*-melianone (right) in combination with the NOEs between C21-H, C18-H3 and C12-H2 observed in 2D NOESY experiments for (**6**), are consistent with the 21(*S*) assignment of (**6**). 3D models have been geometry optimized by molecular dynamics (forcefield: MMFF94, number of steps: 500, algorithm: steepest descent and convergence: 10e-7, run by AvogadroV 1.1.1). Complete $^{13}$C and $^{1}$H δ assignment is shown in table S6.

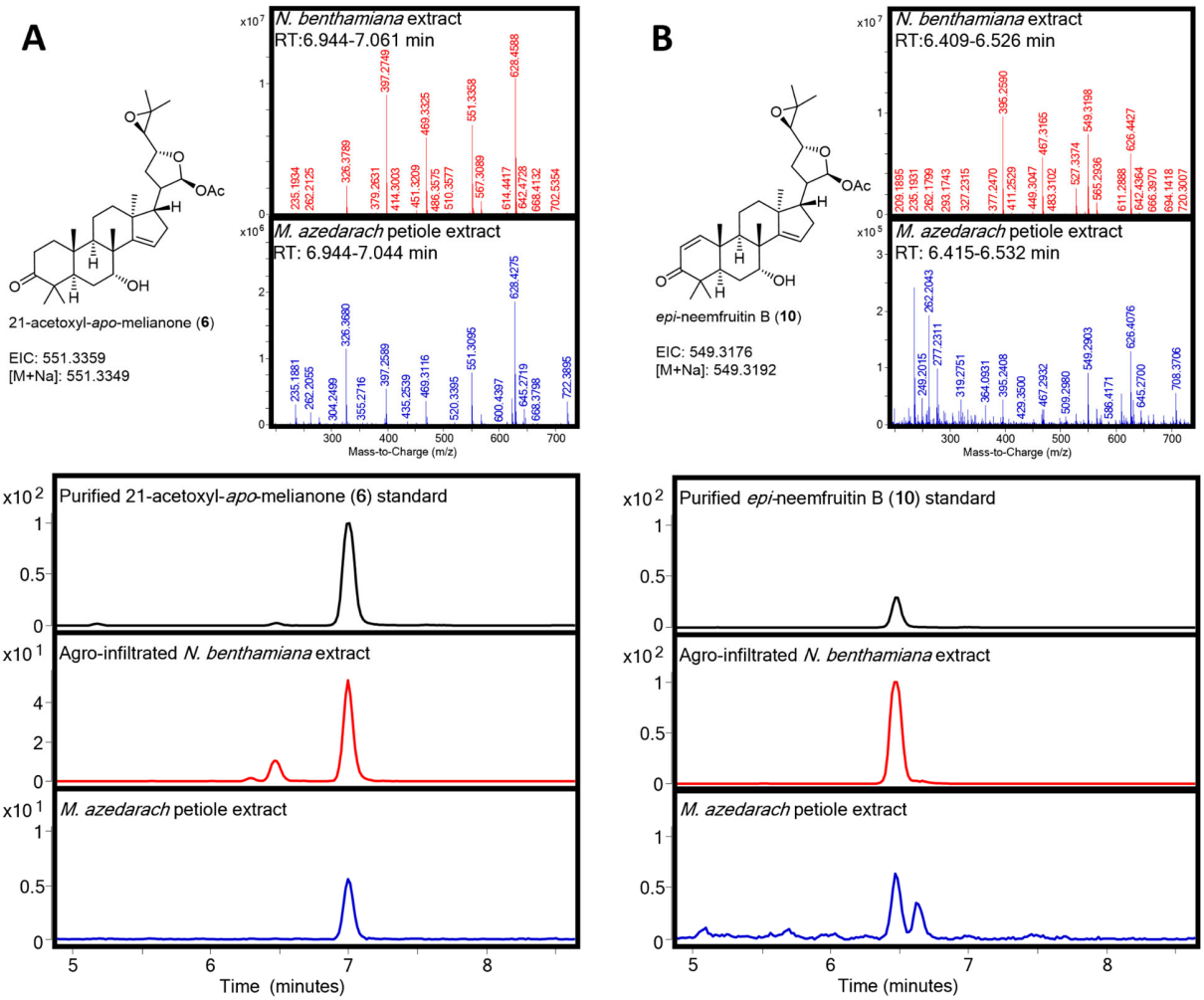



**Fig. S16. Detection of 21(*S*)-acetoxyl-*apo*-melianone (6) and *epi*-neemfruitin B (10) in *Melia azedarach* samples.**

(A) Structure, mass spectra and extracted ion chromatograms (EICs) comparing extracts of *N. benthamiana* expressing 21-acetoxyl-*apo*-melianone (**6**) biosynthetic enzymes (*Ai*OSC1, *Ma*CYP71CD2, *Ma*CYP71BQ5, *Ma*CYP88A108, *Ma*MOI2, *Ma*L21AT and *Ma*SDR) to extracts from *M. azedarach* petiole tissues (individual '11'). EIC of purified 21(*S*)-acetoxyl-*apo*-melianone (**6**) (table S6) is also displayed. EICs displayed are for *m/z* of 511.3349, the calculated mass of [21-acetoxyl-*apo*-melianone+Na]$^+$. (B) Structure, mass spectra and EICs comparing extracts of *N. benthamiana* expressing *epi*-neemfruitin B (**10**) biosynthetic enzymes (the enzymes described in panel (A) with addition of *Ma*CYP88A164 and *Ma*L1AT) to extracts from *M. azedarach* petiole tissues (individual '11'). EIC of *epi*-neemfruitin B (**10**) is also displayed (table S11). EICs displayed are for *m/z* of 549.3192, the calculated mass of [*epi*-neemfruitin B+Na]$^+$. Representative EICs and spectra are displayed (n=3).

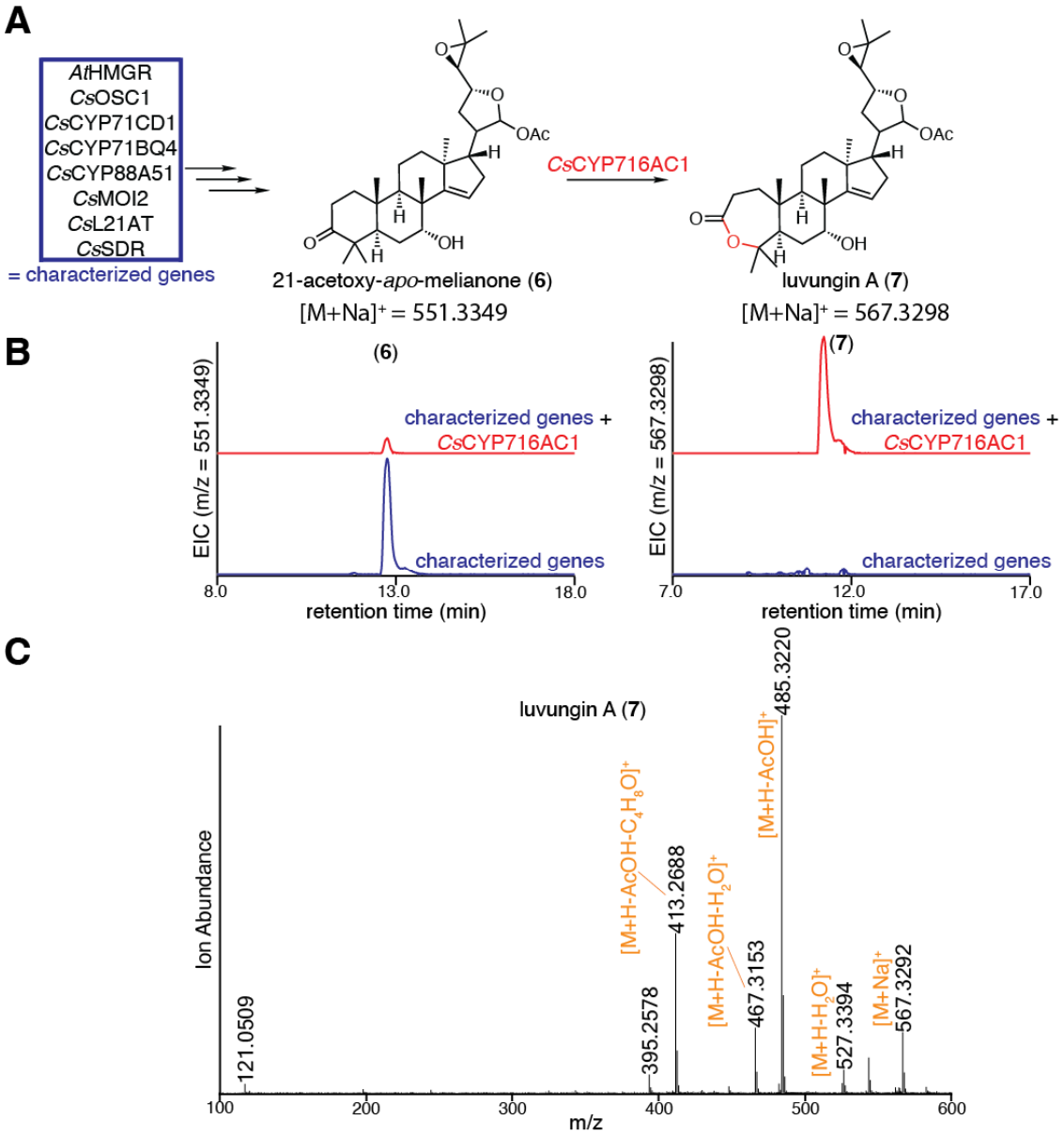



**Fig. S17. Characterization of *Cs*CYP716AC1.**

(A) Predicted function of *Cs*CYP716AC1 in converting 21-acetoxyl-*apo*-melianone (**6**) to luvungin A (**7**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*CYP716AC1 (red). EICs are displayed for *m/z* of 551.3349 (calculated mass for (**6**) $[M+Na]^+$) or 567.3298 (calculated mass for (**7**) $[M+Na]^+$). (C) Mass spectrum of (**7**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Note that $[M+Na]^+$ doesn't fragment well in MSMS and the parent peak $[M+H]^+$ is too low to be useful for MSMS analysis. Representative EICs and mass spectrum are displayed (n=6).

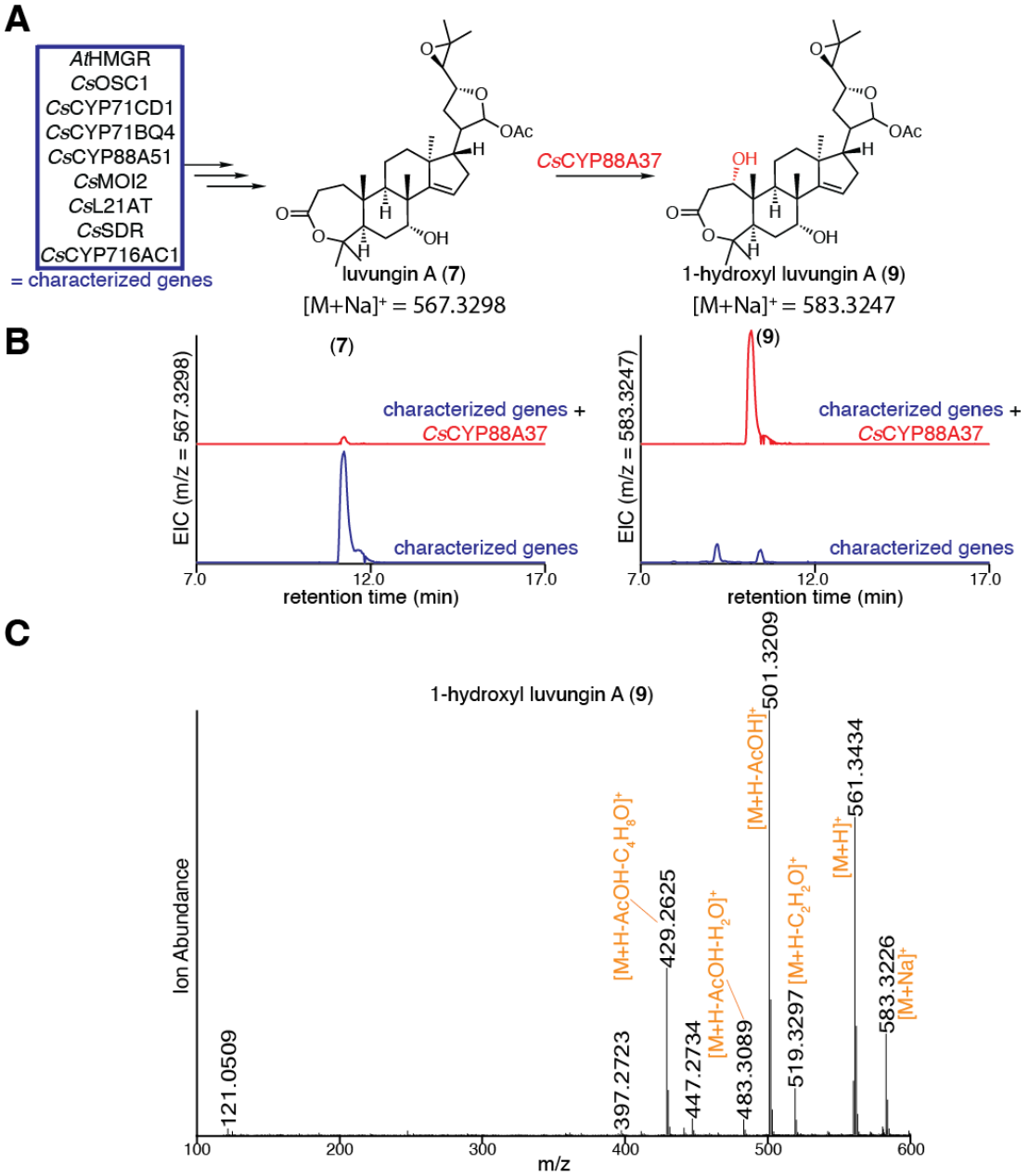



**Fig. S18. Characterization of *Cs*CYP88A37.**

(A) Predicted function of *Cs*CYP88A37 in converting luvungin A (**7**) to 1-hydroxyl luvungin A (**9**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*CYP88A37 (red). EICs are displayed for masses of 567.3298 (calculated mass for (**7**) $[M+Na]^+$) or 583.3247 (calculated mass for (**9**) $[M+Na]^+$). (C) Mass spectrum of (**9**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Proposed formation of the loss of $C_2H_2O$ fragment is shown in fig. S35. Representative EICs and mass spectrum are displayed (n=6).

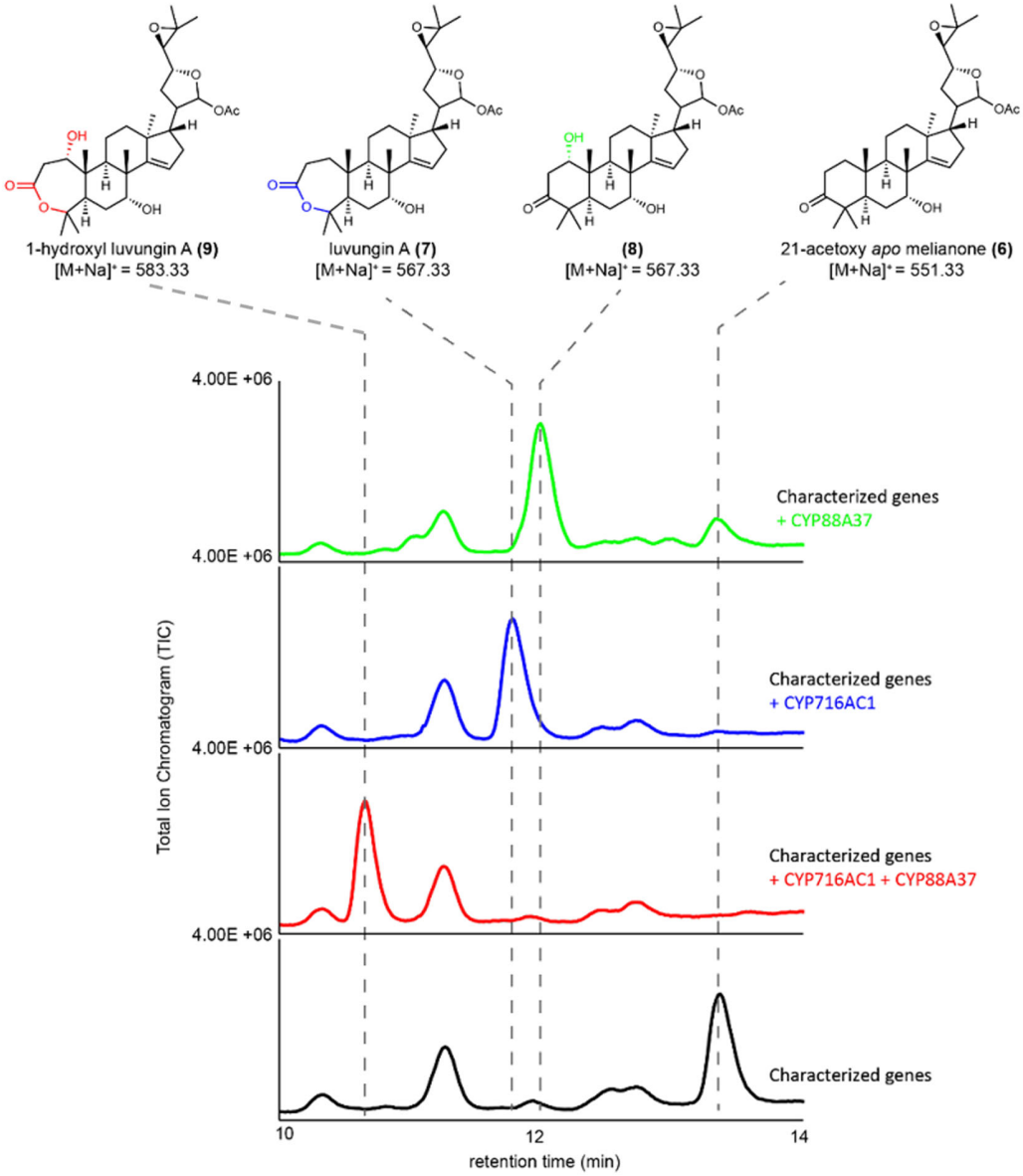



**Fig. S19. Oxidation of 21-acetoxyl-*apo*-melianone (6) by either *Cs*CYP88A37 or *Cs*CYP716AC1.**

Total ion chromatograms (TICs) for extracts of *N. benthamiana* agro-infiltrated with characterized enzymes (*At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2, *Cs*L21AT and *Cs*SDR (black)) in combination with *Cs*CYP88A37 and *Cs*CYP716AC1, either alone (green and blue, respectively), or together (red). *Cs*CYP88A37 or *Cs*CYP716AC1 either act alone on 21-acetoxyl-*apo*-melianone (**6**) to yield (**8**) and (**7**), respectively, or together to yield (**9**). There is incomplete disappearance of (**6**) when *Cs*CYP88A37 is expressed alone. Structures of (**6**), (**7**), and (**9**) are confirmed by NMR while that of (**8**) is proposed based on the characterized function of *Cs*CYP88A37. Representative TICs displayed for the experiments (n=3).

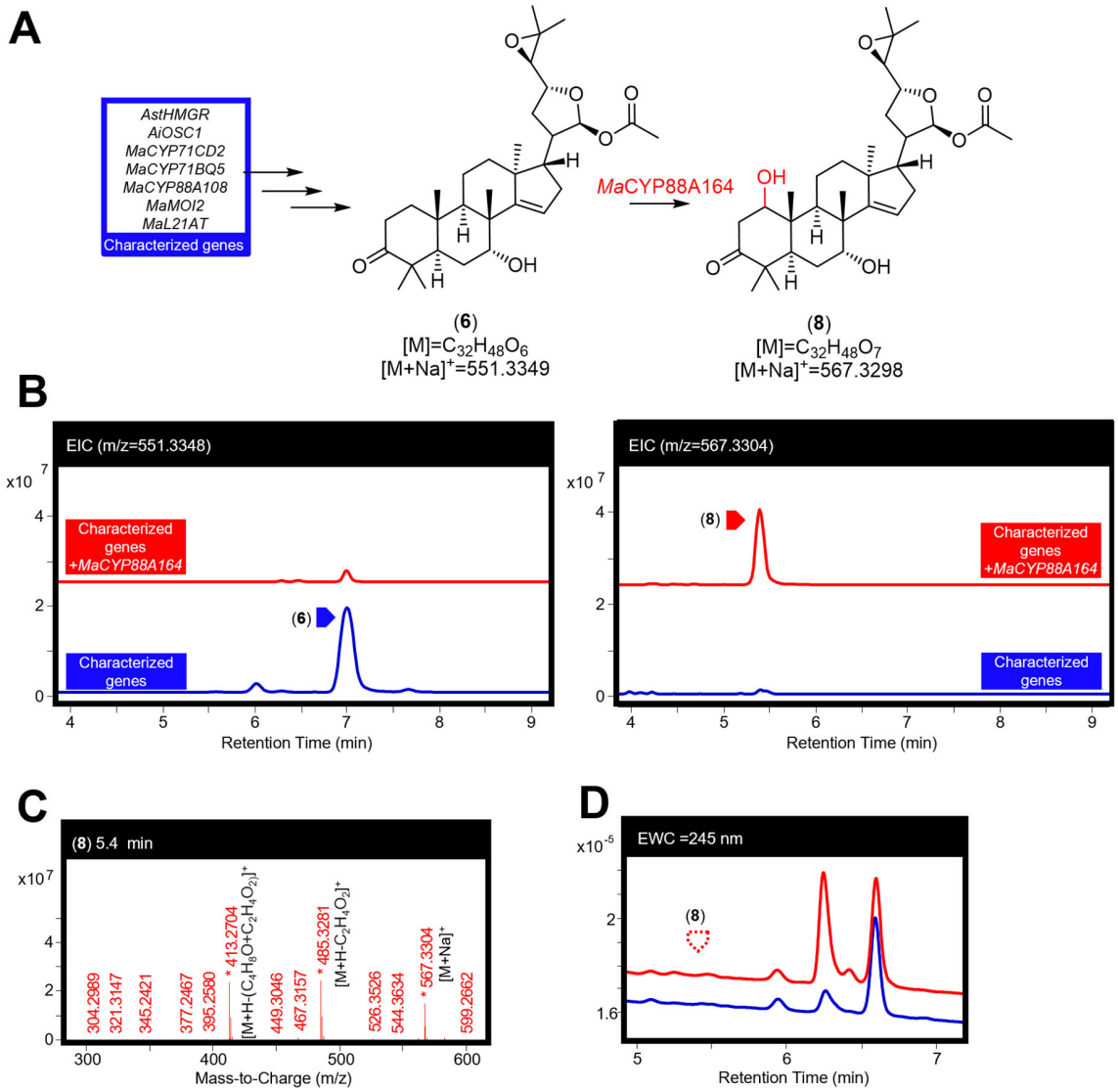



**Fig. S20. Characterization of *Ma*CYP88A164.**
(A) Function of *Ma*CYP88A164 in producing a 1-hydroxyl-21(*S*)-acetoxyl-*apo*-melianone (**8**) (confirmed by NMR of later products (table S11 to table S12)) from 21(*S*)-acetoxyl-*apo*-melianone (**6**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized enzymes (listed in panel A) either alone (blue) or with the addition of *Ma*CYP88A164 (red). The EICs display masses of 551.3348 (observed mass for [(**6**)+Na]$^+$) and 567.3304 (observed mass for [(**8**)+Na]$^+$). (C) Mass spectrum for (**8**) being heterologously produced in *N. benthamiana.* The main observed adduct ([M+Na]$^+$) and fragments (including loss of acetic acid [M+H-$C_2H_4O_2$]$^+$ and loss of the four-carbon epoxide containing fragment and an acetic acid [M+H-($C_4H_8O$+$C_2H_4O_2$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4 nm) for extracts displayed in panel B. Due to the lack of an enone system in (**8**) no UV peak is observed. Representative traces and spectrum are displayed (n=6).

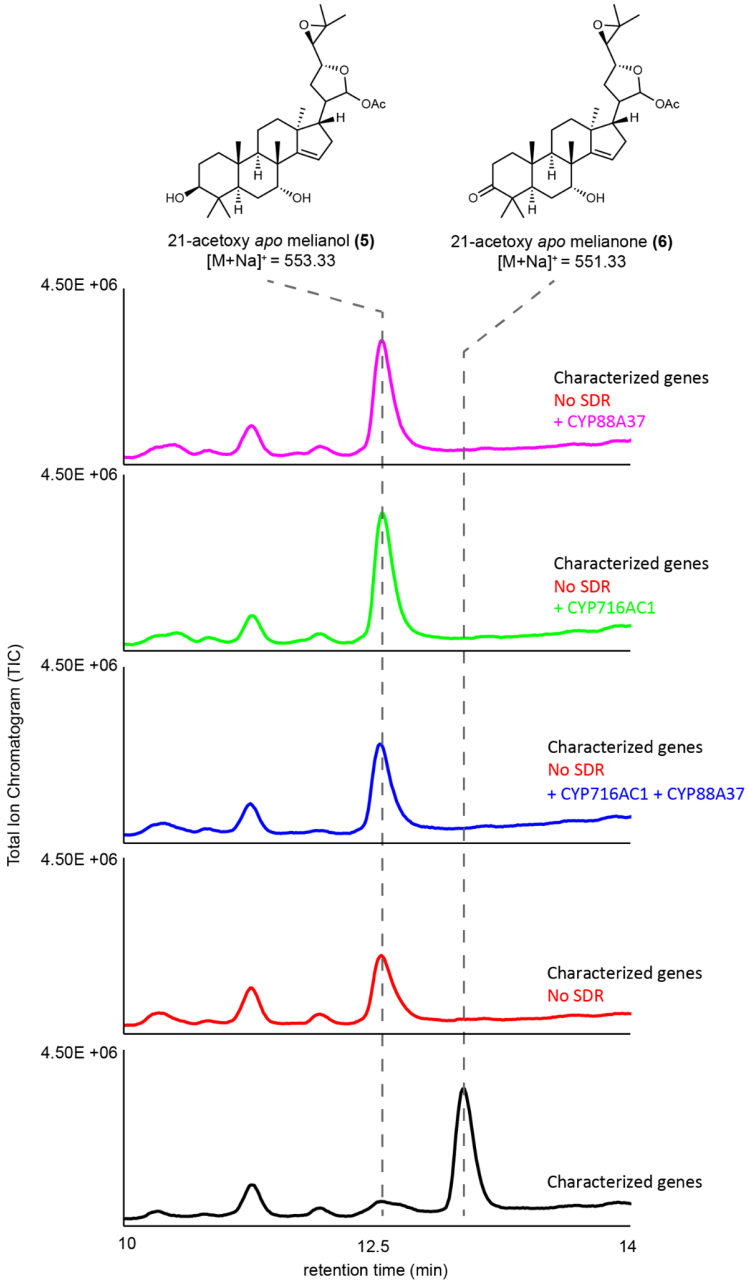



**Fig. S21. Oxidation by *Cs*CYP88A37 or *Cs*CYP716AC1 requires *Cs*SDR.**
Predicted structures and representative total ion chromatograms (TICs). Whilst a clear reduction in (**5**) is observed when *Cs*SDR is co-expressed (converting (**5**) to (**6**)), no conversion is seen by *Cs*CYP88A37 and *Cs*CYP716AC1 in the absence of *Cs*SDR, demonstrating the lack of activity of both CYPs without a C3 ketone substrate. TICs are for extracts of *N. benthamiana* agro-infiltrated with characterized enzymes (*At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2, and *Cs*L21AT) in combination with *CsSDR* (black) or without *Cs*SDR (red), with the addition of *Cs*CYP88A37 or *Cs*CYP716AC1, either alone (pink and green, respectively) or together (blue). Representative TICs are displayed (n=3).

