RESEARCH ARTICLE

The evolution of multi-gene families and metabolic pathways in the evening primroses (*Oenothera*: Onagraceae): A comparative transcriptomics approach

Eunice Kariñho-Betancourt[1]*, David Carlson[2], Jessie Hollister[2], Axel Fischer[3], Stephan Greiner[3○], Marc T. J. Johnson[1○]*

1 Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, Canada, 2 Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, United States of America, 3 Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

○ These authors contributed equally to this work.
* karinho.betancourt@gmail.com (EKB); marc.johnson@utoronto.ca (MTJJ)

Abstract

The plant genus *Oenothera* has played an important role in the study of plant evolution of genomes and plant defense and reproduction. Here, we build on the 1kp transcriptomic dataset by creating 44 new transcriptomes and analyzing a total of 63 transcriptomes to present a large-scale comparative study across 29 *Oenothera* species. Our dataset included 30.4 million reads per individual and 2.3 million transcripts on average. We used this transcriptome resource to examine genome-wide evolutionary patterns and functional diversification by searching for orthologous genes and performed gene family evolution analysis. We found wide heterogeneity in gene family evolution across the genus, with section *Oenothera* exhibiting the most pronounced evolutionary changes. Overall, more significant gene family expansions occurred than contractions. We also analyzed the molecular evolution of phenolic metabolism by retrieving proteins annotated for phenolic enzymatic complexes. We identified 1,568 phenolic genes arranged into 83 multigene families that varied widely across the genus. All taxa experienced rapid phenolic evolution (fast rate of genomic turnover) involving 33 gene families, which exhibited large expansions, gaining about 2-fold more genes than they lost. Upstream enzymes phenylalanine ammonia-lyase (PAL) and 4-coumaroyl: CoA ligase (4CL) accounted for most of the significant expansions and contractions. Our results suggest that adaptive and neutral evolutionary processes have contributed to *Oenothera* diversification and rapid gene family evolution.

Introduction

The evening primrose plant genus *Oenothera* (Onagraceae), has served as a model for addressing numerous problems in plant biology during the past 120 years [1]. Pioneering studies in the late 19th and early 20th centuries using *O. lamarckiana* (= *O. glazioviana*) and other taxa,

contributed to the rediscovery of Mendel's laws of inheritance [2], the formulation of the mutation theory [3–6], helped prove the chromosomal theory of inheritance [7,8], and provided among the earliest examples of genetic self-incompatibility in plant mating [9,10]. The combination of cytogenetics and genetics, first applied to *Oenothera* at the level of the population, contributed to elucidating the role of chromosomal translocations in genome rearrangements, meiotic behavior, and recombination, as well as the role of translocations and speciation [11–14]. In addition, evening primroses are a model system for the study of the genetics, genomics and evolution of plastids in which, for example, the role of the plastids for plant speciation was recognized [15–17], the contribution of the cytoplasmic genetic elements to plant morphology was discovered [18,19], and the theory of selfish cytoplasmic elements developed [20]. In molecular population genetics, the first use of allozyme markers in plants was performed in *Oenothera biennis* [21,22], in which it was shown that populations contained unexpectedly high levels of genetic variation. Finally, *Oenothera* and the Onagraceae at large are among the best studied groups in plant taxonomy and systematics [10,23–34].

Although *Oenothera* has played a fundamental role in plant biology for over a century [1,13], the development of modern genetic and genomic tools has been relatively slow compared to other model systems. Early genomic resources included a complete plastid genome sequence of *O. elata* [35], which was expanded to include all five plastome types in *Oenothera* subsection *Oenothera*, allowing for comparative analysis of chloroplast evolution [36,37]. PCR-based dominant amplified fragment length polymorphisms (AFLPs), codominant simple-sequence repeats (SSR), cleaved amplified polymorphic sequence (CAPS), and single sequence length polymorphism (SSLP) marker systems for both the nuclear genomes and plastomes were developed over a decade ago [38,39], which allowed for the creation of a genetic linkage map in *Oenothera* [40]. Progress in molecular biology and sequencing technology allowed for the creation of early EST libraries [41]. In parallel, cytological techniques have continued to advance, including the development of fluorescence *in situ* hybridization [42,43].

The development of these molecular resources has allowed for recent advances in genetics, evolutionary biology, and systematics, and opened new interdisciplinary directions. For example, these resources allowed the first test in plants of multiple predictions stemming from theory of the evolution of sex [40,44–51]. Recent research also led to important advances in the study of plant defense evolution [52–59], and eco-evolutionary dynamics within communities [60–68]. Further, genomic tools have made it possible to determine the role of nuclear-organellar incompatibilities in speciation [69,70], mechanisms of mutations in plastid genome evolution [71], novel mechanisms in chloroplast gene expression [72], and the identification of chloroplast genes involved in plastid transmission [73]. The advances in cytological methods have further improved our understanding of meiosis, and how translocations in *Oenothera* lead to the formation of chromosomal rings in diakinesis [42,43]. Additionally, metabolomic methods [74] have been combined with phylogenetic and genetic approaches to disentangle the role of secondary compounds in plant defense evolution over micro- and macroevolutionary timescales [53,56,59,75–77]. Finally, molecular research in *Oenothera* has made it possible to refine our understanding of *Oenothera* systematics [33,34,44,45,50,78,79], which resulted in key revisions to the taxonomy of the Onagraceae [25].

Recently, as part of the One Thousand Plant Transcriptomes Initiative (1kp project), more than 1000 species were sequenced, including 18 *Oenothera* species, to provide a robust phylogenomic framework for examining the evolution of green plants [80–82]. Here, we combined the 1kp *Oenothera* dataset with 44 transcriptomes separately generated for an additional 11 species (see [49]) and reassembled 19 transcriptomes to greatly increase the number and quality of the *Oenothera* transcriptome resource. We develop comparative transcriptomic tools

across the genus and show their utility in understanding the genomic and functional diversification.

Oenothera is a monophyletic genus within the Onagraceae family, which is nested within the Myrtales order and the Rosids subclass. The base number of chromosomes is X = N = 7 and evening primrose are usually diploid, but a few species (not studied here) contain polyploid lineages as high as N = 28 [83]. The genus is comprised of 151 species of perennial and annual herbs that are native to North and South America [25]. These species occur throughout boreal, temperate, montane, arid and coastal habitats. *Oenothera* is well known for having variable reproductive and genetic systems. Many species are outcrossing, which often involves genetic self-incompatibility as well as herkogamous flowers that are mainly pollinated by moths and specialized bees [84–86]. Some species are self-pollinating [87], whereas ca. 45 species have a specialized genetic system known as permanent translocation heterozygosity (PTH) [88], which results in plants being functionally asexual because of a loss of recombination and segregation, and a balanced lethal genetic system whereby only heterogametic haploid chromosome sets can form viable offspring [1,13].

Oenothera has served as a model for the study of plant secondary metabolism, particularly as it relates to phenolic metabolism. All members of the genus produce phenolic compounds, including flavonoids and hydrolysable tannins [44,56,59]. Phenolic compounds derived from the amino acid L-phenylalanine via deamination by L-phenylalanine ammonia lyase (PAL). Several classes of phenolics follow the shikimic acid pathway in combination with the mevalonate-acetate route. Different biosynthetic paths of phenolics can be organized into a 'core' phenylpropanoid pathway, including upstream enzymes: PAL, the member of P450s family cinnamate 4-hydroxylase (C4H) and 4-coumarate coenzyme A ligase (4CL), which feeding further specific pathways [89,90]. Complex phenylpropanoids like flavonoids, flavonols and tannins involve specific mid- and downstream enzymes such as chalcone synthase (CHS) and chalcone isomerase (CHI) to yield their true backbone, and flavonoid 3´-hydroxylase (F3´H), flavanone 3-hydroxylase (F3H) and flavonol synthase (FLS) (S1 Fig), which catalyze terminal biosynthetic branches [91,92]. Phenolics are well known stress response metabolites, with induction of phenylpropanoid metabolism and accumulation of phenolic compounds especially observed in response to biotic stresses such as herbivory [93–95]. Evening primrose is consumed by many insect herbivores [46,61], and plants defend themselves with a diverse arsenal of phenolics. Several studies have shown that specific phenolic compounds reduce herbivore performance [77], and herbivores impose selection that drives rapid evolution of these metabolites [52,53]. Investment in secondary metabolites varies biogeographically both within [55] and between species [59]. Moreover, a loss of sex through PTH is associated with reduced levels of defense against generalist insects in *Oenothera* [52].

We used vegetative transcriptomes for a broad taxonomic sampling of *Oenothera* species and subspecies to characterize gene family evolution and gene expression. After providing a thorough transcriptomic overview, we ask the following questions. How have gene families diverged among species and/or clades? How can this transcriptomic analysis be used to examine the evolution of secondary metabolism and defenses in PTH and sexual plants? Answering these questions will contribute to understanding the molecular and phenotypic evolution of this important system.

Materials and methods

Species sampling

The *Oenothera* genus is arranged into two major clades (A and B), which include 18 sections and 18 subsections [25]. We sampled 29 species from seven subclades including diverse

sections and subsections of the genus *Oenothera*. The sampling also comprises five subspecies from two species, for a total of 32 taxa. This included sampling 13 sexual species and 16 functionally asexual PTH species across clades. Species' naming authority, source of seeds, and voucher information associated with the 1kp project and additional transcriptome sampling at the NC State Genomic Facility are provided in S1 Table. We also sampled 3–5 populations from each of 11 species to give insight into genetic variation within species, but are not described here since the focus of this paper was on the comparative context of the transcriptomic analysis. These species provide an excellent platform for addressing evolutionary questions since they maximize diversity across the genus *Oenothera*.

Plant materials

Seeds of 29 *Oenothera* species (32 taxa including subspecies) were started in Petri dishes. All species except for the *Gaura* section (i.e., *O. filiformis*, *O. gaura*, and *O. suffulta*) were germinated in 55 mm Petri dishes on filter paper soaked in 1mL of 0.1% agarose solution in filtered water, at 4 ºC for 18–24 hours and then left in a windowsill that received full sun for 3–4 hours per day. All *Gaura* fruits were soaked in water as described above but the entire indehiscent capsules were planted 1 cm below moist soil and placed under fluorescent lights for germination. Seeds were transplanted to soil when roots and cotyledons emerged from seeds. All plants were planted in 150 mL pots, in sterilized Ready Earth Seedling and Plug Soil Mixture (Sungro Horticulture, Canada) soil and watered *ad libitum* with DI water with 1/3 strength Hoaglands solution. Plants were grown under 16:8, L:D cycle with 25:20˚C (L:D) with fluorescent and incandescent lights. Plants were harvested when the 4th true leaf was 1/4 the area of 3rd true leaf, for all plants. Leaves were flash frozen in liquid N_2 and placed in a -80 ºC freezer. For further details on *Oenothera* culture and propagation (see [96]).

