

RESEARCH

Open Access



# A phylogeny of the evening primrose family (Onagraceae) using a target enrichment approach with 303 nuclear loci

Rick P. Overson<sup>1</sup>, Matthew G. Johnson<sup>2</sup>, Lindsey L. Bechen<sup>3,4</sup>, Sylvia P. Kinoshian<sup>3,5</sup>, Norman A. Douglas<sup>6</sup>, Jeremie B. Fant<sup>3,7</sup>, Peter C. Hoch<sup>8</sup>, Rachel A. Levin<sup>9</sup>, Michael J. Moore<sup>10</sup>, Robert A. Raguso<sup>11</sup>, Warren L. Wagner<sup>12</sup>, Krissa A. Skogen<sup>3,7,13</sup> and Norman J. Wickett<sup>3,7,13\*</sup>

## Abstract

**Background** The evening primrose family (Onagraceae) includes 664 species (803 taxa) with a center of diversity in the Americas, especially western North America. Ongoing research in Onagraceae includes exploring striking variation in floral morphology, scent composition, and breeding system, as well as the role of these traits in driving diversity among plants and their interacting pollinators and herbivores. However, these efforts are limited by the lack of a comprehensive, well-resolved phylogeny. Previous phylogenetic studies based on a few loci strongly support the monophyly of the family and the sister relationship of the two largest tribes but fail to resolve several key relationships.

**Results** We used a target enrichment approach to reconstruct the phylogeny of Onagraceae using 303 highly conserved, low-copy nuclear loci. We present a phylogeny for Onagraceae with 169 individuals representing 152 taxa sampled across the family, including extensive sampling within the largest tribe, Onagreae. Deep splits within the family are strongly supported, whereas relationships among closely related genera and species are characterized by extensive conflict among individual gene trees.

**Conclusions** This phylogenetic resource will augment current research projects focused throughout the family in genomics, ecology, coevolutionary dynamics, biogeography, and the evolution of characters driving diversification in the family.

**Keywords** Phylogenomics, HybSeq, Onagraceae, Evening primrose, Target enrichment, Plant systematics, Phylogenetics

\*Correspondence:

Norman J. Wickett  
nwickett@clemons.edu

<sup>1</sup> Arizona State University, PO Box 875502, Tempe, AZ 85287, USA

<sup>2</sup> Texas Tech University, Box 43131, Lubbock, TX 79409, USA

<sup>3</sup> Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, 1000 Lake Cook Rd, Glencoe, IL 60022, USA

<sup>4</sup> Whitehead Institute for Biomedical Research, 455 Main St, Cambridge, MA 02142, USA

<sup>5</sup> Department of Ecology and Evolutionary Biology, University of Arizona, PO Box 210088, Tucson, AZ 85721, USA

<sup>6</sup> University of Florida, 220 Bartram Hall, Gainesville, FL 32611, USA

<sup>7</sup> Northwestern University, 2205 Tech Dr, Evanston, IL 60208, USA

<sup>8</sup> Missouri Botanical Garden, 4344 Shaw Blvd, St. Louis, MO 63110, USA

<sup>9</sup> Amherst College, 25 East Dr, Amherst, MA 01002, USA

<sup>10</sup> Oberlin College, 119 Woodland St, Oberlin, OH 44074, USA

<sup>11</sup> Cornell University, 215 Tower Rd, Ithaca, NY 14853, USA

<sup>12</sup> Smithsonian Institution, MRC-166, PO Box 37012, Washington, DC 20013, USA

<sup>13</sup> Clemson University, 132 Long Hall, Clemson, SC 29634, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Background**

The evening primrose family (Onagraceae, Myrtales) comprise 664 species of herbs, shrubs, and trees across 22 genera [1], with almost two-thirds of the

species occurring in tribes Epilobieae (2 genera, 173 spp.; Fig. 1H–J) and Onagreae (13 genera, 265 spp.; Figs. 1L–K and 2A–L). Onagraceae have a cosmopolitan distribution, with the majority of species concentrated in the