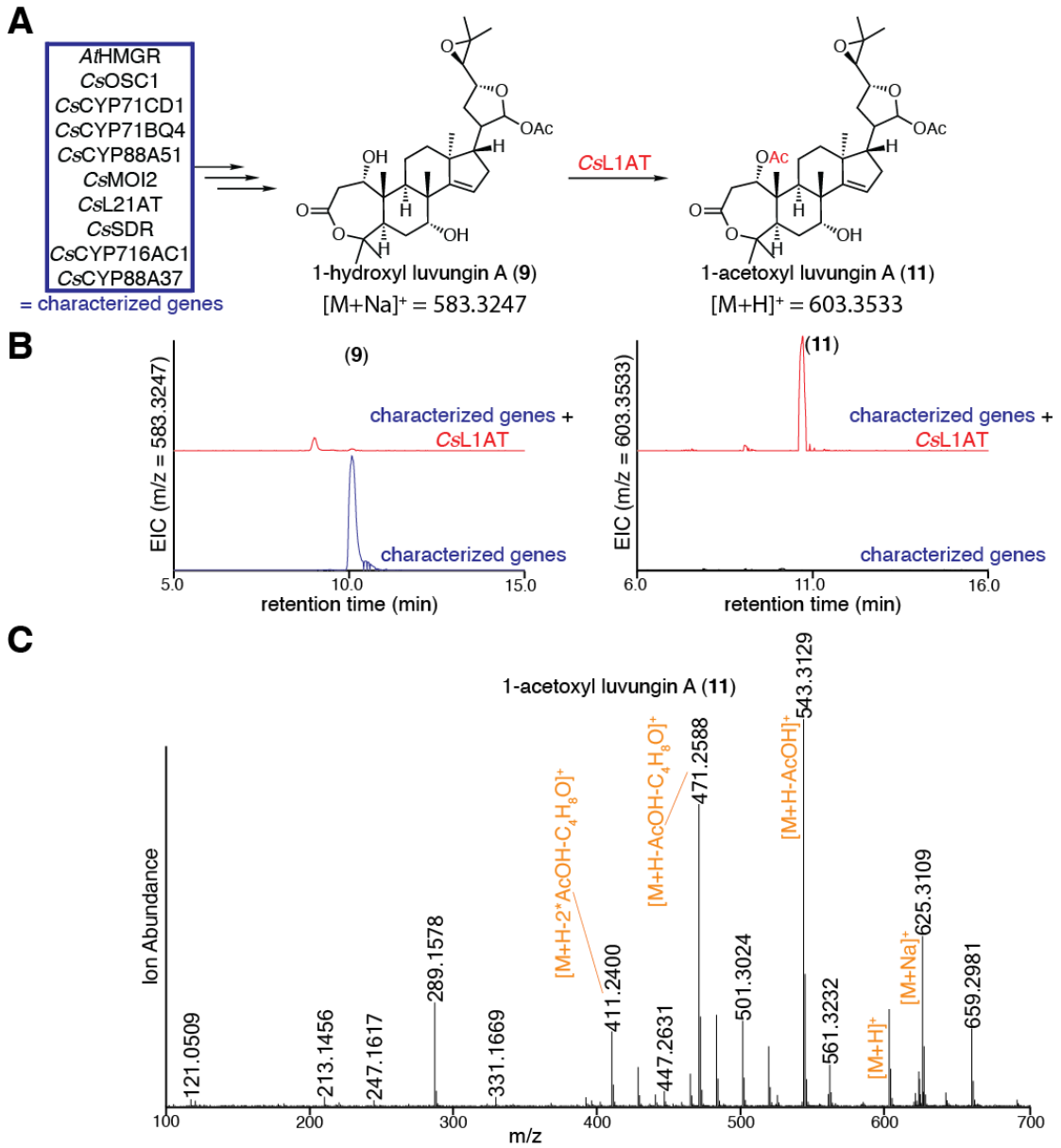



**Fig. S22. Characterization of *Cs*L1AT.**

(A) Predicted function of *Cs*L1AT in converting 1-hydroxyl luvungin A (**9**) to 1-acetoxyl luvungin A (**11**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*L1AT (red). EICs are displayed for masses of 583.3247 (calculated mass for (**9**) $[M+Na]^+$) or 603.3533 (calculated mass for (**11**) $[M+H]^+$). **(C)** Mass spectrum of (**11**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Representative EICs and mass spectrum are displayed (n=6).

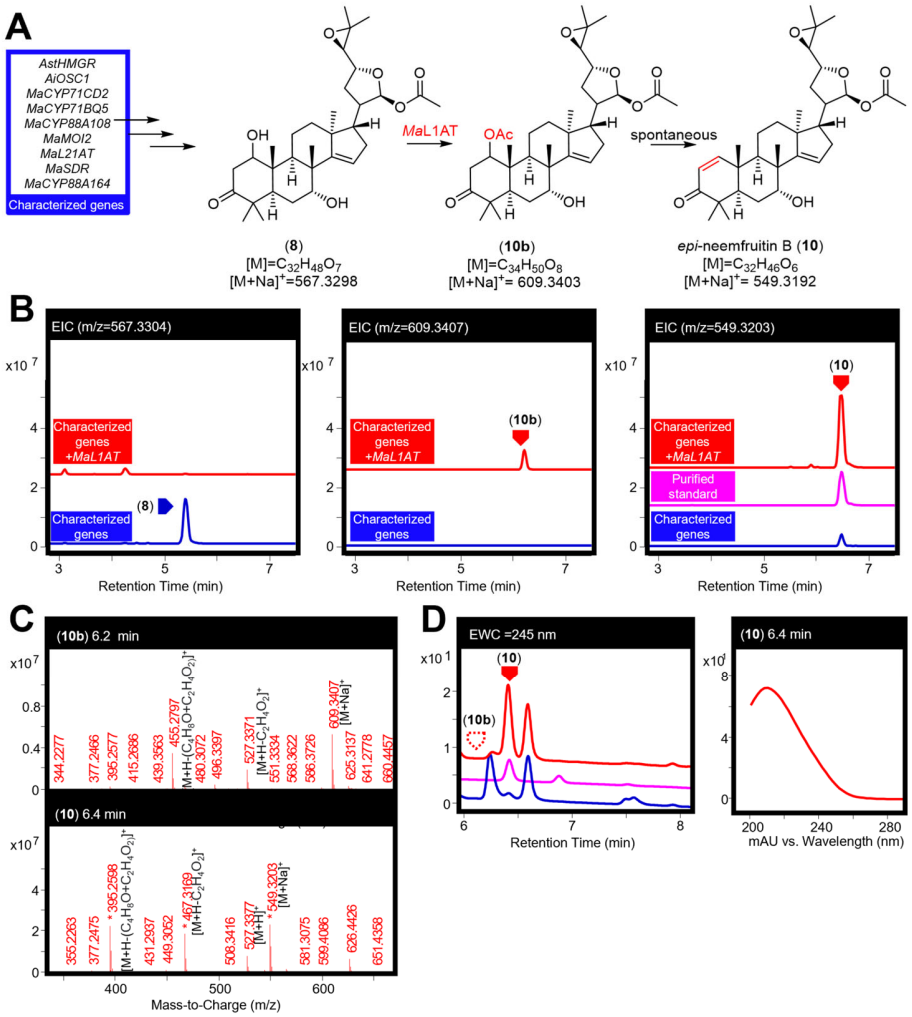



**Fig. S23. Characterisation of *Ma*L1AT.**

(A) Function of *Ma*L1AT in producing *epi*-neemfruitin B (**10**) (confirmed by NMR, table S11 to S12) as a major product, along with 1,21-di-acetoxyl-*apo*-melianone (**10b**), from 1-hydroxyl-21(*S*)-acetoxyl-*apo*-melianone (**8**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue), with *Ma*L1AT (red) or for a purified standard of (**10**) (pink). The EICs display *m/z* of 567.3304 (observed mass for [(**8**)+Na]$^+$), 609.3407 (observed mass for [(**10b**)+Na]$^+$) and 549.3203 (observed mass [(**10**)+Na]$^+$). (C) Mass spectra for (**10b**) and (**10**) being heterologously produced in *N. benthamiana.* The main observed adducts ([M+Na]$^+$ and [M+H]$^+$) and fragments (including loss of acetic acid [M+H-$C_2H_4O_2$]$^+$ and loss the four-carbon epoxide containing fragment and an acetic acid [M+H-($C_4H_8O$+$C_2H_4O_2$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4nm) for extracts displayed in panel B. Due to the lack of an enone system in (**10b**), no UV peak is observed, however (**10**) is UV active and its UV spectrum (mAU) is shown on the right. Standards have been scaled. Representative traces and spectra are displayed (n=6).

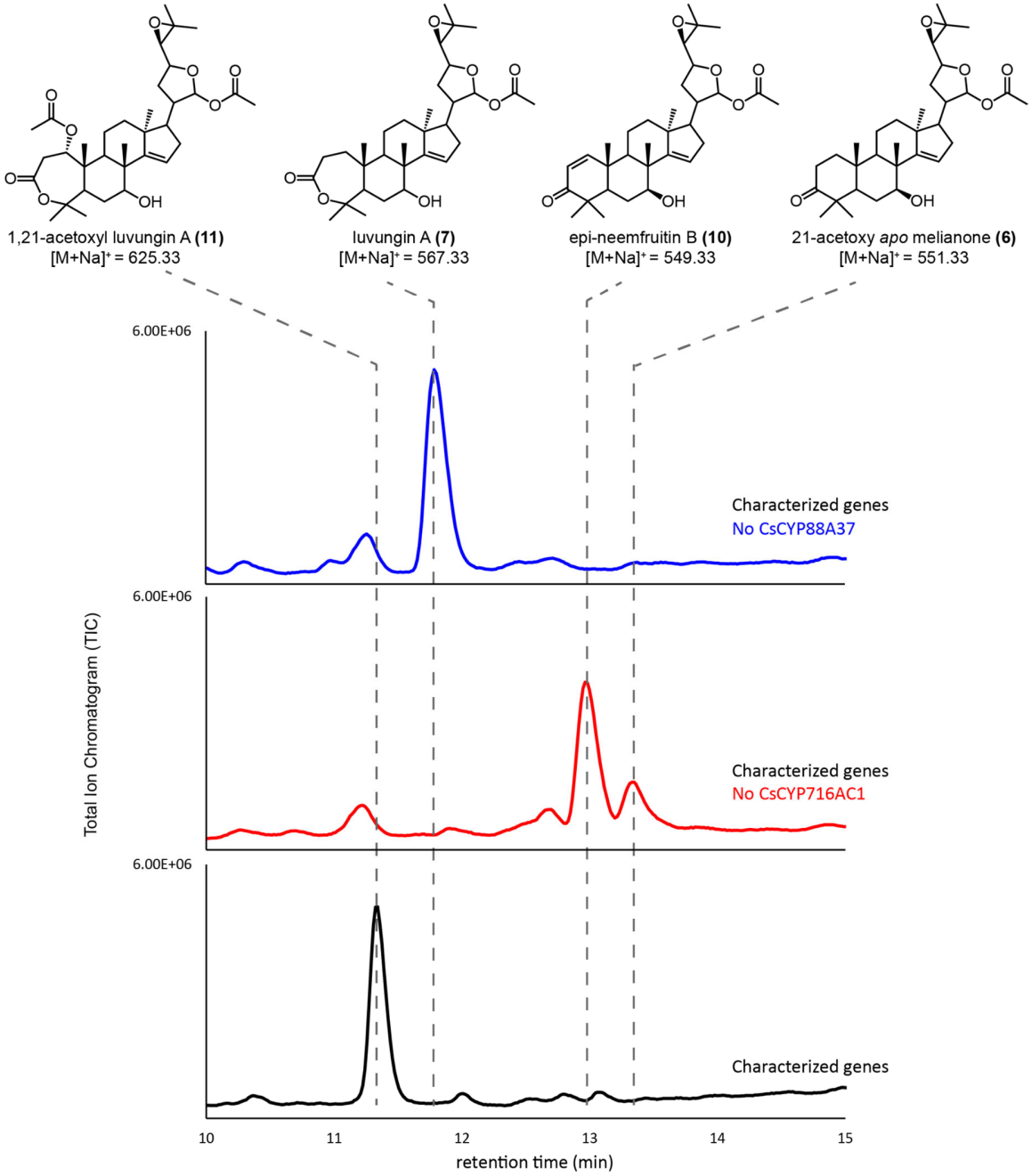



**Fig. S24. Characterization of *Cs*L1AT in the absence of *Cs*CYP716AC1 or *Cs*CYP88A37.** Predicted structures and representative total ion chromatograms (TICs) for extracts of *N. benthamiana* agro-infiltrated with the following enzymes: *At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2, *Cs*L21AT, *Cs*SDR, *Cs*CYP716AC1, *Cs*CYP88A37, and *Cs*L21AT (black), resulting in the production of (**11**). Traces for the same combination of enzymes lacking either *Cs*CYP88A37 (blue) and therefore producing (**7**) or *Cs*CYP716AC1 (red) and therefore producing (**10**) and (**6**) are also shown. This demonstrates that in the absence of *Cs*CYP88A37, no *Cs*L1AT activity is observed. Representative TICs are displayed (n=3).

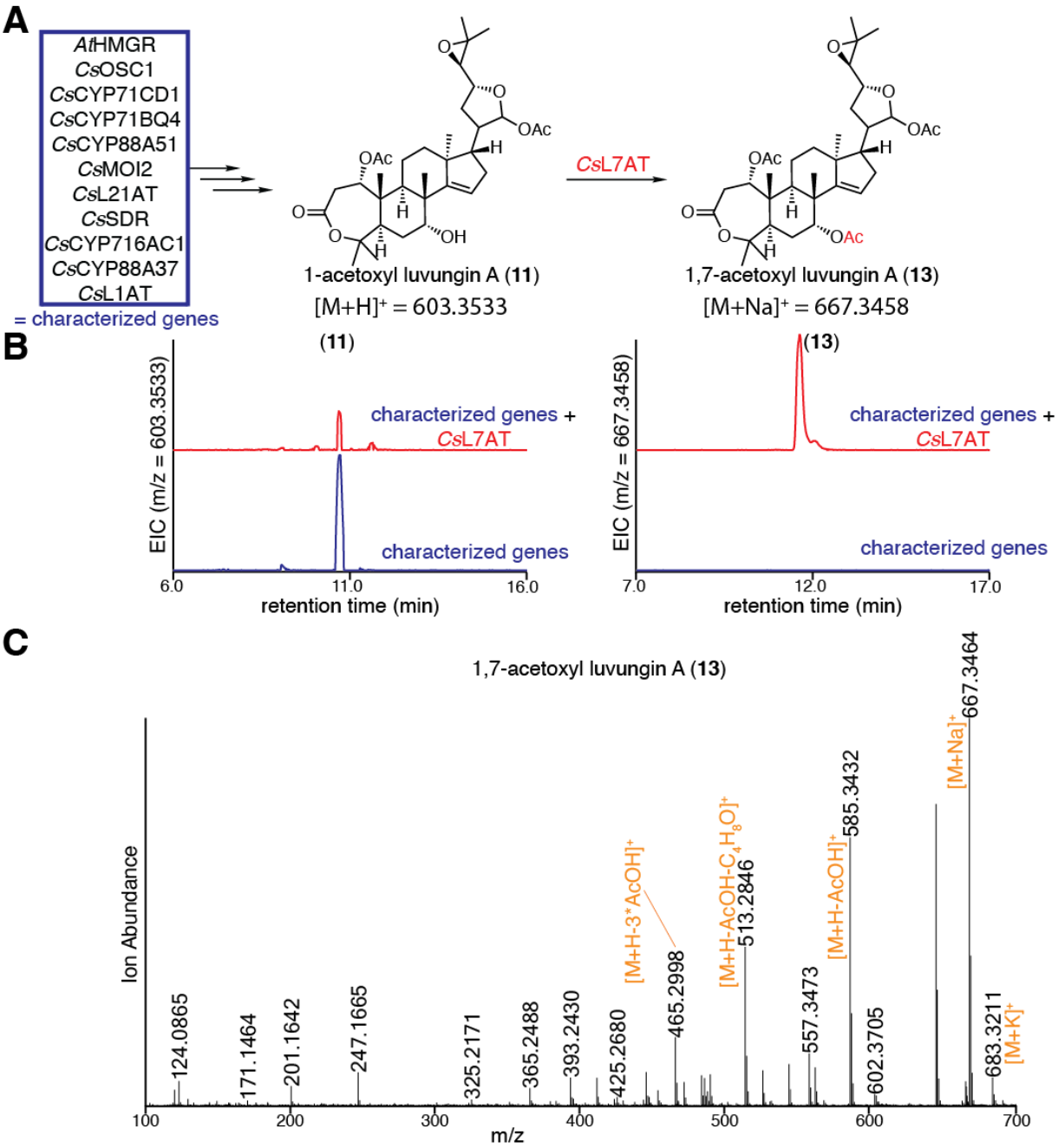



**Fig. S25. Characterization of *Cs*L7AT.**

(A) Predicted function of *Cs*L7AT in converting 1-acetoxyl luvungin A (**11**) to 1,7-acetoxyl luvungin A (**13**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*L7AT (red). EICs are displayed for *m/z* of 603.3533 (calculated mass for (**11**) $[M+H]^+$) or 667.3458 (calculated mass for (**13**) $[M+Na]^+$). (C) Mass spectrum of (**13**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Representative EICs and mass spectrum are displayed (n=6).

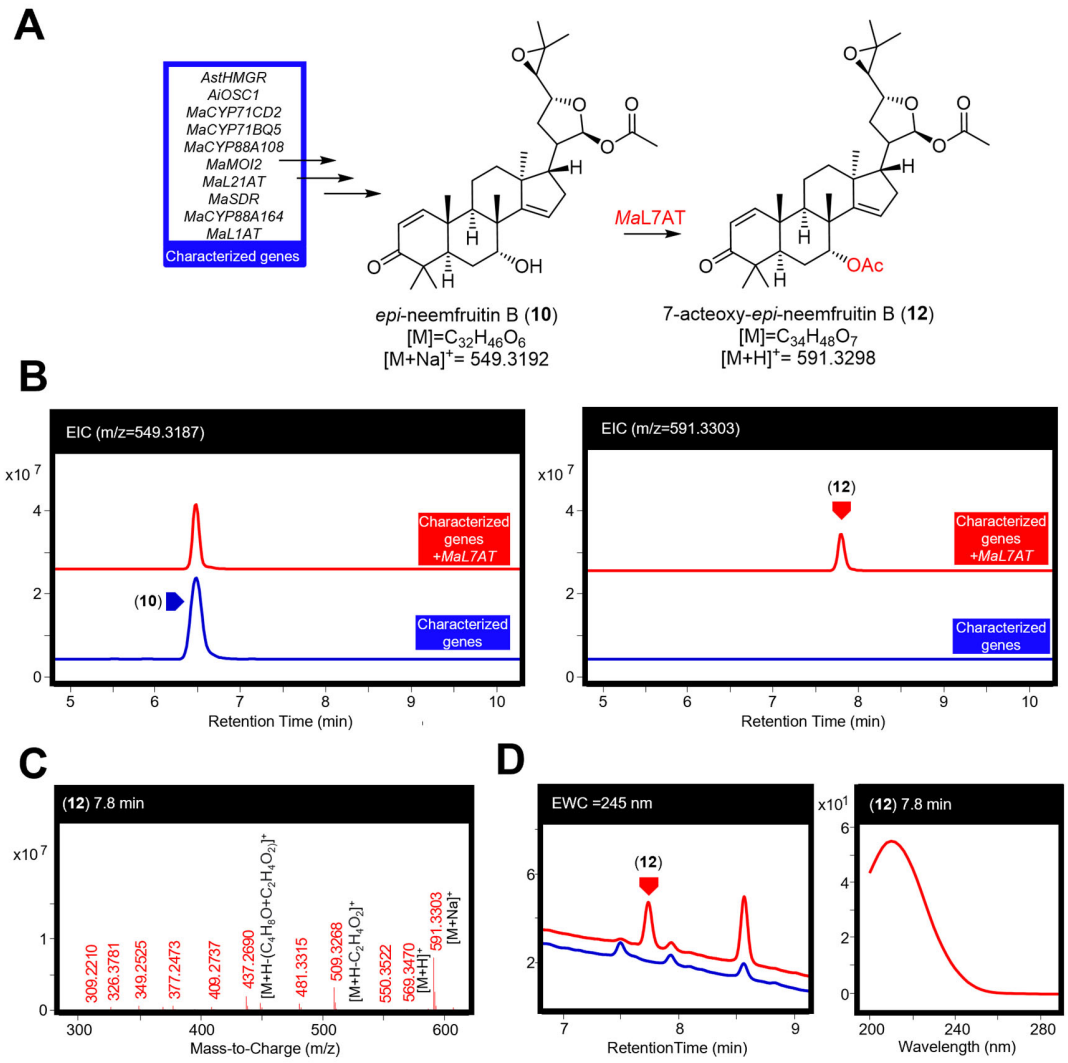



**Fig. S26. Characterization of *Ma*L7AT.**

(A) Function of *Ma*L7AT in producing a 7-acetoxyl-*epi*-neemfruitin B (**12**) (position confirmed by NMR of later product (**14**), table S15) from *epi*-neemfruitin B (**10**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *MaL7AT* (red). The EICs display *m/z* of 549.3187 (observed mass for [(**10**)+Na]$^+$) and 591.3303 (observed mass for [(**12**)+Na]$^+$). (C) Mass spectrum for (**12**) being heterologously produced in *N. benthamiana.* The main observed adducts ([M+H]$^+$ and [M+Na]$^+$) and fragments (including loss of acetic acid [M+H-$C_2H_4O_2$]$^+$ or loss the four-carbon epoxide containing fragment and an acetic acid [M+H-($C_4H_8O$+$C_2H_4O_2$)]$^+$) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4nm) for extracts displayed in panel B. UV spectrum (mAU) of (**12**) being heterologously produced in *N. benthamiana* is shown on the right. Representative traces and spectrum are displayed (n=6).

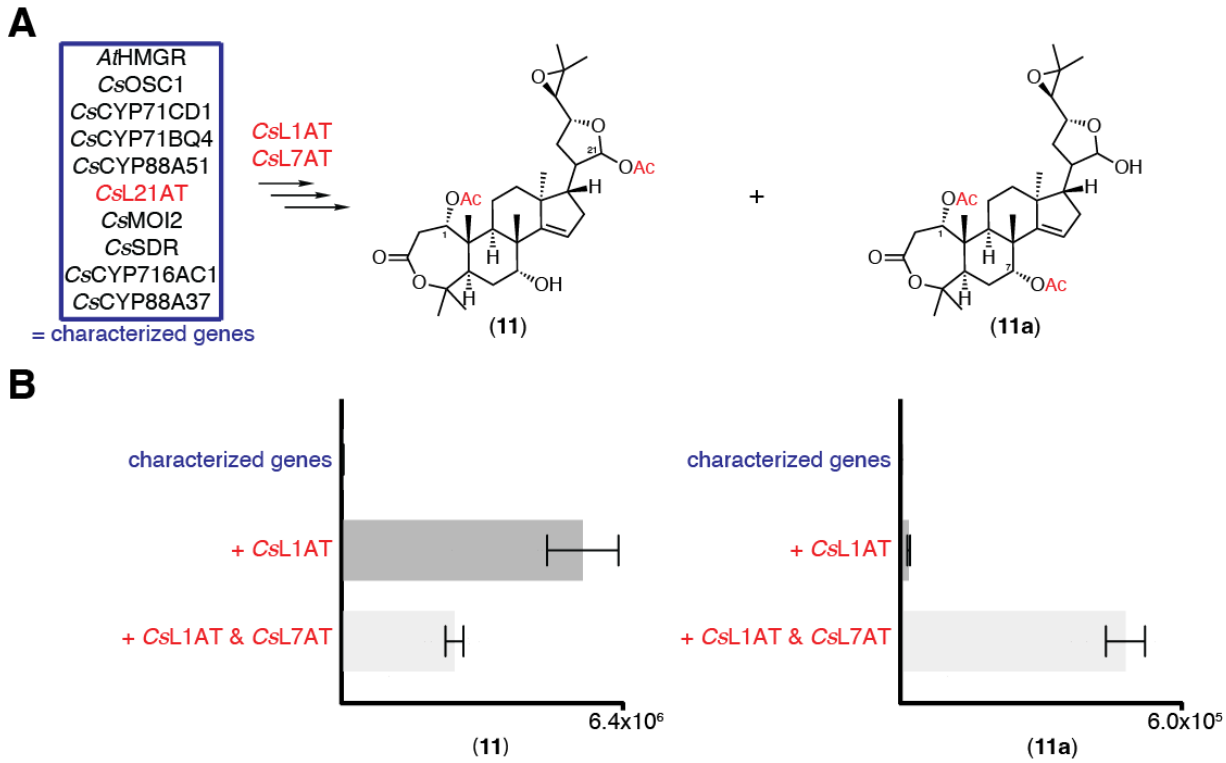



**Fig. S27. Accumulation of 1,21-diacetoxyl (11) and 1,7-diacetoxyl (11a) intermediates.** (A) Structures of the two diacetoxyl protolimonoids, 1,21 diacetoxyl (**11**) and 1,7-diacetoxyl (**11a**)), which are produced when all biosynthetic enzymes for the production of (**13**) (enzymes in the characterized genes box plus *Cs*L1AT and *Cs*L7AT) are co-expressed. Structure of (**11a**) is proposed based on characterized enzymatic functions. (B) Integrated peak area of extracted ion chromatogram (EIC) showing the accumulation of (**11**) and (**11a**) in *N. benthamiana* expressing the characterized genes listed in panel A (blue) with the addition of *Cs*L1AT alone or *Cs*L1AT and *Cs*L7AT (red). Values and error bars represent the mean and the standard error of the mean (n=6).