Sequencing and transcriptome assembly

Total RNA was isolated from leaf tissue of 63 individuals from *Oenothera* spp. following a CTAB/Acid Phenol/Silica Membrane method [protocol 9; 49,80]. This total RNA was then purified into mRNA using the dyna bead mRNA purification kit [80], and employed to prepared 63 TrueSeq libraries for sequencing. Libraries from 50 samples were sequenced under a pair-end mode using Illumina HiSeq as part of the One Thousand Plant Transcriptomes Initiative (1kp project) [80,81], and an additional 13 samples were sequenced under a single-end mode (13 samples) using GAIIx (S1 Table) at the NC State Genomic Sciences Laboratory (https://research.ncsu.edu/gsl/). All RNA-seq reads (SRA project number: PRJEB4922) were newly trimmed and filtered for quality using fastp v.0.20.0 as part of the current paper [97]. The "—cut_tail" flag was used to truncate reads if a 4 bp window fell below an average PHRED quality score of 20. Quality-controlled RNA-seq reads were then assembled into transcripts using Trinity v2.11.0 [98,99] using default settings except for CPU and memory allocation. Assembly metrics were generated using Trinity's 'TrinityStats.pl'.

Functional annotation

Each assembly was functionally annotated using the Trinotate v3.2.1 pipeline [100]. Briefly, we used Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder) to find putative protein-coding sequences (CDS) in each transcriptome. Each set of CDS and translated amino acid sequences were then searched using blastp v2.5.0 with a maximum e-value of 1e-5 [101] and HMMER v3.3's hmmscan with default settings against the Swiss-Prot (last accessed July

13, 2020; [102]) and Pfam (last accessed July 13, 2020; [103]) databases. As part of the Trinotate pipeline, we also used SignalP v4.1 [104] to predict signal peptides, TMHMM v2.0c [105] to find transmembrane helices, and RNAmmer v1.2 [106] to predict RNA genes. These four programs were run using default settings, after which the annotation results were loaded into an SQLite database and summarized with Trinotate's "report" option. To summarize the functional annotation results for each taxon, we counted the number of annotations across the three main GO categories—cellular component, molecular function, and biological process—assigned by Trinotate based on blastp matches between each predicted protein and sequences in the Swiss-Prot database. In addition, we used GOATOOLS v.0.8.2 [107] to further summarize the orthogroup functional annotation results into a set of discrete categories of particular interest. The completeness of the transcriptome gene space was evaluated by using Benchmarking Universal Single-Copy Orthologs (BUSCO), and PLAZA comparative resources. Both BUSCO and PLAZA coreGFs currently define different sets of conserved genes to model the expected gene space at different evolutionary scales within diverse lineages. We performed the BUSCO v.3.0.2 analysis with the transcriptome mode option and the lineage set to embryophyta_odb9 [108], whereas the PLAZA v.4.0 'green plants' coreGFs was set to Rosids [109].

Ortholog identification and phylogenetic inference

A subset of 30 transcriptomes that showed the highest relative completeness within each taxon for each of 30 *Oenothera* species and subspecies were chosen for phylogenetic analysis following the BUSCO-based cutoff of <40% of missing and fragmented genes in [82] (S2 Table). For every Trinity "gene" (i.e. unigene) in each of these assemblies, we retained the longest isoform for downstream analysis. Protein-coding genes were predicted from each single-isoform assembly using Transdecoder and validated by searching them against the Swiss-Prot and Pfam databases as noted above. Next, we used the combined set of predicted proteins to infer orthogroups using OrthoFinder v2.4.0 [110,111]. Orthofinder was run with DIAMOND v.0.9.14 [112] for local protein alignment, MAFFT v7.471 [113,114] to generate multiple sequence alignments, and IQ-TREE v.2.0.3 [115,116] for gene tree inference.

To build a phylogeny for all 30 taxa, we used a subset of 1,017 orthogroups, chosen by OrthoFinder, and consisting entirely of single- copy- genes. These orthogroups were then concatenated and aligned across species with MAFFT v7.471 [113,114]. After determining the best-fitting protein substitution model for our data under the BIC criterion with ModelFinder [117]. We generated two phylogenetic hypotheses for *Oenothera* using two distinct methods. The first analysis was conducted based on a concatenated approach using IQ-TREE [115], while the second inferred phylogenetic relationships followed a quartet-based species tree approach using ASTRAL-MP v5.15.5 [118,119]. For the IQ-TREE tree we performed 1,000 ultrafast bootstrap replicates and used nearest neighbor interchange for bootstrap tree optimization [120]. For the ASTRAL analysis, individual genes were inferred using IQ-TREE as noted above. Each of the 1,017 gene trees were input to ASTRAL for species tree inference, which was run using default settings. On the resulting species tree, the internal branch lengths were presented in coalescent units. The branch support values signify a local posterior probability based on the quartet frequencies. The maximum likelihood phylogenies were rooted using *O. capillifolia* ssp *berlandieri* as outgroup (S2 Fig). The concatenation and coalescence approach resulted in quite congruent species trees, with good support at the basal nodes. Most of the species-level differences occur within subsection *Oenothera*, where neither tree has particularly good branch support values. Since the two different phylogenies were largely

congruent, we employed the concatenation-based tree, which overall showed good node support, for further comparative analysis.

Comparative analysis

Gene family evolution. The output of OrthoFinder was parsed to identify gene families. The species tree based on the concatenated alignments of single-copy orthologs was time-calibrated using the r8s v1.8.1 program [121] with the penalized likelihood method [122]. To provide node age constraints necessary for the time calibration, previously inferred [49] point estimates of divergence time between sister species from four different subsections were used. To assess gene family evolution of the 30 *Oenothera* species and subspecies, we used only the gene families with more than three gene copies per family (23,526) and the species ultrametric tree as inputs to the CAFE v4.2.1106 open access program (Computational Analysis of gene Family Evolution; [121]). The CAFE program uses a stochastic approach to estimate the birth-death (λ) parameter along the provided tree and gene family counts. Lambda describes the probability that any gene will be gained or lost at each node and terminal branch [123], accounting for gene family expansions, contractions. The CAFE software was run using the mode in which the net gain and loss rate are estimated as a single parameter (λ) for each gene family over the whole phylogeny. A strength of the program is that it can detect rapid evolving families, which depict fast rate of genomic turnover (gains and losses per gene per million years). The program estimates the *P* values of gene families in the extant species that are below the threshold. Thus, branches with low *P* values can be regarded as corresponding to gene family expansions and contractions with accelerated evolution rates. On the other hand, a limitation of the program is that it can overestimate gene loss from transcriptome assessment, especially when suboptimal completeness. Hence, we selected the most complete assemblies per species to analyze gene family evolution and employed the longest isoform per gene to avoid redundancies. To ensure transcriptomes were being compared in an equivalent way, all plants were grown in identical conditions and transcriptomes were compared from the same tissue and developmental stage as described above. For the entire analysis, the CAFE overall p-value threshold was kept at its default value (0.01).

Defensive gene families and phenolic evolution. To examine the molecular basis for the evolution of secondary metabolism and plant defenses in PTH and sexual plants, we retrieved gene ontology (GO) annotations to search for defense-related genes including those related to eight major phenolic enzymes; PAL, C4H, 4CL, CHS, CHI, F3´H, F3H and FLS. To investigate the potential functions of the Orthofinder results, we assigned defense-related GO terms to each orthogroup if at least one of the proteins in that orthogroup was annotated with the term. We parsed the GO annotations and the orthogroups database to identify putative defense-related proteins encompassed in specific functional categories, and then grouped them under *umbrella* categories according to GO "biological processes" associated with plant defense. To examine the evolution of phenolic metabolism, we tracked phenolic-related genes of each of the 30 taxa. The count and proportion of *defensive* orthogroups and genes corresponding to specific gene ontology categories were then calculated. To investigate whether the observed gene family evolution is associated with functional diversity of chemical plant defense, we retrieved and summarized gene families associated with phenolic biosynthesis for which the rate of gain/loss was significantly different (rapid evolution; *p*-value threshold = 0.01) among taxa.

The protocols employed for phylogenetic and comparative analyses are deposited at dx.doi.org/10.17504/protocols.io.bwnzpdf6.

Results

de novo transcriptome data and assembly

Based on Illumina short-read sequencing, 63 transcriptomes from 29 *Oenothera* species (32 taxa including subspecies) were generated [80]. On average, 2.5 billion nucleotides per sample were sequenced [49,80]. After quality filtering, we obtained 30.4 million reads per individual on average. For all sequenced individuals, we produced *de novo* assemblies of mRNA transcripts. Individuals had on average 36,909 assembled transcripts of at least 300 bp in length, and 25.4 Mb per-individual average total length of assembled sequence. Predicted proteins from TransDecoder ranged from 13,291 to 28,379. *Oenothera elata* ssp. *hirsutissima* had the largest number of transcripts (49,655) and *O. elata* ssp. *hookeri* had the largest assembly (35M). *Oenothera villosa* ssp. *villosa*, had the lowest number of transcripts (22,547) and assembly length (14 million). (S2 Table). Our 63 *Oenothera* transcriptomes represented 59% of the complete single copy orthologs (BUSCOs) on average, across 29 species PLAZA scores accounted for 86% of conserved gene sets on average, indicating good completeness (S2 Table).

Annotation and gene ontology (GO) analysis

To provide comprehensive annotation of the 63 transcriptomes, we conducted sequence homology searches of the longest isoform using BLAST tools and retained only protein-coding transcripts for downstream analysis. On average, 84% of sequences matched the annotation databases (blastx hits), ranging from 13,752 (61%) annotated genes of *O. villosa* ssp. *villosa*, to 38,200 (98%) of *O. filiformis* (S3 Table). From blasted genes (transcripts with positive hits), 71% of sequences matched with Pfam's database and more than 80% were associated with GO terms. We identify a wide range of GO terms in each assembly, indicating that three functional categories; cell component, molecular function and biological process, were well represented (S3 Fig).

Comparative transcriptomic analyses

Protein coding genes from 30 transcriptomes were used to construct orthogroups (gene families). There were 670,846 genes out of 681,746 (94.8% of total) organized in orthogroups or protein families (S4 Table). The mean and median of gene family size is 25.3 and 36 proteins (G50 = 36), respectively. We found 7,760 families present in all species, whereas 1,071 consisted entirely of single copy-genes.