**Fig. 1** Onagraceae, subfamily Ludwigioideae. **A** *Ludwigia octovalvis*. Flower and immature capsule. Baldwin Co., Alabama (image Warren Wagner in 2003). Subfamily Onagroideae B–L. **B** Tribe Hauyeeae. *Hauya elegans* subsp. *lucida*. Branch with flower at anthesis and in bud. Cultivated at San Diego Zoo (image Wikipedia in 2008). **C, D** Tribe Circaeae. **C.** *Fuchsia inflata*. Inflorescence. Cusco, Peru (image P. Berry in 1978). **D** *Circaea pacifica*. Inflorescence and young fruit. Benton Co Oregon (image G. Carr-4162b in 2006). **E, F** Tribe Lopezieae. **E** *Lopezia racemosa* subsp. *racemosa*. Flower with inflorescence and leaves. Puebla, Mexico (image Jon Amith in 2007). **F** *Megacorax gracielanus*. Chasmogamous flower with bud and stem with leaves. Sierra de Coneto, Durango, Mexico (image M. Socorro Gonzalez-Elizondo in 2021). **G** Tribe Gongylocarpeae. *Gongylocarpus fruticosus*. Stem with flowers and gall-like mature fruit embedded in pith of stem, indehiscent, and inset showing flower closeup. Isla Magdalena, Baja California Sur, Mexico (images Jon Rebman in 2016). Tribe Epilobieae **H–J** **H.** *Chamaenerion angustifolium* subsp. *circumvagum*. Stem with inflorescence with open flowers and maturing capsules. MN (image Peter M. Dziuk, Minnesota Wildflowers in 2002). **I** *Epilobium nankotaizanense*. Habit with flower. Taiwan (image Ching-I Peng in 2008). **J** *Epilobium canum* subsp. *latifolium*. Stem with flowers and some dispersed seeds. Curry Co., OR (image Gerry Carr 6317b in 2015). Tribe Onagreae. **K, L** **K.** *Xylonagra arborea*. Stem with flowers and very young fruit. Cultivated at Missouri Botanical Garden, originally from Baja California, Mexico (image Ching-I Peng in 1982). **L.** *Taraxia ovata*. Habit with flowers. Monterey, CA (image Christian Schwarz)



**Fig. 2** Tribe Onagreae. A–L. **A** *Clarkia rhomboidea*. Stem with flower and immature capsule. Jefferson Co., OR (image Gerry Carr 6468b in 2006). **B** *Chylismiella pterosperma*. Stem with leaves, flower, and buds. Inyo Co., CA (image Larry Blakely in 2001). **C** *Gayophytum diffusum* subsp. *diffusum*. Stem with flowers and buds. Tulare CA (image Peter Raven collection). **D** *Eremothera boothii* subsp. *boothii*. Flowers and buds. Mono Co., CA (image Michael Moore 3260 in 2015). **E** *Camissonia contorta*. Plant with flowers and young capsules. Klickitat Co., WA (image Gerry Carr 0356b in 2009). **F** *Neoholmgrenia andina*. Plant with fading flower and immature capsule. Harney Co., OR (image Gerry Carr 0996b in 2011). **G** *Tetrapteron graciliflorum*. Plant with flower and immature capsule. Marin Co., CA (image David Greenberger in 2018). **H** *Camissoniopsis cheiranthifolia* subsp. *cheiranthifolia*. Stem with flowers and immature capsules. San Mateo Co., CA (image Leslie Flint in 2015). Inset of immature capsule. Monterey, CA (image Steve Rovell in 2016). **I** *Eulobus californicus*. Stem with flowers. San Diego Co., CA (image Ron King in 2021). **J** *Chylismia eastwoodiae*. Flowers and buds with capsules in background. Emery Co., UT (image Rob Raguso in 2001). **K** *Oenothera tetraptera*. Plant with flower (from Oageng Modise) and immature capsule. South Africa (image Behrens in 2021). Inset of dehiscent capsule with seeds clustered inside. South Africa (image Warren Wagner in 2008). **L** *Oenothera toumeyii*. Stem with flowers. Cochise Co., AZ (image Rob Raguso in 1996)

