SUPPLEMENTARY MATERIALS

METHODS

Tissue preparation and library construction

Leaf tissue of *Asclepias syriaca* was sampled from a single individual at the
5   Western Illinois University research farm, raised from seed from a wild population in
McDonough county, Illinois (40.29622ºN, 90.89876ºW; Winthrop B. Phippen *s.n.*, OSC
226164, 226165). DNA was extracted from frozen tissue using the FastDNA Spin Kit
from MPBiomedicals (Santa Ana, CA, USA) following manufacturer's protocols,
modified by the addition of 40 µL 1% polyvinylpyrrolidone and 10 µL β-mercaptoethanol
10   to the lysis solution prior to grinding.

Aliquots of isolated DNA were sheared with a BioRuptor sonicator (Diagenode
Inc., Denville, NJ, USA). Two libraries were prepared following the Illumina protocol for
paired-end libraries (Solexa, Inc, 2006). Ligated fragments were cut from agarose gels
centered around 225 bp and 450 bp, and were amplified through 15 and 14 cycles,
15   respectively, of polymerase chain reaction using Phusion High-Fidelity PCR Master Mix
(New England BioLabs, Ipswich, MA, USA) and standard Illumina primers. Cleaned
product was submitted for sequencing on an Illumina GAII Sequencer (Illumina Inc., San
Diego, CA, USA) at the Center for Genome Research and Biocomputing (CGRB) at
Oregon State University (Corvallis, OR, USA). One lane of the 450 bp library was
20   sequenced with 80 bp paired-end reads, and 5 lanes of the 225 bp library were sequenced
with 120 bp paired-end reads.

Frozen tissue was sent to GlobalBiologics, LLC (Columbia, MO, USA) for DNA
extraction and production of mate-pair libraries using the Illumina Mate Pair Library v2
protocol with average insert sizes of 2750 bp and 3500 bp, and indexed with unique
25   barcode sequences (Bioo Scientific, Austin, TX, USA). Purified DNA was provided to
the CGRB for production of a mate-pair library using the Illumina Nextera protocol with
an average insert size of 2000 bp. The 2000 bp library was sequenced on an Illumina
MiSeq at the CGRB, one of 15 samples pooled on a lane, and sequenced with 76 bp
paired-end reads. The 2750 bp library was sequenced on an Illumina HiSeq 2000

30   sequencer at the CGRB, one of three samples pooled on a lane, and sequenced with 101

bp paired-end reads. The 3500 bp library was sequenced on an Illumina MiSeq at Oregon

Health and Science University (Portland, OR, USA) with 33 bp paired-end reads (Table

2).

Genomic read processing

35   Pairs of reads properly mapping to the *Asclepias* chloroplast or mitochondria,

with three or fewer mismatches between the target and query, were filtered out using

Bowtie 2 v. 2.1.0 (scoring parameter "--score-min L,-6,0"), samtools v. 0.1.18, and

bamtools v. 2.3.0 (Barnett et al., 2013; Langmead & Salzberg, 2012; Li et al., 2009).

Portions of reads matching the Illumina adapter sequences were removed with

40   Trimmomatic v. 0.30 and the "ILLUMINACLIP" option (Bolger et al., 2014). Duplicate

read pairs from the same library were removed using the custom script fastq_collapse.py

(Weitemier, 2014). Paired-end read pairs with sequences that overlapped by ≥7 bp

sharing ≥90% identity were merged using the program FLASH v. 1.2.6 (parameters "-m 7

-M 80 -x 0.10") (Magoč & Salzberg, 2011). The 3' and 5' ends of reads were then

45   trimmed of any bases with a Phred quality score below 30, and any remaining reads less

than 30 bp were removed using Trimmomatic.

        Summary statistics were calculated using a k-mer distribution plot of reads from

the 225 bp insert library after removing chloroplast and mitochondrial reads, but prior to

joining with FLASH. K-mers of 17 bp were counted using BBTools script

50   kmercountexact.sh, and estimates of genome size and heterozygosity were calculated

using the program gce (Bushnell & Rood, 2015; Liu et al., 2013).

RNA-seq library preparation, sequencing, and assembly

        Total RNA was extracted from the individual used for genome sequencing from

leaves and buds separately, by homogenizing approximately 200 mg of fresh frozen

55   tissue on dry ice in a Fast-Prep-24 bead mill. Cold extraction buffer (1.5 mL of 3M

LiCl/8M urea; 1% PVP K-60; 0.1M dithiothreitol; Tai et al., 2004) was added to the

ground tissue. Tissue was then homogenized and cellular debris pelleted at $200 \times g$ for 10

minutes at 4°C. Supernatant was incubated at 4°C overnight. RNA was pelleted by

centrifugation (20,000×*g* for 30 minutes at 4°C) and cleaned using a ZR Plant RNA

60 MiniPrep kit (Zymo Research, Irvine, CA, USA). For each tissue type, an RNA-seq

library was prepared using the Illumina RNA-Seq TruSeq kit v. 2.0 with the

modifications of Parkhomchuk et al. (2009) to allow strand-specific sequencing by dUTP

incorporation.

Libraries were sequenced on an Illumina HiSeq 2000 at the CGRB to yield 101 bp

65 single-end reads. Before further analysis, reads that did not pass the Illumina chastity and

purity filters were removed. Trimmomatic 0.20 (Bolger et al., 2014) was used to trim the

final base of each read, leading and trailing bases with quality scores below Q20, and all

following bases if a sliding window of 5 bp did not have an average quality of at least

Q30. Reads shorter than 36 bp after trimming were excluded.

70 Transcripts were assembled de novo using Trinity (Release 2013-08-14)

(Grabherr et al., 2011) for bud and leaf reads separately, as well as combined into a single

data set using default settings, except for using a minimum contig length of 101 bp. The

same settings were also used to assemble RNA-seq data from leaf tissue of the same *A.

syriaca* individual from a library made using ribosomal RNA subtraction (Straub et al.,

75 2013). Best scoring open reading frames (ORFs) were determined for each library using

the TransDecoder utility provided with Trinity (Haas et al., 2013). Transcripts were

annotated using Mercator (Lohse et al., 2014) and TRAPID (Van Bel et al., 2013).

Comparative transcriptome and gene family evolution analyses in Apocynaceae

80 For a comparative analysis, transcriptomes were obtained for four other species of

Apocynaceae. *Catharanthus roseus* and *Rauvolfia serpentina* transcriptomes were

downloaded from the Medicinal Plant Genomics Resource project database

(http://medicinalplantgenomics.msu.edu; Góngora-Castillo et al., 2012), the

Tabernaemontana elegans transcriptome was downloaded from the PhytoMetaSyn

85 Project database (www.phytometasyn.ca; Xiao et al., 2013), and the *Rhazya stricta*

transcriptome was downloaded from NCBI (Yates et al., 2014). Transcriptomes were

checked for completeness using BUSCO v. 1.22 (Simão et al., 2015). Transcripts of all

species were assigned to reference gene families using TRAPID. Reference gene family assignments for two outgroups, *Coffea canephora* (Denoeud et al., 2014) and *Vitis*

90   *vinifera* (PLAZA v. 2.5; Proost et al., 2009) were obtained.

A phylogenetic tree to use as a framework for comparative analysis was produced using whole plastomes. Four sequences were downloaded from GenBank: *A. syriaca* (NC_022432), *C. roseus* (NC_021423), *R. stricta* (NC_024292), and *V. vinifera* (NC_007957). Plastome sequences for *T. elegans* and *R. serpentina* were assembled from

95   chloroplast transcripts present in the downloaded transcriptomes by iterative mapping of transcripts to the *C. roseus* sequence in Geneious v. 8. Full plastome sequences were aligned using Mafft v. 7.308 (Katoh et al., 2002; Katoh & Standley, 2013). CDS, tRNA, and rRNA genes were extracted from the alignment. Genes with greater than 50% missing data for either *R. serpentina* or *T. elegans* were removed. A maximum likelihood

100   (ML) analysis of the plastid data matrix was conducted in RAxML v. 8.0.26 using the GTR plus gamma model. One thousand bootstrap replicates were performed. The plastid ML tree was converted to an ultrametric tree using divergence times within Apocynaceae based on the analysis of Fishbein et al (In press) and estimates of Wikström et al. (2015) for the *Coffea* split from Apocynaceae and *Vitis* split from Gentianales. In order to

105   examine changes in gene family sizes across Apocynaceae transcriptomes, BadiRate v. 1.35 (Librado et al., 2012) was run using the BDI (birth-death-innovation) stochastic model with a free rate (FR) branch model where each branch can have a different gene turn-over rate. Gains and losses were inferred using Wagner parsimony.