The species phylogeny included seven sections and subsections of *Oenothera* showing two major clades. Clade A comprises all subsections of section *Oenothera*, (i. e., *Oenothera, Munzia, Candela* and *Raimannia*), whereas clade B includes section *Hartmannia* and subsection *Gaura*. *Oenothera capillifolia* ssp. *berlandieri* (section *Calylophus*) rooted the tree at about 1.4 MYA. These data tentatively suggest that Clade A diverged ~ 0.95 MYA from clade B. The rate of gene gain and loss (λ) estimated from the CAFE analysis was 0.0003 genes over time, for the entire tree (Fig 1). Some of the fastest gene family evolution occurred within subsection *Oenothera*. For example, the internal branch with the largest number of rapidly evolving gene families corresponds to the most recent common ancestor of *O. elata* ssp. *hirsutissima* and *O. jamesii*. The terminal branch with the most rapidly evolving gene families and largest significant gene family contractions is the one leading to *O. biennis*; this subclade also included the only two other species that showed more than 4,000 rapidly evolving gene families (*O. grandiflora* and *O. nutans*). *Oenothera elata* ssp. *hookeri* showed the highest number of significant
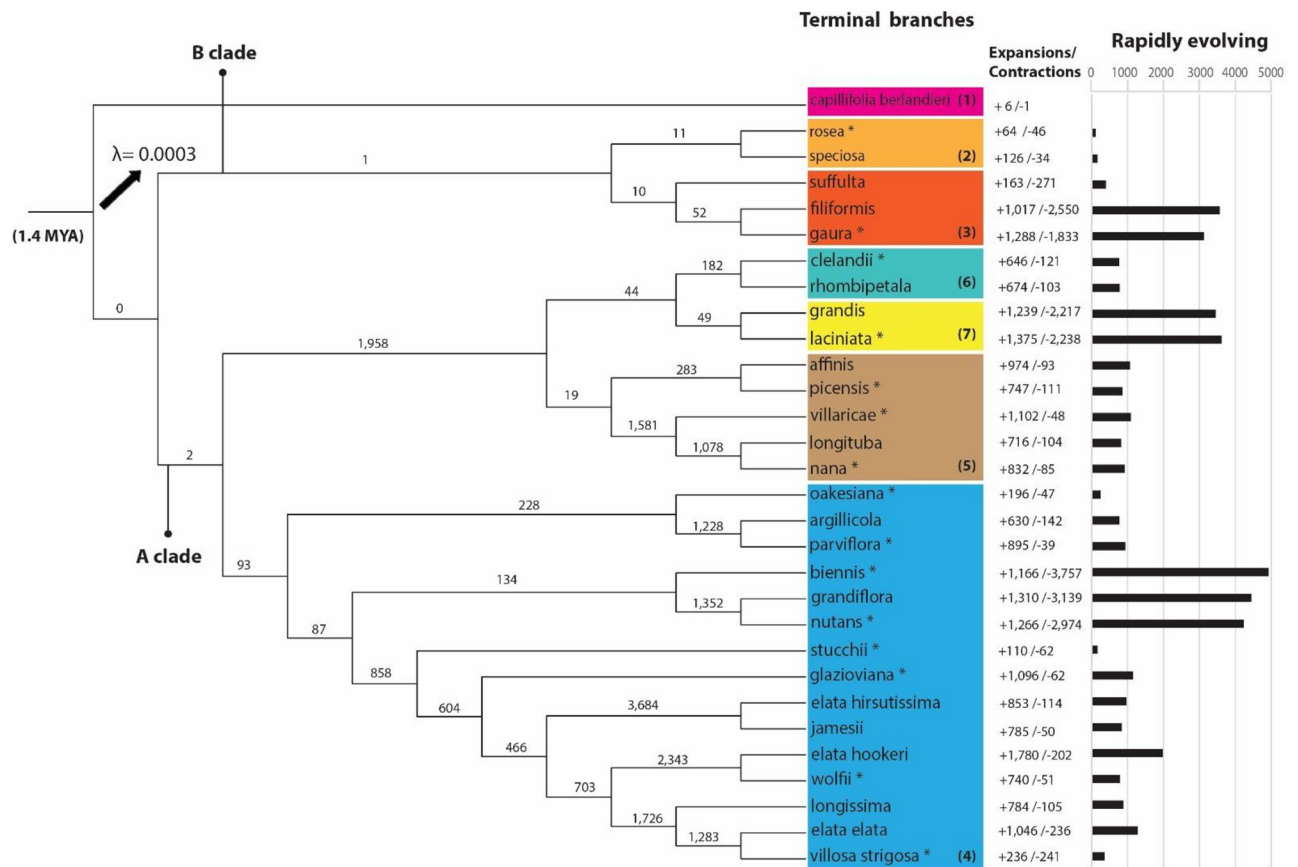
Fig 1. Phylogenetic time tree constructed from concatenated alignment of 1,017 orthogroups consisting entirely of single-copy genes from 30 *Oenothera* **taxa with** *Oenothera capillifolia* **ssp.** *berlandieri* **from subsection** *Calylophus* **(1) as outgroup.** Seven sections/subsections are depicted. Two major clades A and B include: section *Hartmannia* (2), subsection *Gaura* (3), subsection *Oenothera* (4), subsection *Munzia* (5), subsection *Candela* (6) and subsection *Raimannia* (7). Asterisks depict PTH species. The number of significant gene family expansions (+), contractions (-) and rapidly evolving gene families resulted from CAFE analysis are shown on terminal branches. Also, the number of rapidly evolving families are depicted above internal branches. The rate of gene gain and lost (lambda) for the whole tree was 0.0003.

gene family expansions, whereas *O. capillifolia* ssp. *berlandieri* from section *Calylophus* had the least significant expansions and contractions (Fig 1).

Defense-related gene families and phenolics evolution

We identify two major biological processes involved in plant defense and secondary metabolism: 1) regulation of defense response, including response to fungi, bacteria, viruses, nematodes, and insects, and 2) regulation of secondary metabolites and plant hormones, which includes, the phenolic biosynthetic process and hormone-based metabolism. We found 61 GOs related to plant defense and secondary metabolism, from which 74% corresponded to regulation of defense response, and the other 26% belonged to regulation of secondary metabolites and plant hormones (Table 1). Response to fungi and oomyecetes involved the highest number of GO categories (12). Response to insect herbivores and wounding involved 7 GO categories associated with 386 non-exclusive orthogroups (i.e., orthogroups including proteins annotated in diverse GO categories). The shikimate and phenolic biosynthesis showed 8 GO categories, the highest number of categories out of the regulation of secondary metabolism and plant hormones processes. The hormone-based responses, which include the jasmonic

Table 1. Functional categories from 30 *Oenothera* taxa of orthogroups related to defense and secondary metabolism with gene ontology (GO) annotations. For each GO biological process, we provide the sum of annotations related to gene ontology categories, a representative accession number of each category and the number of orthogroups associated with annotations of specific categories.

Biological process	# GO category	GO accession number	# Orthogroups
Regulation of defense response			
Immune response	5	0045087	240
Defense response	5	0006952	1125
Response to fungus and oomycetes	12	0050832	1028
Response to bacteria	9	0042742	1062
Response to virus	4	0051607	110
Response to nematode and other organism	3	0002215	20
Defense response to insect/ herbivore and wounding	7	0009625	386
Regulation of secondary metabolites and plant hormones			
Secondary metabolism	1	0019748	16
Shikimate and phenolic biosynthetic process	8	0009813	201
Hormone-based response (ethylene, jasmonic and salicylic acid metabolism)	7	0009753	724

Note. When a biological process included more than one GO category (gene ontology name), a representative GO accession number was provided.

acid and salicylic acid metabolic pathways, showed the highest number of non-exclusive orthogroups (Table 1). A full list of defense-related GO categories and orthogroups is provided in S5 Table.

To analyze the evolution of phenolic metabolism across 30 *Oenothera* taxa, we quantified gene families with annotations for major phenolic enzymes, PAL, 4CL, C4H, CHS, CHI, F3′H, F3H and FLS, and accounted for those exhibiting rapid evolution. In total, we identified 83 phenolic-related gene families, of which more than half (43) correspond to the "core" phenylpropanoid pathway (PAL, 4CL and C4H). Chalcone enzymes (CHS and CHI) comprised ~10% of phenolic proteins and downstream enzymes involved in the synthesis of flavonoids, coumarins and tannins, among others, represented ca. 40%. The 4CL enzyme showed the highest proportion of phenolic proteins (28%) and the CHS the least proportion (3%) (Fig 2A). Phenolic gene families per species ranged from 28 (*Oenothera elata* ssp. *hookeri* and *O. wolfii*) to 17 (*O. capillifolia* ssp. *berlandieri* and *O. affinis*). Subsection *Oenothera* and *Munzia* comprised species with the largest amount of shared and species-specific phenolic gene families, consistent with their monophyletic relationship within *Oenothera* clade A. We identified 40 species-specific gene families distributed across 21 taxa of the seven subclades. *Oenothera elata* ssp. *hookeri* comprised 6 species-specific gene families, followed by *O. biennis* with 4, and by *O. villaricae*, *O. suffulta*, *O. nana* and *O wolfii* with 3 families by each species. *Oenothera wolfii*, *O. jamesii* and *O. oakesiana* showed the highest number of intersections with other species (23), whereas *Oenothera capillifolia* ssp. *berlandieri* and *O. affinis* showed the least number of shared gene families (14). About 70% of phenolic families occurred in < 10 species, whereas 20% of the families occurred in ≥ 20 species. Three gene families of the "core" phenylpropanoid occurred in all taxa, and two families of the flavonoid branch occurred in all but one taxon. However, except for the three families occurring in the 30 taxa, every gene family shared by at least two taxa showed a unique pattern of intersection, never involving the same species (Fig 2B).