Americas, especially western North America. Almost all genera in the tribes Lopezieae, Gongylocarpeae, Epilobieae, and Onagreae are endemic to or have had their major basal radiation in the Madrean Floristic Region of southwestern North America [2]. Members of *Epilobium* and *Chamaenerion* (the nomenclaturally correct

name for what has previously been referred to by the synonym *Chamerion*) (Fig. 1H–J), for example, have wind-borne seeds and are distributed widely across the world [3, 4]. *Fuchsia* (Fig. 1C), with animal-dispersed berries, most likely arose in South America or southern North America and diversified extensively in the Andean

region but has also colonized New Zealand and Australia (no longer extant), as well as isolated Tahiti [4–8]. Since the mid-twentieth century, the family has been developed as a model system for studying plant evolution [9]. However, a limitation of these previous studies has been the absence of a robust phylogenetic framework within which to examine the evolution of these traits.

Within Onagraceae, there is a wide range of ecological specialization, pollination syndromes, breeding systems, and chromosomal organization, as well as striking inter- and intraspecific variation for floral scent [4]. The family includes lineages with hummingbird pollination as well as lineages of presumably ancestral vespertine anthesis and hawkmoth pollination with multiple evolutionary origins of bee pollination and especially autogamy [4]. Permanent translocation heterozygosity (PTH), which results in the severe attenuation of recombination during meiosis and is extremely rare in plants, occurs in a single species of *Gayophytum*, is quite common in *Oenothera* (46 spp.), and is thought to be a major modulator of the evolutionary and ecological dynamics within *Oenothera* [10, 11]. In addition, polyploidy is common throughout the family, with an estimated 39% of species being polyploid [4]. Despite its modest size, the family has played a major role in evolutionary theory, starting with De Vries' rediscovery of Mendel's laws through experimentation with *Oenothera*, leading to ideas crucial to the development of the Modern Synthesis [12]. More recent research in the group has focused on themes ranging from cytology, embryology, palynology, chemistry, and reproductive and pollination biology [1, 9], chromosome evolution [10, 13–15], and the role that trade-offs in reproductive mode, floral morphology, and floral scent play in driving diversification in the context of plant-insect interactions [16–19].

Onagraceae systematics has a long history of detailed comparative work, with the most recent family-wide treatment [1] synthesizing all available morphological and phylogenetic evidence. The family consists of two subfamilies [1]: Ludwigioideae, comprising *Ludwigia* (82 spp.; Fig. 1A), and Onagroideae, with all remaining taxa. Onagroideae is currently subdivided into six tribes [1]: Hauyae (1 genus, 2 spp.; Fig. 1B), Circaeae (2 genera, 117 spp.; Fig. 1C, D), Lopeziae (2 genera, 23 spp.; Fig. 1E, F), Gongylocarpeae (1 genus, 2 spp.; Fig. 1G), Epilobieae (2 genera, 173 spp.; Fig. 1H–J), and Onagreae (13 genera, 265 spp.; Fig. 1K, L and 2A–L). Phylogenetic evidence based on targeted gene sequencing of plastid DNA [20] and plastid+nuclear DNA [21] confirmed the monophyly of the family and the individual monophyly of those tribes from which multiple species were sampled. Strong support was also found for *Gongylocarpus* (previously embedded within

Onagreae) as sister to Onagreae+Epilobieae, spurring its subsequent elevation to the tribal level [1]. Within Onagreae, Levin et al. [21] demonstrated that *Oenothera* and *Camissonia* were not monophyletic as circumscribed at the time. Thus, Wagner et al. [1] subsequently expanded *Oenothera* to include the former genera *Calylophus*, *Gaura*, and *Stenosiphon*, and divided *Camissonia* into nine genera. Levin et al. [21] additionally found strong support for two deep lineages within *Oenothera*, referred to as lineages “A” and “B”, with the relationships among most genera in Onagreae and most sections within *Oenothera* poorly resolved.