Genomic sequence assembly

110   Processed read-pairs were assembled into contigs using Platanus v. 1.2.1 (Kajitani et al., 2014). Platanus is designed to assemble highly heterozygous diploid genomes, and initially uses several k-mer sizes during assembly. *Asclepias* reads were assembled with an initial k-mer size of 25 bp with a k-mer step increase of 10 bp up to a maximum k-mer of 110 bp. As part of the expectation for heterozygous assembly, Platanus can merge

115   contigs sharing high identity. We allowed contigs sharing 85% identity to be merged (assembly parameters "-k 25 -u 0.15").

Scaffolding was performed with Platanus, setting the paired-end reads as "inward pointing" reads and the mate-pair reads as "outward pointing" reads. Reads were mapped to scaffolds using an initial seed size of 21 bp, one link between contigs was sufficient to

120   align them into a scaffold, and scaffolds sharing 85% identity could be merged (scaffolding parameters "-s 21 -l 1 -u 0.15").

Gaps between scaffolds were closed via local alignment and assembly of reads around the gaps using Platanus. An initial seed size of 21 bp was used to include reads in the mapping around a gap, and a minimum overlap of 21 bp between the newly

125   assembled filler contig and the edges of the scaffold was required to use that contig to fill the gap (gap close parameters "-s 21 -k 21 -vd 21 -vo 21").

Transcripts were mapped to *Asclepias* scaffolds ≥1 kbp using BLAT v. 32x1, and those scaffolds were merged where they were linked by one or more transcripts (Kent, 2002). This was performed with the program Scubat

130   (<https://github.com/elswob/SCUBAT> accessed 12/17/2015) modified so that scaffolds would not be clipped when joined by cap3 v. 02/10/15 (Elsworth, 2012; Huang & Madan, 1999; Tange, 2011).

Contaminant removal

Merged scaffolds were compared against a genomic database of potentially

135   contaminating organisms with the program DeconSeq standalone v. 0.4.3 (Schmieder & Edwards, 2011). Contaminant databases were downloaded from the DeconSeq website representing bacteria, archaea, viruses, 18S rRNA, zebrafish, mouse, and several human genomes (<http://deconseq.sourceforge.net> accessed January 20, 2016). Fungal genomes were obtained from the National Center for Biotechnology Information (NCBI)

140   including *Alternaria arborescens* accession AIIC01, *Aspergillus fumigatus* AAHF01, *Bipolaris maydis* AIHU01, *Botrytis cinerea* assembly GCA_000832945.1, *Cladosprium sphaerospermum* AIIA02, *Fomitopsis pinicola* AEHC02, *Fusarium oxysporum* AAXH01, *Galerina marginata* AYUM01, *Hypoxylon sp.* JYCQ01, *Penicillium expansum* AYHP01, *Rhodotorula graminis* JTAO01, *Saccharomyces cerevisiae* assembly

145   GCA_000146045.2, and *Trichoderma reesei* AAIL02 (Amselem et al., 2011; Firrincieli et

al., 2015; Floudas et al., 2012; Goffeau et al., 1997; Hu et al., 2012; Li et al., 2015; Ma et al., 2010; Martinez et al., 2008; Ng et al., 2012; Nierman et al., 2005; Ohm et al., 2012; Riley et al., 2014; Shaw et al., 2015). The genome of *Solanum lycopersicum* (ITAG 2.4) was downloaded from the Sol Genomics Network (The Tomato Genome Consortium,

150   2012). The fungal and tomato genomes were prepared as DeconSeq databases following the DeconSeq website, including filtering of repeated Ns, removal of duplicate sequences, and indexing with a custom version of BWA released with DeconSeq (Li & Durbin, 2010; <http://deconseq.sourceforge.net> accessed January 20, 2016).

Genomes obtained from the DeconSeq website and the fungal genomes were used

155   as contaminant databases, the tomato genome was used as a retain database. Scaffolds matching one of the contaminant genomes with ≥80% identity along ≥80% of the scaffold length were excluded as contaminants. Those scaffolds matching both a contaminating genome and the tomato genome were retained.

Gene prediction and annotation

160   A library of *Asclepias* repetitive elements was created following guidelines in the MAKER Genome Annotation Pipeline online documentation (Jiang, 2015). The program RepeatModeler v. open-1.0.8 was used to integrate the programs RepeatMasker v. open-4.0.5, rmblastn v. 2.2.28, RECON v. 1.08, Tandem Repeats Finder v. 4.07b, and RepeatScout v. 1.0.5 (Bao & Eddy, 2002; Benson, 1999; Price et al., 2005; Smit et al.,

165   2015). Repeat models initially missing a repeat annotation were compared, using BLAT, against a library of class I and class II transposable elements acquired from the TESeeker website (Kennedy et al., 2010, 2011), and matching sequences provided an annotation. Remaining unannotated models were submitted to the online repeat analysis tool, CENSOR, and provided annotations with a score ≥400 and ≥50% sequence similarity

170   (Kohany et al., 2006). A set of proteins from *Arabidopsis thaliana* was filtered to remove proteins from transposable elements, then compared using BLASTX against the *Asclepias* repeat models. The program ProtExcluder.pl v. 1.1 then used the BLASTX output to remove repeat models and flanking regions matching *Arabidopsis* proteins (Altschul et al., 1990; Jiang, 2015).

175   The set of scaffolds ≥1 kbp were annotated via the online annotation and curation

tool GenSAS v. 4.0 (Humann et al., 2016; Lee et al., 2011), which was used to implement

the following tools for repeat masking, transcript and protein mapping, ab initio gene

prediction, gene consensus creation, and mapping of *Asclepias* predicted proteins:

        Repeats in the assembled sequence were masked via RepeatMasker v. open-4.0.1

180   using the *Asclepias* repeat models and using models developed from dicots more broadly

(Smit et al., 2015).

        Multiple datasets were mapped onto *Asclepias* scaffolds in order to assist with

gene prediction. The best ORFs from assembled *Asclepias* transcripts were mapped using

both BLAT and BLAST (expect < 1e-50, 99% identity). Assembled transcripts from

185   *Calotropis procera,* a member of the same subtribe, Asclepiadinae, were mapped with

BLAT (Kwon et al., 2015). Proteins from *Coffea canephora,* a member of the same order,

Gentianales, were mapped with BLASTX (e<0.0001; Denoeud et al., 2014).

        Genes were predicted using the ab initio tools Augustus v. 3.1.0, SNAP, and

PASA (Haas et al., 2003; Korf, 2004; Stanke et al., 2008). Augustus was run using gene

190   models from *Solanum*, finding genes on both strands, and allowing partial models; SNAP

was run using models from *Arabidopsis thaliana*. PASA was informed by the best ORFs

from assembled *Asclepias* transcripts.

        Multiple lines of evidence were integrated into a gene consensus using

EVidenceModeler (Haas et al., 2008) with the following weights: Augustus, 1; SNAP, 1;

195   *Coffea* proteins, 5; *Asclepias* transcripts (BLAST), 7; *Asclepias* transcripts (BLAT), 7;

Calotropis transcripts, 5; PASA, 7. Consensus gene models were then refined using

PASA, again informed by *Asclepias* transcripts.

        Predicted proteins were compared to the NCBI plant RefSeq database using

BLASTP (expect < 1e-4, BLOSUM62 matrix; Pruitt et al., 2002), as well as being

200   mapped against protein sequences from *Coffea* and *Catharanthus roseus*, a member of a

different subfamily within Apocynaceae (expect < 1e-4; Denoeud et al., 2014; Kellner et

al., 2015). Protein families were classified using the InterPro database and InterProScan

v. 5.8-49.0 (Jones et al., 2014; Mitchell et al., 2015). Transfer RNAs were identified using

tRNAscan-SE v. 1.3.1 (Lowe & Eddy, 1997). Additional open reading frames were found using the getorf tool from the EMBOSS suite, accepting a minimum of 30 bp (Rice et al., 2000).