About 40% of phenolic gene families rapidly evolved according to CAFE analysis. Both PAL and 4CL enzymes accounted for most of the significant phenolic evolution, with 8 and 9 gene families, respectively (Table 2). C4H and FLS involved the least number of rapidly
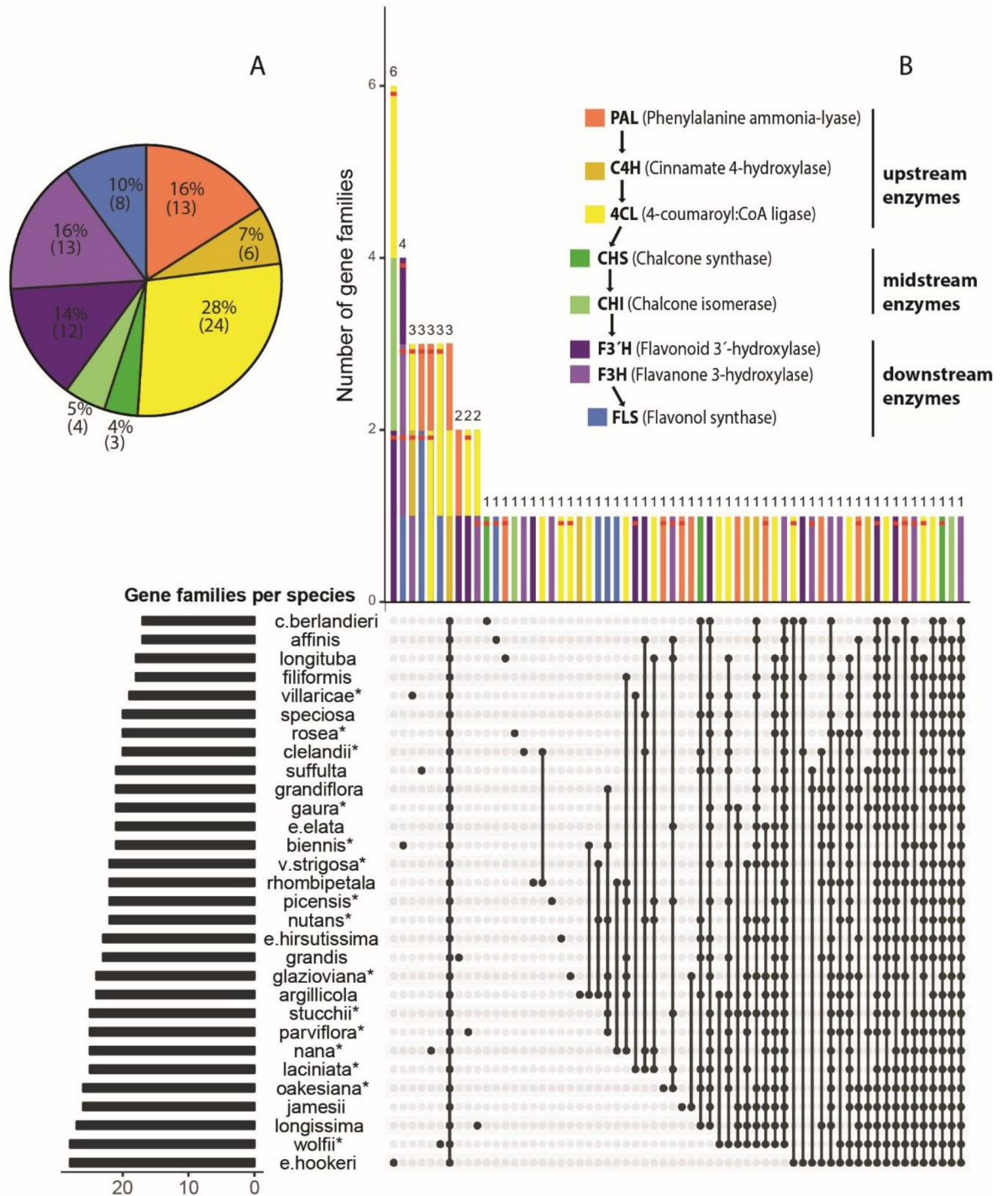
Fig 2. Distribution of phenolic-related genes in 30 *Oenothera* taxa. (A) The relative proportion of 83 gene families related to major enzymes involved in the synthesis of phenolic compounds. (B) shows intersections of major phenolic up- mid- and downstream enzyme related-genes from *Oenothera* transcriptomes. In the upper panel each color bar corresponds to a single or a set of gene families matching to specific pattern of species intersection. Red-striped bars indicate 33 rapidly evolved gene families based on the CAFE analysis. Filled circles in the bottom panel indicate the presence of phenolic-related genes per taxa. Connected circles indicate shared gene families among taxa. Lower left panel indicate the number of gene families corresponding to each filled-circle pattern.

https://doi.org/10.1371/journal.pone.0269307.g002

Table 2. Evolution of phenolic gene families across 30 *Oenothera* taxa as determined by CAFE analysis of seven enzymes; PAL (phenylalanine ammonia-lyase), 4CL (4-coumaroyl: CoA ligase), C4H (Cinnamate 4-hydroxylase), CHS (Chalcone synthase), F3´H (Flavonoid 3´-hydroxylase), F3H (Flavanone 3-hydroxylase) and FLS (Flavonol synthase). We show the total number of gene families that are rapidly evolving (rapid evolution; *p*-value threshold = 0.01), the number of gene families and genes that have experienced expansions contractions. The percentage per each of the seven phenolic enzymes is shown in parentheses.

Enzyme class	RAPIDLY EVOLVING	EXPANSIONS		CONTRACTIONS	
	Gene family	Gene family	Gene gain	Gene family	Gene loss
PAL	8 (24)	7 (22)	20 (16)	3 (19)	4 (10)
4CL	9 (28)	9 (28)	34 (26)	8 (50)	23 (58)
C4H	1 (3)	1 (3)	1 (1)	1 (6)	1 (2)
CHS	2 (6)	2 (6)	6 (5)	0	0
F3´H	5 (15)	5 (16)	23 (18)	2 (13)	9 (23)
F3H	6 (18)	6 (18)	27 (21)	1 (6)	2 (5)
FLS	2 (6)	2 (7)	17 (13)	1 (6)	1 (2)

evolving families, with one and two, respectively. CHI did not show significant evolution when measured as the rate of gene family expansions or contractions.

It is interesting to note that most species-specific gene families experienced rapid evolution. Of the 21 species that showed rapid evolution in species-specific gene families, six species involved ≥ 2 families. *Oenothera biennis* is the only species that showed rapid evolution in almost all its species-specific gene families (three out of four), which are involved in the synthesis of flavonoids (F3´H and F3H). Chalcone synthesis showed contrasting patterns: whereas CHI families showed no rapid evolution, CHS was the only enzyme with a significantly evolved family present in almost all species (28). Only one phenolic family, the last major branching enzyme FLS, showed rapid evolution (Fig 2B).

We also examined gene gain and loss resulting from expansions and contractions of phenolic families. Gene family expansion involved almost all phenolic families (32 out 33), whereas gene family contraction involved about half of families (16). During *Oenothera* diversification, phenolic families gained 128 genes of which 28% corresponded to 4CL. This enzyme also comprised the largest gene loss (58%) within the genus. Branching enzymes F3H and F3´H comprised the second highest number of gained genes with 27 and 18 genes, respectively. Upstream C4H and midstream CHS showed the least gene gain, with one and five genes, respectively. Branching enzymes F3H and FLS, along with C4H loss the least number of genes (≤ 1). CHS families are the only ones exhibiting gene gain and no losses.

Phenolic evolution and the genetic system

We analyzed the distribution of phenolic-related genes between 15 sexual and 15 PTH taxa. Of the 83 gene families with phenolic enzyme annotations, almost half are shared between sexual and PTH plants, while ~25% are specific to sexual and the other 25% to PTH taxa (Fig 3A). On the other hand, of the 33 rapidly evolving phenolic families, about 33% are shared between sexual and PTH species, and 66% are specific to either sexual or PTH species (Fig 3B). We successfully identified 1,568 proteins with annotations of major phenolic enzymes from the total phenolic family count, of which 51% correspond to sexual taxa and the rest to PTH taxa (see S6 Table for species count of phenolic proteins). The relative proportion of enzymes between sexual and PTH plants is highly consistent. 4CL accounted for most of the proteins (27–32%), followed by PAL (~ 18%), F3´H and F3H (~ 17%), CHS (< 15%) C4H (~ 6%) CHI (~ 2%) and FLS (1%) (Fig 3C). Rapidly evolving families comprised 535 proteins, from which 54% correspond to sexual plants and the rest to PTH. Although the amount of rapidly evolving proteins was mostly consistent between sexual and PTH taxa, proteins of specific phenolic enzymes
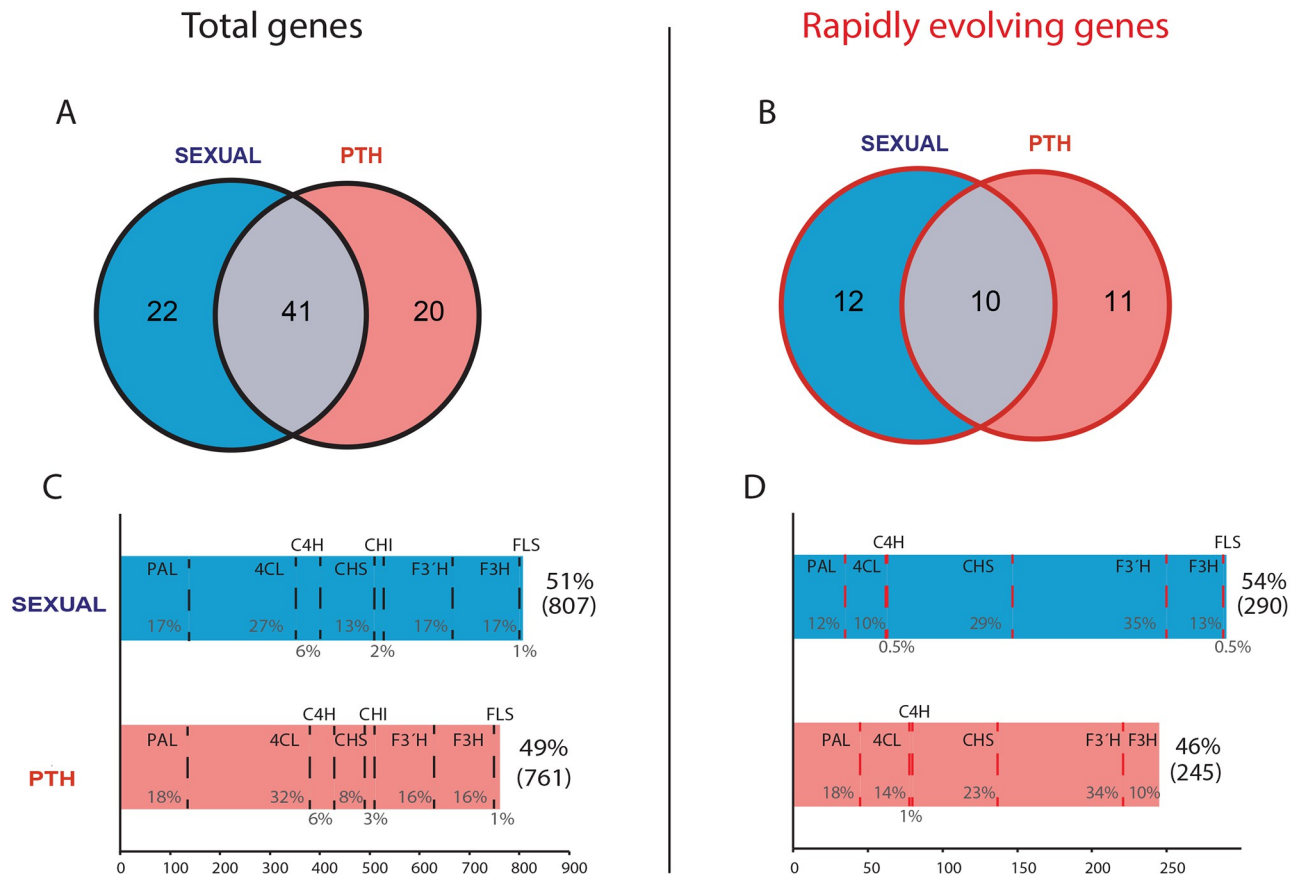
Fig 3. Distribution of phenolic-related genes of 15 sexual and 15 PTH *Oenothera* taxa. Venn diagrams show the intersection between sexual and PTH plants including (A) 83 phenolic-related gene families and (B) 33 rapidly evolving families based on CAFE analysis. Bar charts depict the summary of transcripts related to phenolic enzymes stemming from (C) total gene family count and (D) rapidly evolving gene families.

differed in their rate of evolution (expansion/contraction) between the two reproductive types. Whereas 4CL proteins were the most widely distributed in the total gene families, when non-rapidly evolving families were excluded, CHS and F3´H became the dominant enzymes, accounting for more than a half of the phenolic proteins in both sexual and PTH taxa. It is important to note that CHS is the enzyme that involves the least number of phenolic gene families (3), but at the same time comprised between 29% and 23% of all proteins that evolve significantly, in sexual and PTH taxa, respectively. Although the overall proportion of phenolic-related proteins was higher in sexual plants, PTH taxa showed a higher number of genes related to upstream enzymes involving 429 and 80 proteins in both the total count and rapidly evolving gene families, respectively (Fig 3C and 3D).