Subsequently, Johnson et al. [17] inferred phylogenetic relationships of Onagraceae, with a focus on tribe Onagreae. They incorporated data from Levin et al. [21] while also expanding species sampling and adding two additional nuclear markers. In agreement with Levin et al. [21], Johnson et al. [17] found support for the monophyly of Onagreae, Epilobieae, and the recently erected Gongylocarpeae, as well as the previously detected lineages A and B within *Oenothera*. However, several conflicting hypotheses of relationships exist between the two analyses. For example, Levin et al. [21] found moderate support for the monotypic Baja California endemic *Xylonagra arborea* as sister to the rest of Onagreae. In contrast, the analysis of Johnson et al. [17] indicated weak support for this species being nested well within the tribe, with weak support for *Taraxia* as sister to the remaining members of Onagreae. These differences between the two studies suggest potential conflict among gene trees or other analytical constraints.

### Goals of the study

Several phylogenetic relationships within the subfamily Onagroideae remain unresolved. The individual monophyly of the subfamily's six tribes as currently circumscribed appears strongly supported by morphology and DNA ([1] and references therein), but relationships among them are not fully resolved. Additionally, most relationships within the species-rich Onagreae are equivocal, suggesting rapid diversification in this group. Here we employ targeted enrichment of 303 nuclear genes to: (1) elucidate relationships among tribes within Onagroideae, (2) understand relationships among genera in tribe Onagreae and within *Oenothera*, and (3) examine support for the monophyly of genera and historically difficult to resolve clades by exploring levels of conflict among gene trees. The phylogenetic resource provided here will be valuable for understanding biogeographic patterns in Onagraceae, as well as comparative studies ranging from trait evolution to comparative genomics to community ecology.

## Results and discussion

### Target capture and phylogenomic datasets

We used a target capture array of 322 low-copy nuclear protein-coding genes [22, 23] designed using transcriptomes of *O. serrulata* and *O. capillifolia* subsp. *capillifolia* (sect. *Calylophus*), from the 1KP Project [24]. The array uses 120-bp RNA probes to hybridize with genomic DNA fragments prior to amplification and sequencing. Attrition of target loci due to unknown causes in the laboratory, as well as subsequent bioinformatic quality filtering, resulted in a final dataset of 303 loci successfully extracted from Illumina MiSeq libraries prepared for 143 Onagraceae taxa, plus four outgroups. Specimen collection details including voucher information and determination can be found in Table S1 (Additional file 1).

Although the probe sequences were designed from two species of *Oenothera* sect. *Calylophus*, we did not observe any relationship between target recovery and phylogenetic distance to that section within *Oenothera* (Fig. S1a, Additional file 2) or between target recovery and sample age (Fig. S1b, Additional file 2). To confirm whether genes had paralogous copies in any taxa, we assessed the presence of multiple gene contigs assembled for a gene within each sample using the paralog finding scripts distributed with HybPiper. The distribution of putative paralogs suggests that these duplicate copies are largely recent in origin, potentially impacting some species level relationships (particularly within *Clarkia*) but not the higher-level relationships that are the focus of this study. Our dataset was further supplemented by orthologous 1KP Project transcriptome sequences from 21 species, primarily from *Oenothera* sect. *Oenothera*. The final dataset included 168 accessions (Table S1, Additional file 1). We detected an average of 272 genes across all samples. The number of genes recovered varied from 109 to 309 (mean 272, median 298), resulting in gene sequence matrices ranging from 24 to 99% taxon occupancy (mean 83%, median 85%).

We used the individual gene alignments to make two estimations of the species phylogeny by doing the following: 1) concatenating the gene alignments and inferring a species tree using RAxML [25, 26]; and 2) constructing individual gene phylogenies using RAxML and reconstructing the species tree using ASTRAL [27] (Fig. 3), a summary gene tree/species tree method consistent with the multispecies coalescent model.