Some predicted proteins were missing one or more exons, either because they were fragmented on the ends of scaffolds or, rarely, transcript evidence predicted exons with non-canonical splice sites. The predicted coding sequence produced by GenSas for some of these proteins was out of frame. In these cases the coding sequence was translated under all reading frames and a translation lacking internal stop codons was selected, if available.

An estimate of the completeness of the assembled gene space was calculated using the program BUSCO v. 1.22 and a set of 956 conserved single copy plant genes (Simão et al., 2015). BUSCO was run independently on the set of coding sequences returned following gene prediction as well as on the assembled scaffolds ≥1 kbp using Augustus gene prediction with *Solanum* models. Predicted genes from *Asclepias*, *Catharanthus*, *Coffea*, and *Vitis vinifera* (obtained from the PLAZA 3.0 database) were clustered into orthogroups using OrthoFinder v. 0.7.1 (Emms & Kelly, 2015; Proost et al., 2015; The French-Italian Public Consortium for Grapevine Genome Characterization, 2007).

Gene analyses

The P5βR region (PLAZA v. 2.5 gene family HOM000752) was identified in assembled scaffolds with BLAT (Kent, 2002), using the P5βR sequences from *Asclepias curassavica* (ADG56538; Bauer et al., 2010) and *Catharanthus roseus* (KJ873882-KJ873887; Munkert et al., 2015) as references. A maximum likelihood tree was constructed from peptide sequences of two *A. syriaca* regions with high identity to P5βR, six *Catharanthus* P5βR sequences, the *A. curassavica* sequence, sequences from *Digitalis purpurea* and *D. lantata* (ACZ66261, AAS76634), representing P5βR and P5βR2 paralogs, respectively, and a sequence from *Picea sitchensis* (ABK24388). P5βR sequence alignments were performed using MUSCLE, as implemented in Geneious v. 9.1.5, with a maximum of 10 iterations (Edgar, 2004; Kearse et al., 2012). The optimal

models of amino acid substitution, rate variation among sites, and equilibrium frequencies were inferred using the Akaike and Bayesian information criteria, as

235   implemented in the online tool PhyML 3.0, which was also used to infer trees under those models and incorporating aBayes support values (Anisimova et al., 2011; Guindon et al., 2010; Guindon & Gascuel, 2003).

SNP finding and targeted enrichment probe development

The Platanus genome assembler uses a de Bruijn graph approach for contig

240   assembly (Kajitani et al., 2014). Certain types of branches in this graph, known as "bubbles," may be caused by heterozygosity and are saved by the program for use in later assembly stages. Here, saved bubbles were filtered to identify those likely to represent heterozygous sites in low-copy regions of the genome.

The program CD-HIT-EST v. 4.5.4 was used to cluster any bubbles sharing ≥90%

245   identity, which were removed, leaving only unique bubbles (Li & Godzik, 2006). Unique bubbles were mapped against the set of *Asclepias* scaffolds ≥1 kbp using BLAT at minimum identity thresholds of 90% and 95% (Kent, 2002). A set of 4000 SNP probes developed from a preliminary study using a similar approach, but from a different genome assembly, were mapped against the assembly presented here with a 90% identity

250   threshold (Weitemier et al., 2014). Custom scripts were written to select one appropriate bubble from each scaffold <10 kbp, and up to two bubbles from scaffolds ≥10 kbp, up to a total of 20,000 bubbles. Bubbles mapping only once within the ≥90% identity mapping analysis were selected first, progressively adding bubbles that either mapped to ≤4 locations in the ≥90% identity mapping or mapped to ≤3 locations in the ≥95% identity

255   mapping. Bubble sequences were trimmed to 80 bp, centered around the SNP site where possible. Potential SNP probes were further analyzed by MYcroarray (Ann Arbor, MI, USA) and excluded if they were predicted to anneal in a solution hybridization reaction to >10 locations within the *Asclepias* genome at 62.5-65°C or >2 locations above 65°C. Twenty thousand RNA oligos suitable for targeted enrichment, matching 17,684

260   scaffolds, were produced by MYcroarray.

Linkage mapping population

Mature follicles were collected from the open pollinated plant that was the subject of genome sequencing. Approximately 100 seeds from six follicles collected from four stems of this plant (1, 3, 1, and 1 follicle per stem) were germinated and grown at

265   Oklahoma State University. Due to the pollination system of *Asclepias*, seeds in a fruit are almost certainly fertilized by a single pollen donor (Sparrow & Pearson, 1948; Wyatt & Broyles, 1990), meaning up to six paternal parents are represented among the 96 mapping offspring.

Seeds were surface sterilized in 5% bleach and soaked for 24 hr in distilled water.

270   The testa was nicked opposite from the micropylar end and the seeds germinated on moist filter paper, in petri dishes, in the dark, at room temperature. Germination occurred within 4-7 days, and seedlings where planted into MetroMix 902 media in plug trays when radicles attained a length of 2-3 cm. Seedlings were again transplanted to 3-inch deep pots following the expansion to two sets of true leaves. Seedlings were grown under

275   high intensity fluorescent lights in a controlled environment chamber at 14 hr daylength at approximately 27˚C. Plants were grown for approximately 90 days, harvested, and rinsed in distilled water, and frozen at -80˚C. DNA was extracted from roots, shoots, or a combination of roots and shoots using the FastDNA® kit (MP Biomedicals, Santa Ana, California) and Thermo Savant FastPrep® FP120 Cell Disrupter (Thermo Scientific,

280   Waltham, MA, USA). DNA quantity and quality were visualized using agarose gel electrophoresis and quantified with a Qubit® fluoremeter (Invitrogen, Carlsbad, CA, USA) and Quant-iT™ DNA-BR Assay Kit.

Ninety-six Genomic DNA samples were diluted as necessary with ultrapure water to obtain approximately 3 µg in 100 ul and sheared on a Bioruptor UCD-200 (Diagenode)

285   at low power for 12 cycles of 30 s on/30 s off. Several samples required sonication for 5-10 additional cycles to achieve a high concentration of fragments at the target size of 300-400 bp. Illumina-compatible, dual-indexed libraries were produced with the TruSeq® HT kit (Illumina), each with a unique barcode.

Barcoded libraries were pooled by equal DNA mass in three groups of 32

290   samples. These were enriched for targeted SNP regions using RNA oligos and following

MYcroarray MYbaits protocol v. 3.00. Enriched pools were then themselves evenly

pooled and sequenced with 150 bp paired-end reads on an Illumina HiSeq 3000 at the

CGRB, producing 120.3 Mbp of sequence data (NCBI short read archive: SRX2163716-

SRX2163811).

295   ***Linkage analyses***

Reads were processed using Trimmomatic v. 0.33 to remove adapter sequences,

bases on the ends of reads with a Phred quality score below three, and clipping once a

sliding window of 4 bp fell below an average quality score of 17 (Bolger et al., 2014).

Processed reads for each sample were mapped onto the assembled scaffolds using

300   bowtie2 with "sensitive" settings and a maximum fragment length of 600 bp (Langmead

& Salzberg, 2012). Reads from the 225 bp insert library of the sequenced individual were

also mapped back onto assembled scaffolds using the same settings. Mappings for all

individuals and the parent were combined using samtools v. 0.1.16 and SNPs called using

the bcftools "view" command (Li et al., 2009).

305   The file containing all variants was converted to a format suitable for the R

package OneMap, using a custom perl script (Tennessen, 2015), retaining only sites

heterozygous in one parent, the maternal sequenced individual. In this filtering the minor

genotype abundance (either heterozygote or homozygote) needed to be at least 24 across

90 samples, loci could have up to 30% missing individuals, and potential genotypes

310   within individuals were ignored if their Phred probability score was 15 or above (i.e., of

the three possible genotypes AA, Aa, aa, one should be most probable with a low Phred

score and the other two less probable with Phred scores above 15). This retained a subset

of SNPs where the maternal parent was heterozygous and the paternal parents for all

offspring were homozygous for the same allele.

