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

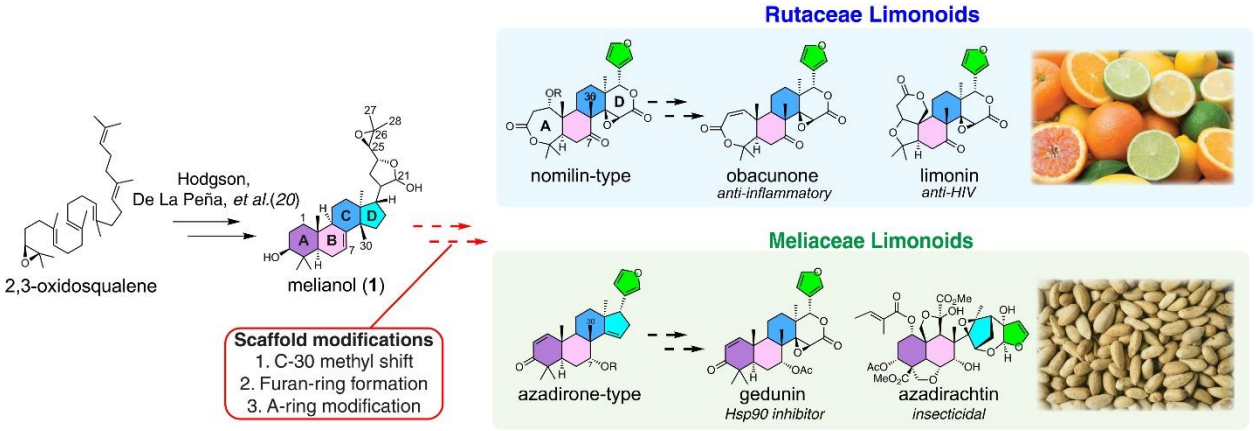


Fig. 1. Structures of Rutaceae and Meliaceae limonoids and proposed biosynthetic pathway. We previously characterized three conserved enzymes from both *Citrus* and *Melia* species that catalyze the formation of the protolimonoid melianol (**1**) from 2,3-oxidosqualene (*20*). Additionally, conserved 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.