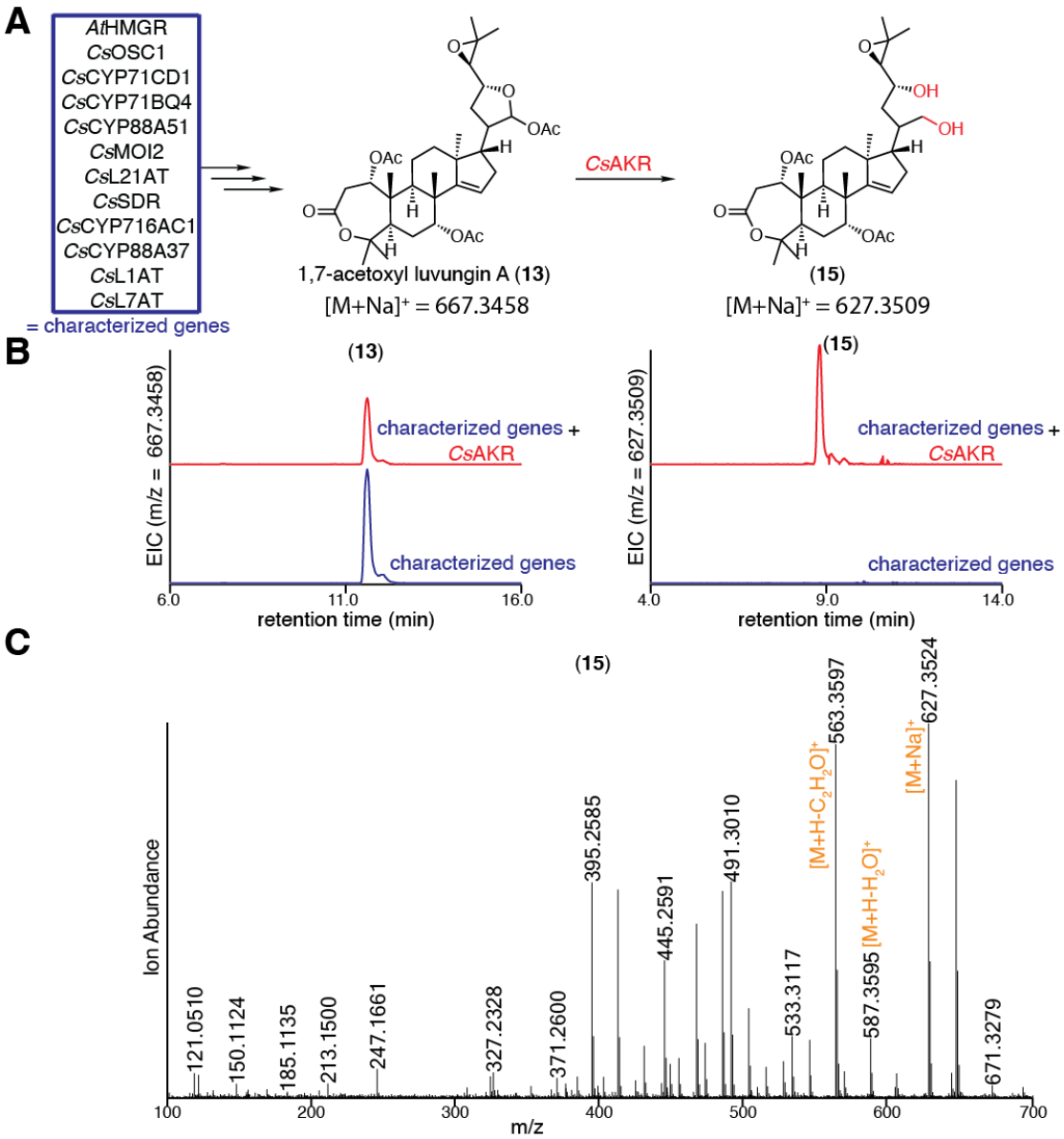



**Fig. S28. Characterization of *Cs*AKR.**

(A) Predicted function of *Cs*AKR in converting 1,7-acetoxyl luvungin A (**13**) to (**15**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*AKR (red). The EICs display *m/z* of 667.3458 (calculated mass for (**13**) $[M+Na]^+$) or 627.3509 (calculated mass for (**15**) $[M+Na]^+$). (C) Mass spectrum of (**15**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Proposed formation of the loss of $C_2H_2O$ fragment is shown in fig. S35. Representative EICs and mass spectrum are displayed (n=6).

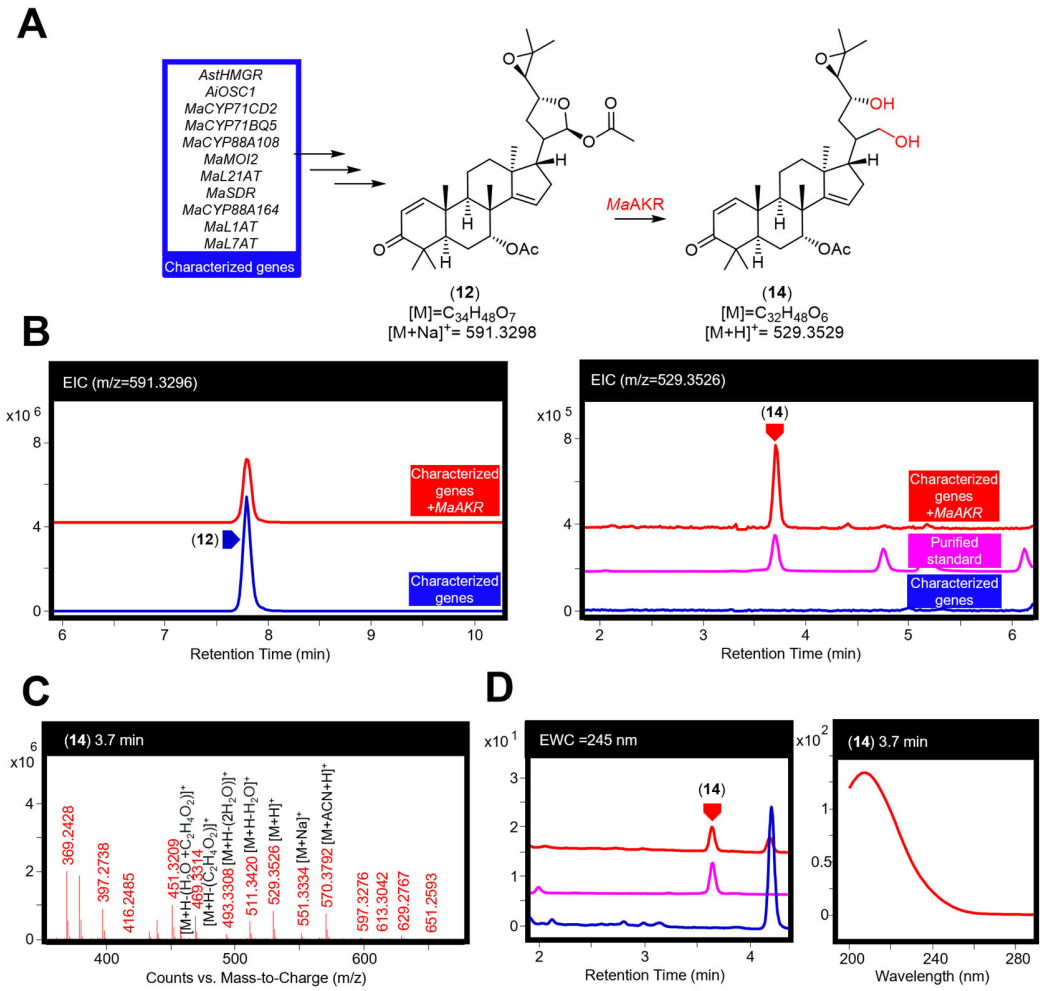



**Fig. S29. Characterization of *Ma*AKR.**

(A) Function of *Ma*AKR in producing the 21,23 diol (**14**) (confirmed by NMR, table S15) from 7-acetoxyl-*epi*-neemfruitin B (**12**), although substrate for enzymatic transformation is likely an earlier non-C21-acetylated intermediate. (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue), with the addition of *Ma*AKR (red) or for a purified standard of (**14**) (pink). The EICs display *m/z* of 591.3296 (observed mass for [(**12**)+Na]$^+$) and 529.3526 (observed mass of [(**14**)+H]$^+$). (C) Mass spectrum for (**14**) being heterologously produced in *N. benthamiana.* The main observed adducts ([M+H]$^+$, [M+ACN+H]$^+$and [M+Na]$^+$) and fragments (including loss of water molecules [M+H-$H_2O$]$^+$, acetic acid [M+H-$C_2H_4O_2$]$^+$ or a combination of both) are labeled. The fragments are consistent with the presence of a diol, rather than the precursor hemi-acetal ring. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4 nm) for extracts displayed in panel B. UV spectrum (mAU) of (**14**) being heterologously produced in *N. benthamiana* is shown on the right. Traces of standards have been scaled. Representative traces and spectra are displayed (n=6).

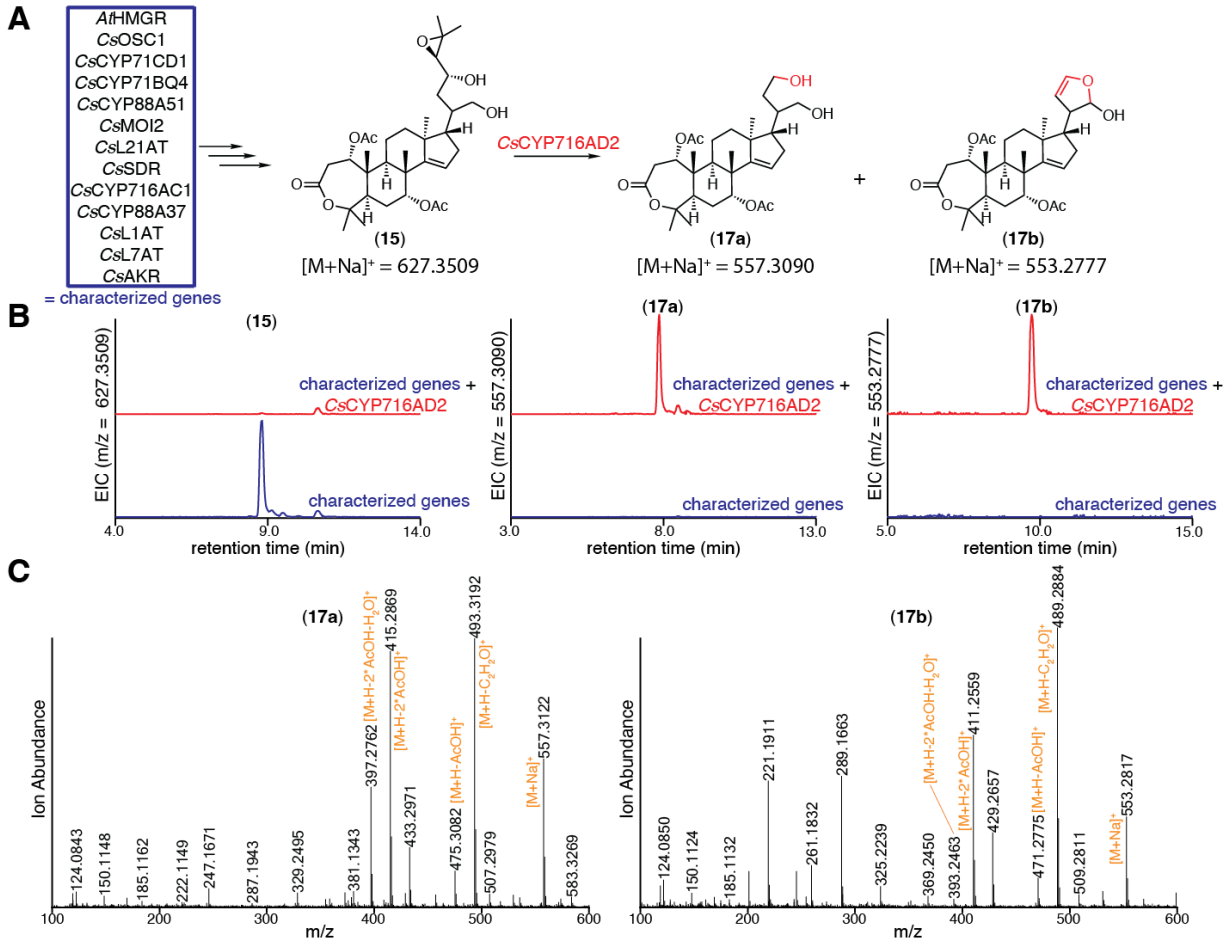



**Fig. S30. Characterization of *Cs*CYP716AD2.**

(A) Predicted function of *Cs*CYP716AD2 in converting (**15**) to (**17a**) and (**17b**), both top features identified by XCMS online (*76*). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*CYP716AD2 (red). EICs are displayed for *m/z* of 627.3509 (calculated mass for (**15**) $[M+Na]^+$), 557.3090 (calculated mass for (**17a**) $[M+Na]^+$) and 553.2777 (calculated mass for (**17b**) $[M+Na]^+$). (C) Mass spectra of (**17a**) and (**17b**) heterologously produced in *N. benthamiana* are shown in panel B, with major adducts and fragments labeled. Proposed formation of the loss of $C_2H_2O$ fragment is shown in fig. S35. Representative EICs and mass spectra are displayed (n=6).

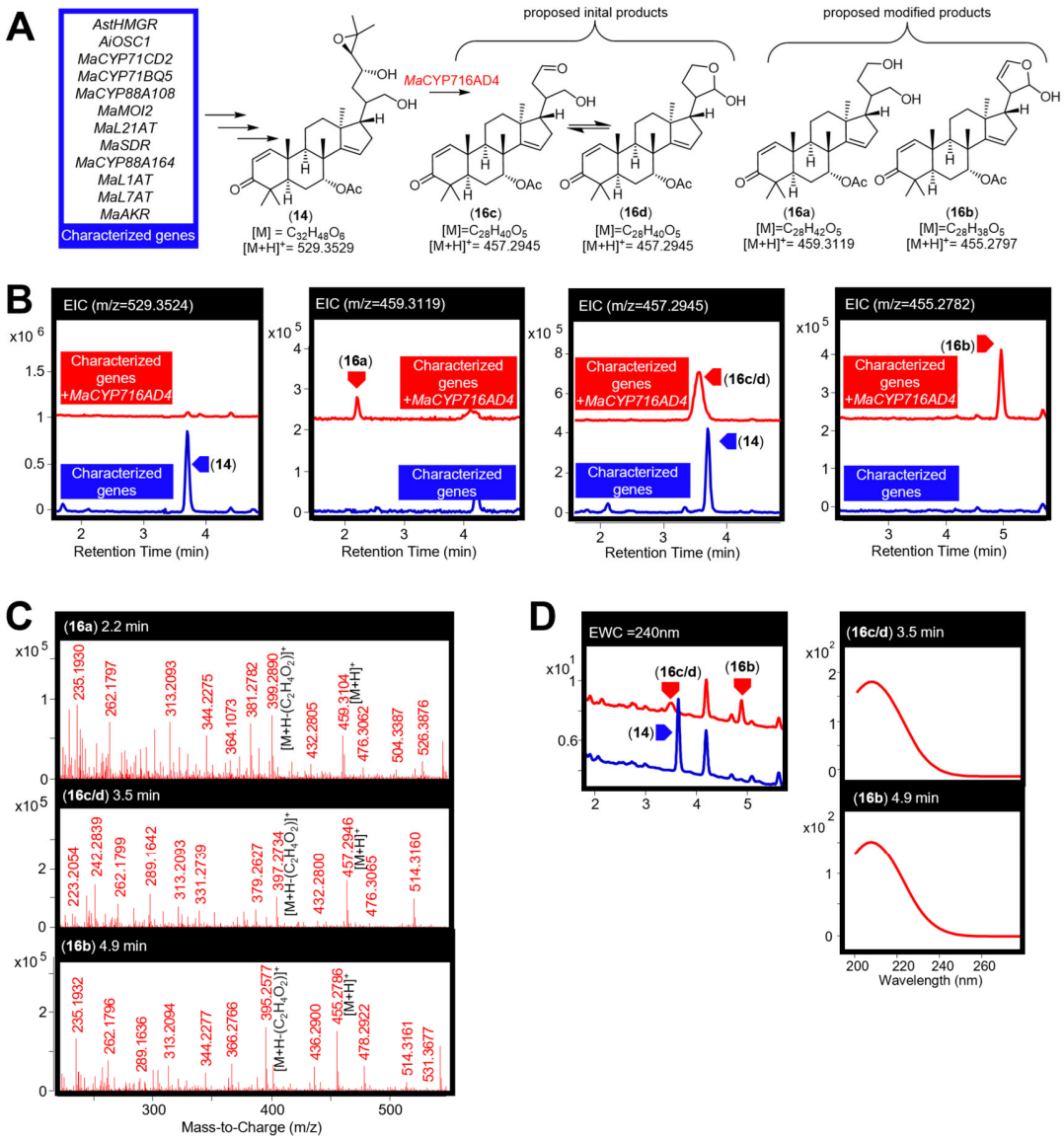



**Fig. S31. Characterization of *Ma*CYP716AD4.**

(A) Predicted function of *Ma*CYP716AD4 in converting (**14**) to (**16**) (proposed mechansim is shown in fig. S32). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Ma*CYP716AD4 (red). EICs are displayed for *m/z* of 529.3524 (observed mass of [(**14**)+H]$^+$), 459.3119 (observed mass for [(**16a**)+H]$^+$), 457.2945 (calculated mass for [(**16c/d**)+H]$^+$)) and of 455.2782 (observed mass for [(**16b**)+H]$^+$). (C) Mass spectra for (**16a-d**) being heterologously produced in *N. benthamiana,* the main observed adduct ([M+H]$^+$) and fragment (loss of acetic acid [M+H-$C_2H_4O_2$]$^+$, consistent with the loss of C7 acetoxy group) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4 nm) for extracts displayed in panel B and UV spectra (mAU) of (**16b-d**). (**16a**) was not observed on EWC likely due to low abundance. Representative traces and spectra are given (n=6).

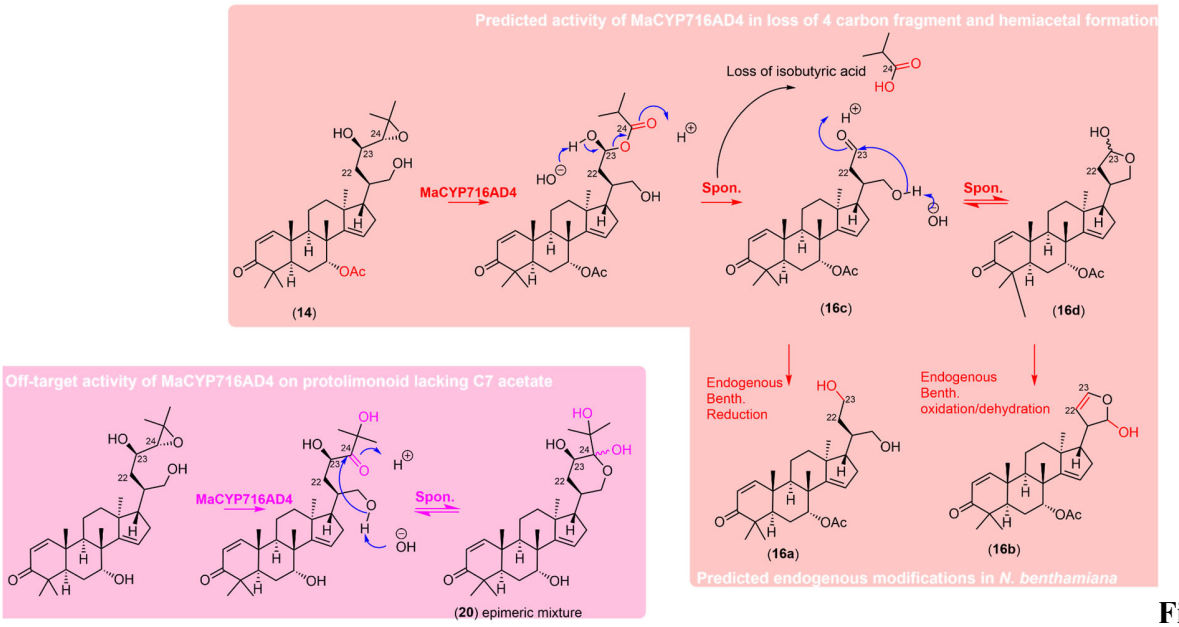



**Fig. S32. Hypothetical scheme for the reaction of CYP716ADs via a Baeyer-Villiger type mechanism.**

Proposed reactions explaining the occurrence of the observed products of *Ma*CYP716AD4 (**16a-d**) (red background), along with the occurrence of side-product (**20**) (NMR confirmed; table S20) when *Ma*CYP716AD4 is expressed in the absence of *Ma*L7AT (pink background). *Ma*CYP716AD4 is speculated to act via a Baeyer-Villiger mechanism. This would involve the enzyme converting the epoxide of (**14**) to a ketone at C-24 prior to the introduction of an ester. The resulting product may then be spontaneously cleaved, with loss of isobutyric acid, resulting in a C-23 aldehyde product (**16c**) which could spontaneously form a 5-membered hemi-acetal ring (**16d**). Endogenous enzymes in *N. benthamiana* could feasibly reduce/oxidize the initial products (**16c/d**) to (**16a/b**). Evidence to support this hypothesis comes from the purification and structural analysis of a six-membered hemiacetal product of *Ma*CYP716AD4 (**20**) that is produced only in the absence of *Ma*L7AT (fig. S44, table S20). Although this exact product has not been isolated from nature before, protolimonoids with similar E-rings have been reported (*77*). This 6-membered hemiacetal product suggests that *Ma*CYP716AD4 first converts the epoxide to a C-24 ketone. In the absence of C-7 *O*-acetylation, the C-24 ketone is hydroxylated at C-25 instead of undergoing a Baeyer-Villiger oxidation by *Ma*CYP716AD4, perhaps due to different substrate positioning in the active site. The proposed mechanism can likely be extended to *Cs*CYP716AD2 as it shows a similar side product in the *Cs*L7AT dropout experiment (fig. S45).

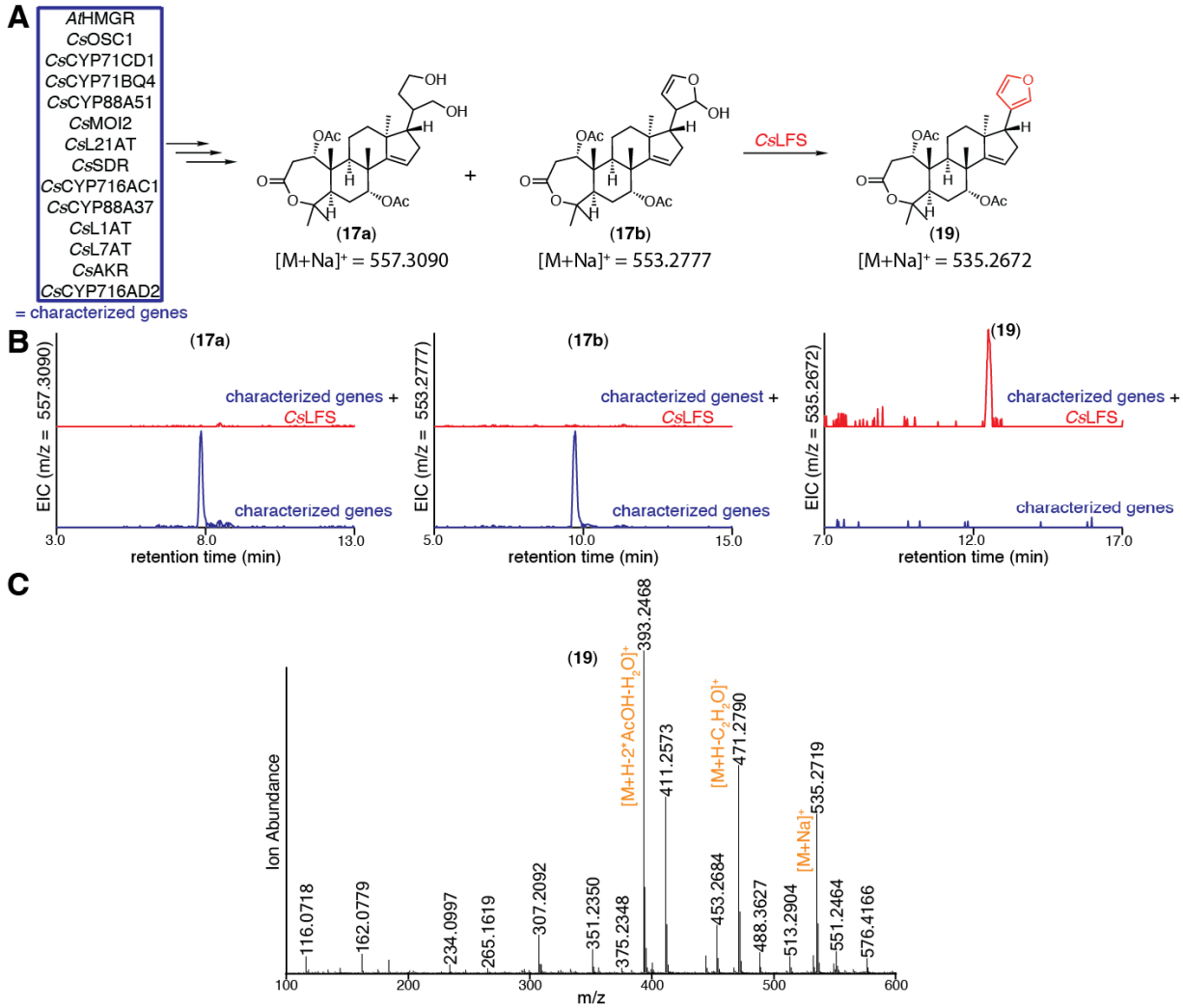



**Fig. S33. Characterization of *Cs*LFS.**

(A) Predicted function of *Cs*LFS in converting (**17a**) and (**17b**) to (**19**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Cs*LFS (red). The EICs display *m/z* of 557.3090 (calculated mass for (**17a**) $[M+Na]^+$), 553.2777 (calculated mass for (**17b**) $[M+Na]^+$), or 535.2672 (calculated mass for (**19**) $[M+Na]^+$). (C) Mass spectrum of (**19**) being heterologously produced in *N. benthamiana,* as shown in panel B, with major adducts and fragments labeled. Proposed formation of the loss of $C_2H_2O$ fragment is shown in fig. S35. Representative EICs and mass spectrum are displayed (n=6).