315   A subset of 22 full siblings, those from the follicle producing the most offspring,

were filtered in the same manner, thereby retaining SNPs heterozygous in only the

maternal or the paternal parent. Filtering in this set required a minor genotype abundance

of at least five, loci could have up to four missing individuals, and genotypes with Phred

probabilities of 20 or above were ignored (i.e., the final genotype calls are more certain

320   because alternative genotypes are less likely).

SNP sets were clustered into linkage groups in R v. 3.2.2 using the package

OneMap v. 2.0-4 (Margarido et al., 2007; R Core Team, 2014). One SNP from each

scaffold was selected from SNPs among the full set of individuals, and were grouped

using a logarithm of odds (LOD) threshold of 8.4. This clustered SNP loci into eleven

325   clear groups, referred to here as the core linkage groups.

From the full sibling set of SNPs, those held on the same scaffold and with

identical genotypes across individuals (i.e., in perfect linkage) were grouped, and SNPs

on different scaffolds in perfect linkage with no missing data were grouped. This was

performed separately for loci where either the maternal or paternal parent was

330   heterozygous. These loci were clustered into groups using LOD scores 6.1, 6.0, and 5.5.

Each of these groupings produced hundreds of groups, but each contained about 22

groups that were substantially larger than the others.

A custom R script was used to combine the linkage group identity of scaffolds in

the core linkage groups with scaffolds and groups in the sibling sets. For example,

335   scaffold A could be assigned to a linkage group if it was in perfect linkage in the sibling

set with scaffold B, and scaffold B was also present in the core linkage groups. If

multiple scaffolds were perfectly linked, but associated with different core linkage

groups, no unknown scaffolds would be assigned unless the most common core linkage

group was three times as common as the next core group.

340   Linkage groupings in the sibling sets could be assigned to core linkage groups

based on the membership of the scaffolds they contained. If the markers indicating that a

sibling group should belong to a certain core linkage group were ten times as common as

markers supporting a second most common assignment, then the sibling group was

assigned to the core group, and all unknown scaffolds it contained also assigned to that

345   group. (E.g., sibling group A contains ten scaffolds known to be on core linkage group 1,

one scaffold known to be on core linkage group 2, and one unknown scaffold; sibling

group A is assigned to core linkage group 1 and the unknown scaffold is similarly assigned.)

This process was performed iteratively, progressively assigning scaffolds to core
350   linkage groups. It was performed first with the sibling set grouped with LOD 6.1, then the grouping with LOD 6.0, finally the grouping with LOD 5.5.

Synteny within Gentianales

Scaffolds found within the core linkage groups were mapped to *Coffea* coding sequences (BLASTN, expect < 1, best hit chosen) and mapped to their location on *Coffea*
355   pseudochromosomes. Six *Asclepias* linkage groups contained scaffolds that mapped preferentially to a single *Coffea* pseudochromosome, which in turn had *Asclepias* scaffolds mapping preferentially from that linkage group. From these six linkage groups one marker was selected for every 1 Mbp segment of the *Coffea* chromosome. Recombination fractions were measured among these loci using OneMap (retaining
360   "safe" markers with THRES=5) and converted to cM using the Kosambi mapping function.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/doi: 10.1016/S0022-2836(05)80360-2

Amselem, J., Cuomo, C. A., van Kan, J. A. L., Viaud, M., Benito, E. P., Couloux, A., … Dickman, M. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics, 7,* e1002230. https://doi.org/10.1371/journal.pgen.1002230

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, *60*, 685–699. https://doi.org/10.1093/sysbio/syr041

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research, 12,* 1269–1276. https://doi.org/10.1101/gr.88502

Barnett, D., Garrison, E., Marth, G., & Stromberg, M. (2013). *bamtools*. Retrieved from https://github.com/pezmaster31/bamtools

Bauer, P., Munkert, J., Brydziun, M., Burda, E., Müller-Uri, F., Gröger, H., … Kreis, W. (2010). Highly conserved progesterone 5β-reductase genes (P5βR) from 5β-cardenolide-free and 5β-cardenolide-producing angiosperms. *Phytochemistry, 71,* 1495–1505. https://doi.org/10.1016/j.phytochem.2010.06.004

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research, 27,* 573–580.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30,* 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bushnell, B., & Rood, J. (2015). *BBTools*. Retrieved from https://sourceforge.net/projects/bbmap/

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., … Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science, 345,* 1181–1184. https://doi.org/10.1126/science.1255274

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Elsworth, B. (2012). *Unearthing the genome of* Lumbricus rubellus (Ph.D. dissertation). The University of Edinburgh, Edinburgh, Scotland, U.K. Retrieved from https://www.era.lib.ed.ac.uk/bitstream/handle/1842/7596/Elsworth2013.pdf

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 1–14. https://doi.org/10.1186/s13059-015-0721-2

Firrincieli, A., Otillar, R., Salamov, A., Schmutz, J., Khan, Z., Redman, R. S., … Doty, S. L. (2015). Genome sequence of the plant growth promoting endophytic yeast

Rhodotorula graminis WP1. *Frontiers in Microbiology*, *6*, 978. https://doi.org/10.3389/fmicb.2015.00978

Fishbein, M., Livshultz, T., Straub, S. C. K., Simões, A. O., Boutte, J., McDonnell, A., & Foote, A. D. (In press). Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. *American Journal of Botany*.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R. A., Henrissat, B., … Hibbett, D. S. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, *336*, 1715–1719. https://doi.org/10.1126/science.1221748

Goffeau, A., Aert, R., Agostini-Carbone, M. L., Ahmed, A., Aigle, M., Alberghina, L., … Zumstein, E. (1997). The yeast genome directory. *Nature*, *387*, 5–5.

Góngora-Castillo, E., Fedewa, G., Yeo, Y., Chappell, J., DellaPenna, D., & Buell, C. R. (2012). Genomic approaches for interrogating the biochemistry of medicinal plant species. In David A. Hopwood (Ed.), *Methods in Enzymology* (Vol. Volume 517, pp. 139–159). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780124046344000073

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. https://doi.org/10.1038/nbt.1883

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*, 307–321. https://doi.org/10.1093/sysbio/syq010

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*, 696–704. https://doi.org/10.1080/10635150390235520

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., … White, O. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*, 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*, 1–22. https://doi.org/10.1186/gb-2008-9-1-r7

Hu, J., Chen, C., Peever, T., Dang, H., Lawrence, C., & Mitchell, T. (2012). Genomic characterization of the conditionally dispensable chromosome in *Alternaria*

arborescens provides evidence for horizontal gene transfer. *BMC Genomics*, *13*, 171. https://doi.org/10.1186/1471-2164-13-171

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, *9*, 868–877. https://doi.org/10.1101/gr.9.9.868

Humann, J., Ficklin, S. P., Lee, T., Cheng, C.-H., Jung, S., Wegrzyn, J., … Main, D. (2016). GenSAS v4.0: A web-based platform for structural and functional genome annotation and curation. In *Plant and Animal Genome XXIV*. San Diego, California.