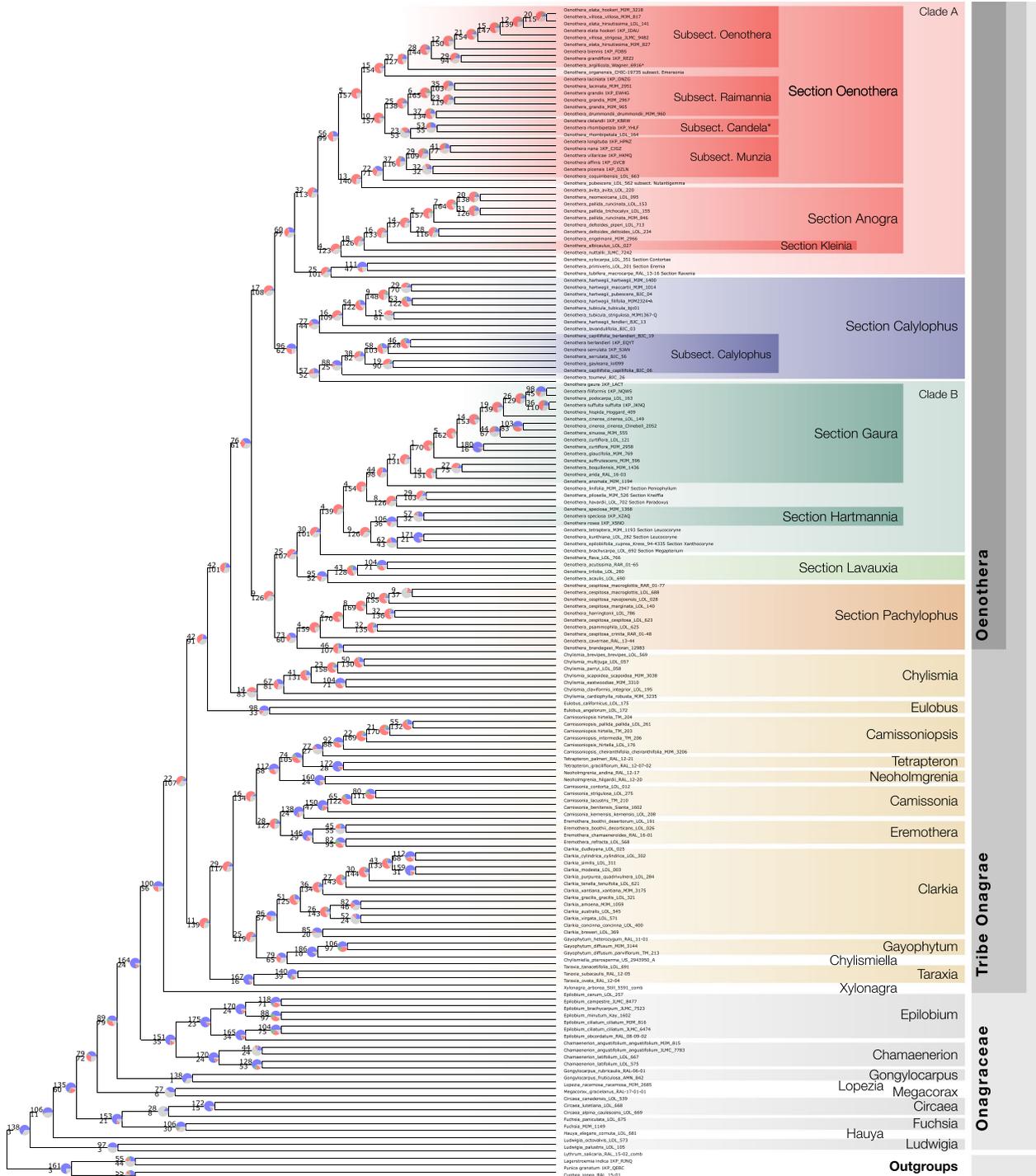
### Examining gene tree conflict

Large, genome-scale datasets, such as the ones obtained via target capture, have been shown to have very high support via traditional metrics including

bootstrapping and posterior probability [28]. Nodes that are maximally supported may still have evidence of conflicting signals among gene trees, which can be further explored by summarizing support for each bipartition across many gene trees. We used PhyParts [29] to assess the number of gene trees concordant with and significantly conflicting with the ASTRAL species phylogeny. As PhyParts requires rooted gene trees, this analysis was done on a reduced set of 206 gene trees that had adequate sampling in our outgroups. Throughout the discussion we will refer to the level of gene tree concordance and conflict accordingly: PhyP = 143/15, referring to the total number of gene trees out of 206 that agree with (143) and disagree with (15) the corresponding topology in the species tree. Note that not all gene trees will be concordant or conflicting; some may be uninformative for a specific bipartition.