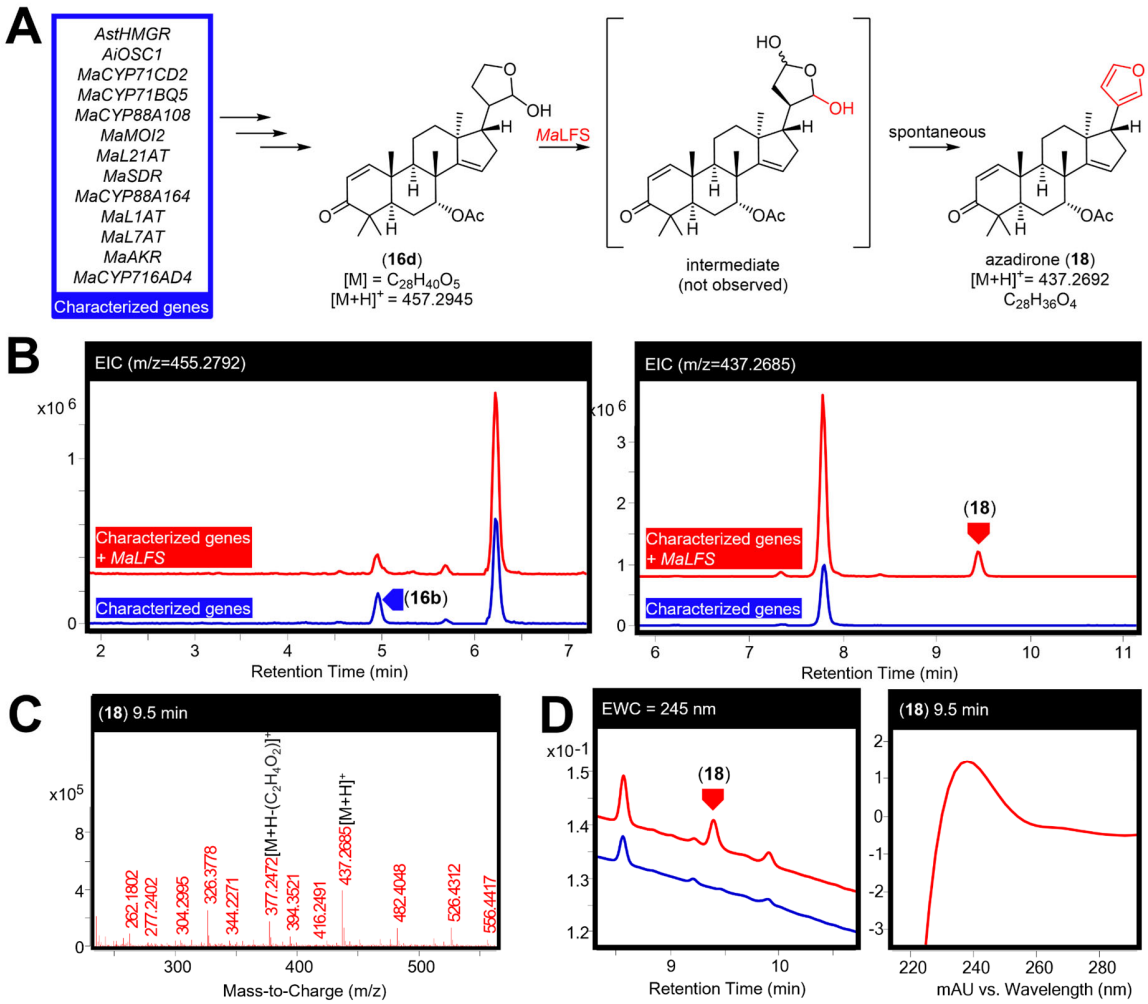



**Fig. S34. Characterisation of *MaLFS*.**
(A) Predicted function of *Ma*LFS in converting (**16d**) to azadirone (**18**). (B) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the characterized genes (listed in panel A) either alone (blue) or with the addition of *Ma*LFS (red). The EICs display *m/z* of 455.2792 (observed mass of [(**16b**)+H]$^+$) and of 437.2687 (observed mass of [(**18**)+H]$^+$). (C) Mass spectrum for (**18**) being heterologously produced in *N. benthamiana.* The main observed adduct ([M+H]$^+$) and fragment (loss of acetic acid [M+H-$C_2H_4O_2$]$^+$, consistent with the loss of C7 acetoxy group and the literature (*74*)) are labeled. (D) Extracted wavelength chromatograms (EWCs) of 245 nm (width of 4 nm) for extracts displayed in Panel B. UV spectrum (mAU) of azadirone (**18**) being heterologously produced in *N. benthamiana* is shown on the right. Representative traces and spectra are displayed (n=6).

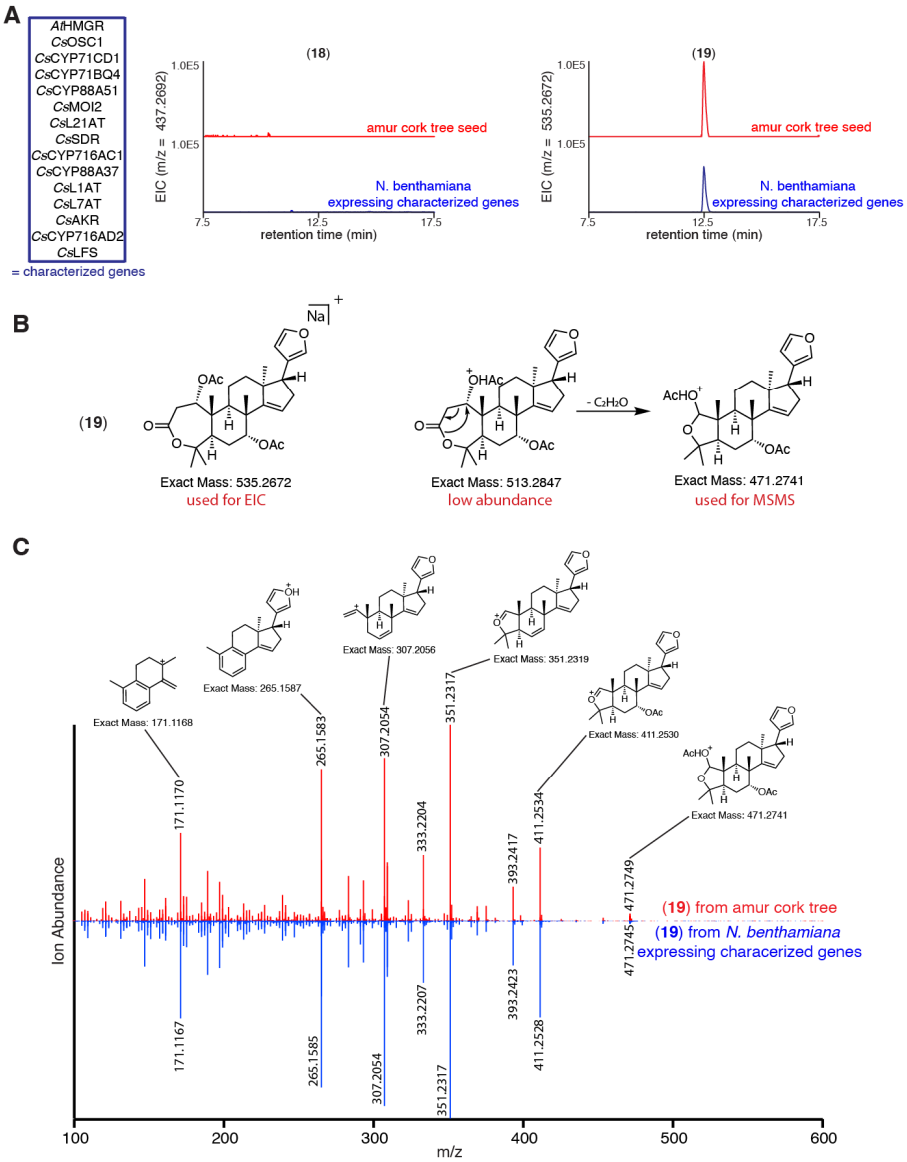



**Fig. S35. Detection of kihadalactone A (19) but not azadirone (18) in agro-infiltrated *N. benthamiana* extracts and amur cork tree seeds.**

(A) Extracted ion chromatograms (EICs) for both *Phellodendron amurense* (amur cork tree; Rutaceae plant) seed extracts (red) and *N. benthamiana* extracts agro-infiltrated with the combinations of genes outlined in the blue box (blue). EICs display *m/z* of $[M+H]^+$ = 437.2692 (calculated mass of (**18**)) and $[M+Na]^+$=535.2672 (calculated mass of (**19**)). (B) Structure and exact mass of (**19**) sodium adduct used for EIC and the proposed reaction scheme for the formation of the most abundance fragmentation peak (m/z 471.2741, loss of $C_2H_2O$) used for MSMS fragmentation. (C) MSMS spectra of (**19**) in *P. amurense* extract (red) compared with (**19**) being heterologously produced in *N. benthamiana* (blue) using genes listed in panel A. Proposed structures of major fragmentation peaks are shown. Collision energy 20eV is used in the MSMS. Representative EICs and mass spectra are displayed, n=3 biological replicates.

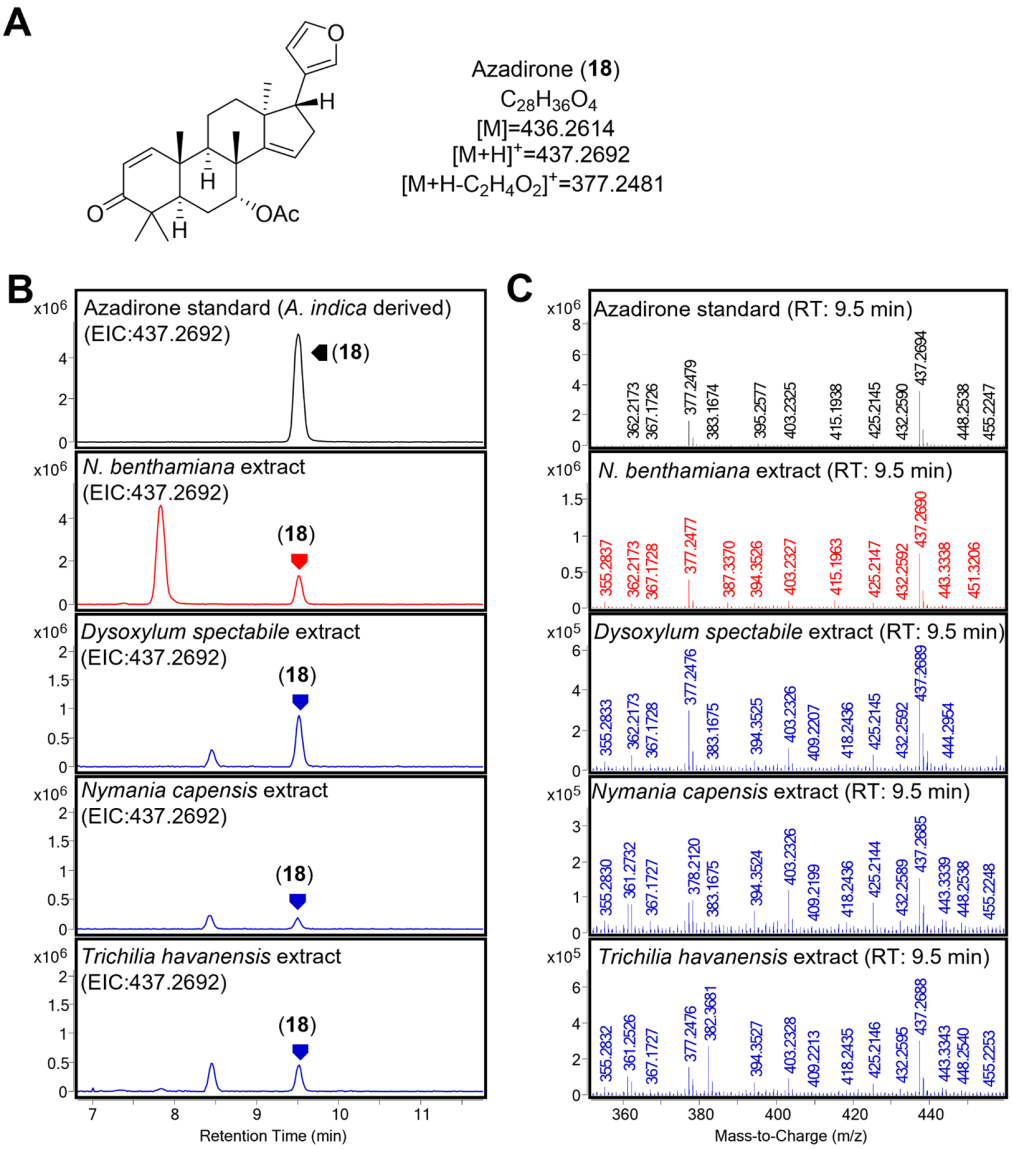



**Fig. S36. Azadirone (18) in agro-infiltrated *N. benthamiana* and Meliaceae extracts.** (A) Structure, formula and exact masses of observed adducts of azadirone (**18**), which, as well as being identified in agro-infiltrated *N. benthamiana* extracts, was identified in three Meliaceae species (*Trichilia havanensis* (*78*), *Dysoxylum spectabile* and *Nymania capensis*) sourced from Kew Gardens. (B) Extracted ion chromatograms (EICs) comparing an analytical standard of azadirone (**18**) (black, purified from *A. indica* leaf powder (table S17)), to extract from *N. benthamiana* expressing azadirone (**18**) biosynthetic enzymes (*Ai*OSC1*, Ma*CYP71CD2, *Ma*CYP71BQ5, *Ma*CYP88A108*, Ma*MOI2, *Ma*L21AT, *Ma*SDR, *Ma*CYP88A164, *Ma*L1AT, *Ma*L7AT, *Ma*AKR, *Ma*CYP716AD4 and *Ma*LFS (red)) and extracts of the three Meliaceae species identified as containing azadirone (**18**) (blue). (C) Mass spectra of the azadirone (**18**) peak corresponding to each of the extracts displayed in panel B.

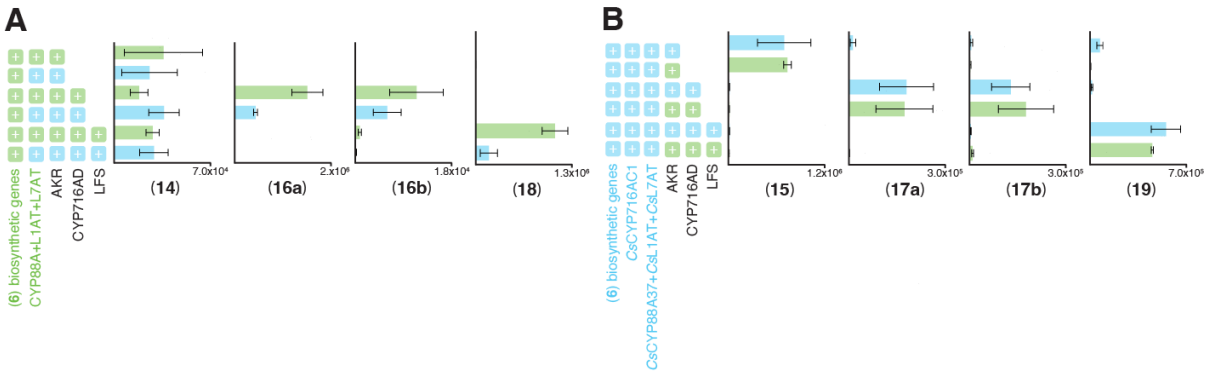



**Fig. S37. Compatibility of *C. sinensis* and *M. azedarach* pathways.**
(A) Integrated peak area of extracted ion chromatograms (EICs) for four of the final products in the *M. azedarach* pathway (**14**, **16a**, **16b**, **18**) being produced by heterologous expression in *N. benthamiana,* either exclusively with enzymes from *M. azedarach* (green), or instead using the relevant *C. sinensis* homologs (blue) (B) Integrated peak area of extracted ion chromatograms (EICs) for the last four *C. sinensis* pathway products (**15**, **17a**, **17b**, **19**) being produced by heterologous expression in *N. benthamiana,* either exclusively with enzymes from *C. sinensis* (blue), or with the relevant *M. azedarach* homologs (green). Biosynthetic enzymes from *M. azedarach* for production of (**6**) are as follows; *Ma*OSC1, *Ma*CYP71CD2, *Ma*CYP71BQ5, *Ma*CYP88A108, *Ma*MOI2, *Ma*L21AT and *Ma*SDR. Biosynthetic enzymes from *C. sinensis* for production of (**6**) are as follows; *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2, *Cs*L21AT and *Cs*SDR. Additional enzymes used are listed in the figure. CYP88A refers to either *Ma*CYP88A164 or *Cs*CYP88A37; CYP716AD refers to either *Ma*CYP716AD4 or *Cs*CYP716AD2. Values and error bars represent the mean and the standard error of the mean; n=3 biological replicates.

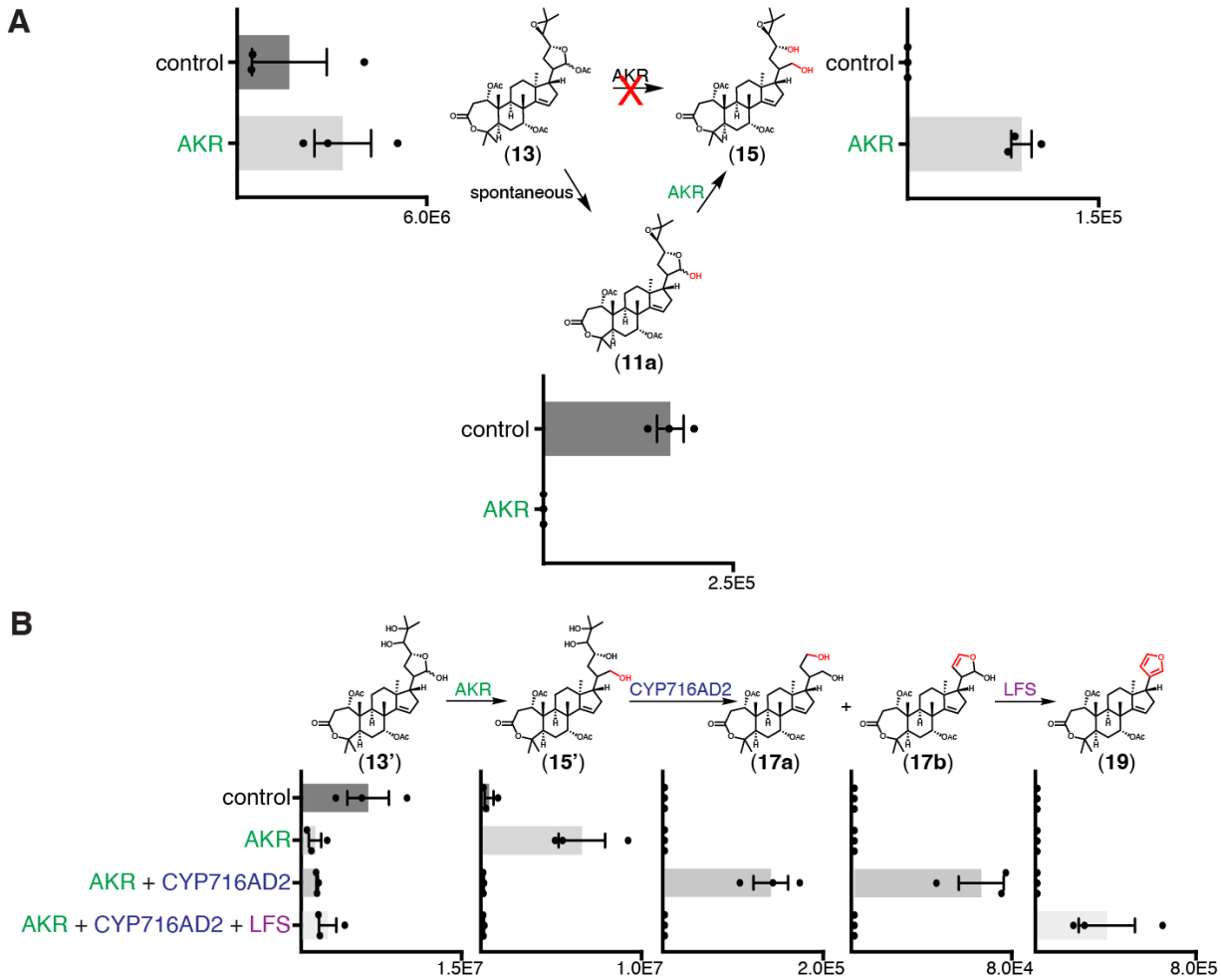



**Fig. S38. Characterization of *Cs*AKR through *in planta* feeding of (13) and (13').**
(A) *N. benthamiana* agro-infiltrated with either induction buffer (control) or *Cs*AKR (AKR) both co-infiltrated with a 50 uM solution of (**13**). Integrated peak areas from extracted ion chromatograms (EICs) for (**13**), (**15**) or the spontaneously formed 1,7-diacetoxyl (**11a**), demonstrating that (**11a**) is the more likely the substrate of *Cs*AKR rather than (**13**). (B) *N. benthamiana* agro-infiltrated with either induction buffer alone (control) or one of the following combinations: *Cs*AKR, *Cs*AKR + *Cs*CYP716AD2 or *Cs*AKR + *Cs*CYP716AD2 + *Cs*LFS. The control and combination were each co-infiltrated with a 200 uM solution of (**13'**). Integrated peak areas from EICs for (**13'**), (**15'**), (**17a-b**), (**19**) show that (**13'**) can be reduced by *Cs*AKR to yield (**15'**), which can be further processed by *Cs*CYP716AD2 and *Cs*LFS to form (**19**). All enzymes shown in the figure are from *C. sinensis.* Values and error bars represent the mean and the standard error of the mean (n=3).

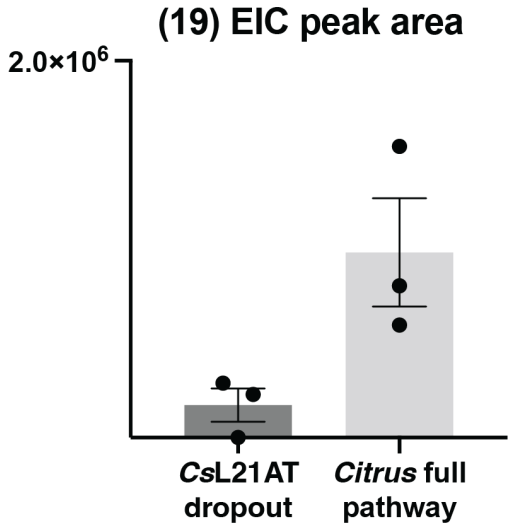



**Fig. S39. *Cs*L21AT increases yield of (19).**
Integrated peak area of extracted ion chromatograms (EICs) for kihadalactone A (**19**) with the full (**19**) pathway heterologously expressed in *N. benthamiana* (*Citrus* full pathway) or the full pathway without *Cs*L21AT (*Cs*L21AT dropout). The full pathway includes the following enzymes: *At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2, *Cs*SDR, *Cs*CYP88A37, *Cs*CYP716AC1, *Cs*L21AT, *Cs*L1AT, *Cs*L7AT, *Cs*AKR, *Cs*CYP716AD2 and *Cs*LFS. Values and error bars represent the mean and the standard error of the mean; n=3 biological replicates.