Jiang, N. (2015, September 28). Repeat Library Construction-Basic. Retrieved October 7, 2015, from http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction-Basic

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., … Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, *24*, 1384–1395. https://doi.org/10.1101/gr.170720.113

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

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., … O'Connor, S. E. (2015). Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal*, *82*, 680–692. https://doi.org/10.1111/tpj.12827

Kennedy, R. C., Unger, M. F., Christley, S., Collins, F. H., & Madey, G. R. (2010, November 17). TESeeker. Retrieved September 30, 2015, from https://www3.nd.edu/~teseeker/download.html

Kennedy, R. C., Unger, M. F., Christley, S., Collins, F. H., & Madey, G. R. (2011). An automated homology-based approach for identifying transposable elements. *BMC Bioinformatics*, *12*, 1–10. https://doi.org/10.1186/1471-2105-12-130

Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, *12*, 656–664. https://doi.org/10.1101/gr.229202

Kohany, O., Gentles, A. J., Hankus, L., & Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, *7*, 1–7. https://doi.org/10.1186/1471-2105-7-474

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 1–9. https://doi.org/10.1186/1471-2105-5-59

Kwon, C. W., Park, K.-M., Kang, B.-C., Kweon, D.-H., Kim, M.-D., Shin, S. W., … Chang, P.-S. (2015). Cysteine protease profiles of the medicinal plant *Calotropis procera* R. Br. revealed by de novo transcriptome analysis. *PLoS ONE, 10*, e0119328. https://doi.org/10.1371/journal.pone.0119328

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Lee, T., Peace, C., Jung, S., Zheng, P., Main, D., & Cho, I. (2011). GenSAS — An online integrated genome sequence annotation pipeline. *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, *4*, 1967–1973. https://doi.org/10.1109/BMEI.2011.6098712

Li, B., Zong, Y., Du, Z., Chen, Y., Zhang, Z., Qin, G., … Tian, S. (2015). Genomic characterization reveals insights into patulin biosynthesis and pathogenicity in *Penicillium* species. *Molecular Plant-Microbe Interactions*, *28*, 635–647. https://doi.org/10.1094/MPMI-12-14-0398-FI

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Librado, P., Vieira, F. G., & Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics*, *28*, 279–281. https://doi.org/10.1093/bioinformatics/btr623

Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., … Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *ArXiv:1308.2012v1 [q-Bio]*. Retrieved from http://arxiv.org/abs/1308.2012

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., … Usadel, B. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment*, *37*, 1250–1258. https://doi.org/10.1111/pce.12231

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964. https://doi.org/10.1093/nar/25.5.0955

Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., … Rep, M. (2010). Comparative genomics reveals mobile

pathogenicity chromosomes in *Fusarium*. *Nature*, *464*, 367–373. https://doi.org/10.1038/nature08850

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Margarido, G. R. A., Souza, A. P., & Garcia, A. A. F. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas*, *144*, 78–79. https://doi.org/10.1111/j.2007.0018-0661.02000.x

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., … Brettin, T. S. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*, *26*, 553–560. https://doi.org/10.1038/nbt1403

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., … Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, *43*, D213–D221. https://doi.org/10.1093/nar/gku1243

Munkert, J., Pollier, J., Miettinen, K., Van Moerkercke, A., Payne, R., Müller-Uri, F., … Goossens, A. (2015). Iridoid synthase activity is common among the plant progesterone 5β-reductase family. *Molecular Plant*, *8*, 136–152. https://doi.org/10.1016/j.molp.2014.11.005

Ng, K. P., Yew, S. M., Chan, C. L., Soo-Hoo, T. S., Na, S. L., Hassan, H., … Yee, W.-Y. (2012). Sequencing of *Cladosporium sphaerospermum*, a Dematiaceous fungus isolated from blood culture. *Eukaryotic Cell*, *11*, 705–706. https://doi.org/10.1128/EC.00081-12

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., … Denning, D. W. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*, *438*, 1151–1156. https://doi.org/10.1038/nature04332

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., … Grigoriev, I. V. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens*, *8*, e1003037. https://doi.org/10.1371/journal.ppat.1003037

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., … Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*, *37*, e123–e123. https://doi.org/10.1093/nar/gkp596

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, *21*, i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Proost, S., Bel, M. V., Sterck, L., Billiau, K., Parys, T. V., Peer, Y. V. de, & Vandepoele, K. (2009). PLAZA: A Comparative Genomics Resource to Study Gene and Genome Evolution in Plants. *The Plant Cell Online*, *21*, 3718–3731. https://doi.org/10.1105/tpc.109.071506

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inzé, D., Mueller-Roeber, B., & Vandepoele, K. (2015). PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Research*, *43*, D974–D981. https://doi.org/10.1093/nar/gku986

Pruitt, K., Brown, G., Tatusova, T., & Maglott, D. (2002). The Reference Sequence (RefSeq) Database. In J. McEntyre & J. Ostell (Eds.), *The NCBI Handbook* (p. Chapter 18). Bethesda, MD: National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK21091/

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Riley, R., Salamov, A. A., Brown, D. W., Nagy, L. G., Floudas, D., Held, B. W., … Grigoriev, I. V. (2014). Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 9923–9928. https://doi.org/10.1073/pnas.1400592111

Schmieder, R., & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE*, *6*, e17288. https://doi.org/10.1371/journal.pone.0017288

Shaw, J. J., Berbasova, T., Sasaki, T., Jefferson-George, K., Spakowicz, D. J., Dunican, B. F., … Strobel, S. A. (2015). Identification of a fungal 1,8-cineole synthase from *Hypoxylon* sp. with specificity determinants in common with the plant synthases. *The Journal of Biological Chemistry*, *290*, 8511–8526. https://doi.org/10.1074/jbc.M114.636159

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker. Retrieved October 7, 2015, from http://www.repeatmasker.org/

Solexa, Inc. (2006). Protocol for whole genome sequencing using Solexa technology. *BioTechniques Protocol Guide 2007*, 291.

Sparrow, F. K., & Pearson, N. L. (1948). Pollen compatibility in *Asclepias syriaca*. *Journal of Agricultural Research*, *77*, 187–199.

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Straub, S. C. K., Cronn, R. C., Edwards, C., Fishbein, M., & Liston, A. (2013). Horizontal transfer of DNA from the mitochondrial to the plastid genome and its

subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution*, *5*, 1872–1885. https://doi.org/10.1093/gbe/evt140

Tai, H. H., Pelletier, C., & Beardmore, T. (2004). Total RNA isolation from *Picea mariana* dry seed. *Plant Molecular Biology Reporter*, *22*, 93–93. https://doi.org/10.1007/BF02773357

Tange, O. (2011, February). GNU Parallel: The command-line power tool. ;;*Login: The USENIX Magazine*, *36*, 42–47.

Tennessen, J. A. (2015). *MakeOnemapFormatFromVcfNoParentsTwoGenos.pl*. Perl. Retrieved from https://github.com/jacobtennessen/MiSCVARS/blob/master/MakeOnemapFormat FromVcfNoParentsTwoGenos.pl

The French-Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*, 463–467. https://doi.org/10.1038/nature06148

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*, 635–641. https://doi.org/10.1038/nature11119

Van Bel, M., Proost, S., Van Neste, C., Deforce, D., Van de Peer, Y., & Vandepoele, K. (2013). TRAPID: an efficient online tool for the functional and comparative analysis of de novoRNA-Seq transcriptomes. *Genome Biology*, *14,* R134.

Weitemier, K. (2014). fastq_collapse.py (Version 0.03). Retrieved from https://github.com/kweitemier/fastq_collapse

Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences*, *2*, 1400042. https://doi.org/10.3732/apps.1400042

Wikström, N., Kainulainen, K., Razafimandimbison, S. G., Smedmark, J. E. E., & Bremer, B. (2015). A Revised Time Tree of the Asterids: Establishing a Temporal Framework For Evolutionary Studies of the Coffee Family (Rubiaceae). *PLOS ONE, 10*, e0126690. https://doi.org/10.1371/journal.pone.0126690

Wyatt, R., & Broyles, S. B. (1990). Reproductive biology of milkweeds (*Asclepias*): Recent advances. In S. Kawano (Ed.), *Biological approaches and evolutionary trends in plants* (pp. 255–272). San Diego, California: Academic Press, Inc.

Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C. J. S., Chakrabarty, R., … Sensen, C. W. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology*, *166*, 122–134. https://doi.org/10.1016/j.jbiotec.2013.04.004

Yates, S. A., Chernukhin, I., Alvarez-Fernandez, R., Bechtold, U., Baeshen, M., Baeshen, N., … Mullineaux, P. M. (2014). The temporal foliar transcriptome of the perennial C3 desert plant *Rhazya stricta* in its natural environment. *BMC Plant Biology*, *14*, 2.
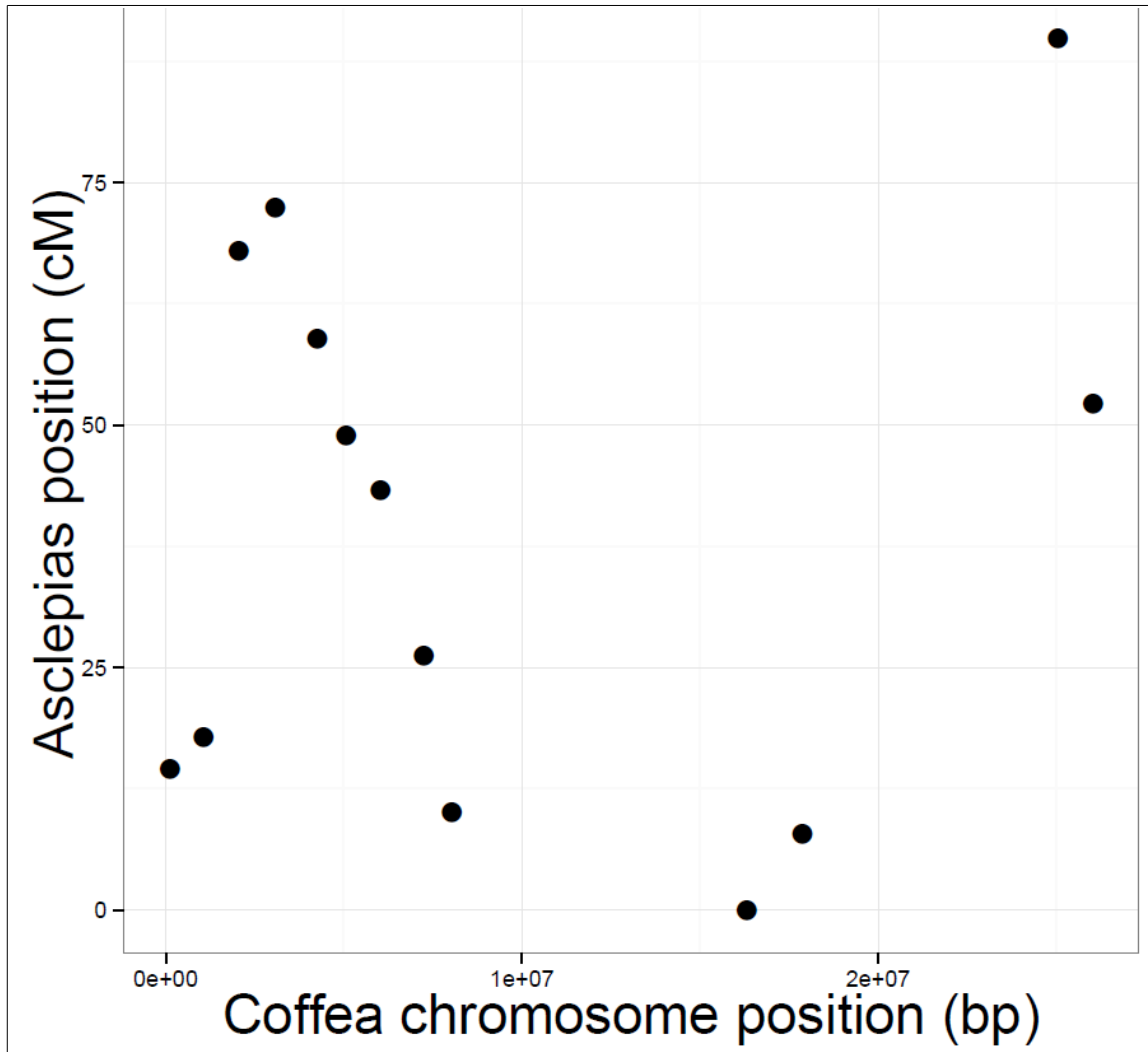
Figure S1: Synteny between *Asclepias* linkage group 2 and *Coffea* pseudochromosome 10.

A subset of scaffolds from *Asclepias* linkage group 2 mapped to their positions on *Coffea canephora* pseudochromosome 10, and ordered along the y-axis by recombination distance within *Asclepias*.
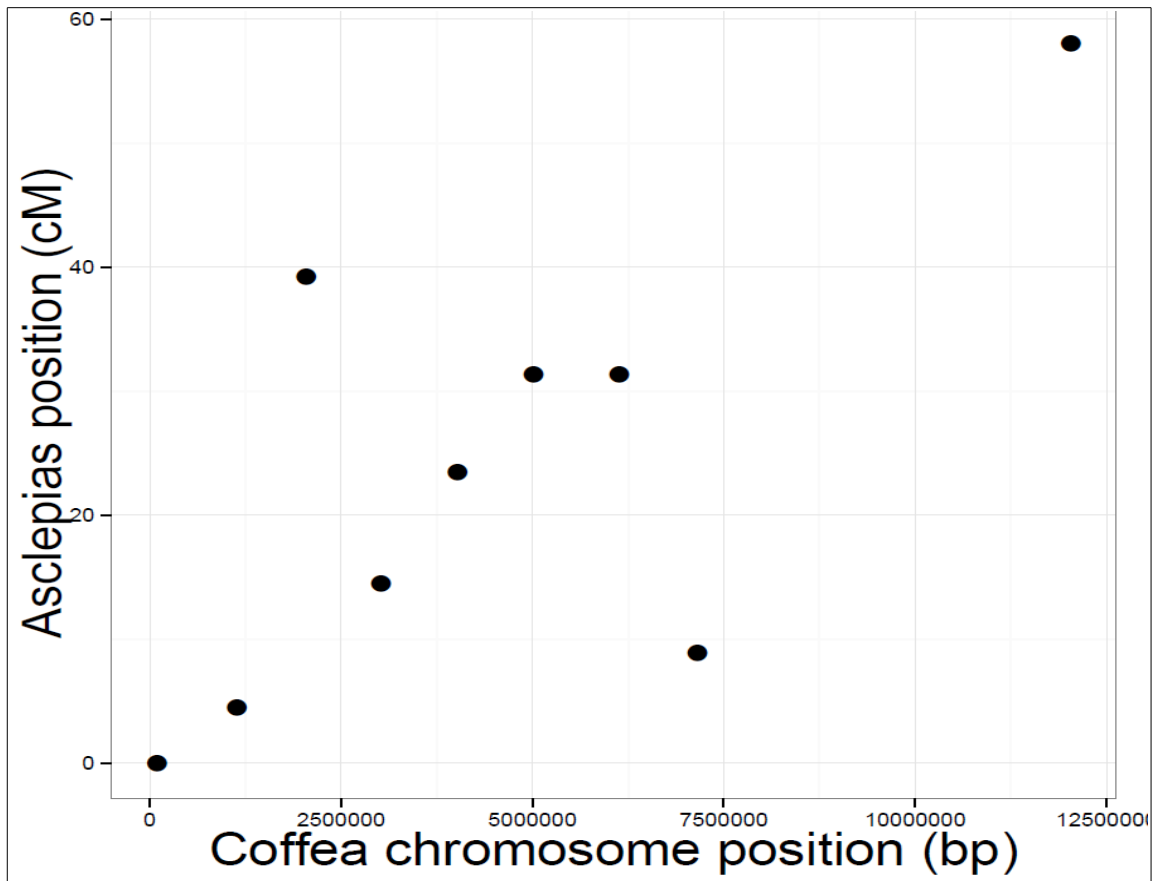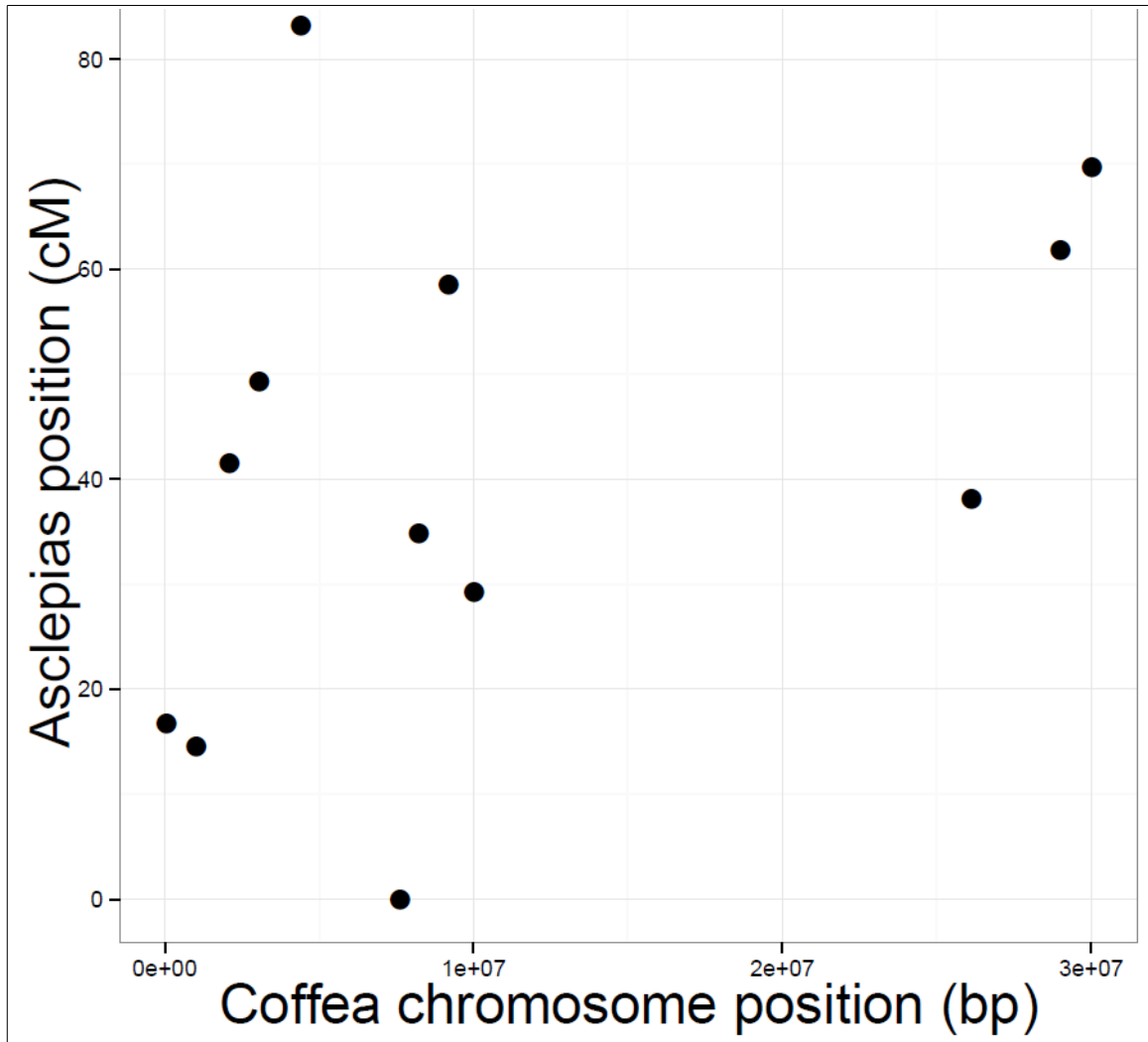
Figure S2: Synteny between *Asclepias* linkage group 8 and *Coffea* pseudochromosome 3.