Discussion

Our study analyzes genomic evolution in the genus *Oenothera*, and address gene family diversification with special focus on phenolic metabolism. Two results from our analysis are particularly important for answering our research questions. First, we found a large number of gene families exhibiting rapid evolution, yet there was large heterogeneity in gene family evolution across the genus, with section *Oenothera* exhibiting most of the largest expansions and contractions, and significant evolutionary changes. Second, when we focused on molecular

evolution of phenolic enzymes, which are involved in secondary metabolism including defense, we observed that ca. 40% of the phenolic gene families exhibited rapid evolution during diversification, involving genes encoding for up-, mid-, and downstream enzymes. We discuss the importance of these results for the macroevolution of *Oenothera*, the evolution of phenolic metabolism and plant defense.

Transcriptomic resource

We described 63 *de novo* assembled leaf transcriptomes from 29 *Oenothera* species, encompassing the natural variation in the genetic system (PTH vs sexual), geographic distribution, reproductive and defensive chemical traits, among others, across seven subclades. Our new assemblies increased the number of annotated genes by 20% compared to a previous assessment [49], providing a wide view of functional diversity in protein-coding genes. BUSCO gene sets indicate fairly complete transcriptomes given that a single tissue was sampled, yet these varied over 2-fold among taxa. The use of the PLAZA comparative resource showed higher values of relative completeness and less variation among species. Altogether these scores indicate the quality of RNA samples, sequencing, assembly, and gene prediction We believe this transcriptomic resource will be useful in addressing many problems in plant biology, including the systematics of *Oenothera*, the functional roles and divergence of proteins, the role of gene family diversification in molecular, phenotypic and species diversification, the genomic consequences of sexual reproduction and hybridization, and the evolution of plant defence.

Major patterns of gene family evolution

There was large variation in gene family evolution across *Oenothera*. We found three major patterns: 1) across the genus, gene expansions were more common than contractions; 2) the most rapidly evolving gene families accounted for larger contractions than expansions; and 3) subsection *Oenothera* comprised species with the largest rapid evolution. Leebens-Mack et al. [82] addressed gene family evolution in the 1kp project, across green plants (Viridiplantae), including representative *Oenothera* species and sister clades, but did not detect significant expansion/contractions on major gene families within Rosids. Wide scale examination of gene family evolution at large phylogenetic scales is likely to overlook finer-scale evolutionary processes. Our study is the first in addressing gene family evolution not only in *Oenothera* but also within the Onagraceae, documenting the genomic variability of 32 different taxa (29 species) and the evolutionary dynamics to which protein families have been subject to during the last 1.4 MYA.

The large variation in the number of gene family expansions and contractions that we observed may be influenced by different factors, such as gene duplication, *de novo* gene creation, gene loss, and changes in environmental conditions [124,125]. Studies have shown a key role of gene duplication for the evolution of new genes and gene families [126,127], leading to increased protein yield and neofunctionalization during divergence [128]. In *Oenothera*, hybridization plays an important role in speciation, simultaneously causing the fixation of diverged alleles in the heterozygous state when hybridization is accompanied by functionally asexual PTH reproduction [13,50]. Heterozygous PTH genotypes involve heterogeneous gene duplications (combining divergent alleles of a single locus), which creates new material for long-term genetic innovation, contributing to gene family proliferation. Although in our study the number of taxa exhibiting larger gene family expansions was similar between sexual and PTH species, notably, the functionally asexual species *O. biennis* showed disproportionate gene family contractions accounting for the highest rapid evolution across the genus. On the other hand, sexual *O. elata* ssp. *hookerii* accounted for the largest gene gains. These results

showing that rates of gene family evolution are overlapping between PTH and sexual taxa, even though the mechanisms causing gene gains and losses might be distinct. Nonetheless, gene loss and contractions should be taken with caution due to limitations of transcriptomic analyses.

Oenothera subclades showed contrasting divergence of gene families. Subsections *Gaura*, *Raimannia*, and *Oenothera* exhibited the greatest gene expansions (>3,000 gene families change over time in at least one species per clade), whereas subsection *Calylophus* and section *Hartmannia* exhibited the slowest gene expansion (<200 gene families change over time by species). Subsection *Oenothera* is the largest group included in this study; it also has the most rapid evolution across the genus. This group of 11 species native to North America has expanded its range within the last 400 years to other continents, including Europe [1,32,129,130]. Rapid climate change following deglaciation also led to rapid range expansions and new environmental conditions for these species [45,57]. All these factors may have imposed both strong and varying selection, coupled with prominent neutral evolutionary processes through drift and gene flow, which might have contributed to rapid gene family evolution.

Genomic evolution of plant defense and phenolic metabolism

We assessed the evolution of genes implicated in plant defense, secondary metabolism and more specifically phenolic biosynthesis. Response to pathogens comprised the most annotated proteins, followed by response to wounding, insect herbivores and other invertebrates. A review of pathogenesis ontology showed a large set of GO terms corresponding with our annotation, implicated in response to oomycetes [131]. The regulation of secondary metabolism of defense-related compounds (phenolic compounds and phytohormones) was less extensive, comprising about a third of the GO terms of the regulation of plant defense, in agreement with annotations of secondary metabolism in species of the same biosynthetic pathway [132]. Most annotated proteins with these GO categories are ubiquitous in plants, depicting common functional complexes resulting from species interactions and life history. This set of GO terms provides a solid base to further compare and contrast the molecular underpinnings of plant defense and secondary metabolism.

Phenolic metabolism involved 1,568 common genes arranged into 83 gene families that vary widely across the genus *Oenothera*. The genes encoding for the eight major phenolic enzymes occurred in all clades except Section *Calylophus* (which lacked CHI in our assembly), although they were differentially spread among taxa. In our study, upstream enzymes PAL, 4CL and C4H of the core phenylpropanoid complex comprised most of the phenolic proteins (51%) occurring in *Oenothera*, followed by downstream enzymes F3´H, F3H and FLS implicated in branching synthesis of final compounds, which covered 40% of phenolic families, and by midstream CHS and CHI, which channelize precursors into branches and had the smallest fraction of phenolic families covering about 10% of proteins. Each of these up-, mid-, and downstream enzymatic complexes also show an uneven distribution among species. For instance, species such as *O. filiformis* and *O. jamesii* had the largest PAL- and 4CL-related genes, with 17 and 27 proteins, respectively, but a relatively low count of mid- and downstream proteins. Our results show that orthologous genes of phenolic metabolism are well conserved across *Oenothera* and are consistent with comparative studies documenting the uneven size of multigene phenolic families across angiosperms [133–140]. Differences in the presence and copy number of phenolic genes across species documented in our study, gives clues as to the evolutionary dynamics of the genus and the evolution of phenolic metabolism and their enzymatic complexes.

All taxa experienced rapid evolution of phenolic proteins, involving 33 gene families, with an average of six families per species (Table 2). Overall, rapidly evolving phenolic families experienced greater expansions than contractions, gaining about 2-fold more genes than they lost. Upstream enzymes PAL and 4CL of the core phenylpropanoid complex comprised most of the rapidly evolving gene families, both enzymes contained the largest expansion and the latter exhibited the largest contraction (40% of 4CL genes were lost) during diversification. Our results are consistent with the expectation that upstream enzymes of greatest control over flux in metabolic pathways experience disproportionate evolutionary changes [141–144]. In addition, consistent with expectation of rapid evolution of downstream complexes as subject to reduce selective constraint [145,146], we found that branching enzymes F3′H, F3H and FLS accounted for more than half of genes gained [67], involving about 40% of gene families. Notably, chalcone genes that accounted for the smallest gene family set, were the only ones experiencing expansions and no contractions.

Variation across phenolic complexes of rapidly evolving genes also showed contrasting patterns among *Oenothera* taxa and clades, ranging from 3 up to 10 families per species. One significantly evolved gene family (Orthogroup ID: OG0002469) of CHS metabolism, occurred in almost all taxa (except for *O. biennis* and *O. clelandii*), whereas most rapid evolution focused on species-specific families. *Oenothera biennis* and *O. elata* ssp. *hookeri* of subsection *Oenothera* comprised most of the rapid evolution of species-specific families. In fact, this subsection comprised species with the largest rapid phenolic evolution, with five taxa *O. oakesiana*, *O. elata* ssp. *hirsutissima*, *O. elata* ssp. *hookeri*, *O. jamesii*, *O. wolfii* and *O. longissima* having ≥ 9 rapidly evolving phenolic families. These divergent patterns can be associated with changes in ecological factors such as climate and the biotic environment. Plant phenolics play a key role as defense mechanism against herbivores [147–150]. In *Oenothera*, it has been shown that the level of phenolic-based defenses increased in cold climates at higher latitudes [59], where most of the rapidly evolving species occurred (e.g., *O. oakesiana* and *O. wolfii* grow in regions of mean annual temperature below 11˚C). This suggests that adaptive response to the abiotic and biotic environment may have influenced the proliferation of phenolic families during *Oenothera* diversification. These adaptive responses are likely to influence fine-tuning changes along the biosynthetic machinery. For instance, *O. biennis* which produces a wide range of flavonoids and hydrolysable tannins unique in the plant kingdom and subject to ongoing natural selection [52,53], experienced the largest evolutionary changes of exclusive gene families involved in the synthesis of flavonoids (F3′H and F3H).