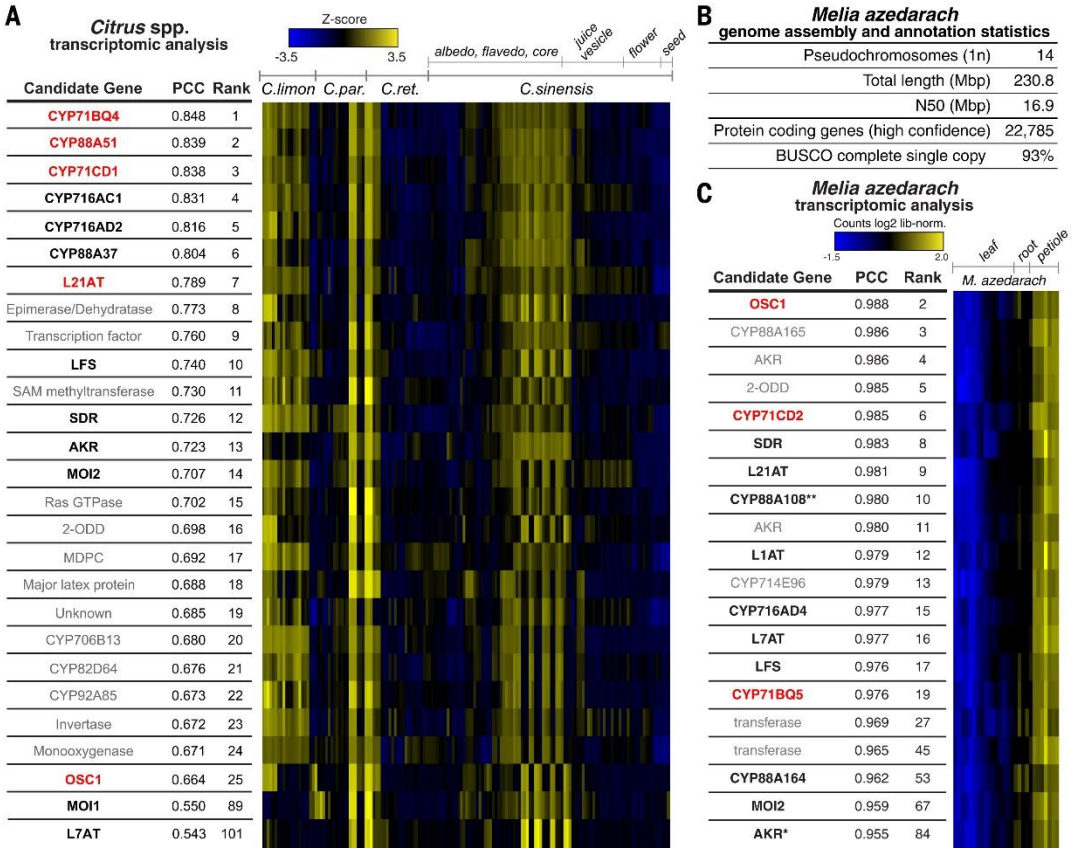


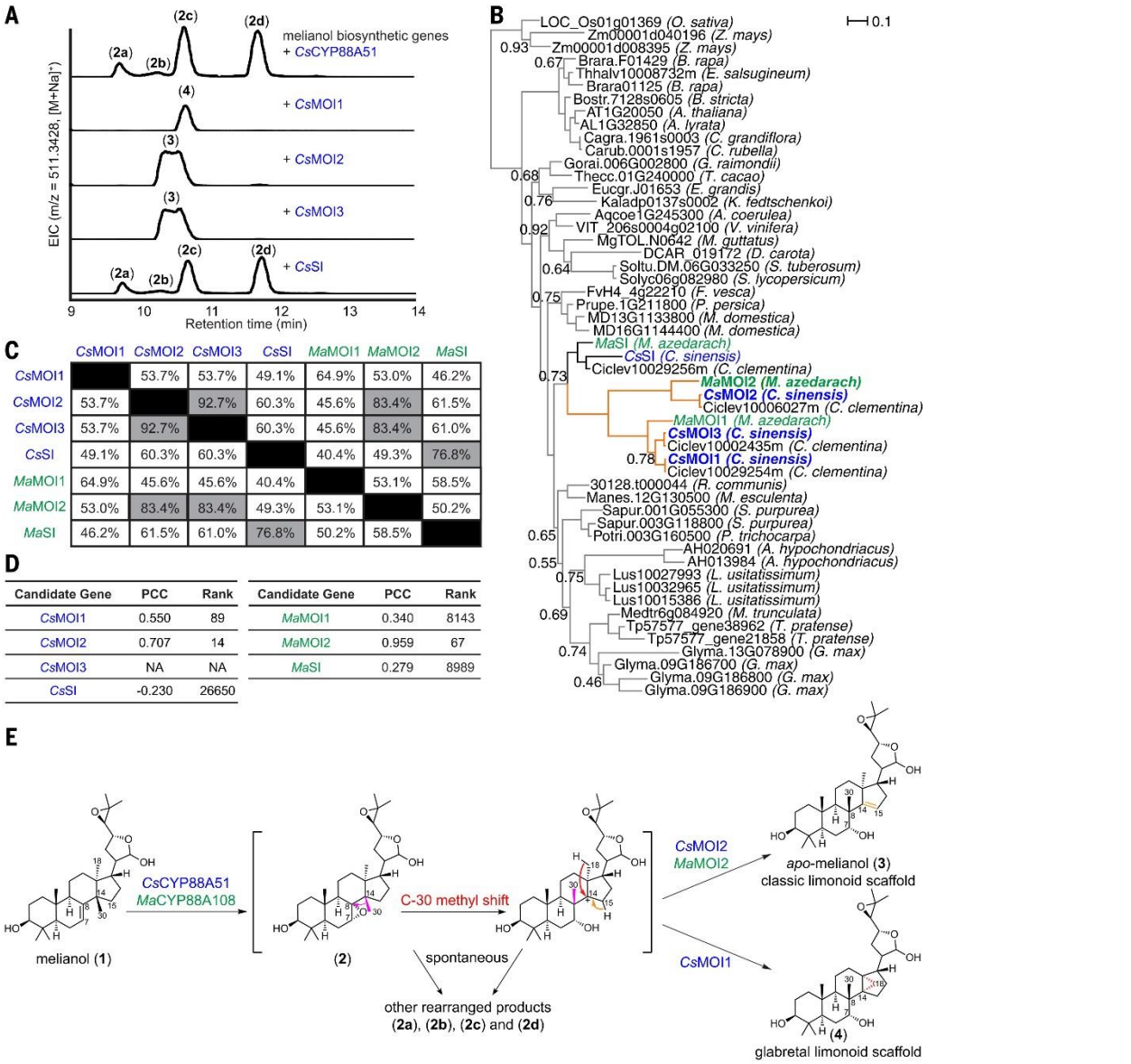
Fig. 2. Genomic and transcriptomic analysis of *Citrus* and *Melia* resources.