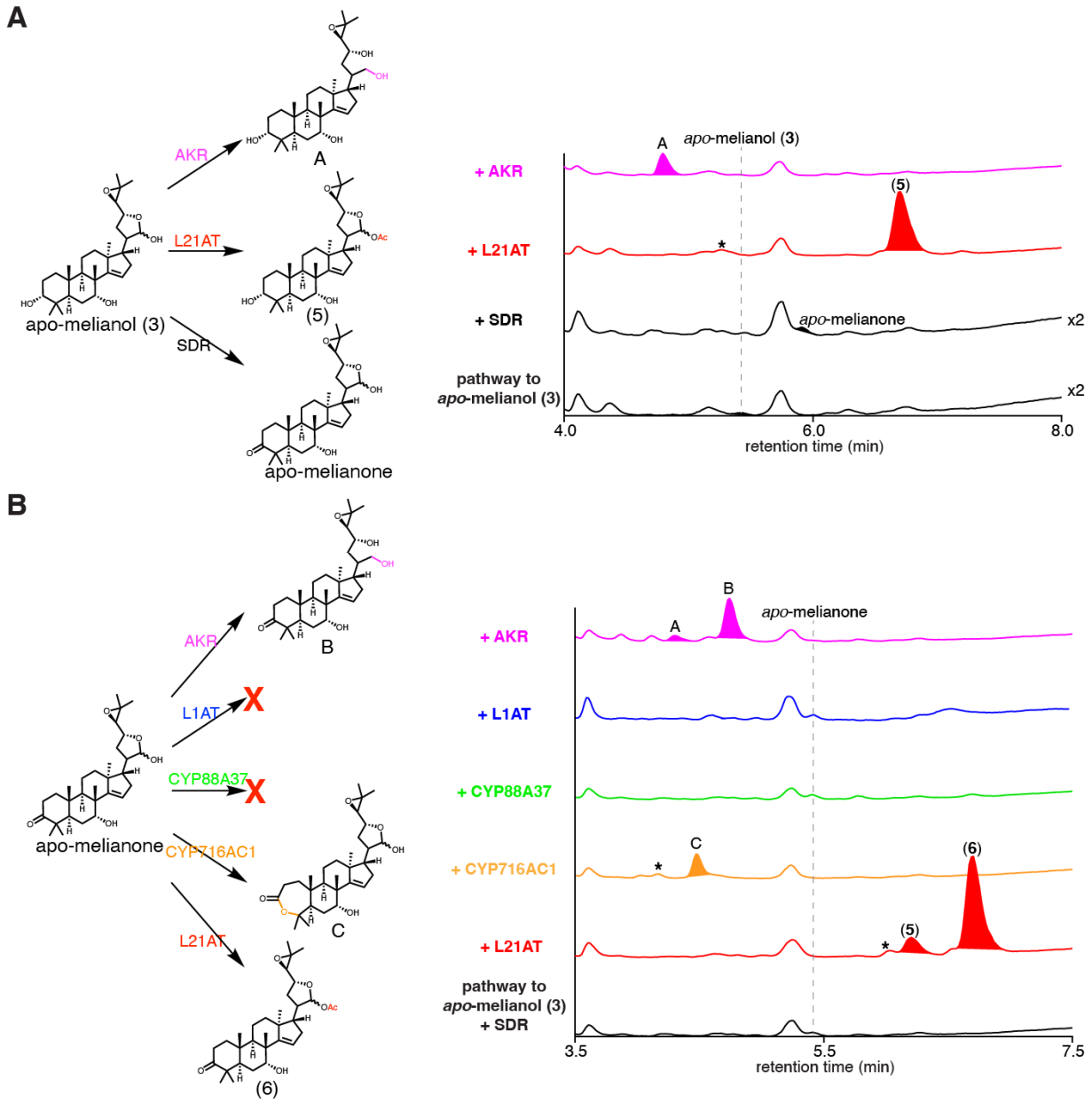



**Fig. S40. Partial construction of *Citrus* limonoid metabolic network.**

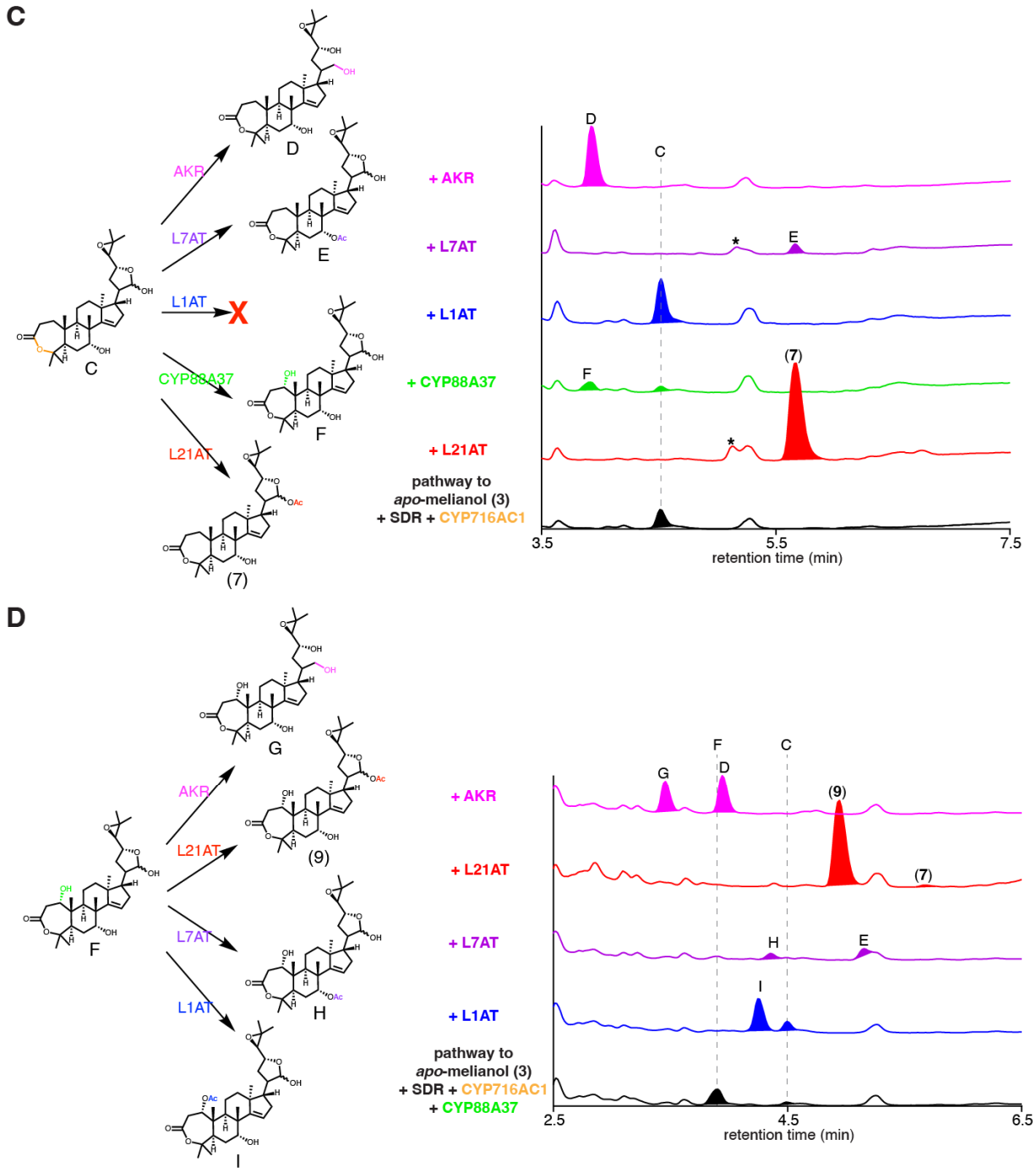



**Fig. S40. Partial construction of *Citrus* limonoid metabolic network (continued).**

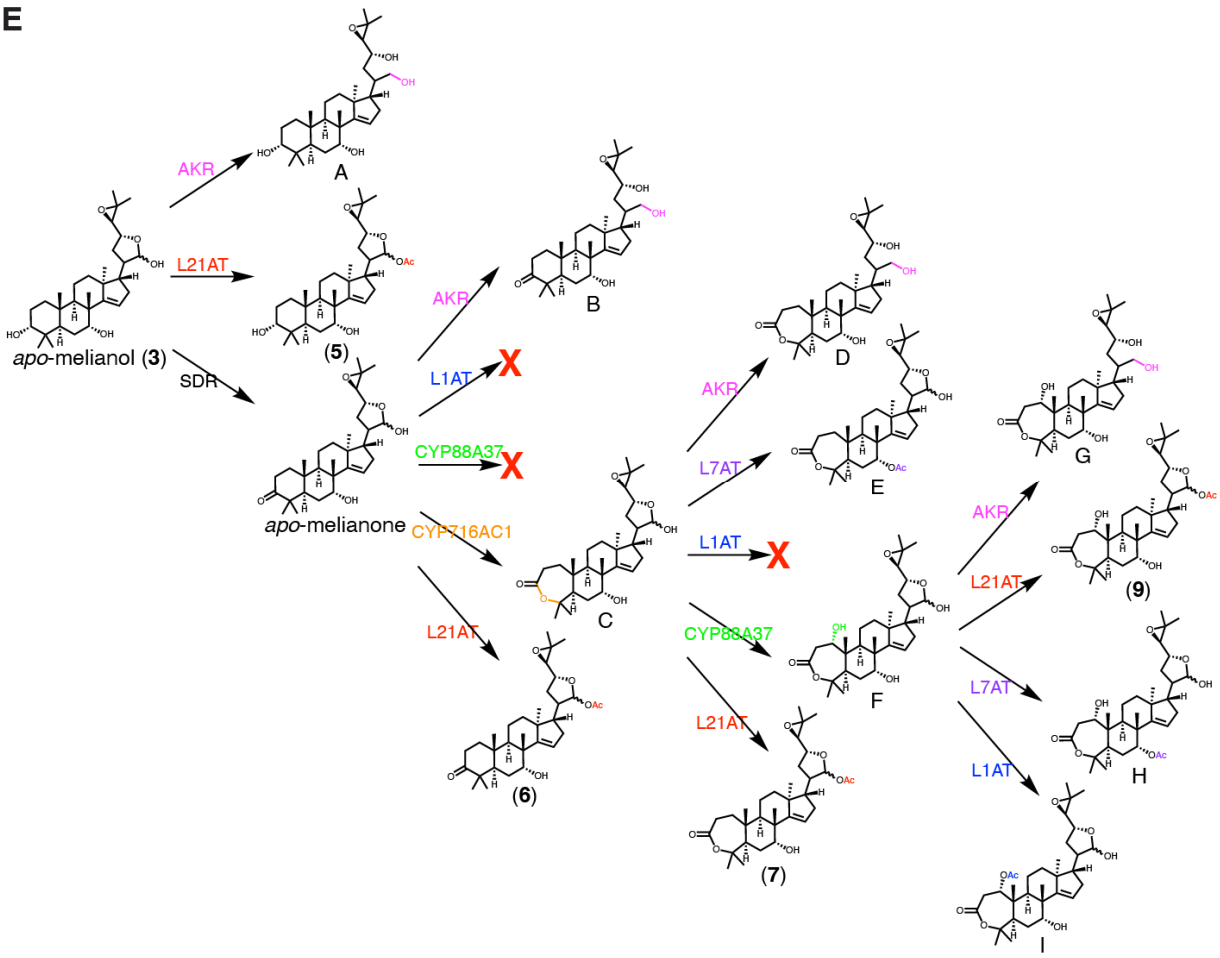
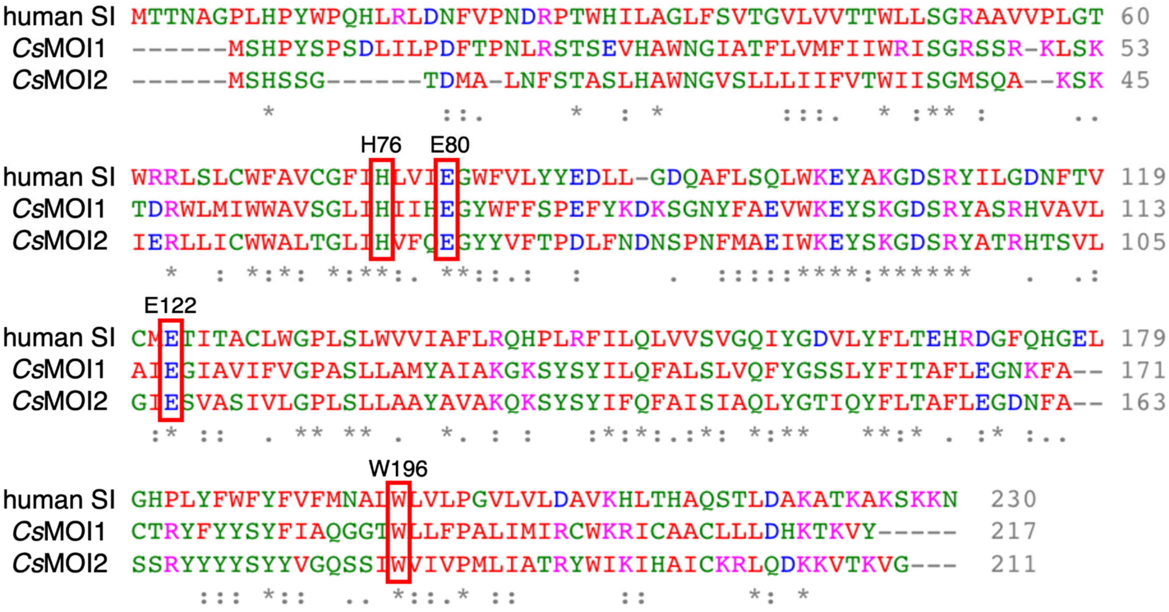



**Fig. S40. Partial construction of *Citrus* limonoid metabolic network (continued).** Total ion chromatograms (TICs) of *N. benthamiana* extracts agro-infiltrated with one of the following enzymes sets (A) *apo*-melianol (*At*HMGR, *Cs*OSC1, *Cs*CYP71CD1, *Cs*CYP71BQ4, *Cs*CYP88A51, *Cs*MOI2), (B) *apo*-melianol enzymes with the addition of *Cs*SDR, (C) *apo*-melianol enzymes with the addition of *Cs*SDR and *Cs*CYP716AC1, (D) *apo*-melianol enzymes with the addition of *Cs*SDR, *Cs*CYP716AC1 and CYP88A37. Alongside these, TICs for each gene set with the addition of a selection of genes (including *Cs*SDR, *Cs*CYP88A37, *Cs*CYP716AC1, *Cs*AKR, *Cs*L21AT, *Cs*L7AT, *Cs*L1AT) are displayed to demonstrate how the pathway can function as a metabolic network. Newly identified products are labeled A-I, and proposed structures (based on the characterized enzymatic transformation of each enzyme in this study) and reaction schemes are given on the left of each panel. This demonstrates that *Citrus* protolimonoids can generally be accepted by multiple biosynthetic enzymes to yield their corresponding products. Asterisk indicates new products that are likely the result of endogenous *N. benthamiana* enzymes acting on limonoid molecules. This analysis indicates the following: (A) *Apo*-melianol (**3**) can be the substrate of *Cs*AKR, *Cs*L21AT and *Cs*SDR, (B) *Apo*-melianone can be the substrate of *Cs*AKR, *Cs*CYP716AC1 and *Cs*L21AT but not *Cs*L1AT and *Cs*CYP88A37, (C) Product C, the product of *Cs*SDR and *Cs*CYP716AC1 acting on (**3**), can be

**Fig. S40. Partial construction of *Citrus* limonoid metabolic network (continued).**
the substrate of *Cs*AKR, *Cs*L7AT, *Cs*CYP88A37 and *Cs*L21AT and (D) Product F, the product of *Cs*SDR, *Cs*CYP716AC1 and *Cs*CYP88A37 acting on (**3**), can be the substrate of by *Cs*AKR, *Cs*L21AT, *Cs*L7AT and *Cs*L1AT. (E) A model of the *Citrus* limonoid metabolic network supported by data for individual metabolic steps shown in panel (A) to (D). Note that this is not a comprehensive network with all potential pathways; only a subset of all possible pathways were investigated as illustrated in the data here.



```
human SI MTTNAGPLHPYWPQHLRLDNFVPNDRPTWHILAGLFSVTGVLVVTTWLLSGRAAVVPLGT 60
CsMOI1   ------MSHPYSPSDLILPDFTPNLRSTSEVHAWNGIATFLVMFIIWRISGRSSR-KLSK 53
CsMOI2   ------MSHSSG------TDMA-LNFSTASLHAWNGVSLLLIIFVTWIISGMSQA--KSK 45
                 *          ::.     *  : *       :::.  * :** :     ..

human SI WRRLSLCWFAVCGFIHLVIEGWFVLYYEDLL-GDQAFLSQLWKEYAKGDSRYILGDNFTV 119
CsMOI1   TDRWLMIWWAVSGLIHIIHEGYWFFSPEFYKDKSGNYFAEVWKEYSKGDSRYASRHVAVL 113
CsMOI2   IERLLICWWALTGLIHVFQEGYYVFTPDLFNDNSPNFMAEIWKEYSKGDSRYATRHTSVL 105
           *  : *:*: *:**:. **::.:  :     .  :::::****:******   .  .:

human SI CMETITACLWGPLSLWVVIAFLRQHPLRFILQLVVSVGQIYGDVLYFLTEHRDGFQHGEL 179
CsMOI1   AIEGIAVIFVGPASLLAMYAIAKGKSYSYILQFALSLVQFYGSSLYFITAFLEGNKFA-- 171
CsMOI2   GIESVASIVLGPLSLLAAYAVAKQKSYSYIFQFAISIAQLYGTIQYFLTAFLEGDNFA-- 163
          :* ::  . ** ** .  *. : :   :*:*:.:*: *:**   **:* . :* :..

human SI GHPLYFWFYFVFMNALWLVLPGVLVLDAVKHLTHAQSTLDAKATKAKSKKN   230
CsMOI1   CTRYFYYSYFIAQGGTWLLFPALIMIRCWKRICAACLLLDHKTKVY-----   217
CsMOI2   SSRYYYYSYYVGQSSIWVIVPMLIATRYWIKIHAICKRLQDKKVTKVG---   211
             ::: *::  .. *::.* ::      ::      *: *
```

**Fig. S41. Alignment indicating the conserved active site residues between human sterol isomerase, *Cs*MOI1 and *Cs*MOI2.**

The active site of human sterol isomerase (SI) has previously been studied through protein crystal structural analysis and substrate docking (*79*). The residues H76, E80, E122, and W196 (highlighted in red boxes) of human SI were each proposed to be key in stabilizing the carbocation intermediate during isomerization. The conservation of these residues in *Cs*MOI1/2 suggests a similar isomerization mechanism via the formation of a carbocation. To determine how two different types of rearrangements are controlled by *Cs*MOI1 and *Cs*MOI2, despite their conserved active site residues, will require further study on the binding pocket of these enzymes. The protein sequences were aligned through the online Clustal Omega tool.

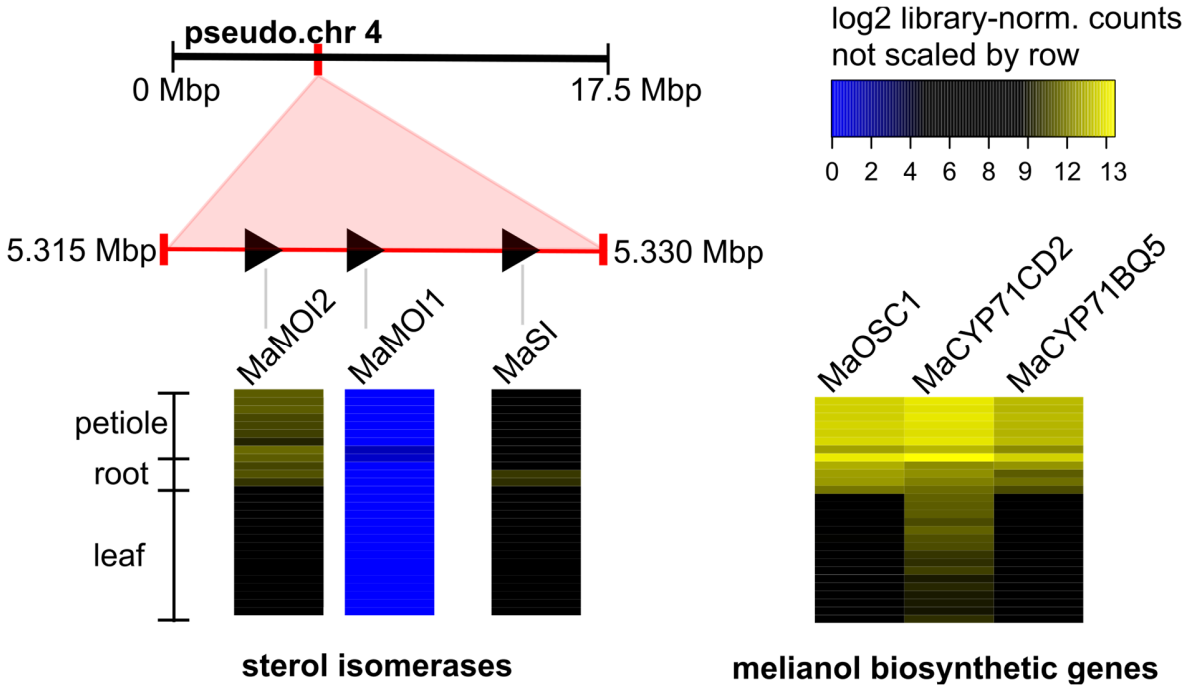



**Fig. S42. Genomic location and expression patterns of sterol isomerases in *M. azedarach*.** The expression pattern and genomic location (on pseudo-chromosome 4) of all sterol isomerase (SI) candidates (Interpro: IPR007905 (Emopamil-binding protein)) in the *M. azedarach* genome. SIs that have melianol oxide isomerase activity when tested by agro-mediated expression in *N. benthamiana* with melianol biosynthetic genes and *Ma*CYP88A108 have been renamed MOI, , along with *Ma*MOI1, due to sequence similarity to *Cs*MOI1. Gene IDs are provided (table S10). The expression pattern of melianol biosynthetic genes is shown on the right for comparative purposes. Heatmap was constructed using library normalized $\log_2$ read counts in Heatmap3 V1.1.1 (*44*) (with no scaling by row). A protein coding version of *MaMOI1* was not amplifiable based on the *M. azedarach* genome annotation.

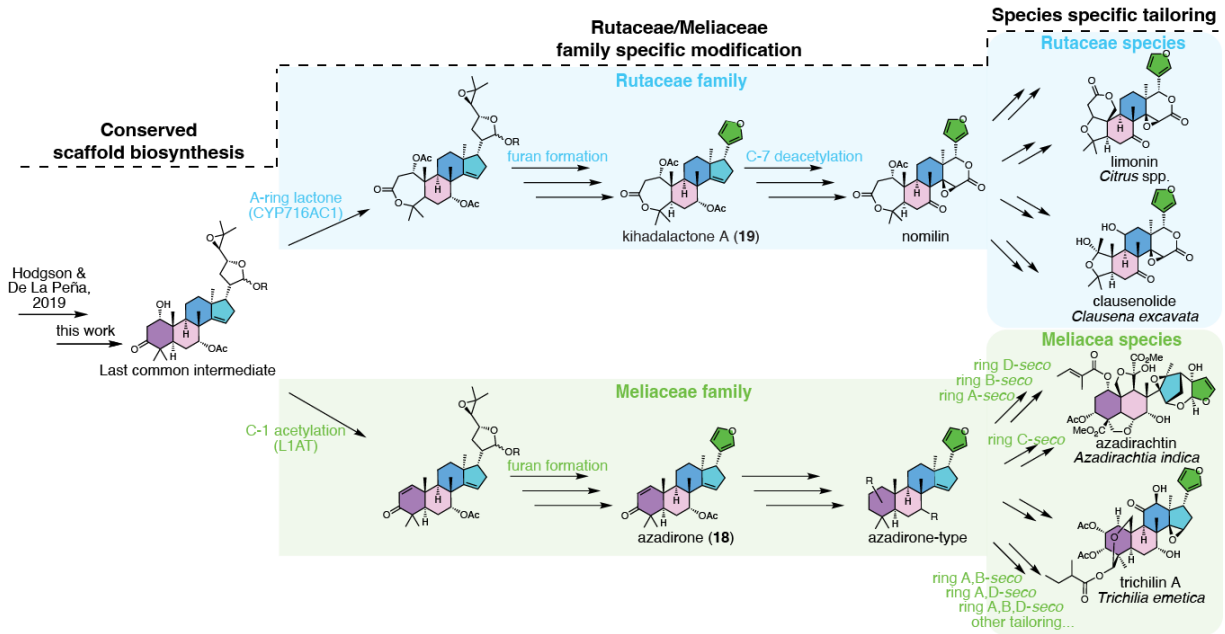



**Fig. S43. Proposed limonoid biosynthetic pathway in Rutaceae and Meliaceae plants.** Protolimonoid core scaffold biosynthesis is shared between the two families, forming the last common biosynthetic intermediate, which is structurally similar to (**6**). The pathway diverges with Rutaceae and Meliaceae family specific modifications, notably the A-ring lactone formation by *Cs*CYP716AC1 to yield nomilin- and azadirone-type biosynthetic intermediate. The Melia and Citrus pathways likely go through the biosynthetic intermediates azadirone (**18**) and kihadalactone A (**19**), respectively, as we have shown that C-7 *O*-acetylation is a prerequisite for furan formation in both pathways. The nomilin- and azadirone-type intermediates can undergo further species-specific tailoring to form structurally diverse limonoids, many of which are species specific (species of isolation is indicated below the molecule name).

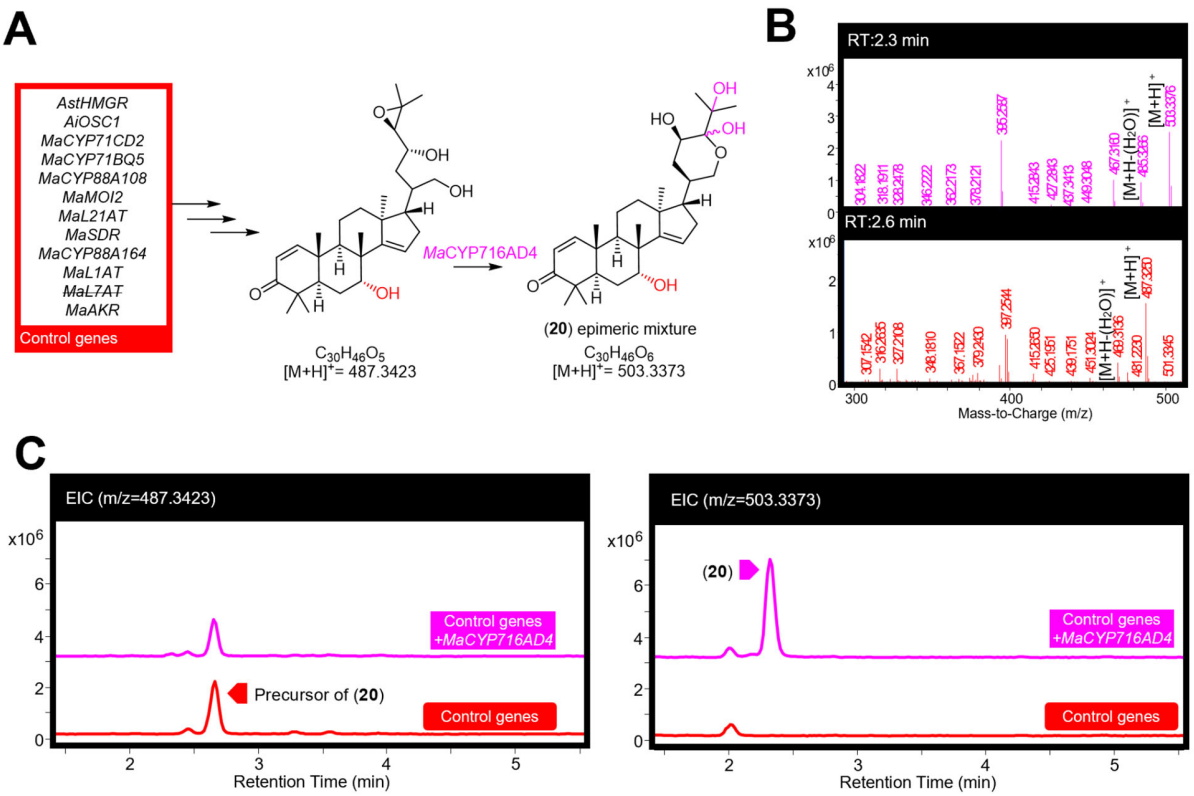



**Fig. S44. *Ma*CYP716AD4 side-product (20) formed in the absence of C-7-*O*-acetoxyl group.** (A) Proposed off-target function of *Ma*CYP716AD4 in producing the side-product (**20**) (NMR confirmed, table S20). Predicted mechanism is expanded upon in fig. S32. (B) Mass spectra for (**20**) (pink) and its precursor (red), being heterologously produced in *N. benthamiana,* displaying the main observed adduct ($[M+H]^+$) and water loss fragment ($[M+H-H_2O]^+$). (C) Extracted ion chromatograms (EICs) for extracts of *N. benthamiana* agro-infiltrated with the control genes (importantly lacking *MaL7AT*) listed in panel A either alone (red) or with the addition of *Ma*CYP716AD4 (pink). EICs display *m/z* of 487.3423 (calculated mass of the precursor of (**20**) $[M+H]^+$) and 503.3373 (calculated mass of (**20**) $[M+H]^+$). Representative EICs and mass spectra are displayed (n=3).

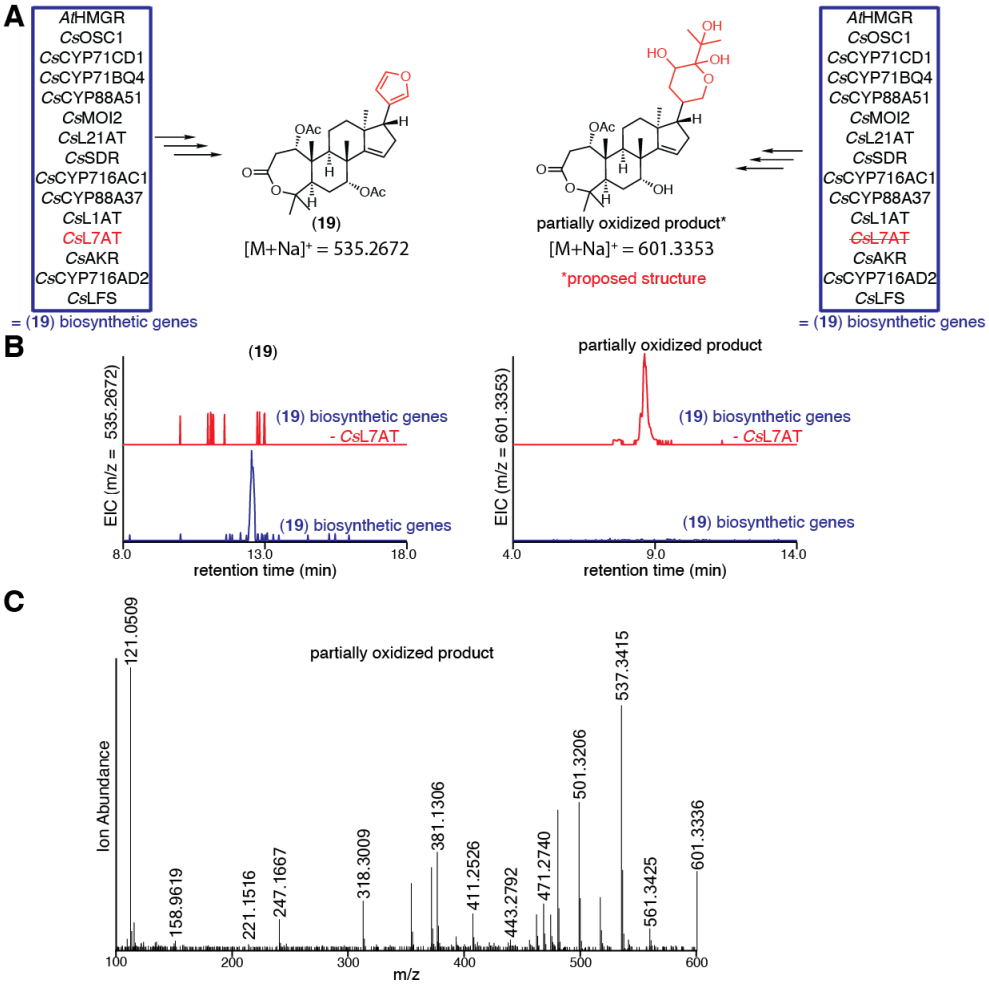



**Fig. S45. *Cs*L7AT is required for furan formation.**
(A) Predicted structures when *Cs*L7AT is either included or omitted from the set of co-expressed biosynthetic enzymes required for the production of kihadalactone A (**19**). The proposed structure resembles that of (**20**) (table S20), which was purified from heterologous expression of *M. azedarach* enzymes for (**16**), in the absence of *Ma*L7AT (fig. S44). (B) Extracted ion chromatograms (EICs) of *N. benthamiana* extracts agro-infiltrated with the combinations of genes outlined in panel A, either with (blue) or without (red) *Cs*L7AT. EICs display *m/z* of $[M+Na]^+$=535.2672 (calculated mass of (**19**)) and $[M+Na]^+$ = 601.3353 (calculated mass for proposed partially oxidized product). Neither (**19**) nor the corresponding C-7 deacetylated limonoid product was observed in the absence of *Cs*L7AT, however, a new peak of 601.3353 appeared, corresponding to a partially oxidized limonoid presumed to be formed by *Cs*CYP716AD2. (C) Mass spectrum of the observed partially oxidized product being heterologously produced in *N. benthamiana,* as shown in panel B. Representative EICs and mass spectrum are displayed (n=6).