Although empirical evidence across *Oenothera* has confirmed classic predictions of decreased levels of defense with reduced sexual reproduction [44] (albeit increased diversity of flavonoid metabolites [48], our results showed less clear differences between functionally asexual PTH and sexual plants. We found that although sexual and PTH plants are well differentiated having about 30% (total) and 50% (rapidly evolving) specific gene families each, both groups had almost the same number of phenolic proteins. However, we found differences in the relative proportion of proteins encoding different enzymes among sexual and PTH plants. For instance, PTH plants had ~20% more rapidly evolving genes of upstream enzymes than sexual plants. In contrast, sexual plants had ~30% more rapidly evolving mid- and downstream genes than asexual plants. This suggests that although PTH and sexual plants share a common gene catalog, the mating system may contribute to molecular changes in specific phenolic complexes.

Conclusion

The transcriptomic resource developed here provides a rich and powerful tool for the comparative study of plant biology, especially as it relates to systematics, the study of secondary

metabolism, plant sexual reproduction and genomic evolution more generally. We analyzed the molecular evolution of secondary metabolism and plant defense. By identifying orthologous genes across taxa combined with phylogenetic and gene family evolution analysis, we revealed a variable number of molecular changes resulting in genomic expansions and contractions o that have also shaped biosynthetic phenolic genes. Integration of pathway-level genomic, transcriptomic data and phylogenetic approaches provides a clearer framework to illustrate differences among species. The molecular resource we present are intended to provide a further step to build a predictive gene framework for understanding the evolutionary forces driving population, community and species-level diversity in the historically significant model organism *Oenothera*.

Supporting information

S1 Fig. Schematic of the general phenylpropanoid, isoflavonoid, flavonoid pathways.
(DOCX)

S2 Fig. Maximum likelihood species trees inferred from 1,017 orthogroups consisting entirely of single-copy genes from 30 *Oenothera* taxa with *O. capillifolia* spp *berlandieri* as outgroup.
(DOCX)

S3 Fig. Functional diversity across 30 *Oenothera* taxa, according to gene ontology (GO) nomenclature.
(DOCX)

S1 Table. Taxonomic, sexual system and collection data from 63 *Oenothera* samples used for transcriptomes assembly.
(XLSX)

S2 Table. Overview of the RNA-seq assembly of 63 *Oenothera* transcriptomes and summarized of BUSCO and PLAZA scores.
(XLSX)

S3 Table. Annotation stats of blasted annotated genes based on the longest isoform from 63 *Oenothera* individuals.
(XLSX)

S4 Table. OrthoFinder statistics for orthogroup construction of 30 *Oenothera* taxa.
(DOCX)

S5 Table. Gene count for each gene ontology category and associated orthogroups identified across 30 *Oenothera* transcriptomes.
(XLSX)

S6 Table. Count of phenolic proteins in 30 *Oenothera* taxa.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Eunice Kariñho-Betancourt, Marc T. J. Johnson.

Data curation: Eunice Kariñho-Betancourt.

Formal analysis: Eunice Kariñho-Betancourt, David Carlson.

Funding acquisition: Marc T. J. Johnson.

Investigation: Eunice Kariñho-Betancourt, Marc T. J. Johnson.

Methodology: Eunice Kariñho-Betancourt, Jessie Hollister, Stephan Greiner.

Supervision: Stephan Greiner, Marc T. J. Johnson.

Writing – original draft: Eunice Kariñho-Betancourt.

Writing – review & editing: David Carlson, Jessie Hollister, Axel Fischer, Stephan Greiner, Marc T. J. Johnson.

References

1. Harte C. *Oenothera*: *Contributions of a Plant to Biology*. Springer-Verlag; 1994.

2. De Vries H. Sur la fécondation hybride de l'albumen. Comptes rendus l'Académie des Sci. 1899; 129: 973–975.

3. De Vries H. Die Mutationstheorie. 1901; 752.

4. Campos L, Schwerin A. Making Mutations: Objects, Practices, Contexts. Workshop at the Max Planck Institute for the History of Science, Berlin. (Max Planck Institute for the History of Science; 2010.

5. Nei M, Nozawa M. Roles of mutation and selection in speciation: From Hugo de Vries to the modern genomic era. Genome Biol Evol. 2011; 3: 812–829. https://doi.org/10.1093/gbe/evr028 PMID: 21903731

6. Nei M. Mutation-Driven Evolution. Oxford University Press; 2013.

7. Darlington CD. The evolution of genetic systems. Cambridge University Press; 1939.

8. Darlington CD. The Evolution of Genetic Systems: Contributions of Cytology to Evolutionary Theory. In: Mayr E, Provine W, editors. Evolutionary Synthesis. Perspectives on the Unification of Biology. Cambridge Uiversity Press, 1980. pp 70–80.

9. Emerson S. The genetics of self-incompatibility in *Oenothera rhombipetala*. Genetica. 1937; 36: 190–202.

10. Stubbe W, Raven PH. A genetic contribution to the taxonomy of *Oenothera* sect. Oenothera (including subsections Euoenothera, Emersonia, Raimannia and Munzia). Plant Syst Evol. 1979; 133: 39–59.

11. Cleland RE. Chromosome structure in *Oenothera* and its effect on the evolution of the genus. Cytologia. 1957; 22: 5–19.

12. Cleland RE. The cytogenetics of *Oenothera*. In: Caspri EW, Thoday JM, editors. Advances in Genetics. Academic Press; 1963. pp 147–237.

13. Cleland RE. Oenothera cytogenetics and evolution. Academic Press; 1972.

14. Holsinger KE Ellstrand NC. The evolution and ecology of permanent translocation heterozygotes. Am J. 1984; 124: 48–71.

15. Cleland RE. Plastid behaviour of the North American Euoenotheras. Planta. 1962; 57: 699–712.

16. Stubbe W. The rôle of the plastome in evolution of the genus *Oenothera*. Genetica. 1964; 35: 28–33.

17. Greiner S, Rauwolf UWE, Meurer J, Herrmann RG. The role of plastids in plant speciation. Mol. Ecol. 2011; 20: 671–691. https://doi.org/10.1111/j.1365-294X.2010.04984.x PMID: 21214654

18. Schwemmle J, Haustein E, Sturm A, Binder M. Genetische und zytologische Untersuchungen an Eu-Oenotheren: Teil I bis VI. Z Indukt Abstammungs- Vererbungsl, 1938; 358–800.

19. Kirk JTO, Tilney-Bassett RAE. The plastids. Their Chemistry, Structure, Growth and Inheritance. Amsterdam, New York, Oxford: Elsevier; 1978.

20. Grun P. Cytoplasmic Genetics and Evolution. New York: Columbia University Press; 1976.

21. Levin DA. Genic heterozygosity and protein polymorphism among local populations of *Oenothera biennis*. Genetics. 1975; 79: 477–491. https://doi.org/10.1093/genetics/79.3.477 PMID: 17248679

22. Levin DA, Howland GP, Steiner E. Protein polymorphism and genic heterozygosity in a population of the permanent translocation heterozygote, *Oenothera biennis*. Proc Natl Acad Sci U.S.A. 1972; 69: 1475–1477. https://doi.org/10.1073/pnas.69.6.1475 PMID: 4504363

23. Lewis H, Szweykowski J. The genus Gayophytum (Onagraceae). Brittonia. 1964; 16: 343–391.

24. Wagner WL. Systematics of *Oenothera Sections Contortae*, *Eremia*, and *Ravenia (Onagraceae)*. Syst Bot. 2005; 30: 332–356.

25. Wagner WL, Hoch PC, Raven PH. Revised classification of the Onagraceae. Syst Bot Monogr. 2007; 83: 1–222.

26. Wagner WL, Krakos KN, Hoch PC, Taxonomic changes in *Oenothera* sections Gaura and Calylophus (onagraceae). PhytoKeys. 2013; 28: 61–72.

27. Munz PA. Onagraceae. Flora of North America. II; 1965. pp. 1–278.

28. Raven PH. 1964. The generic subdivision of Onagraceae, tribe Onagreae. *Brittonia* 16, 276–288.

29. Raven PH, Gregory DP. A revision of the genus *Gaura* (Onagraceae). Memoirs of the Torrey Botanical Club; 1972. pp. 1–96.

30. Raven PH, Dietrich W, Stubbe W. An outline of the systematics of *Oenothera* subsect. Euoenothera (Onagraceae). Syst Bot. 1979; 4: 242–252.

31. Dietrich W, Wagner WL. Systematics of *Oenothera* section *Oenothera* subsection Raimannia and subsection Nutantigemma (Onagraceae). Syst Bot Monogr. 1988; 24: 1–91.

32. Dietrich W, Wagner WL, Raven PH. Systematics of *Oenothera* section *Oenothera* subsection *Oenothera* (Onagraceae). Syst Bot Monogr. 1997; 50: 12–34.

33. Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, et al. Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. Am J Bot. 2003; 90: 107–115. https://doi.org/10.3732/ajb.90.1.107 PMID: 21659085

34. Levin RA, Wagner WL, Hoch PC, Hahn WJ, Rodriguez A, Baum DA, et al. Paraphyly in tribe Onagreae: insights into phylogenetic relationships of Onagraceae based on nuclear and chloroplast sequence data. Syst Bot. 2004; 29: 147–164.

35. Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, et al. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera plastomes*. Mol Genet Genom. 2000; 263: 581–585. https://doi.org/10.1007/pl00008686 PMID: 10852478

36. Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. Nucleic Acids Res. 2008; 36: 2366–2378. https://doi.org/10.1093/nar/gkn081 PMID: 18299283

37. Greiner S, Wang X, Herrmann G, Rauwolf U, Mayer K, Haberer G, et al. The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. Mol Biol Evol. 2008; 25: 2019–2030. https://doi.org/10.1093/molbev/msn149 PMID: 18614526

38. Larson EL, Bogdanowicz SM, Agrawal AA, Johnson MTJ, Harrison RG. Isolation and characterization of polymorphic microsatellite loci in common evening primrose (*Oenothera biennis*). Mol Ecol Resour. 2008; 8: 434–436. https://doi.org/10.1111/j.1471-8286.2007.01983.x PMID: 21585813

39. Rauwolf U, Golczyk H, Meurer J, Herrmann RG, Greiner S. Molecular marker systems for oenothera genetics. Genetics. 2008; 180: 1289–1306. https://doi.org/10.1534/genetics.108.091249 PMID: 18791241

40. Rauwolf U, Greiner S, Mráček J, Rauwolf M., Golczyk H, Mohler V, et al. Uncoupling of sexual reproduction from homologous recombination in homozygous *Oenothera* species. Heredity. 2011; 107: 87–94. https://doi.org/10.1038/hdy.2010.171 PMID: 21448231

41. Mráček J, Greiner S, Cho WK, et al. Construction, database integration, and application of an *Oenothera* EST library. Genomics. 2006; 88: 272–280.