A subset of scaffolds from *Asclepias* linkage group 8 mapped to their positions on *Coffea canephora* pseudochromosome 3, and ordered along the y-axis by recombination distance within *Asclepias*.
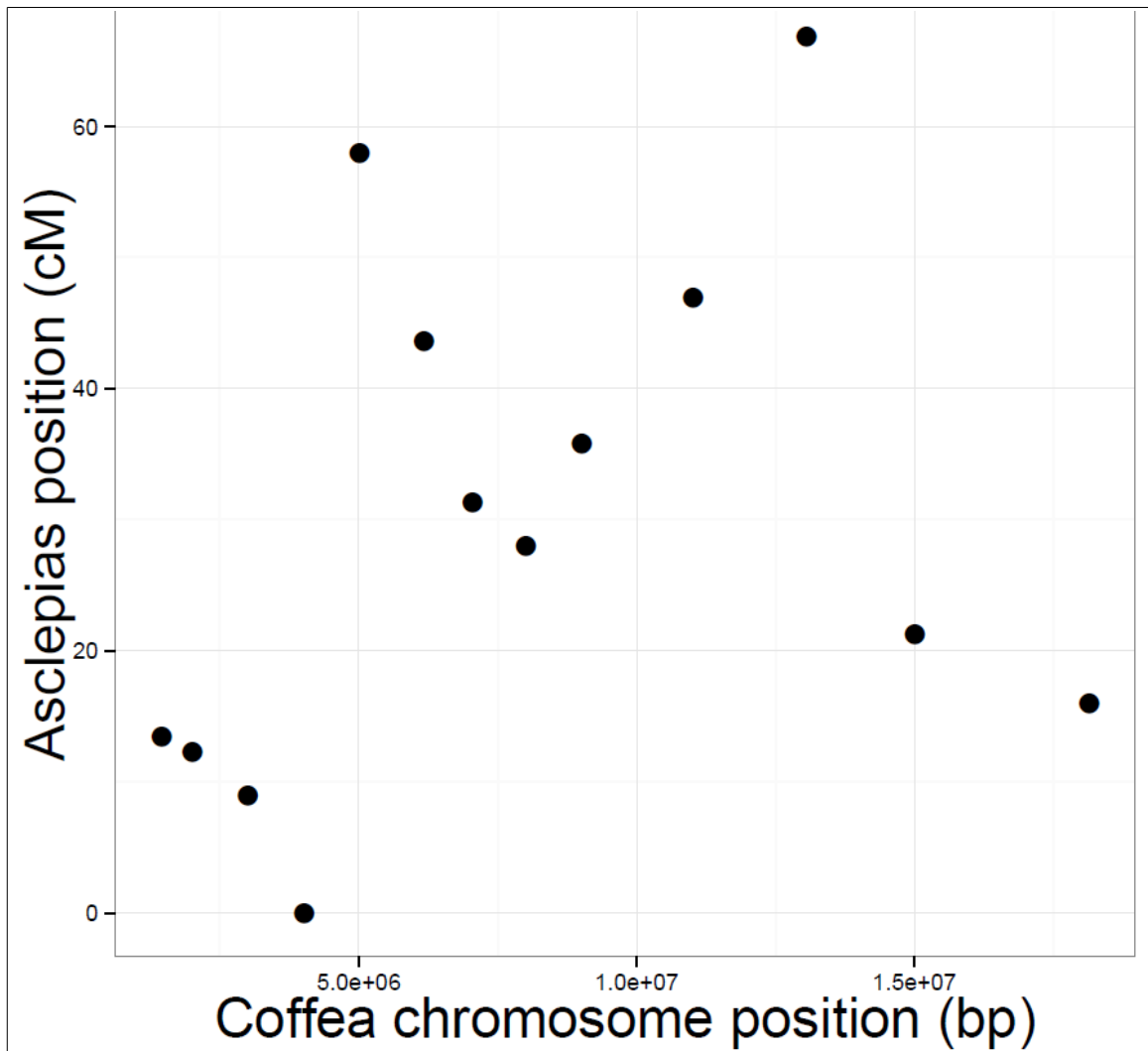
Figure S3: Synteny between *Asclepias* linkage group 4 and *Coffea* pseudochromosome 8.

A subset of scaffolds from *Asclepias* linkage group 4 mapped to their positions on *Coffea canephora* pseudochromosome 8, and ordered along the y-axis by recombination distance within *Asclepias*.

Figure S4: Synteny between *Asclepias* linkage group 6 and *Coffea* pseudochromosome
6.