We further explored the level of support among the gene trees for the monophyly of key clades (Fig. 4) using DiscoVista [30], a software that creates visualizations of discordance in phylogenomic datasets. We also examined the position of historically difficult clades within Onagraceae, comparing the summary (ASTRAL) topology to the two alternative, rooted quartet trees for each focal node (Fig. 5). A dominant summary topology with the two alternative topologies in relatively equal frequency is consistent with speciation in the presence of incomplete lineage sorting (ILS); if all three topologies are present in roughly equal frequencies, this suggests that significant levels of ILS and gene tree estimation error may prevent the accurate resolution of the node for the given data [30].

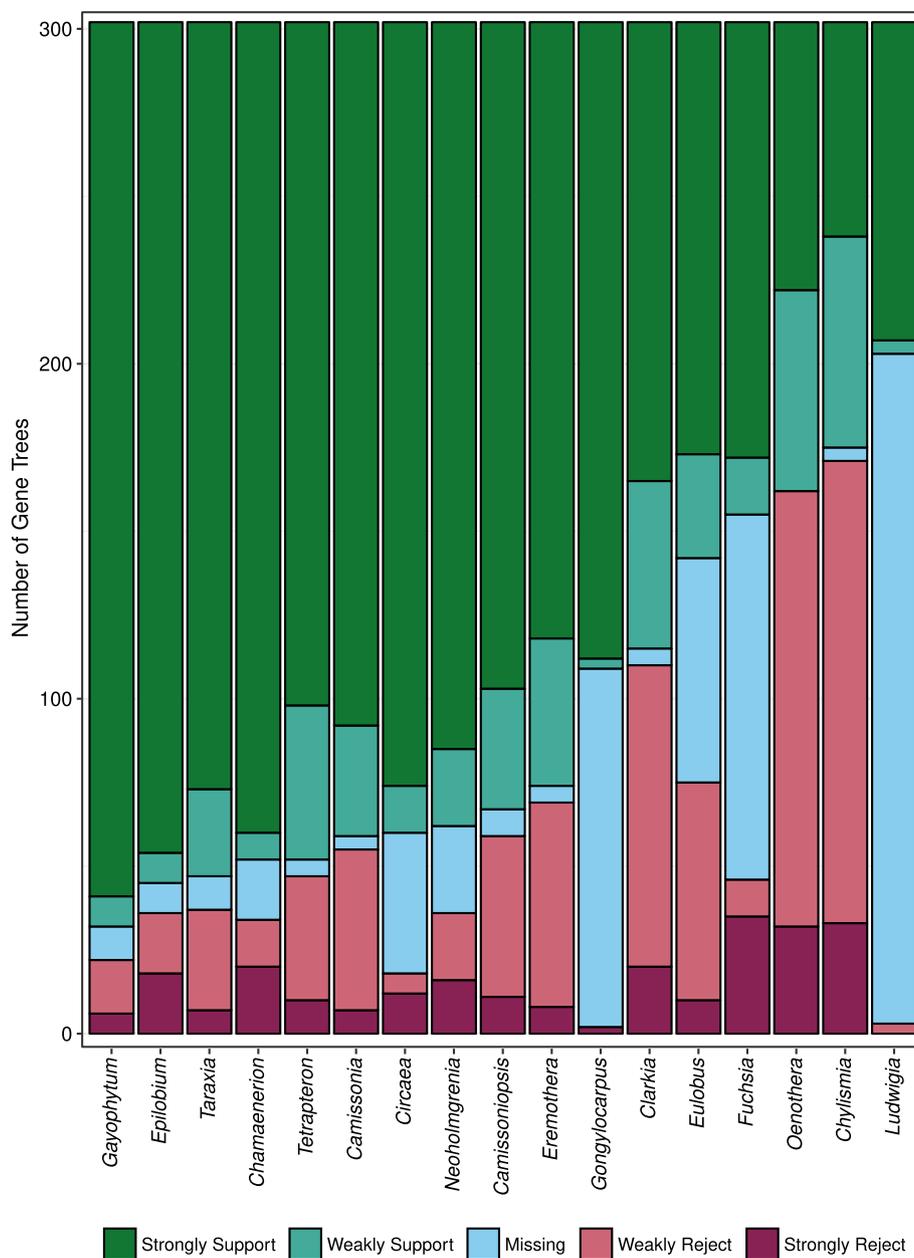
Our target enrichment approach, therefore, allowed for the construction of a large dataset with minimal missing data (10.24% gaps combined across all trimmed alignments) for loci that contain sufficient variability to be informative from the family to species level. In many cases, areas of the Onagraceae phylogeny that disagreed among previous studies were resolved with higher confidence. In other cases, our results reveal that conflicting prior studies may reflect high levels of gene tree discordance in several key nodes, and that conflict is not restricted to shallow phylogenetic scales [31, 32]. Coalescent theory predicts that in many cases, a gene tree that is concordant with the true speciation history may be less likely than conflicting gene trees [33, 34], and demonstrates that these branches may occur anywhere on the tree, not just in more recent nodes (see Fig. 5f for a clear example of this). This phenomenon is referred to as an anomaly zone, where a set of short branches in a phylogenetic tree can result in the most common gene tree topology differing from the true species tree topology [35].