**Supplementary Tables**

**Table S1. Summary of *M. azedarach* genome assembly and annotation.**

| ***Melia azedarach* genome assembly statistics** | |
|---|---|
| Number of contigs | 346 |
| Largest contig | 20,704,184 |
| Total length | 230,865,674 |
| GC (%) | 32.21 |
| N50 | 16,923,081 |
| N75 | 14,637,465 |
| L50 | 7 |
| L75 | 10 |
| Ns per 100 kbp | 9.44 |
| ***Melia azedarach* genome annotation statistics** | |
| **Genes** | |
| Total number of genes | 26,738 |
| Protein coding (high) | 22,785 |
| Transposable element (high) | 1,250 |
| Predicted (low) | 230 |
| Protein coding (low) | 1,651 |
| Transposable element (low) | 822 |
| **Transcripts** | |
| Transcripts per gene | 1.16 |
| Total number of transcripts | 31,048 |
| **CDS** | |
| Transcript mean size CDS (bp) | 1,309.11 |
| Min CDS | 78 |
| Max CDS | 15,903 |
| CDS mean size (bp) | 245.97 |
| Exon mean size (bp) | 312.11 |
| Exons per transcript | 5.71 |
| Total exons | 177,227 |
| Monoexonic transcripts | 5,473 |
| **cDNA** | |
| Transcript mean size cDNA (bp) | 1,781.55 |
| Min cDNA | 114 |
| Max cDNA | 16,537 |
| Intron mean size (bp) | 392.02 |
| 5UTR mean size (bp) | 186.24 |
| 3UTR mean size (bp) | 286.21 |

**Table S1. Summary of *M. azedarach* genome assembly and annotation (continued).**

| BUSCO- assessment | | |
|---|---|---|
| | ***Melia azedarach*** | ***Arabidopsis thaliana*** |
| Complete genes (single-copy) | 1,339 | 1,416 |
| Complete genes (2 copies) | 46 | 11 |
| Complete genes (3+ copies) | 7 | 4 |
| Fragmented genes | 20 | 5 |
| Missing genes | 28 | 4 |

*M. azedarach* pseudo-chromosome level genome statistics were generated by QUAST V.4.6.3 (*80*) and are based on contigs of size ≥500 bp. Statistics for *M. azedarach* annotation generated by the Earlham Institute. Genes are classified as either: protein coding, predicted (limited homology support <30%) or transposable element (>40% overlap with interspersed repeats). Genes were assigned a confidence classification of high or low based on their ability to meet specified criteria (>80% coverage to reference proteins or >60% protein coverage with >40% of the structure supported by transcriptome data). Statistics for coding sequences (CDS) and complementary DNA (cDNA) as also included. BUSCO (Benchmarking Universal Single-Copy Orthologs) (*24*) assessment of protein annotation of *M. azedarach* and gold standard *Arabidopsis thaliana*, performed by the Earlham Institute. The genome assembly and the annotation for assembled pseudo-chromosomes have been submitted to NCBI under the BioProject number PRJNA906622.

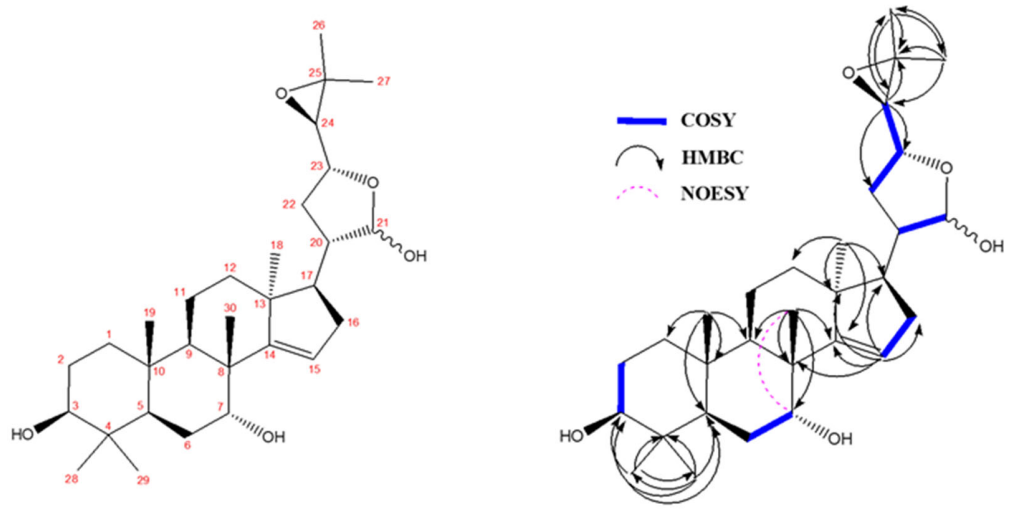
**Table S2. Summary of paired end reads generated for *M. azedarach* RNA-seq.**

| Sample | Rep. | Lane 1 | Lane2 | Total (per rep.) | Total (per sample) |
|---|---|---|---|---|---|
| *M. azedarach* 'individual 11' Upper Leaf | 1A | 7,312,258 | 7,957,007 | 15,269,265 | 78,676,364 |
| | 1B | 12,440,818 | 13,367,863 | 25,808,681 | |
| | 1C | 9,677,158 | 10,402,466 | 20,079,624 | |
| | 1D | 8,501,858 | 9,016,936 | 17,518,794 | |
| *M. azedarach* 'individual 11' Lower Leaf | 2A | 14,706,042 | 15,713,081 | 30,419,123 | 95,833,402 |
| | 2B | 9,952,003 | 10,506,690 | 20,458,693 | |
| | 2C | 9,995,844 | 10,724,057 | 20,719,901 | |
| | 2D | 11,759,629 | 12,476,056 | 24,235,685 | |
| *M. azedarach* 'individual 11' Petiole | 3A | 11,225,462 | 12,293,851 | 23,519,313 | 82,662,893 |
| | 3B | 8,518,447 | 9,151,386 | 17,669,833 | |
| | 3C | 8,723,766 | 9,267,735 | 17,991,501 | |
| | 3D | 11,360,248 | 12,121,998 | 23,482,246 | |
| *M. azedarach* 'individual 11' Root | 4A | 12,795,130 | 13,497,456 | 26,292,586 | 107,736,216 |
| | 4B | 9,430,278 | 10,235,484 | 19,665,762 | |
| | 4D | 14,075,197 | 14,780,596 | 28,855,793 | |
| | 4F | 15,951,734 | 16,970,341 | 32,922,075 | |
| *M. azedarach* 'individual 02' Upper Leaf | 5A | 8,425,596 | 8,942,230 | 17,367,826 | 86,473,401 |
| | 5B | 7,483,256 | 7,905,622 | 15,388,878 | |
| | 5C | 15,588,782 | 16,252,245 | 31,841,027 | |
| | 5D | 10,597,294 | 11,278,376 | 21,875,670 | |
| *M. azedarach* 'individual 02' Lower Leaf | 6A | 7,570,717 | 7,949,090 | 15,519,807 | 83,790,681 |
| | 6B | 15,757,443 | 16,754,196 | 32,511,639 | |
| | 6C | 8,341,297 | 8,628,045 | 16,969,342 | |
| | 6D | 9,074,779 | 9,715,114 | 18,789,893 | |
| *M. azedarach* 'individual 02' Petiole | 7A | 15,145,250 | 16,073,522 | 31,218,772 | 100,692,927 |
| | 7B | 11,317,371 | 12,034,239 | 23,351,610 | |
| | 7C | 12,710,000 | 13,530,865 | 26,240,865 | |
| | 7D | 9,595,536 | 10,286,144 | 19,881,680 | |

Numbers of paired-end reads are reported per lane, replicate and sample. Petiole samples include rachis. Raw RNA-seq reads have been deposited on NCBI under the BioProject PRJNA906055.

**Table S3. $^{13}C$ & $^{1}H$ δ assignments of *apo*-melianol (3) produced using heterologously expressed genes from *M. azedarach* (C-21 epimeric mixture)**

**Carbon numbering scheme and selected COSY, HMBC and NOESY**

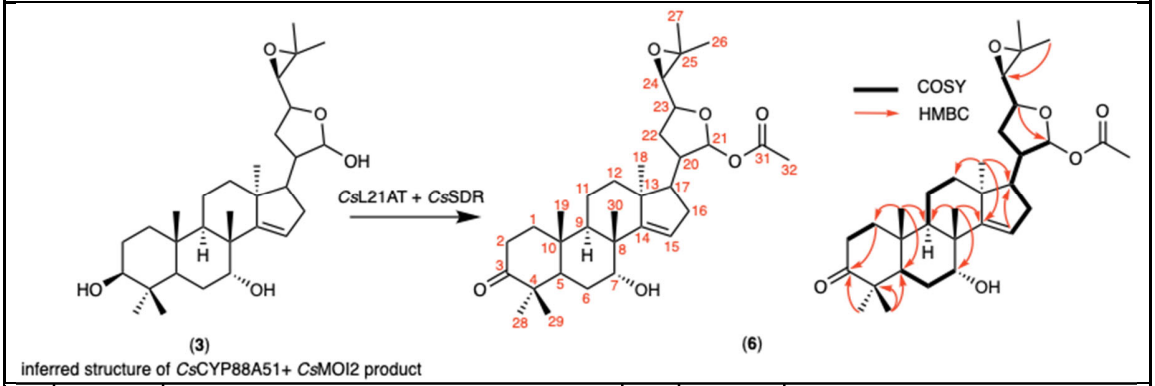



| C | $^{13}C$ δ (150 MHz) | | $^{1}H$ δ (600 MHz) | | C | $^{13}C$ δ (150 MHz) | | $^{1}H$ δ (600 MHz) | |
|---|---|---|---|---|---|---|---|---|---|
| **14** | 162.53 | 162.20 | / | | **1** | 37.91 | 37.88 | 1.61 (1H, m)<br>1.04 (1H, m) | |
| **15** | 119.52 | 119.04 | 5.47 (1H, m) | 5.46 (1H, m) | **10** | 37.58 | 37.57 | / | |
| **21** | 102.38 | 97.59 | 5.39 (1H, m) | 5.38 (1H, m) | **16** | 35.07 | 34.73 | 2.18 (2H, m) | 2.12 (2H, m) |
| **3** | 78.77 | 78.74 | 3.28 (1H, dd J= 11.3, 4.5) | | **22** | 34.71 | 31.35 | 2.09 (1H, m)<br>1.41 (1H, m) | 2.02 (1H, m)<br>1.74 (1H, m) |
| **23** | 78.37 | 77.24 | 3.91 (1H, m) | 3.96 (1H, m) | **12** | 33.06 | 32.80 | 1.82 (1H, m)<br>1.52 (1H, m) | 1.98 (1H, m)<br>1.46 (1H, m) |
| **7** | 72.33 | 72.32 | 3.93 (1H, m) | | **28** | 27.67 | | 0.99 (3H, s) | |
| **24** | 67.62 | 65.21 | 2.83<br>(1H, d J= 7.5) | 2.70<br>(1H, d J= 7.5) | **30** | 27.65 | 27.63 | 1.058 (3H, s) | 1.056 (3H, s) |
| **25** | 58.09 | 57.27 | / | | **2** | 27.14 | | 1.64 (1H, m)<br>1.58 (1H, m) | |
| **17** | 57.51 | 52.73 | 1.74 (1H, m) | 2.01 (1H, m) | **26** | 25.02 | 24.92 | 1.33 (3H, s) | 1.32 (3H, s) |
| **20** | 47.80 | 45.48 | 2.39 (1H, m) | 1.49 (1H, m) | **6** | 23.68 | 23.66 | 1.85 (1H, m)<br>1.74 (1H, m) | |
| **13** | 47.09 | 46.70 | / | | **18** | 19.79 | 19.45 | 1.02 (3H, s) | 1.09 (3H, s) |
| **5** | 46.53 | 46.49 | 1.50 (1H, m) | 1.48 (1H, m) | **27** | 19.35 | 19.21 | 1.32 (3H, s) | |
| **8** | 44.26 | 44.24 | / | | **11** | 16.34 | 16.29 | 1.69 (1H, m)<br>1.52 (1H, m) | |
| **9** | 41.78 | 41.74 | 1.93 (1H, m) | 1.91 (1H, m) | **29** | 15.45 | | 0.79 (3H, s) | |
| **4** | 38.36 | | / | | **19** | 15.41 | 15.39 | 0.89 (3H, s) | |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined.

**Table S4. $^{13}C$ & $^{1}H$ δ assignments of (6) produced using heterologously expressed genes from *C. sinensis.***

**Carbon numbering scheme and selected COSY and HMBC**

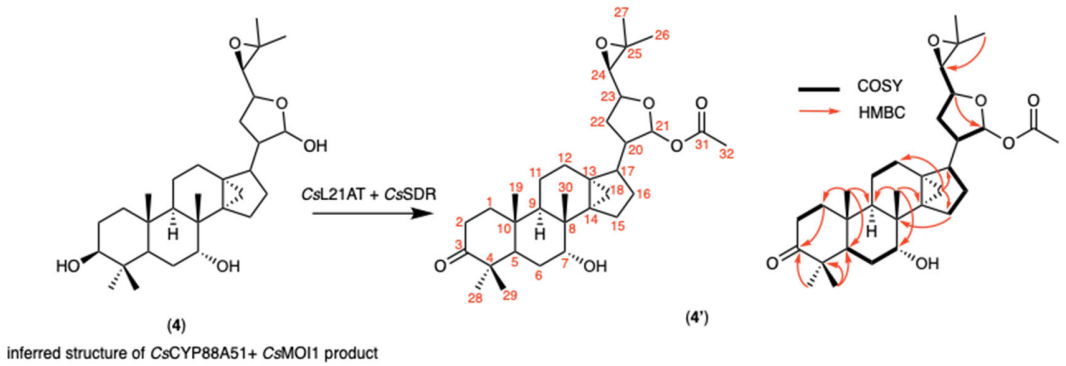



| C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | | C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | |
|---|---|---|---|---|---|---|---|
| **1** | 38.58 | 1.49 (1H, m) | 1.82 (1H, m) | **17** | 52.73 | 1.91 (1H, dt J = 8.1, 10.4) | |
| **2** | 34.03 | 2.41 (1H, ddd J = 3.8, 7.5, 15.8) | 2.51 (1H, ddd J = 7.5, 10.3, 15.7) | **18** | 19.82 | 1.02 (3H, s) | |
| **3** | 217.31 | / | | **19** | 15.04 | 0.98 (3H, s) | |
| **4** | 47.00 | / | | **20** | 44.33 | 2.35 (1H, dddd J = 4.1, 7.0, 10.9, 12.5) | |
| **5** | 46.71 | 2.07 (1H, m) | | **21** | 96.70 | 6.24 (1H, d J = 4.1) | |
| **6** | 24.92 | 1.77 (1H, m) | 1.82 (1H, m) | **22** | 31.50 | 1.68 (1H, m) | 2.07 (1H, m) |
| **7** | 72.05 | 3.95 (1H, appt t = 2.8) | | **23** | 79.86 | 3.91 (1H, dt J = 9.9, 7.2) | |
| **8** | 44.17 | / | | **24** | 66.81 | 2.65 (1H, d J = 7.5) | |
| **9** | 40.91 | 1.99 (1H, dd J = 7.6, 12.0) | | **25** | 57.25 | / | |
| **10** | 37.30 | / | | **26** | 19.45 | 1.27 (3H, s) | |
| **11** | 16.38 | 1.54 (1H, m) | 1.68 (1H, m) | **27** | 25.03 | 1.31 (3H, s) | |
| **12** | 32.45 | 1.29 (1H, m) | 1.59 (1H, m) | **28** | 21.28 | 1.03 (3H, s) | |
| **13** | 46.62 | / | | **29** | 26.35 | 1.08 (3H, s) | |
| **14** | 161.70 | / | | **30** | 27.36 | 1.08 (3H, s) | |
| **15** | 119.72 | 5.49 (1H, dd J = 1.9, 3.4) | | **31** | 170.05 | / | |
| **16** | 35.19 | 2.2 (2H, m) | | **32** | 21.61 | 2.04 (3H, s) | |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. Literature comparison is also given (table S7).

**Table S5. $^{13}C$ & $^{1}H$ δ assignments of (4') produced using heterologously expressed genes from *C. sinensis.***

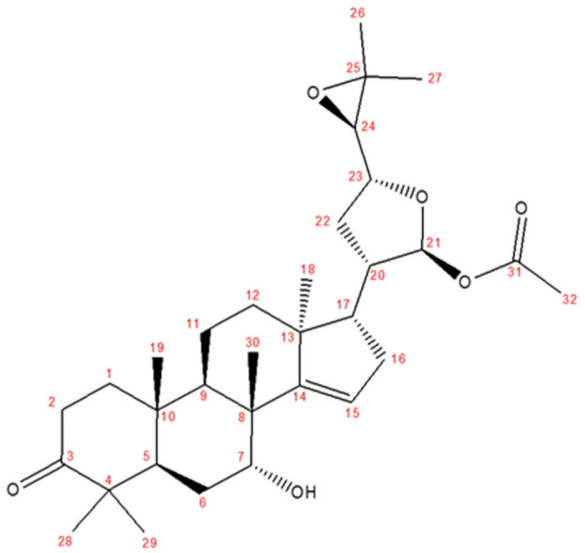
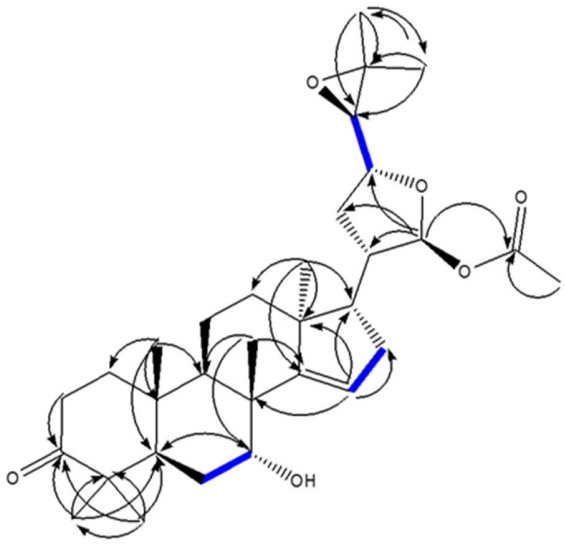
**Carbon numbering scheme and selected COSY and HMBC**




| C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | | C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | |
|---|---|---|---|---|---|---|---|
| **1** | 39.24 | 1.42 (1H, dt J = 13.2, 8.6) | 1.78 (1H, dt J = 13.2, 8.6) | **17** | 44.89 | 2.08 (1H, m) | |
| **2** | 33.91 | 2.43 (2H, dd J = 6.2, 8.6) | | **18** | 13.61 | 0.44 (1H, d J = 5.0) | 0.69 (1H, d J = 5.0) |
| **3** | 217.68 | / | | **19** | 15.74 | 0.92 (3H, s) | |
| **4** | 46.78 | / | | **20** | 48.03 | 2.02 (1H, m) | |
| **5** | 45.69 | 2.07 (1H, m) | | **21** | 97.55 | 6.25 (1H, d J = 3.8) | |
| **6** | 25.5 | 1.64 (2H, m) | | **22** | 30.64 | 1.63 (1H, m) | 2.01 (1H, m) |
| **7** | 73.88 | 3.76 (1H, appt t J = 2.7) | | **23** | 79.91 | 3.86 (1H, ddd J = 6.2, 7.5, 9.6) | |
| **8** | 36.88 | / | | **24** | 66.78 | 2.64 (1H, d J = 7.6) | |
| **9** | 42.96 | 1.32 (1H, m) | | **25** | 57.21 | / | |
| **10** | 36.86 | / | | **26** | 24.99 | 1.29 (3H, s) | |
| **11** | 16.66 | 1.26 (1H, m) | 1.33 (1H, m) | **27** | 19.40 | 1.24 (3H, s) | |
| **12** | 25.26 | 1.64 (1H, m) | 1.72 (1H, m) | **28** | 26.71 | 1.06 (3H, s) | |
| **13** | 28.75 | / | | **29** | 21.09 | 0.99 (3H, s) | |
| **14** | 38.75 | / | | **30** | 19.53 | 1.03 (3H, s) | |
| **15** | 26.32 | 1.56 (1H, dd J = 8.3, 12.6) | 1.92 (1H, m) | **31** | 170.01 | / | |
| **16** | 27.56 | 0.93 (1H, m) | 1.68 (1H, m) | **32** | 21.66 | 2.06 (3H, s) | |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined.

**Table S6. $^{13}C$ & $^{1}H$ δ assignments of 21(*S*)-acetoxyl-*apo*-melianone (6) produced using heterologously expressed genes from *M. azedarach.***

| Carbon numbering scheme and selected COSY and HMBC | | | | | | |
|---|---|---|---|---|---|---|
|  | | |  | | | |
| **C** | **$^{13}C$ δ (150 MHz)** | **$^{1}H$ δ (600 MHz)** | **C** | **$^{13}C$ δ (150 MHz)** | **$^{1}H$ δ (600 MHz)** | |
| **3** | 217.27 | / | **1** | 38.46 | 1.85 (1H, m) | 1.51 (1H, m) |
| **31** | 169.97 | / | **10** | 37.19 | / | |
| **14** | 161.62 | / | **16** | 35.09 | 2.22 (2H, m) | |
| **15** | 119.61 | 5.51 (1H, m) | **2** | 33.93 | 2.53 (1H, m) | 2.43 (1H, m) |
| **21** | 96.59 | 6.26 (1H, d J= 4.1) | **12** | 32.32 | 1.61 (1H, m) | 1.30 (1H, m) |
| **23** | 79.75 | 3.93 (1H, ddd, J= 7.6, 7.8, 10) | **22** | 31.40 | 2.08 (1H, m) | 1.70 (1H, m) |
| **7** | 71.95 | 3.96 (1H, appt t, J= 2.9) | **30** | 27.26 | 1.10 (3H, s) | |
| **24** | 66.71 | 2.67 (1H, d, J= 7.6) | **28** | 26.25 | 1.10 (3H, s) | |
| **25** | 57.17 | / | **26** | 24.93 | 1.33 (3H, s) | |
| **17** | 52.62 | 1.93 (1H, m) | **6** | 24.80 | 1.84 (1H, m) | 1.79 (1H, m) |
| **4** | 46.91 | / | **32** | 21.52 | 2.06 (3H, s) | |
| **13** | 46.61 | / | **29** | 21.18 | 1.05 (3H, s) | |
| **5** | 46.51 | 2.08 (1H, m) | **18** | 19.71 | 1.03 (3H, s) | |
| **20** | 44.21 | 2.37 (1H, m) | **27** | 19.35 | 1.29 (3H, s) | |
| **8** | 44.07 | / | **11** | 16.27 | 1.71 (1H, m) | 1.56 (1H, m) |
| **9** | 40.80 | 2.01 (1H, m) | **19** | 14.95 | 1.00 (3H, s) | |

NMR spectra were recorded in $CDCl_3$, referenced to TMS and characterization was performed following the general considerations outlined. The compound was assigned as the C21(*S*) epimer on the basis of observed NOEs (fig. S15), also consistent with the literature (table S7).

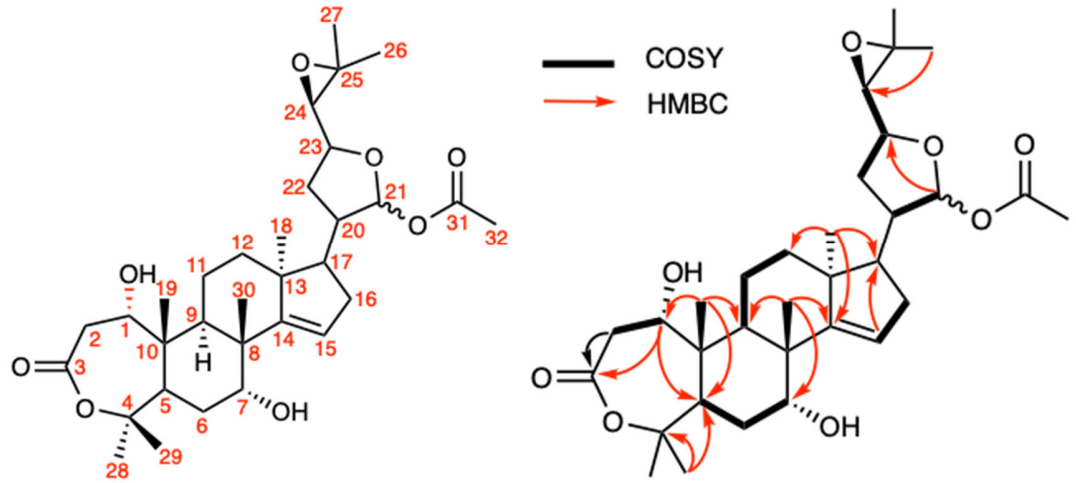
**Table S7. $^{13}C$ δ comparison with the literature for 21(*S*)-acetoxyl-*apo*-melianone (6).**

| C | Literature* | *M. azedarach* (rounded) | Δ Literature to *M. azedarach* | *C. sinensis* (rounded) | Δ Literature to *C. sinensis* |
|---|---|---|---|---|---|
| **3** | **217.2** | 217.3 | 0.1 | 217.31 | 0.11 |
| **31** | **170** | 170 | 0 | 170.05 | 0.05 |
| **14** | **161.5** | 161.6 | 0.1 | 161.7 | 0.2 |
| **15** | **119.6** | 119.6 | 0 | 119.72 | 0.12 |
| **21** | **96.6** | 96.6 | 0 | 96.7 | 0.1 |
| **23** | **79.7** | 79.7 | 0 | 79.86 | 0.16 |
| **7** | **71.9** | 71.9 | 0 | 72.05 | 0.15 |
| **24** | **66.7** | 66.7 | 0 | 66.81 | 0.11 |
| **25** | **57.1** | 57.2 | 0.1 | 57.25 | 0.15 |
| **17** | **52.6** | 52.6 | 0 | 52.73 | 0.13 |
| **4** | **46.9** | 46.9 | 0 | 47 | 0.1 |
| **5** | **46.5** | 46.5 | 0 | 46.71 | 0.21 |
| **13** | **46.5** | 46.6 | 0.1 | 46.62 | 0.12 |
| **20** | **44.2** | 44.2 | 0 | 44.33 | 0.13 |
| **8** | **44** | 44.1 | 0.1 | 44.17 | 0.17 |
| **9** | **40.8** | 40.8 | 0 | 40.91 | 0.11 |
| **1** | **38.5** | 38.5 | 0 | 38.58 | 0.08 |
| **10** | **37.1** | 37.2 | 0.1 | 37.3 | 0.2 |
| **16** | **35.1** | 35.1 | 0 | 35.19 | 0.09 |
| **2** | **33.9** | 33.9 | 0 | 34.03 | 0.13 |
| **12** | **32.3** | 32.3 | 0 | 32.45 | 0.15 |
| **22** | **31.3** | 31.4 | 0.1 | 31.5 | 0.2 |
| **30** | **27.2** | 27.3 | 0.1 | 27.36 | 0.16 |
| **28** | **26.2** | 26.2 | 0 | 26.35 | 0.15 |
| **6** | **24.9** | 24.8 | -0.1 | 24.92 | 0.02 |
| **26** | **24.9** | 24.9 | 0 | 25.03 | 0.13 |
| **32** | **21.5** | 21.5 | 0 | 21.61 | 0.11 |
| **29** | **21.1** | 21.2 | 0.1 | 21.28 | 0.18 |
| **18** | **19.7** | 19.7 | 0 | 19.82 | 0.12 |
| **27** | **19.3** | 19.4 | 0.1 | 19.45 | 0.15 |
| **11** | **16.3** | 16.3 | 0 | 16.38 | 0.08 |
| **19** | **14.9** | 14.9 | 0 | 15.04 | 0.14 |

Comparison of the $^{13}C$ δ values for (**6**) from literature (100 mHZ) and this work (150 mHZ), for (**6**) purified from heterologous expression of *M. azedarach* and *C. sinensis* enzymes. All NMRs were performed in $CDCl_3$. Asterix (*) refers to literature assignment present in (*32*). Full assignment of (**6**) purified from heterologous expression of *M. azedarach* (table S6) and *C. sinensis* (table S4) enzymes are provided.