42. Golczyk H, Musiał K, Rauwolf U, Meurer J, Herrmann RG, Greiner S. Meiotic events in *Oenothera*—a non-standard pattern of chromosome behavior. Genome. 2008; 51: 952–958. https://doi.org/10.1139/G08-081 PMID: 18956028

43. Golczyk H, Massouh A, Greiner S. Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. Plant Cell. 2014; 26: 1280–1293. https://doi.org/10.1105/tpc.114.122655 PMID: 24681616

44. Johnson MTJ, Smith SD, Rausher MD. Plant sex and the evolution of plant defenses against herbivores. Proc Natl Acad Sci USA. 2009; 106: 18079–18084. https://doi.org/10.1073/pnas.0904695106 PMID: 19617572

45. Johnson MTJ, Smith SD, Rausher MD. Effects of plant sex on range distributions and allocation to reproduction. New Phytol. 2010; 186: 769–779 https://doi.org/10.1111/j.1469-8137.2010.03201.x PMID: 20180909

46. Johnson MTJ. The contribution of evening primrose (*Oenothera biennis*) to a modern synthesis of evolutionary ecology. Popul. Ecol. 2011; 53: 9–21.

47. Hersch-Green EI. Myburg H, Johnson MTJ. Adaptive molecular evolution of a defense gene in sexual but not functionally asexual evening primroses. J Evol Biol. 2012; 25: 1576–1586. https://doi.org/10.1111/j.1420-9101.2012.02542.x PMID: 22587337

48. Godfrey RM, Johnson MTJ. Effects of functionally asexual reproduction on quantitative genetic variation in the evening primroses (*Oenothera*, Onagraceae). Am J Bot. 2014; 101: 1906–1914. https://doi.org/10.3732/ajb.1400226 PMID: 25366856

49. Hollister JD, Greiner S, Wang W, Wang J, Zhang Y, Wong GKS, et al. Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. Mol Biol Evol. 2015; 32: 896–905. https://doi.org/10.1093/molbev/msu345 PMID: 25534028

50. Hollister JD, Greiner S, Johnson MTJ, Wright SI. Hybridization and a loss of sex shape genome-wide diversity and the origin of species in the evening primroses (*Oenothera*, Onagraceae). New Phytol. 2019; 224: 1372–1380. https://doi.org/10.1111/nph.16053 PMID: 31309571

51. Maron JL, Johnson MTJ, Hastings AP, Agrawal AA. Fitness consequences of occasional outcrossing in a functionally asexual plant (*Oenothera biennis*). Ecology. 2018; 99: 464–473. https://doi.org/10.1002/ecy.2099 PMID: 29205317

52. Johnson MTJ, Agrawal AA, Maron JL, Salminen JP. Heritability, covariation and natural selection on 24 traits of common evening primrose (*Oenothera biennis*) from a field experiment. J Evol Biol. 2009; 22: 1295–1307. https://doi.org/10.1111/j.1420-9101.2009.01747.x PMID: 19490388

53. Agrawal AA, Hastings AP, Johnson MTJ, Maron JL, Salminen JP. Insect herbivores drive real-time ecological and evolutionary change in plant populations. Science. 2012; 338: 113–116. https://doi.org/10.1126/science.1225977 PMID: 23042894

54. Anstett DN, Naujokaitis-Lewis I, Johnson MTJ. Latitudinal gradients in herbivory on *Oenothera biennis* vary according to herbivore guild and specialization. Ecology. 2014; 95: 2915–2923.

55. Anstett DN, Ahern JR, Glinos J, Nawar N, Salminen JP, Johnson MTJ. Can genetically based clines in plant defense explain greater herbivory at higher latitudes? Ecol Lett. 2015; 18: 1376–1386. https://doi.org/10.1111/ele.12532 PMID: 26482702

56. Johnson MTJ, Ives AR, Ahern J, Salminen JP. Macroevolution of plant defenses against herbivores in the evening primroses. New Phytol. 2014; 203: 267–279. https://doi.org/10.1111/nph.12763 PMID: 24634986

57. Anstett DN, Chen W, Johnson MTJ. Latitudinal gradients in induced and constitutive resistance against herbivores. J Chem Ecol. 2016; 42: 772–781. https://doi.org/10.1007/s10886-016-0735-6 PMID: 27501815

58. Puentes A, Johnson MTJ. Tolerance to deer herbivory and resistance to insect herbivores in the common evening primrose (*Oenothera biennis*). J Evol Biol. 2016; 29: 86–97. https://doi.org/10.1111/jeb.12764 PMID: 26395768

59. Anstett DN, Ahern JR, Johnson MTJ, Salminen JP. Testing for latitudinal gradients in defense at the macroevolutionary scale. Evolution. 2018; 72: 2129–2143. https://doi.org/10.1111/evo.13579 PMID: 30101976

60. Johnson MTJ, Agrawal AA. Plant genotype and environment interact to shape a diverse arthropod community on evening primrose (*Oenothera biennis*). Ecology. 2005; 86: 874–885.

61. Johnson MTJ, Agrawal AA. Covariation and composition of arthropod species across plant genotypes of evening primrose, Oenothera biennis. Oikos. 116, 941–956 (2007).

62. Johnson MTJ, Lajeunesse MJ, Agrawal AA. Additive and interactive effects of plant genotypic diversity on arthropod communities and plant fitness. Ecol Lett. 2006; 9: 24–34. https://doi.org/10.1111/j.1461-0248.2005.00833.x PMID: 16958865

63. Johnson MTJ. Bottom-up effects of plant genotype on aphids, ants, and predators. Ecology. 2008; 89: 145–154. https://doi.org/10.1890/07-0395.1 PMID: 18376556

64. Johnson MTJ, Dinnage R, Zhou AY, Hunter MD. Environmental variation has stronger effects than plant genotype on competition among plant species. J Ecol. 2008; 96: 947–955.

65. Johnson MTJ, Vellend M, Stinchcombe JR. Evolution in plant populations as a driver of ecological changes in arthropod communities. Philos Trans R Soc B. 2009; 364: 1593–1605. https://doi.org/10.1098/rstb.2008.0334 PMID: 19414473

66. Carmona D, Johnson MTJ. The genetics of chutes and ladders: a community genetics approach to tri-trophic interactions. Oikos. 2016; 125: 1657–1667.

67. Fitzpatrick CR, Mikhailitchenko AV, Anstett DN, Johnson MTJ. The influence of range-wide plant genetic variation on soil invertebrate communities. Ecography. 2018; 41: 1135–1146.

68. Fitzpatrick CR, Keller SR. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. Ecol Lett. 2015; 18: 1–16. https://doi.org/10.1111/ele.12376 PMID: 25270536

69. Greiner S, Bock R. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays. 2013; 35: 354–365. https://doi.org/10.1002/bies.201200137 PMID: 23361615

70. Zupok A, Kozul D, Schöttler MA, Niehörster J, Garbsch F, Liere K, et al. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. Plant Cell. 2021; 33: 2583–2601. https://doi.org/10.1093/plcell/koab155 PMID: 34048579

71. Massouh A, Schubert J, Yaneva-Roder L, Ulbricht-Jones ES, Zupok A, Johnson MTJ, et al. Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. Plant Cell. 2016; 28: 911–929. https://doi.org/10.1105/tpc.15.00879 PMID: 27053421

72. Malinova I, Zupok A, Massouh A, Schöttler MA, Meyer EH, Yaneva-Roder L, et al. 2021. Correction of frameshift mutations in the atpB gene by translational recoding in chloroplasts of *Oenothera* and tobacco. Plant Cell; 33: 1682–1705. https://doi.org/10.1093/plcell/koab050 PMID: 33561268

73. Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, et al. Chloroplast competition is controlled by lipid biosynthesis in evening primroses. Proc Natl Acad Sci USA. 2019; 116: 5665–5674. https://doi.org/10.1073/pnas.1811661116 PMID: 30833407

74. Karonen M, Parker J, Agrawal A, Salminen JP. First evidence of hexameric and heptameric ellagitannins in plants detected by liquid chromatography/electrospray ionisation mass spectrometry. Rapid Commun Mass Spectrom. 2010; 24: 3151–3156. https://doi.org/10.1002/rcm.4756 PMID: 20941762

75. Parker JD, Salminen JP, Agrawal AA. Evolutionary potential of root chemical defense: genetic correlations with shoot chemistry and plant growth. J Chem Ecol. 2012; 38: 992–995. https://doi.org/10.1007/s10886-012-0163-1 PMID: 22790783

76. Lemoine NP, Doublet D, Salminen JP, Burkepile DE, Parker JD. Responses of plant phenology, growth, defense, and reproduction to interactive effects of warming and insect herbivory. Ecology. 2017; 98: 1817–1828. https://doi.org/10.1002/ecy.1855 PMID: 28403543

77. Anstett DN, Cheval I, D'Souza C, Salminen JP, Johnson MTJ. Ellagitannins from the Onagraceae decrease the performance of generalist and specialist herbivores. J Chem Ecol. 2019; 45: 86–94. https://doi.org/10.1007/s10886-018-1038-x PMID: 30511298

78. Hoggard GD, Kores PJ, Molvray M, Hoggard RKO. The phylogeny of Gaura (Onagraceae) based on ITS, ETS and TrnL-F sequence data. Am J Bot. 2004; 91: 139–148. https://doi.org/10.3732/ajb.91.1.139 PMID: 21653370

79. Krakos KN, Reece JS, Raven PH. Molecular phylogenetics and reproductive biology of *Oenothera* section Kneiffia (Onagraceae). Syst Bot. 2014; 39: 523–532.

80. Johnson MTJ, Carpenter EJ, Tian Z, et al. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS ONE. 2012; 7: e50266. https://doi.org/10.1371/journal.pone.0050226 PMID: 23185583

81. Carpenter EJ, Matasci N, Ayyampalayam S, et al. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience*. 2019; 8: giz126. https://doi.org/10.1093/gigascience/giz126 PMID: 31644802

82. Leebens-Mack JH, Barker MS, Carpenter EJ, et al. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019; 574: 679–685. https://doi.org/10.1038/s41586-019-1693-2 PMID: 31645766

83. Straley GB. Systematics of Oenothera Sect. Kneiffia (Onagraceae). Annals of the Missouri Botanical Garden 64, no. 3. Missouri Botanical Garden Press. 1977; 74: 381–424.

84. Clinebell RR, Crowe A, Gregory DP, Hoch PC. Pollination ecology of Gaura and Calylophus (Onagraceae, tribe Onagreae) in western Texas, USA. Ann Missouri Bot Gard. 2004; 91: 369–400.