**Fig. 3** Best-scoring ASTRAL-II tree (displayed as a cladogram) based on 303 input best-scoring maximum likelihood gene trees from RAxML.

Numbers above branches represent the number of gene trees in concordance with a particular clade in the species tree (blue in pie chart), whereas those below indicate gene trees in conflict (red/green in pie chart). Pie charts were constructed using 206 genes, and additionally represent the portion of conflicting gene trees that support the main alternative (green), those that support all remaining alternatives (red), and finally those with no information (gray), which could be due to bootstrap support < 50% for the branch in question or missing data (Fig. S1, Additional file 2).

The ASTRAL tree and concatenated ML tree are represented with branch lengths (with the exception of tip branch lengths) in Figure S2 (Additional file 2) and individual tree files are provided in the Dryad repository that accompanies this study



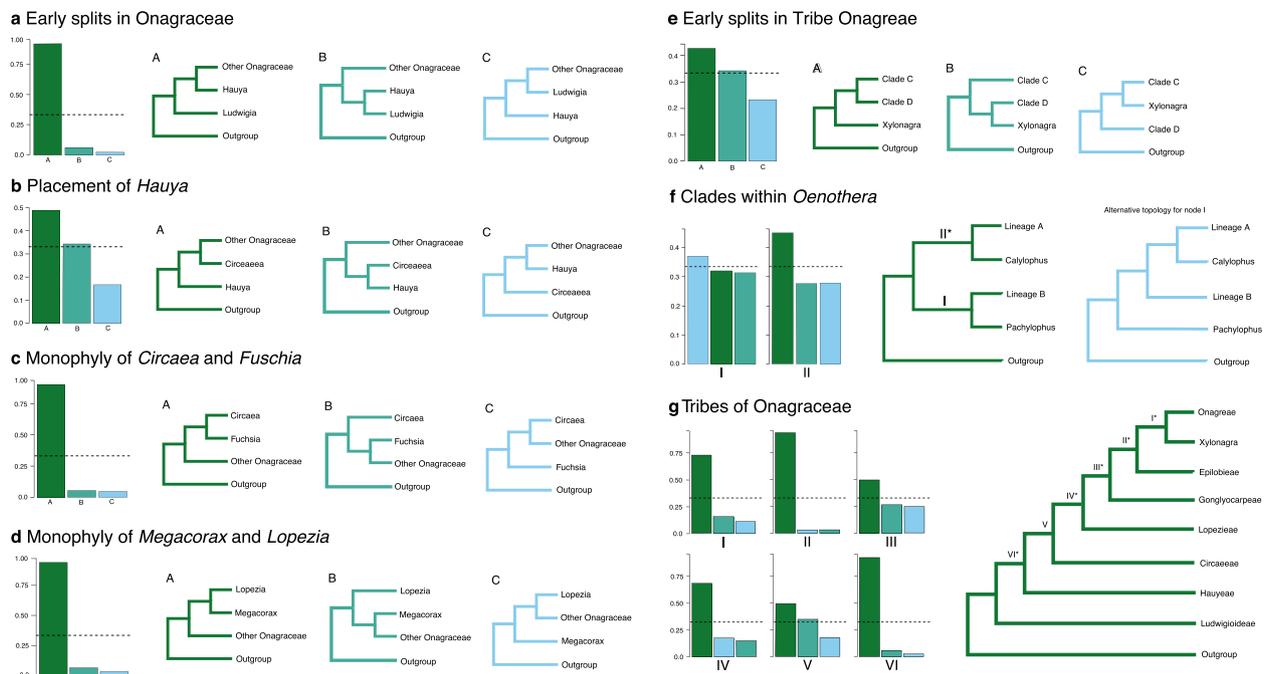
**Fig. 4** Distribution of support for the monophyly of 17 of 22 genera (x-axis) in Onagraceae among 303 gene trees using DiscoVista. Gene trees (y-axis) are shown to strongly support (dark green), weakly support (teal), weakly reject (pink), or strongly reject (maroon) genera. Shown in blue are gene trees that do not contain any data for a certain genus