**Table S8. $^{13}C$ & $^{1}H$ δ assignments of 1-hydroxyl luvungin A (9) produced using heterologously expressed genes from *C. sinensis.***

**Carbon numbering scheme and selected COSY and HMBC**

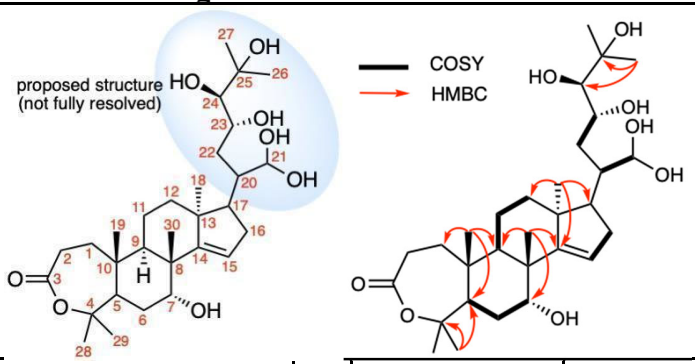



| C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | | C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | |
|---|---|---|---|---|---|---|---|
| **1** | 68.71 | 3.72 (1H, d J = 7.5) | | **17** | 52.53 | 1.90 (1H, m) | |
| **2** | 39.12 | 2.90 (1H, dd J = 7.5, 15.5) | 3.22 (1H, d J = 15.5) | **18** | 19.20 | 1.04 (3H, s) | |
| **3** | 170.12 | / | | **19** | 15.53 | 1.05 (3H, s) | |
| **4** | 86.37 | / | | **20** | 44.32 | 2.37 (1H, m) | |
| **5** | 41.65 | 2.70 (1H, d = 12.3) | | **21** | 96.76 | 6.24 (1H, d J = 4.0)* | |
| **6** | 27.19 | 1.82 (1H, m) | 1.97 (1H, m) | **22** | 31.48 | 1.55 (1H, m) | 2.07 (1H, m) |
| **7** | 71.53 | 3.87 (1H, br) | | **23** | 79.87 | 3.92 (1H, dt J = 10.1, 7.2) | |
| **8** | 43.88 | / | | **24** | 66.87 | 2.66 (1H, d J = 7.6) | |
| **9** | 33.61 | 2.73 (1H, dd J = 7.7, 11.5) | | **25** | 57.41 | / | |
| **10** | 45.47 | / | | **26** | 25.05 | 1.32 (3H, s) | |
| **11** | 16.43 | 1.45 (1H, m) | 1.85 (1H, m) | **27** | 19.45 | 1.28 (3H, s) | |
| **12** | 32.57 | 1.26 (1H, m) | 1.62 (1H, m) | **28** | 34.39 | 1.46 (3H, s) | |
| **13** | 46.72 | / | | **29** | 23.78 | 1.46 (3H, s) | |
| **14** | 161.92 | / | | **30** | 28.05 | 1.10 (3H, s) | |
| **15** | 119.70 | 5.48 (1H, t J = 2.4) | | **31** | 172.43 | / | |
| **16** | 35.23 | 2.21 (2H, m) | | **32** | 21.64 | 2.04 (3H, s) | |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. 1-hydroxyl-luvungin A (**9**) was purified as a pair of C-21 epimers in a ratio of ~5:1. The most significant difference between the two spectra was the $^{1}H$ δ of C-21 (marked with *). The minor epimer showed a $^{1}H$ δ of 6.28 ppm (d, J = 3.2). The absolute stereochemistry of the epimers were not resolved.

**Table S9. $^{13}C$ & $^{1}H$ δ partial assignments of degraded luvungin A (7) produced using heterologously expressed genes from *C. sinensis.***

**Carbon numbering scheme and selected COSY and HMBC**




| C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | | C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) |
|---|---|---|---|---|---|---|
| **1** | 37.58 | 1.76 (2H, m) | | **16** | 35.04 | 2.17 (2H, m) |
| **2** | N/A | N/A | | **17** | 52.92 | N/A |
| **3** | 161.37 | / | | **18** | 20.09 | 1.02 (3H, s) |
| **4** | 85.98 | / | | **19** | 16.42 | 1.10 (3H, s) |
| **5** | 46.31 | 2.37 (1H, m) | | **20** | N/A | 2.15 (1H, m) |
| **6** | N/A | 1.81 (1H, m) | 1.88 (1H, m) | **21** | 97.09 | 5.28 (1H, s) |
| **7** | 71.58 | 3.90 (1H, s) | | **22** | N/A | 1.90 (2H, m) |
| **8** | 43.91 | / | | **23** | 78.50 | 4.46 (1H, m) |
| **9** | 41.41 | 2.65 (1H, m) | | **24** | 75.22 | 3.19 (1H, m) |
| **10** | 40.21 | / | | **25** | 73.17 | / |
| **11** | N/A | 1.58 (1H, m) | 1.75 (1H, m) | **26** | 26.71 | 1.26 (3H, s) |
| **12** | 33.17 | N/A | | **27** | 26.64 | 1.29 (3H, s) |
| **13** | 46.29 | / | | **28** | 32.05 | 1.49 (3H, s) |
| **14** | 161.12 | / | | **29** | 26.00 | 1.43 (3H, s) |
| **15** | 120.04 | 5.48 (1H, s) | | **30** | 26.86 | 1.09 (3H, s) |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. Proposed structure of the degraded product of luvungin (**7**) is shown, the blue shaded area indicates the uncertain structural moiety. Partial assignment of the degraded product was achieved through comparison with the complete NMR assignment of (**9**) (table S8). While the exact functional groups on C-21 and C-23~25 couldn't be fully resolved by the NMR due to overlapped signals and low signal intensities, the higher $^{13}C$ δ of C-24,25 (75.22 and 73.17 ppm) compared to those in (**9**) (66.87 and 57.41 ppm) suggested that the C-24,25 epoxide was opened. HMBC correlation from C-28/29 to C-4 and the presence of C-3 ketone ($^{13}C$ δ = 161.12 ppm) were two key pieces of evidence supporting the A-ring lactone structure, which was further corroborated by the complete assignment of (**9**) (Table S8). N/A indicates incomplete assignment due to poor signal or signal overlap.

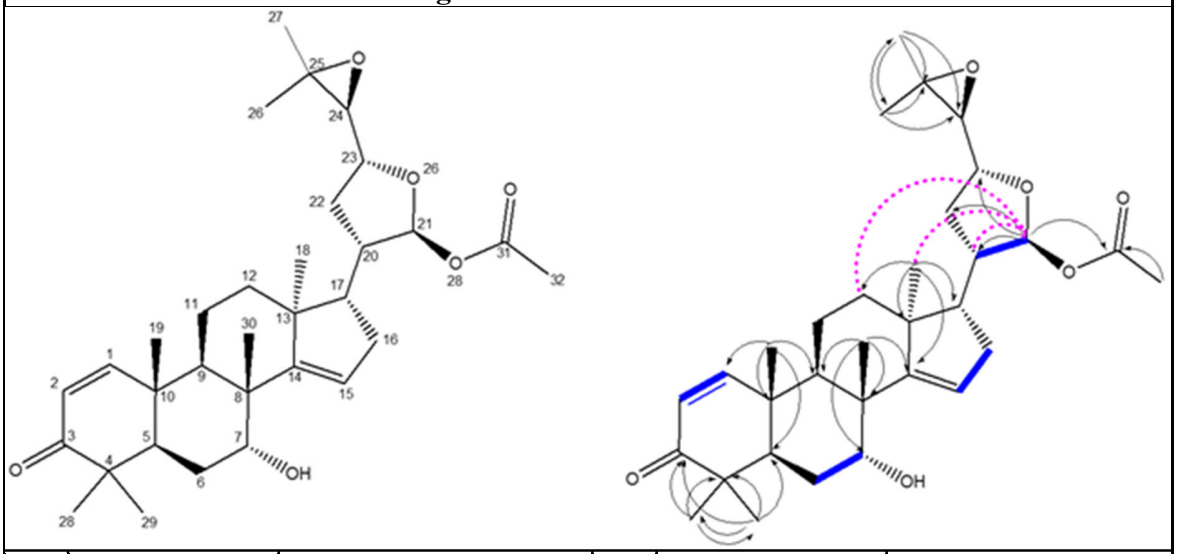
**Table S10. Gene ID of active *Melia azedarach* limonoid biosynthetic genes in this study.**

| # | Name | GeneID (*M. azedarach* genome) | GenBank (genome) | GenBank (transcriptome) |
|---|---|---|---|---|
| **1** | *MaOSC1** | MELAZ155640_EIv1_0159960.1 | | MK803261 |
| **2** | *MaCYP71CD2* | MELAZ155640_EIv1_0070910.1 | | MK803271 |
| **3** | *MaCYP71BQ5* | MELAZ155640_EIv1_0148050.1 | | MK803264 |
| **4** | *MaCYP88A108* | MELAZ155640_EIv1_0061960.1 | OP947595 | MK803265 |
| **5** | *MaMOI2**** | MELAZ155640_EIv1_0192980.1 | OP947596 | |
| **6** | *MaL21AT* | MELAZ155640_EIv1_0142070.1 | OP947597 | |
| **7** | *MaSDR* | MELAZ155640_EIv1_0198190.1 | OP947598 | |
| **9** | *MaCYP88A164* | MELAZ155640_EIv1_0061950.1 | OP947599 | |
| **10** | *MaL1AT* | MELAZ155640_EIv1_0164450.1 | OP947600 | |
| **11** | *MaL7AT* | MELAZ155640_EIv1_0235630.1 | OP947601 | |
| **12** | *MaAKR*** | MELAZ155640_EIv1_0165520.1 | | OP947602 |
| **13** | *MaCYP716AD4* | MELAZ155640_EIv1_0052990.1 | OP947603 | |
| **14** | *MaLFS* | MELAZ155640_EIv1_0015190.1 | OP947604 | |
| | *MaMOI1* | MELAZ155640_EIv1_0192990.1 | | |
| | *MaSI* | MELAZ155640_EIv1_0193000.1 | | |
| | Closest *CsCYP716AC1* seq**** | MELAZ155640_EIv1_0122250.1 | | |

Gene name and relevant ID from *M. azedarach* genome (or transcriptome data) for all functional *M.azedarach* genes (numbered in order of reported occurrence) described in this study as well as the additional sterol isomerases and cytochrome p450s mentioned. Asterisks denote the following. (*) indicates that *MaOSC1* is the *Melia azedarach* version of a tirucalla-7,24-dien-3β-ol synthase (previously characterized (*20*)), however the *A. indica* version (AiOSC1, GenBank:MK803262 (*20*)) was used for all experimental work in this paper. (**) indicates *MaAKR* was identified as a candidate based on sequence similarity to *Cs*AKR, however is truncated in the *M. azedarach* genome annotation (potentially accounting for its lower ranking than other functional genes (Fig. 2C)), a full-length copy (TRINITY_DN15268_c1_g3_i2.p1, table S20) was identified in a transcriptome assembly constructed from *M. azedarach* petiole RNA-seq data. (***) indicates that the functional sequence for *MaMOI2,* which was cloned and used in this study, contained the first intron in addition to the exons. Due to its functionality in *N. benthamiana* it is assumed this intron is spliced out *in planta* to achieve functionality, as the resultant protein without splicing would be truncated. The full cloned sequence with intron indicated is available (table S24). (****) indicates that this gene is truncated and not co-expressed (PCC: -0.137, Rank:15335). GenBank accession numbers are given for all functional genes discussed in this paper, for sequences derived from the *M. azedarach* genome as well as transcriptomic resources, either newly generated or from pre-existing work (*20*).

**Table S11. $^{13}C$ & $^{1}H$ δ assignments of *epi*-neemfruitin B (10) produced using heterologously expressed genes from *M. azedarach.***

**Carbon numbering scheme and selected COSY and HMBC**



| C | $^{13}C$ δ (150 MHz) | $^{1}H$ δ (600 MHz) | C | $^{13}C$ δ (150 MHz) | $^{1}H$ δ (600 MHz) |
|---|---|---|---|---|---|
| **3** | 205.18 | / | **20** | 44.32 | 2.38 (1H, m) |
| **31** | 170.13 | / | **10** | 40.34 | / |
| **14** | 161.60 | / | **9** | 36.61 | 2.20 (1H, m) |
| **1** | 158.14 | 7.10 (1H, d J= 10.2) | **16** | 35.23 | 2.23 (2H, m) |
| **2** | 125.71 | 5.83 (1H, d J= 10.2) | **12** | 32.41 | 1.68 (1H, m)<br>1.36 (1H, m) |
| **15** | 119.85 | 5.52 (1H, m) | **22** | 31.52 | 2.09 (1H, m)<br>1.71 (1H, m) |
| **21** | 96.72 | 6.27 (1H, d J= 4.1) | **30** | 27.87 | 1.13 (3H, s) |
| **23** | 79.90 | 3.93 (1H, m) | **29** | 27.27 | 1.16 (3H, s) |
| **7** | 71.61 | 3.99 (1H, m) | **26** | 25.07 | 1.33 (3H, s) |
| **24** | 66.83 | 2.67 (1H, d J= 7.60) | **6** | 24.39 | 1.88 (2H, m) |
| **25** | 57.35 | / | **32** | 21.67 | 2.07 (3H, s) |
| **17** | 52.79 | 1.95 (1H, m) | **28** | 21.64 | 1.09 (3H, s) |
| **13** | 46.72 | / | **18** | 19.92 | 1.03 (3H, s) |
| **8** | 44.90 | / | **27** | 19.49 | 1.29 (3H, s) |
| **5** | 44.64 | 2.39 (1H, m) | **19** | 19.04 | 1.16 (3H, s) |
| **4** | 44.34 | / | **11** | 16.42 | 1.96 (1H, m)<br>1.70 (1H, m) |

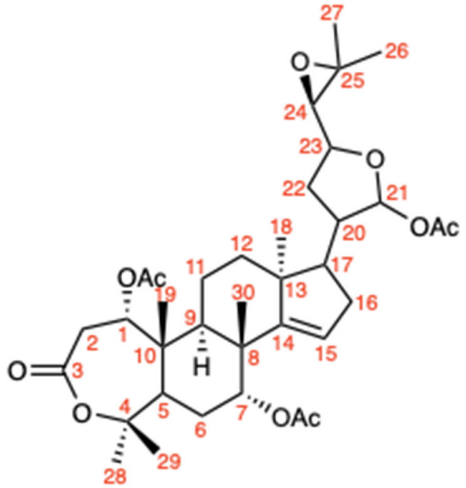
NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. Opposite stereochemistry at C21 to previously reported neemfruitin B assigned due to NOEs observed between C21-H and C18-H3 and C12-H2. This is consistent to those observed for 21(*S*)-acetoxyl-*apo*-melianone (**6**) (table S6, fig. S15) and different to those reported for neemfruitin B (fig. S10) (*33*).

**Table S12. $^{13}C$ δ comparison with the literature for *epi*-neemfruitin B (10) to neemfruitin B.**

| C | *epi*-neemfruitin B this work (as reported) | *epi*-neemfruitin B this work (rounded) | neemfruitin B literature* (as reported) | Δ |
|---|---|---|---|---|
| **3** | 205.18 | 205.2 | **205.8** | 0.6 |
| **31** | 170.13 | 170.1 | **170.7** | 0.6 |
| **14** | 161.60 | 161.6 | **161.9** | 0.3 |
| **1** | 158.14 | 158.1 | **161.8** | 3.7 |
| **2** | 125.71 | 125.7 | **127.2** | 1.5 |
| **15** | 119.85 | 119.9 | **119.9** | 0.0 |
| **21** | 96.72 | 96.7 | **96.8** | 0.1 |
| **23** | 79.90 | 79.9 | **80.3** | 0.4 |
| **7** | 71.61 | 71.6 | **72** | 0.4 |
| **24** | 66.83 | 66.8 | **67** | 0.2 |
| **25** | 57.35 | 57.3 | **57.8** | 0.5 |
| **17** | 52.79 | 52.8 | **53.1**** | 0.3 |
| **13** | 46.72 | 46.7 | **46.7**** | 0.0 |
| **8** | 44.90 | 44.9 | **44.8** | -0.1 |
| **5** | 44.64 | 44.6 | **44.7** | 0.1 |
| **4** | 44.34 | 44.3 | **44.7** | 0.4 |
| **20** | 44.32 | 44.3 | **44.3** | 0.0 |
| **10** | 40.34 | 40.3 | **40.5** | 0.2 |
| **9** | 36.61 | 36.6 | **36.8** | 0.2 |
| **16** | 35.23 | 35.2 | **35.4** | 0.2 |
| **12** | 32.41 | 32.4 | **33.4** | 1.0 |
| **22** | 31.52 | 31.5 | **32.3** | 0.8 |
| **30** | 27.87 | 27.9 | **27.4** | -0.5 |
| **29** | 27.27 | 27.3 | **27.1** | -0.2 |
| **26** | 25.07 | 25.1 | **25.9** | 0.8 |
| **6** | 24.39 | 24.4 | **24** | -0.4 |
| **32** | 21.67 | 21.7 | **23.2** | 1.5 |
| **28** | 21.64 | 21.6 | **21.3** | -0.3 |
| **18** | 19.92 | 19.9 | **21.2** | 1.3 |
| **27** | 19.49 | 19.5 | **19.5** | 0.0 |
| **19** | 19.04 | 19.0 | **18.9** | -0.1 |
| **11** | 16.42 | 16.4 | **16.8** | 0.4 |

Comparison of $^{13}C$ δ values for neemfruitin B from the literature and for *epi*-neemfruitin B (**10**) from this work. Asterisks refer to the following: (*) literature assignment present in (*33*) and (**) values believed to be mis-assigned in literature. Full-assignment of *epi*-neemfruitin B (**10**) is available (table S11).

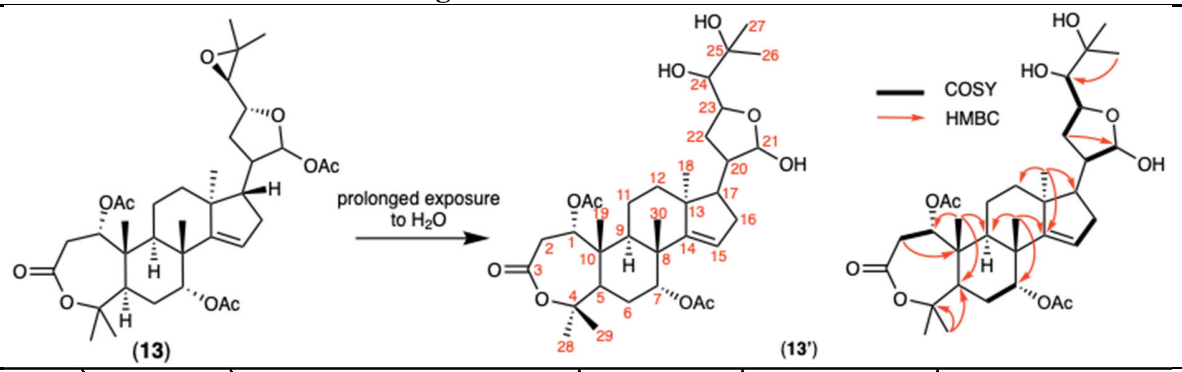
**Table S13. $^1$H δ assignments of L7AT product (13) produced using heterologously expressed genes from *C. sinensis.***

| Carbon numbering scheme | | | |
|---|---|---|---|
|  | | | |
| **C** | **$^1$H δ (ppm, J in Hz)** | **C** | **$^1$H δ (ppm, J in Hz)** |
| **1** | 4.77 (1H, d J = 5.9) | **19** | 1.15 (3H, s) |
| **2** | 3.15 (2H, m) | **20** | 2.31 (1H, dddd J = N/A) |
| **3** | / | **21** | 6.22 (1H, d J = 4.1) |
| **4** | / | **22** | 1.68 (1H, m)<br>2.05 (1H, m) |
| **5** | 2.53 (1H, m) | **23** | 3.9 (1H, dt J = 10.1, 7.0) |
| **6** | 1.88 (1H, m)<br>1.94 (1H, d m) | **24** | 2.65 (1H, d J = 7.7) |
| **7** | 5.16 (1H, m) | **25** | / |
| **8** | / | **26** | 1.28 (3H, s) |
| **9** | N/A | **27** | 1.32 (3H, s) |
| **10** | / | **28** | 1.39 (3H, s) |
| **11** | N/A | **29** | 1.49 (3H, s) |
| **12** | N/A | **30** | 1.14 (3H, s) |
| **13** | / | **-OCO<u>C</u>H$_3$** | 1.99 (3H, s) |
| **14** | / | **-OCOCH$_3$** | 2.04 (3H, s) |
| **15** | 5.28 (1H, d J = 2.2) | **-OCOCH$_3$** | 2.09 (3H, s) |
| **16** | 2.13 (2H, m) | **-O<u>C</u>OCH$_3$** | / |
| **17** | 1.88 (1H, m) | **-O<u>C</u>OCH$_3$** | / |
| **18** | 0.97 (3H, 2) | **-O<u>C</u>OCH$_3$** | / |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. N/A indicates incomplete assignment due to poor signal or signal overlap.

**Table S14. $^{13}C$ & $^{1}H$ δ assignments of (13'), degradation product of (13) produced using heterologously expressed genes from *C. sinensis.***

**Carbon numbering scheme and selected COSY and HMBC**

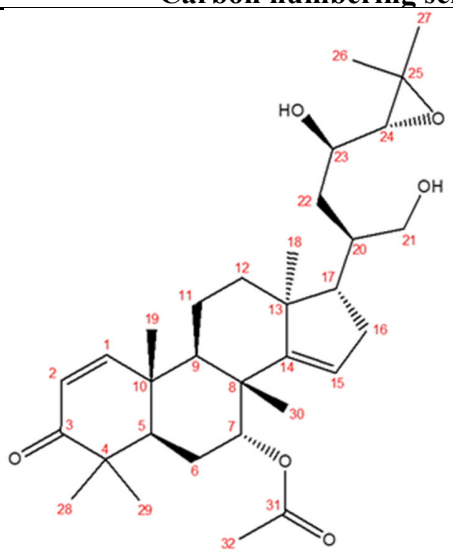


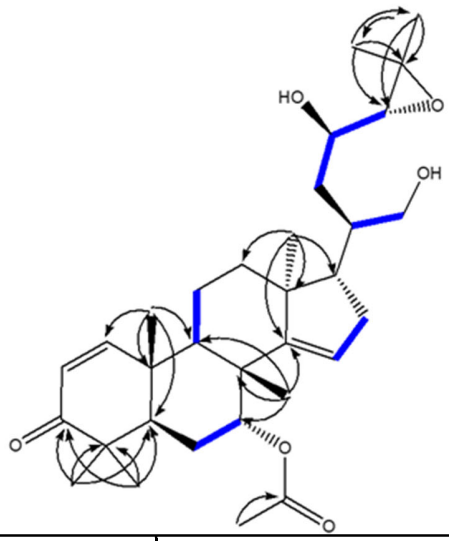

| C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) | C | $^{13}C$ δ (ppm) | $^{1}H$ δ (ppm, J in Hz) |
|---|---|---|---|---|---|
| **1** | 71.16 | 4.83 (1H, dd J = 1.5, 6.0) | **18** | 19.55 | 0.95 (3H, s) |
| **2** | 35.08 | 3.15 (2H, m) | **19** | 15.34 | 1.14 (3H, s) |
| **3** | 170.51 | / | **20** | 44.92 | 2.09 (1H, m) |
| **4** | 85.74 | / | **21** | 96.94/102.87* | 5.26 (1H, d J = 3.8) |
| **5** | 44.26 | 2.51 (1H, dd J = 2.3, 12.9) | **22** | 30.16 | 1.87 (1H, m)<br>1.98 (1H, m) |
| **6** | 26.46 | 1.86 (1H, m)<br>1.94 (1H, m) | **23** | 78.72 | 4.47 (1H, t J = 8.4) |
| **7** | 74.43 | 5.14 (1H, m) | **24** | 75.16 | 3.15 (1H, m) |
| **8** | 41.96 | / | **25** | 73.68 | / |
| **9** | 35.75 | 2.53 (1H, dd J = 7.2, 11.7) | **26** | 26.86 | 1.25 (3H, s) |
| **10** | 44.39 | / | **27** | 26.81 | 1.28 (3H, s) |
| **11** | 16.45 | 1.48 (2H, m) | **28** | 34.58 | 1.39 (3H, s) |
| **12** | 33.51 | 1.54 (1H, m)<br>1.75 (1H, dd J = 9.0, 12.3) | **29** | 23.76 | 1.49 (3H, s) |
| **13** | 46.37 | / | **30** | 27.48 | 1.14 (3H, s) |
| **14** | 159.02 | / | **-OCOCH$_3$** | 21.00 | 1.98 (3H, s) |
| **15** | 119.22 | 5.25 (1H, m) | **-OCO<u>C</u>H$_3$** | 21.24 | 2.10 (3H, s) |
| **16** | 35.21 | 2.02 (1H, m)<br>2.12 (1H, m) | **O<u>C</u>OCH$_3$** | 170.02 | / |
| **17** | 52.78 | 1.94 (1H, m) | **O<u>C</u>OCH$_3$** | 170.17 | / |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. (*) indicates value from C-21 epimers.