A subset of scaffolds from *Asclepias* linkage group 6 mapped to their positions on *Coffea*
canephora pseudochromosome 6, and ordered along the y-axis by recombination distance
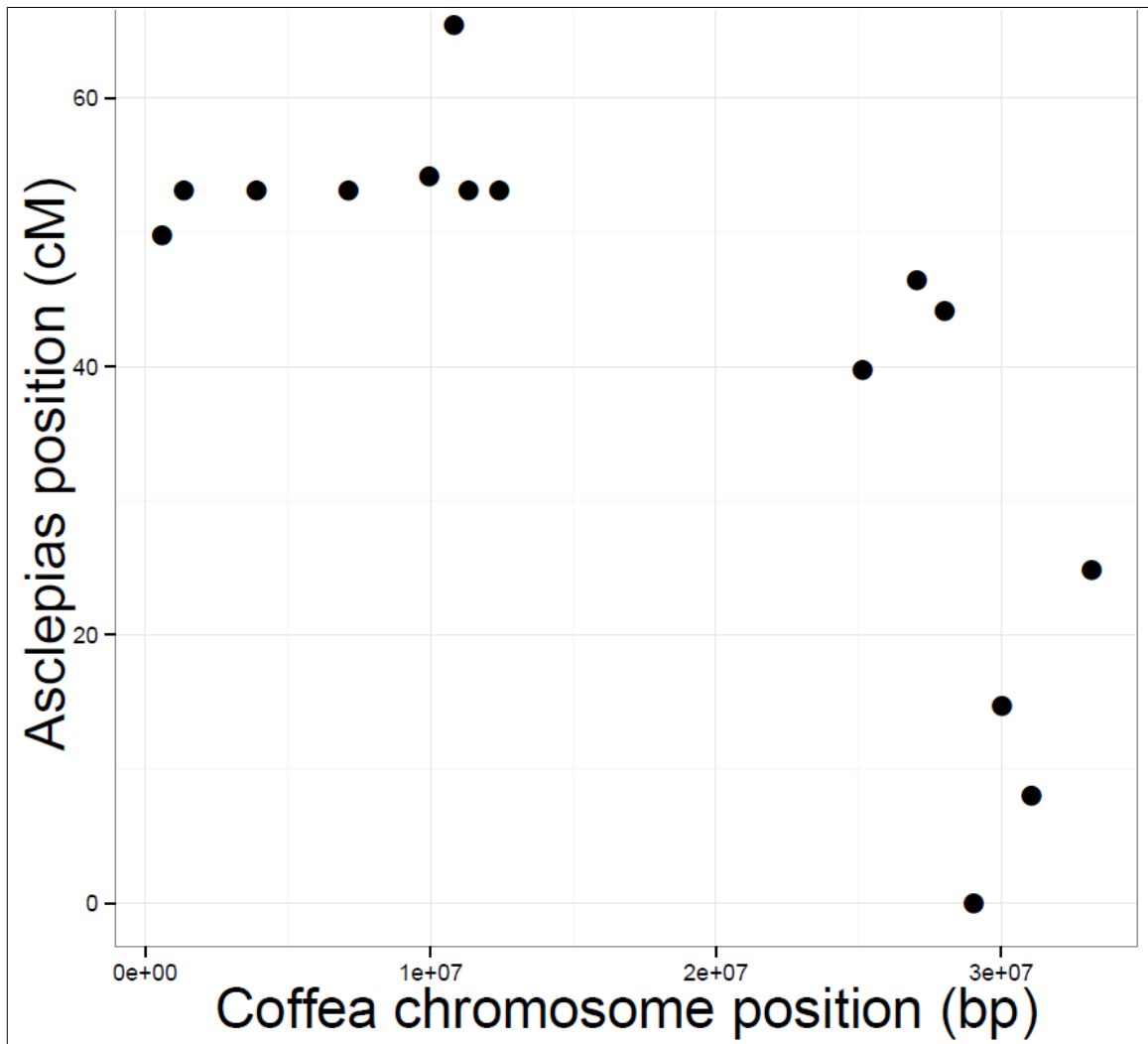within *Asclepias*.

Figure S5: Synteny between *Asclepias* linkage group 7 and *Coffea* pseudochromosome 11.

A subset of scaffolds from *Asclepias* linkage group 7 mapped to their positions on *Coffea canephora* pseudochromosome 11, and ordered along the y-axis by recombination distance within *Asclepias*.

Figure S6: Synteny between *Asclepias* linkage group 9 and *Coffea* pseudochromosome 1.

A subset of scaffolds from *Asclepias* linkage group 9 mapped to their positions on *Coffea canephora* pseudochromosome 1, and ordered along the y-axis by recombination distance within *Asclepias*.
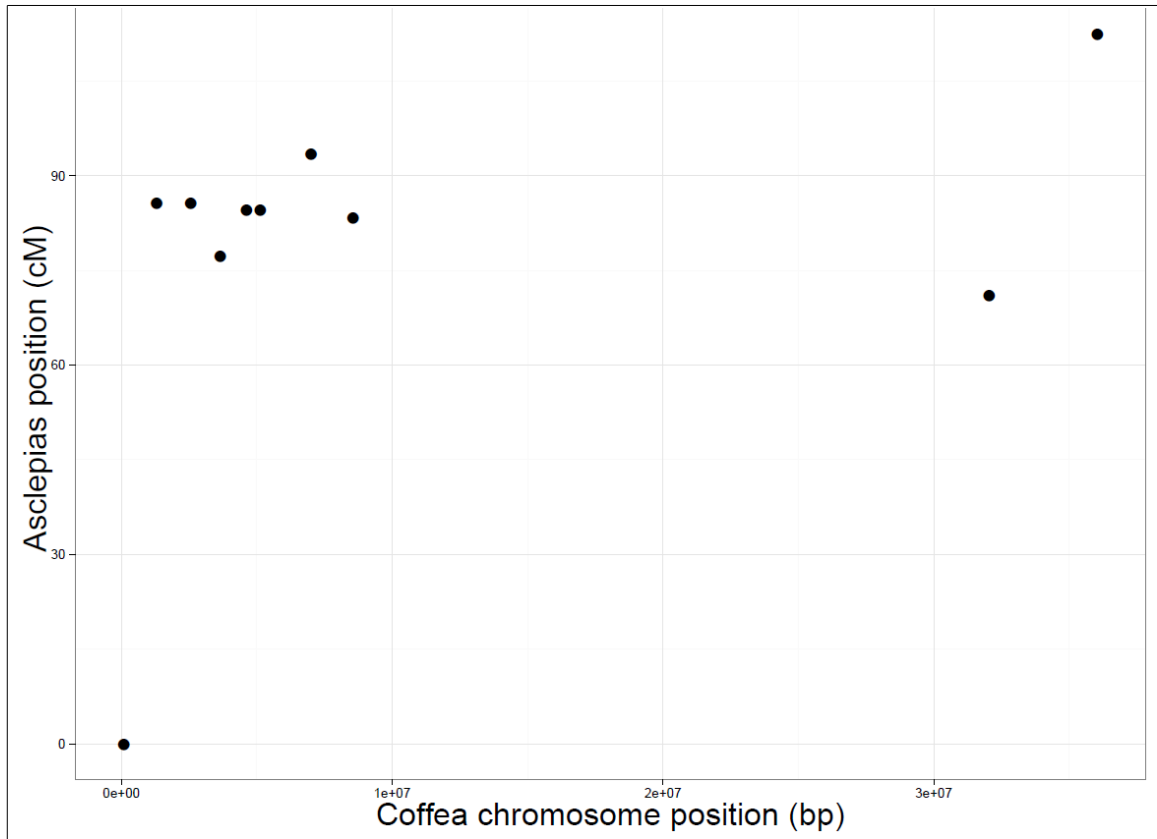
Table S1: Shared orthogroups among *Asclepias, Catharanthus, Coffea,* and *Vitis*.

Values along the diagonal are the number of orthogroups found within that genus.

	Asclepias	Catharanthus	Coffea	Vitis
Asclepias	9,837	9,275	9,439	8,753
Catharanthus		12,709	12,111	11,072
Coffea			13,233	11,688
Vitis				12,117

Table S2: Genomic coordinates of *Asclepias* P5βR genes.

Under **Scaffold ID** the linkage group of the scaffold is preceded by "LG." **Called:** Whether the gene prediction consensus accurately predicted the correct exons. The prediction that failed did predict a gene product, but included exons from adjacent genes. Ψ-progesterone 5β-reductase was accurately predicted to not produce a product.

Gene	Scaffold ID	Start	Stop	Called
Progesterone 5β-reductase 1	LG11_scaffold_m502	31537	33459	No
Ψ-progesterone 5β-reductase	LG11_scaffold_m502	33942	34876	NA
Progesterone 5β-reductase 6	LG00_scaffold217668	1305	136	Yes