(**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*)).

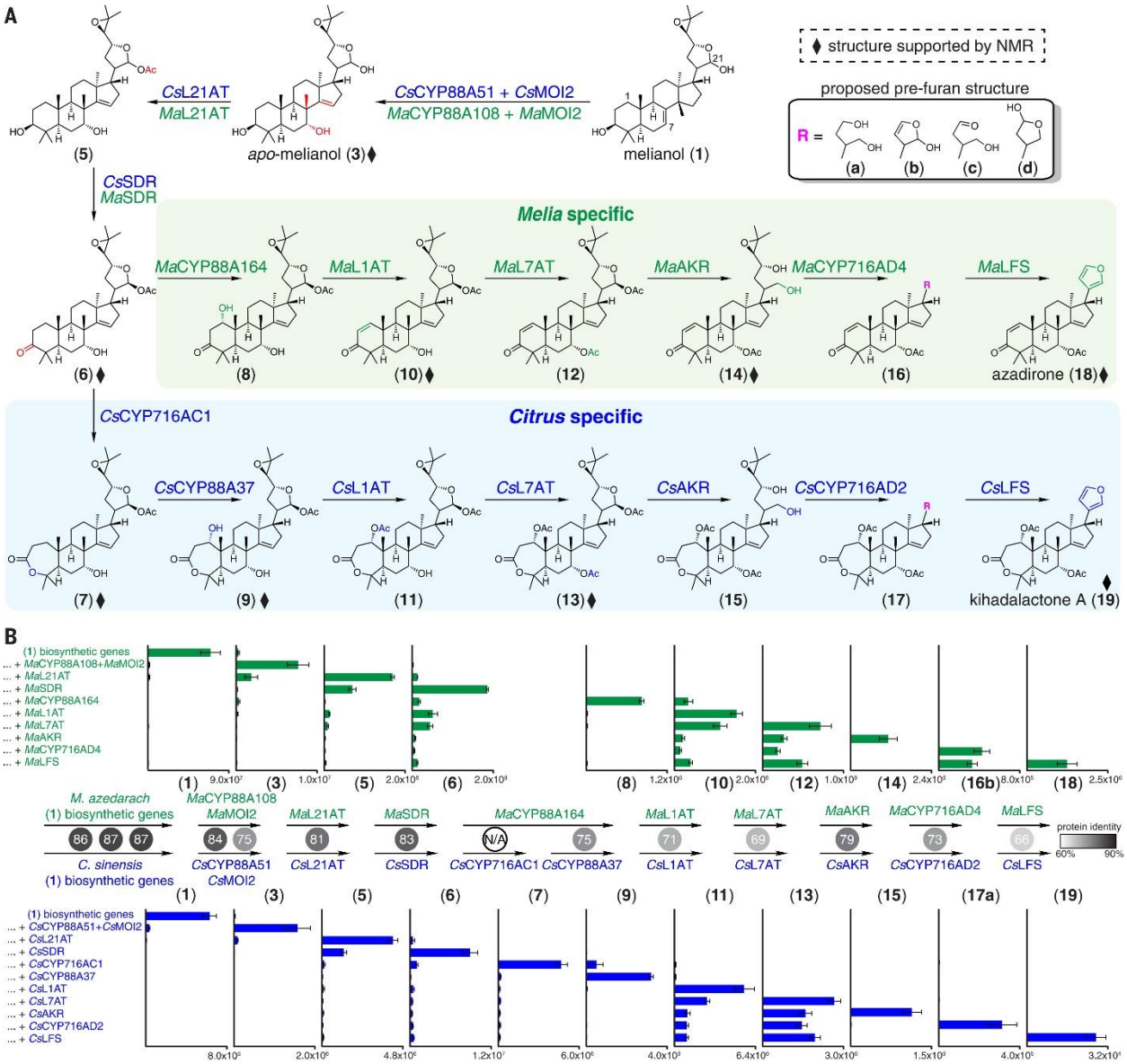


Fig. 4. Complete biosynthetic pathway to azadirone (18) and kihadalactone A (19). (**A**) Gene sets that lead to the production of azadirone (**18**) and kihadalactone A (**19**) in *N. benthamiana* leaves. Genes from *Citrus* are shown in blue and those from *Melia* are shown in green. The arrow reflects 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).