**Table S15. $^{13}C$ & $^{1}H$ δ assignments of AKR product (14) produced using heterologously expressed genes from *M. azedarach.***

**Carbon numbering scheme and selected COSY and HMBC**

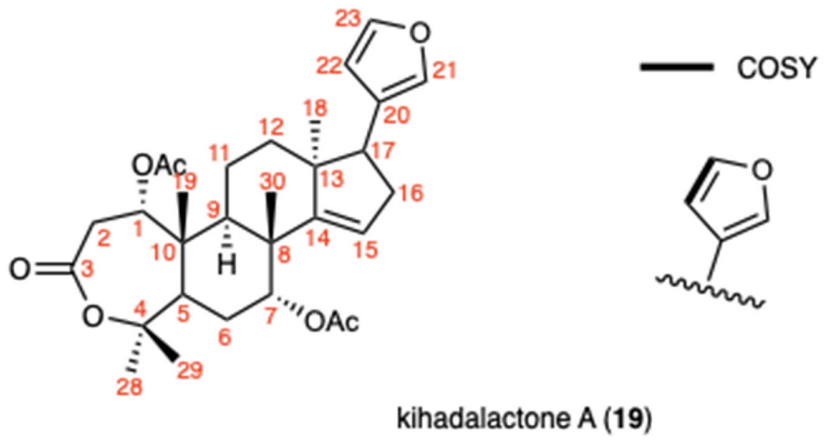



| C | $^{13}C$ δ (150 MHz) | $^{1}H$ δ (600 MHz) | C | $^{13}C$ δ (150 MHz) | $^{1}H$ δ (600 MHz) |
|---|---|---|---|---|---|
| **3** | 203.06 | / | **20** | 41.34 | 1.81 (1H, m) |
| **31** | 169.22 | / | **10** | 39.83 | / |
| **14** | 160.00 | / | **9** | 39.07 | 2.07 (1H, m) |
| **1** | 157.23 | 6.69 (1H, d J= 10.2) | **22** | 36.17 | 1.60 (1H, m)<br>1.47 (1H, m) |
| **2** | 125.88 | 5.90 (1H, d J= 10.2) | **16** | 35.43 | 2.08 (1H, m)<br>1.86 (1H, m) |
| **15** | 119.19 | 5.34 (1H, m) | **12** | 34.87 | 1.84 (1H, m)<br>1.52 (1H, m) |
| **7** | 74.54 | 5.29 (1H, m) | **30** | 27.39 | 0.91 (3H, s) |
| **23** | 71.28 | 3.48 (1H, m) | **29** | 27.34 | 1.15 (3H, s) |
| **24** | 67.86 | 2.64 (1H, d J= 7.9) | **26** | 24.87 | 1.09 (3H, s) |
| **21** | 64.58 | 3.94 (1H, dd J= 11.0, 3.1)<br>3.54 (1H, dd J= 11.0, 6.5) | **6** | 24.15 | 1.67 (1H, m)<br>1.57 (1H, m) |
| **25** | 58.85 | / | **28** | 21.42 | 1.00 (3H, s) |
| **17** | 56.20 | 1.73 (1H, m) | **32** | 20.83 | 1.65 (3H, s) |
| **13** | 46.86 | / | **18** | 20.05 | 1.00 (3H, s) |
| **5** | 46.65 | 2.19 (1H, dd J= 13.1, 2.5) | **27** | 19.36 | 1.10 (3H, s) |
| **4** | 44.26 | / | **19** | 18.98 | 0.81 (3H, s) |
| **8** | 43.01 | / | **11** | 16.96 | 1.52 (1H, m)<br>1.31 (1H, m) |

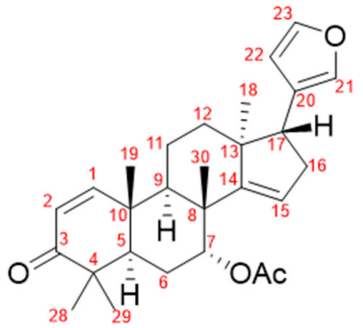
NMR spectra were recorded in benzene-$d_6$, referenced to 7.16 and 128.06, following the general considerations outlined.

**Table S16. $^1$H δ assignments of the furan moiety for kihadalactone A (19) produced using heterologously expressed genes from *C. sinensis.***

| Carbon numbering scheme and selected COSY | | | |
|---|---|---|---|
|   | | | |
| **C** | **$^1$H δ (ppm, J in Hz)** | **$^1$H δ literature** | **Δ** |
| **21** | 7.19 (1H, m) | 7.23 | -0.04 |
| **22** | 6.25 (1H, m) | 6.26 | -0.01 |
| **23** | 7.37 (1H, appt t J = 1.7) | 7.37 | 0.00 |

NMR spectra were recorded in $CDCl_3$, referenced to TMS, and characterization was performed following the general considerations outlined. While complete $^1$H δ assignment of kihadalactone A (**19**) was hampered by its low yield and co-eluting impurities, the signature furan moiety for limonoids was clearly distinguishable from other peaks on NMR $^1$H spectrum, and the assignment for furan moiety is shown here. The chemical shifts and coupling constant are consistent with literature values (*34*), supporting the presence of (**19**).

**Table S17. $^{13}C$ δ comparison with literature values for azadirone (18)**

| Carbon numbering scheme | | | |
|---|---|---|---|
|  | | | |
| **C** | **Azadirone this work (as reported)** | **Azadirone literature (as reported)** | **Δ** |
| **3** | 204.64 | 204.58 | 0.06 |
| **7αCOCH$_3$** | 170.16 | 170.11 | 0.05 |
| **14** | 158.85 | 158.78 | 0.07 |
| **1** | 158.2 | 158.18 | 0.02 |
| **23** | 142.59 | 142.52 | 0.07 |
| **21** | 139.71 | 139.63 | 0.08 |
| **2** | 125.5 | 125.41 | 0.09 |
| **20** | 124.59 | 124.52 | 0.07 |
| **15** | 119.11 | 119.01 | 0.1 |
| **22** | 111.06 | 111 | 0.06 |
| **7** | 74.52 | 74.42 | 0.1 |
| **17** | 51.63 | 51.53 | 0.1 |
| **13** | 47.18 | 47.1 | 0.08 |
| **5** | 46.15 | 46.05 | 0.1 |
| **4** | 44.16 | 44.07 | 0.09 |
| **8** | 42.81 | 42.73 | 0.08 |
| **10** | 39.96 | 39.87 | 0.09 |
| **9** | 38.66 | 38.55 | 0.11 |
| **16** | 34.38 | 34.3 | 0.08 |
| **12** | 32.99 | 32.89 | 0.1 |
| **30** | 27.35 | 27.28 | 0.07 |
| **28** | 27.07 | 26.99 | 0.08 |
| **6** | 23.81 | 23.73 | 0.08 |
| **18** | 21.32 | 21.26 | 0.06 |
| **7αCOCH$_3$** | 21.18 | 21.13 | 0.05 |
| **29** | 20.64 | 20.56 | 0.08 |
| **19** | 19.08 | 19.02 | 0.06 |
| **11** | 16.51 | 16.43 | 0.08 |

Comparison of $^{13}C$ δ values for azadirone (**18**), isolated for *A. indica* leaf powder, in this work (150 mHZ) with the literature assignment (100 mHZ) (*81*).

**Table S18. Gene ID/Accession numbers of active *Citrus* limonoid biosynthetic genes and other *Citrus* genes in this study.**

| | Gene Name | Gene ID | NCBI accession number |
|---|---|---|---|
| **1** | *CsOSC1* | XM_006468053 | |
| **2** | *CsCYP71CD1* | XM_006467236 | |
| **3** | *CsCYP71BQ4* | XM_006469432 | |
| **4** | *CsCYP88A51* | XM_006485364 | OQ091247 |
| **5** | *CsMOI1* | XM_006478528 | OQ091248 |
| **6** | *CsMOI2* | XM_006494479 | OQ091249 |
| **7** | *CsMOI3* | XM_006471624 | |
| **8** | *CsL21AT* | XM_006482023 | OQ091241 |
| **9** | *CsSDR* | XM_006481636 | OQ091238 |
| **10** | *CsCYP716AC1* | XM_006464942 | OQ091239 |
| **11** | *CsCYP88A37* | XM_006485365 | OQ091240 |
| **12** | *CsL1AT* | XM_006478966 | OQ091242 |
| **13** | *CsL7AT* | Cs1g05840.1 | OQ091243 |
| **14** | *CsAKR* | XM_006492221 | OQ091244 |
| **15** | *CsCYP716AD2* | XM_006494121 | OQ091245 |
| **16** | *CsLFS* | Cs5g20040.1 | OQ091246 |
| **17** | *CsSI* | XM_006478527 | |

The 12 genes cloned and characterized from *C. sinensis* with gene ID either from NCBI BioProject PRJNA86123 (*82*) or NICCE (*22*). All newly characterized genes have been deposited and accession numbers are given.

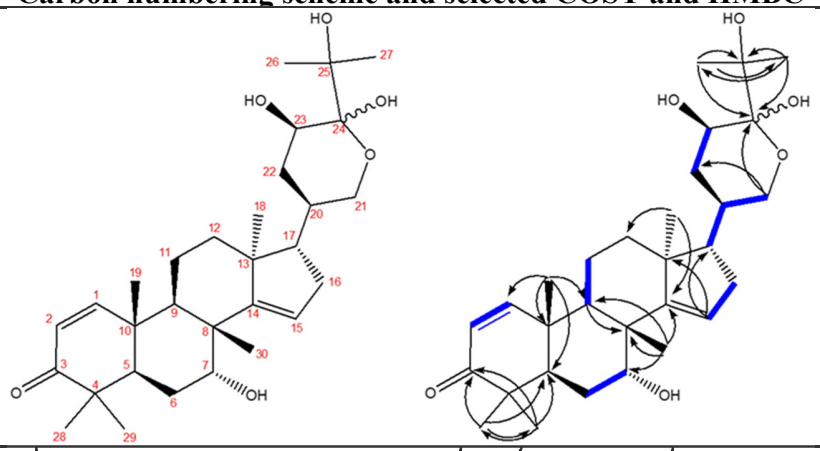
**Table S19. Full length CDS and peptide sequence of MaAKR (transcriptome derived).**

| | |
|---|---|
| **CDS** | >MaAKR<br>ATGGCGAAAACAGTGAGCATTCCTTCTGTAACCCTAGGCTCAACAGGCATAACCA<br>TGCCCCTTGTTGGGTTCGGAACGGTGGAATATCCTTTATGTGAATGGTTTAAAGA<br>CGCCGTTCTCCATGCAATCAAACTCGGATACAGACACTTCGATACTGCTTCAACT<br>TACCCTTCAGAACAGCCTCTTGGTGAAGCCATCACCGAAGCTCTCCGCCTCGGCC<br>TCATAAAATCCCGCGACGAGCTCTTCATCACTTCCAAGCTCTGGCTCACCGATTC<br>CTTCCCTGACCGCGTCATCCCGGCGCTGAAGAAATCTCTCAAGAATATGGGATTG<br>GAGTACTTGGATTGTTATCTGATTCATTTTCCGGTGTGTTTGATTCCGGAGGCGA<br>CGTATCCGGTGAAGAAGGAGGATATTCGTCCGATGGATTTTGAGGGTGTGTGGGC<br>TGCAATGGAGGAATGTCAAAAGCTTGGTCTTACCAAAACCATTGGAGTAAGCAAC<br>TTTACTGCCAAAAAACTCGAGAGGATACTTGCTACTGCAAAAATCCTTCCGGCTG<br>TCAATCAGGTGGAGATGAACCCAGTATGGCAACAAAAGAAGCTGAGGCAGTTTTG<br>TGAAGAAAAAGGCATACATTTCTCAGCTTTCTCTCCATTAGGAGCCGTAGGAACA<br>GACTGGGGACATAATCGAGTCATGGAATGTGAGGTGCTGAAAGAGATTGCAAAAG<br>CTAAAGGAAAATCACTTGCTCAGATTGCAATCCGTTGGGTTTACCAACAAGGAGT<br>GAGTGTGATTACAAAGAGCTTTAACAAACAAAGAATGGAAGAGAACCTGGACATA<br>TTTGACTGGAAGTTGACTCCTGAAGAGCTACACAAGATTGATCAAATTCCACAGT<br>ATAGAGGAAGTCGTGGTGAGACTTTTGTTTCAGAAAATGGTCCTTACAAAACTCT<br>TGAAGAAATGTGGGACGGAGAGATTTAA |
| **peptide** | >MaAKR<br>MAKTVSIPSVTLGSTGITMPLVGFGTVEYPLCEWFKDAVLHAIKLGYRHFDTAST<br>YPSEQPLGEAITEALRLGLIKSRDELFITSKLWLTDSFPDRVIPALKKSLKNMGL<br>EYLDCYLIHFPVCLIPEATYPVKKEDIRPMDFEGVWAAMEECQKLGLTKTIGVSN<br>FTAKKLERILATAKILPAVNQVEMNPVWQQKKLRQFCEEKGIHFSAFSPLGAVGT<br>DWGHNRVMECEVLKEIAKAKGKSLAQIAIRWVYQQGVSVITKSFNKQRMEENLDI<br>FDWKLTPEELHKIDQIPQYRGSRGETFVSENGPYKTLEEMWDGEI* |

Coding sequence (cds) of cloned and full-length version of *MaAKR* (GenBank: OP947602), which was identified as a candidate based on sequence similarity to *Cs*AKR, however was truncated in the *M. azedarach* genome annotation. Therefore the full-length copy identified above, was sourced in a transcriptome assembly constructed *de novo* from *M. azedarach* petiole RNA-seq data (table S2) using trinity (*65*, *66*) .

**Table S20. $^{13}C$ & $^{1}H$ δ assignments of *Ma*CYP716AD4 side-product (C24 epimeric mixture) (20) produced using heterologously expressed genes from *M. azedarach.***

**Carbon numbering scheme and selected COSY and HMBC**




| C | $^{13}C$ δ (150 Mhz) | | $^{1}H$ δ (600 MHz) | | C | $^{13}C$ δ (150 Mhz) | | $^{1}H$ δ (600 Mhz) | |
|---|---|---|---|---|---|---|---|---|---|
| **3** | 203.33 | | / | | **10** | 40.13 | 40.11 | / | / |
| **14** | 161.57 | 162.01 | / | / | **9** | 37.04 | 37.15 | 2.11 (1H, m) | |
| **1** | 157.00 | 157.04 | 6.64 (1H, d J=10.2) | 6.66 (1H, d J=10.2) | **16** | 34.02 | 35.03 | 1.96 (1H, m)<br>1.58 (1H, m) | 2.06 (1H, m)<br>1.68 (1H, m) |
| **2** | 125.98 | 125.97 | 5.92 (1H, d J=10.2) | 5.91 (1H, d J=10.2) | **12** | 34.14 | 34.54 | 1.74 (2H, m) | 1.79 (1H, m) |
| **15** | 119.78 | 119.91 | 5.09 (1H, brd J=2.4) | 5.11 (1H, brd J=2.4) | **22** | 33.52 | 32.50 | 1.68 (2H, m) | 1.79 (1H, m)<br>1.67 (1H, m) |
| **24** | 97.77 | 96.39 | / | / | **20** | 30.21 | | 2.20 (1H, m) | |
| **25** | 76.19 | 76.90 | / | / | **29** | 27.58 | | 1.33 (3H, s) | 1.32 (3H, s) |
| **7** | 71.77 | 71.90 | 3.74 (1H, brm)* | | **30** | 27.54 | | 0.85 (3H, s) | 0.84 (3H, s) |
| **23** | 67.77 | 64.23 | 3.87 (1H, m)** | 3.98 (1H, aptq J=5.5)*** | **6** | 24.80 | 24.81 | 1.79 (1H, m)<br>1.59 (1H, m) | |
| **21** | 65.50 | 62.24 | 3.81 (1H, dd J=11.4, 5.0)<br>3.60 (1H, t J= 11.4) | 3.90 (1H, dd J=11.5, 2.5)<br>3.52 (1H, brd J=11.5) | **26** | 24.74 | | 1.43 (3H, s) | 1.32 (3H, s) |
| **17** | 57.36 | 52.15 | 1.22 (1H, m) | 1.98 (1H, m) | **27** | 23.28 | 24.32 | 1.18 (3H, s) | 1.28 (3H, s) |
| **13** | 46.79 | 46.72 | / | / | **28** | 21.77 | 21.74 | 1.11 (3H, s) | 1.10 (3H, s) |
| **5** | 44.92 | 44.98 | 2.58 (1H, brdd J=13.0, 2.2) | | **18** | 19.35 | 19.40 | 0.86 (3H, s) | 0.68 (3H, s) |
| **8** | 44.86 | | / | | **19** | 18.95 | | 0.87 (3H, s) | 0.85 (3H, s) |
| **4** | 44.43 | | / | | **11** | 16.51 | 16.60 | 1.51 (1H, m)<br>1.30 (1H, m) | |

NMR spectra were recorded in benzene-$d_6$, referenced to 7.16 and 128.06, following the general considerations outlined. Isolated product is a C24 epimeric mixture ca. 125 : 68 ratio. The δ for most abundant epimer is reported where a difference is observed. Asterisks indicate the following COSY coupling to OH; (*) δ 1.88, (**) δ 2.85 and (***) δ 2.61.

**Table S21. List of primer pairs used to clone genes from *C. sinensis*.**

| Gene | Use | Primer Sequence (5' to 3') |
|---|---|---|
| *CsCYP88A51* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGATTCGAATTTTTTGTGG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCATCCGACCCTAATGACTTTTGC |
| *CsMOI1* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGAGTCATCCATATTCG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCAATAAACTTTGGTCTTG |
| *CsMOI2* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGAGCCATTCATCTGGG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCAACCAACCTTGGTCACC |
| *CsMOI3* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGAGTCATCCCTATTCGCC |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCAATAAACTTTGCTCTTGTGGTC |
| *CsL21AT* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGATCTCCAAATCACCTGC |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCAAAATATGCTTGGATTAGGGGAAG |
| *CsSDR* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGAACGGCCCTTCCTCTG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTACTTGATAAGACCGTAAGCCC |
| *CsCYP716AC1* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGAATTCATTATCCTTTCCTTACTTCTTC |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTAATTGTTGGGATAGAGGCGAACTGG |
| *CsCYP88A37* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGAGTTAGATTTCTCATGG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTACTTGAACCCGACTACTTTTGC |
| *CsL1AT* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGAGATCAATAACGTTTCTTCAG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTAAATTAAGCTTGTATCAATAGAAGC |
| *CsL7AT* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGAGCCTGAAATACTTTCCATAG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTACCACAATGGGCATGGATC |
| *CsAKR* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGGGACGGCCATTCCAGAG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTAAATTTCTCCATCCCATATTTCCTCCACAGTTCT |
| *CsCYP716AD2* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGAGCTCCTCCTCCTCC |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***CTAATTCTCATAGGCATAGGGATAGAGG |
| *CsLFS* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGGCTGATCATTCAACAGTAAATGG |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TTAAACAGCTTTGTTGTCTTTCAC |
| *CsSI* | Fwd | ***ATTCTGCCCAAATTCGCGACCGGT***ATGAGCCATCCGTATGTGC |
| | Rev | ***GAAACCAGAGTTAAAGGCCTCGAG***TCAGCGAACTTTATTCTTCTTCTGC |

Nucleotides emphasized in bold and italics consist of the 5' overlaps designed for Gibson assembly using pEAQ-HT vector. All other nucleotides represent sequences that hybridize to the gene of interest.

**Table S22. List of primer pairs used to clone genes from *M. azedarach.***

| Target | Use | Sequence |
|---|---|---|
| candidate CYP88A165 | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTTAGATATCTTGTGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTTTCATTTGAGCTTGATGACTTT |
| candidate AKR | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGTGCAGTGCCTGAG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAACTCTGCATCAAGCTG |
| candidate 2-ODD | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAGAACGGATTGATGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATATTTTGTGACGTCTATTAC |
| MaSDR | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAACAGTTATTCATCCGCG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTAATTGATAAGATTATAAGCTTTC |
| MaL21AT | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATCTCCGAATCACTTCC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAAGTATGGTGGGATTAGG |
| MaCYP88A108 * | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTAAATTTCCTGTGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGAAGTTCTTGACCTTGATG |
| candidate AKR | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAGCTTTGCATCTTGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAACTCTGCATCAAGCTG |
| MaL1AT | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTCAAGATTGTTTCTTC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAATTATGCTTGTATCAACAGAGG |
| candidate CYP714E96 | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGCACTTCTTTAACTTTGGGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAATCCTCTTGACATGGAG |
| MaCYP716AD4 | Fwd | GGGGACCACTTTGTACAAGAAAGCTGGGTATTATTTGTTGTAGGGATATAGGCG |
| | Rev | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTCTTCCTACCC |
| MaL7AT | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCCTGAAATAATTTCC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCACATGGGACTTGGG |

**Table S22. List of primer pairs used to clone genes from *M. azedarach* (continued).**

| | | |
|---|---|---|
| MaLFS | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGATCATCTGACTGC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTATGCTTTCTTTCCCACAG |
| candidate transferase | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAATCAAAATTATTTC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAATCTAGCCTTTTTTGAC |
| candidate transferase | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAATGGAAATC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGTTGGAAGAAGC |
| MaCYP88A164 | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGCTCAGATTTGTTGTGG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCATTTAAGCTTAACGATTCTTGC |
| MaMOI2 | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGCGACTCATCATCTG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGCGAACTTTGGTCTTG |
| MaAKR (genome) | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGCTAAAGACGATTG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCATTCAGGAGTCAAC |
| MaAKR (transcriptome) | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGAAAACAGTG |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAATCTCTCCGTCCC |
| MaMOI1 | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGCCATCCATATTCG |
| | Rev | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGACTAACCATCCATACG |
| MaSI | Fwd | GGGGACCACTTTGTACAAGAAAGCTGGGTATTAATTGGTCTTACACTTC |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGCGAACTTTGGTC |
| pDNR207 (attL sites) | Fwd | TCGCGTTAACGCTAGCAT |
| | Rev | GTAACATCAGAGATTTTGAGACAC |
| pEAQ-HT-DEST1 (attB sites) | Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTA |
| | Rev | GGGGACCACTTTGTACAAGAAAGCTGGGTA |

All primers used in this study for the cloning of genes from *M. azedarach.* The gene or target and use (forward or reverse) are all listed. Asterix (*) indicates a gene previously cloned (*20*), and here re-cloned due to extended 5' coding sequence in the new *M. azedarach* genome.

**Table S23. Isolera™ Prime fractionation conditions for purification of products of heterologously expressed *M. azedarach* enzymes.**

| | Cartridge/phase | Solvent system | Gradient (B %) | CV | Yield of product (mg) |
|---|---|---|---|---|---|
| (**3**) | SNAP Ultra 50g (Normal) | A: Hexane<br>B: Ethyl acetate | 6-100% | 28 | 200 |
| | SNAP KP-Sil 25g (Normal) | A: Dichloromethane<br>B: Methanol | 0-10% | 106 | 130 |
| | SNAP Ultra 10g (x2) (Normal) | A: Dichloromethane<br>B: Methanol | 4-5% | 74 | 40 |
| (**6**) | SNAP Ultra 50g (Normal) | A: Hexane<br>B: Ethyl acetate | 6-100% | 33 (x2) | 980 |
| | SNAP KP-Sil 25g (Normal) | A: Hexane<br>B: Ethyl acetate | 25-55% | 117 | 570 |
| | Sfär Silica D Duo 25g (Normal) | A: Hexane<br>B: Ethyl acetate | 28% | 46 | 470 |
| | | | 28-38% | 22 | |
| | Sfär Silica D Duo 25g (Normal) | A: Dichloromethane<br>B: Methanol | 0-5% | 73 | 295 |
| | | | 5-7% | 17 | |
| | SNAP Ultra 10g (Normal) | A: Dichloromethane<br>B: Methanol | 5% | 9 | 220 |
| (**10**) | Sfär Silica D Duo 200g (Normal) | A: Hexane<br>B: Ethyl acetate | 6-100% | 19 | 300 |
| | SNAP KP-Sil 10g (Normal) | A: Hexane<br>B: Ethyl acetate | 50-75% | 200 | 70 |
| (**14**) | Sfär Silica D Duo 200g (Normal) | A: Hexane<br>B: Ethyl acetate | 6-100% | 19.3 | 200 |
| (**18**) | Sfär Silica D Duo 200g (Normal) | A: Hexane<br>B: Ethyl acetate | 5%<br>5-10%<br>10%<br>10-15%<br>15%<br>15-20%<br>20-25%<br>25% | 1.5<br>1.5<br>1.5<br>1.5<br>1.5<br>1.5<br>1.5<br>1.5 | 400 |
| (**20**) | Sfar C18 D- Duo 120g (Reverse) | A: Water<br>B: Acetonitrile | 30-100<br>100 | 13.4<br>1.2 | - |

Details of conditions used for Isolera™ Prime fractionation including: phase, column, solvent system, percentage gradient of solvent B, column volume (CV) and dry weight of resulting extract (yield). All samples were dry-loaded onto Isolera™ Prime (Biotage) using Silica gel or Celite® (Sigma-Aldrich) for normal or reverse phase, respectively.

**Table S24. Full length cloned nucleotide sequence of MaMOI2**

| | |
|---|---|
| **Cloned nucleotide sequence** | >MaMOI2<br>ATGAGCGACTCATCATCTGTTCCCGTGGATTTTGTGCTAAACTTCTCAACTG<br>CCGCCTTGCATGCTTGGAATGGCCTCAGTTTATTCTTAATCGTCTTCATCTC<br>CTGGTTTATCTCCG***GTATGTCTGCTTATTAATCTATTAAGTACACTTCGTAT***<br>***ATAATTCTACCTCAATCATATGTAGTTTATTGTTTGACGTGTATATCATATA***<br>***TCTACATATATATACGTTTGCATGAATTGATCATTGCTTGCAG***GGTTGACAC<br>AGGCGAAAACAAAAATGGACAGAGTGGTATTATGCTGGTGGGCTCTCACTGG<br>CCTTATTCATGTCTTTCAAGAGGGTTATTATGTTTTCACTCCAGATTTATTT<br>AAAGACGATTCTCCTAATTTTATGGCTGAAATTTGTAAGTACAATATACACA<br>TATGTGTGTATATATACATATATATATATGATTCACAATATTTATTATCTAA<br>AGAAATGGGATATATATAAATTAAACATAAACCTGCAGGGAAAGAATACAGC<br>AAAGGTGATTCAAGATATGCAACAAGACACACTTCAGTTCTTACCATCGAAT<br>CGATGGCTTCAGTTGTTCTGGGACCTCTTAGCCTTCTAGCAGCGTATGCTTT<br>AGCTAAAGCGAAGTCATACAACTACATTCTTCAGTTTGGAGTCTCAATTGCG<br>CAGCTGTATGGGGCTTGTCTATATTTCCTAAGTGCTTTCCTGGAGGGGGATA<br>ATTTTGCTTCTTCTCCGTATTTTTACTGGGCATATTACGTTGGACAAAGTAG<br>CATCTGGGTTATAGTACCAGCACTCATAGCTATACGTTGCTGGAAAAAAATC<br>AATGCTATTTGCTATCTTCAAGACAAGAAGAACAAGACCAAAGTTCGCTGA |

The sequence (generated by sanger sequencing) of the cloned version of *MaMOI2*, which differs from predicted sequence due to the retention of the first intron (table S10), which is assumed to be removed by splicing in *N. benthamiana* to achieve correct coding sequence (GenBank: OP947596). Intron is highlighted in bold italics.

**Captions for Data S1**

**Data S1. NMR spectra for all isolated compounds**
Copies of 1D NMR (including $^{1}H$, $^{13}C$ and DEPT-135 NMR) and 2D NMR (including DEPT-edited-HSQC, HMBC, COSY and NOESY or ROESY) spectra for the products isolated from heterologous expression in *N. benthamiana* of limonoid biosynthetic genes from *C. sinensis* ((**6**), (**4'**), (**9**), (**13**), (**13'**) and (**19**)) and from *M. azedarach* ((**3**), (**6**), (**10**), (**14**) and (**20**)). Along with the $^{13}C$ NMR spectra of (**18**) isolated from *A. indica.*