85. von Arx M, Goyret J, Davidowitz G, Raguso RA. Floral humidity as a reliable sensory cue for profitability assessment by nectar-foraging hawkmoths. Proc Natl Acad Sci USA. 2012; 109: 9471–9476. https://doi.org/10.1073/pnas.1121624109 PMID: 22645365

86. Krakos KN, Fabricant SA. Generalized versus specialized pollination systems in *Oenothera* (Onagraceae). J Pollinat Ecol. 2014; 14: 235–243.

87. Ellstrand NC, Levin DA. Recombination system and population structure in *Oenothera*. Evolution. 1980; 34: 923–933. https://doi.org/10.1111/j.1558-5646.1980.tb04030.x PMID: 28581144

88. Johnson MTJ, Fitzjohn RG, Smith SD, Rausher MD, Otto SP. Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. Evolution. 65, 2011; 3230–3240. https://doi.org/10.1111/j.1558-5646.2011.01378.x PMID: 22023588

89. Winkel-Shirley B. Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. Physiol. Plant. 1999; 107: 142–149.

90. Kalinowska M, Bielawska A, Lewandowska-Siwkiewicz H, Priebe W, Lewandowski W. Apples: Content of phenolic compounds vs. variety, part of apple and cultivation model, extraction of phenolic compounds, biological properties. Plant Physiol. Biochem. 2014; 84, 169–188. https://doi.org/10.1016/j.plaphy.2014.09.006 PMID: 25282014

91. Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MS, Wang L. The phenylpropanoid pathway and plant defence—a genomics perspective. Mol. Plant Pathol. 2002; 3: 371–390. https://doi.org/10.1046/j.1364-3703.2002.00131.x PMID: 20569344

92. Falcone-Ferreyra ML, Rius S, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci. 2012; 3: 222. https://doi.org/10.3389/fpls.2012.00222 PMID: 23060891

93. Hartley SE, Firn RD. Phenolic biosynthesis, leaf damage, and insect herbivory in birch (*Betula pendula*). J Chem Ecol. 1989; 15: 275–283. https://doi.org/10.1007/BF02027789 PMID: 24271442

94. Schaller A. Induced plant resistance to herbivory. Berlin Germany, Springer; 2008.

95. Lone R, Shuab R, Kamili AN. Plant Phenolics in Sustainable Agriculture. Springer Nature; 2020.

96. Greiner S, Köhl K. Growing evening primroses (*Oenothera*). Front Plant Sci. 2014; 5: 1–12.

97. Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34: i884–i890. https://doi.org/10.1093/bioinformatics/bty560 PMID: 30423086

98. Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8: 1494–1512. https://doi.org/10.1038/nprot.2013.084 PMID: 23845962

99. Grabherr MG, Haas BJ, Yassour M, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011; 29: 644–652.

100. Bryant DM, Johnson K, DiTommaso T, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 2017; 18: 762–776. https://doi.org/10.1016/j.celrep.2016.12.063 PMID: 28099853

101. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

102. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003; 31: 365–370. https://doi.org/10.1093/nar/gkg095 PMID: 12520024

103. Bateman A, Coin L, Durbin R, et al. The Pfam protein families database. Nucleic Acids Res. 2004; 32: D138–D141. https://doi.org/10.1093/nar/gkh121 PMID: 14681378

104. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8: 785–786. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

105. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001; 305: 567–580. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

106. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007; 35: 3100–3108. https://doi.org/10.1093/nar/gkm160 PMID: 17452365

107. Klopfenstein DV, Zhang L, Pedersen BS, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018; 8: 1–17. https://doi.org/10.1038/s41598-018-28948-z PMID: 30022098

108. Simão FA, Waterhouse RM, Loannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologos. Bioinformatics 2015; 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717

109. Van Bel M, Diels T, Vancaester E, et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. 2018; 46: D1190–D1196. https://doi.org/10.1093/nar/gkx1002 PMID: 29069403

110. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16: 157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

111. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20: 1–14. https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

112. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12: 59–60. https://doi.org/10.1038/nmeth.3176 PMID: 25402007

113. Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005; 33: 511–518. https://doi.org/10.1093/nar/gki198 PMID: 15661851

114. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

115. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–274. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

116. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020; 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015 PMID: 32011700

117. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–589. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

118. Mirarab S, Bayzid SM, Warnow T. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science. 2014; 364: 1250463. https://doi.org/10.1126/science.1250463 PMID: 25504728

119. Yin J, Zhang C, Mirarab S. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. 2019; 35: 3961–3969.

120. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018; 35: 518–522. https://doi.org/10.1093/molbev/msx281 PMID: 29077904

121. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006; 22: 1269–1271. https://doi.org/10.1093/bioinformatics/btl097 PMID: 16543274

122. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 2002; 19: 101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974 PMID: 11752195

123. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013; 30: 1987–1997. https://doi.org/10.1093/molbev/mst100 PMID: 23709260

124. Guo YL. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. Plant J. 2013; 73: 941–951. https://doi.org/10.1111/tpj.12089 PMID: 23216999

125. McLysaght A, Hurst LD. Open questions in the study of de novo genes: what, how and why. Nat Rev Genet. 2016; 17: 567–579. https://doi.org/10.1038/nrg.2016.78 PMID: 27452112

126. Milesi P, Weill M, Lenormand T, Labbé P. Heterogeneous gene duplications can be adaptive because they permanently associate overdominant alleles. Evol Lett. 2017; 1: 169–180. https://doi.org/10.1002/evl3.17 PMID: 30283647

127. Milesi P, Assogba BS, Atyame CM, Pocquet N, Berthomieu A, Unal S, et al. The evolutionary fate of heterogeneous gene duplications: a precarious overdominant equilibrium between environment, sublethality and complementation. Mol. Ecol. 2018; 27: 493–507. https://doi.org/10.1111/mec.14463 PMID: 29230902

128. Lenormand T, Guillemaud T, Bourguet D, Raymond M. Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. Evolution. 1998; 52: 1705–1712. https://doi.org/10.1111/j.1558-5646.1998.tb02250.x PMID: 28565319

129. Rostański K, Rostański A, Gerold-Śmietańska I, Wąsowicz P. Evening-primroses (*Oenothera*) occuring in Europe. W. Szafer Institute of Botany, PAS, Katowice-Kraków. 2010.

130. Woźniak-Chodacka M. A revision of taxonomic relation between *Oenothera royfraseri* and *O. turoviensis* (sect. *Oenothera*, subsect. *Oenothera*; Onagraceae) based on multivariate analyses of morphological characters. Phytotaxa. 2020; 435: 164–180.

131. Meng S, Torto-Alalibo T, Chibucos MC, Tyler BM, Dean RA. Common processes in pathogenesis by fungal and oomycete plant pathogens, described with Gene Ontology terms. BMC Microbiol. 2009; 9: 1–11. https://doi.org/10.1186/1471-2180-9-S1-S7 PMID: 19278555

132. Kariñho-Betancourt E, Hernández-Soto P, Rendón-Anaya M, Calderón-Cortés N, Oyama K. Differential expression of genes associated with phenolic compounds in galls of *Quercus castanea* induced by *Amphibolips michoacaensis*. J Plant Interact. 2019; 14: 177–186.

133. Butland SL, Chow ML, Ellis BE. A diverse family of phenylalanine ammonia-lyase genes expressed in pine trees and cell cultures. Plant Mol Biol. 1998; 37: 15–24. https://doi.org/10.1023/a:1005941228246 PMID: 9620261

134. Wei XX, Wang XQ. Evolution of 4-coumarate: coenzyme A ligase (4CL) gene and divergence of *Larix* (Pinaceae). Mol Phylogenet Evol. 2004; 31: 542–553. https://doi.org/10.1016/j.ympev.2003.08.015 PMID: 15062793

135. Tohge T, Nishiyama Y, Hirai MY, et al. Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J. 2005; 42: 218–235. https://doi.org/10.1111/j.1365-313X.2005.02371.x PMID: 15807784

136. Li L, Lu S, Chiang V. A genomic and molecular view of wood formation. Annu Rev Plant Biol. 2006; 25: 215–233.

137. Hyun MW, Yun YH, Kim JY, Kim SH. Fungal and plant phenylalanine ammonia-lyase. Mycobiology. 2011; 39: 257–265. https://doi.org/10.5941/MYCO.2011.39.4.257 PMID: 22783113

138. Ngaki MN, Louie GV, Philippe RN, Manning G, Pojer F, Bowman ME, et al. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature. 2012; 485: 530–533. https://doi.org/10.1038/nature11009 PMID: 22622584

139. Tohge T, Watanabe M, Hoefgen R, Fernie AR. The evolution of phenylpropanoid metabolism in the green lineage. Crit Rev Biochem Mol Biol. 2013; 48: 123–152. https://doi.org/10.3109/10409238.2012.758083 PMID: 23350798

140. Lavhale SG, Kalunke RM, Giri AP. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. Planta. 2018; 248: 1063–1078. https://doi.org/10.1007/s00425-018-2965-z PMID: 30078075

141. Eanes WF. Analysis of selection on enzyme polymorphisms. Annu Rev Ecol Evol Syst. 1999; 30: 301–326.

142. Watt WB, Dean AM. Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. Annu Rev Genet. 2000; 34: 593–622. https://doi.org/10.1146/annurev.genet.34.1.593 PMID: 11092840

143. Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, et al. Adaptive evolution of metabolic pathways in *Drosophila*. Mol Bio. Evol. 2007; 24: 1347–1354. https://doi.org/10.1093/molbev/msm057 PMID: 17379620

144. Wright KM, Rausher MD. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics. 2010; 184: 483–502. https://doi.org/10.1534/genetics.109.110411 PMID: 19966064

145. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 1999; 16: 266–274. https://doi.org/10.1093/oxfordjournals.molbev.a026108 PMID: 10028292

146. Rausher MD, Lu Y, Meyer K. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol. 2008; 67: 137–144. https://doi.org/10.1007/s00239-008-9105-5 PMID: 18654810

147. Levin DA. The origin of reproductive isolating mechanisms in flowering plants. Taxon. 1971; 20: 91–113.

148. Treutter D. Significance of flavonoids in plant resistance: a review. Environ Chem Lett. 2006; 4: 147–157.

149. Agrawal AA, Fishbein M, Halitschke R, Hastings AP, Rabosky DL, Rasmann S. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proc Natl Acad Sci USA. 2009; 106: 18067–18072. https://doi.org/10.1073/pnas.0904862106 PMID: 19805160

150. War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, et al. Mechanisms of plant defense against insect herbivores. Plant Signal Behav. 2012; 7: 1306–1320. https://doi.org/10.4161/psb.21663 PMID: 22895106