**Major splits within Onagraceae**

The topologies from both the RAXML + ASTRAL analysis and the best-scoring maximum likelihood tree from the concatenated supermatrix are very similar, with relationships between all tribes and genera identical between the two analyses (Figure S2, Additional file 2). All tribal and generic relationships outside Onagraceae received 100% support from the ASTRAL and concatenated analyses

(Figs. 2, 3 and 4). However, lower-level relationships, especially within the more heavily sampled Onagraceae, are often defined by branches with shorter lengths in our ML analysis, as well as increased gene conflict revealed by PhyParts (Fig. 3).

All analyses strongly support the relationship of the pantropical *Ludwigia* (subfamily Ludwigioideae) as sister to the rest of the family (PhyP=143/15; Fig. 5a), which



**Fig. 5** Alternative topologies for clades within Onagraceae. The relative frequency of gene trees is shown for quartets of taxa (bar graphs). Green bars indicate the RAxML-ASTRAL topology, teal and light blue bars indicate alternative topologies. Dotted lines indicate 1/3 of the total gene trees for each quartet; roughly equal proportions for the three possible topologies indicate a polytomy within the present dataset. **a** Early-diverging Onagraceae; **b** Placement of *Hauya*; **c** Monophyly of *Circaea* and *Fuchsia*; **d** Monophyly of *Megacorax* and *Lopezia*; **e** Early-diverging Onagreae, where clades X and Y refer to the following groups: clade X comprises *Camissoniopsis*, *Eremothera*, *Camissonia*, *Tetrapteron*, and *Neoholmgrenia*; clade Y comprises *Clarkia*, *Chylismiella*, and *Gayophytum*; **f** Clades within *Oenothera*, the quartet at node I is ((Lineage B, sect. *Pachylophus*), Lineage A + sect. *Calylophus*) Outgroup) and node II is ((Lineage A, sect. *Calylophus*), (Lineage B + sect. *Pachylophus*) Outgroup), the mode common alternative topology for node I is shown

has been established in previous phylogenetic studies [20, 21, 36–38]. *Ludwigia* also has well-documented morphological autapomorphies, which are floral tube absence, pollen in most sections shed as tetrads (occasionally polyads), a nectary at the base of a stamen, and ovule arche-sporium single-celled, and outer integument dermal [39, 40]. Previous morphological [41] and molecular phylogenetic ([21, 38], p. 200) evidence support the monophyly of *Hauya*, but its relationship to other tribes has been difficult to resolve. This moth- or potentially bat-pollinated group of two species (Fig. 1B) ranging from central Mexico to Costa Rica has been hypothesized to possess morphological synapomorphies that closely align *Hauya* with members of *Clarkia*, as well as *Oenothera* sections *Calylophus* and *Gaura* [41, 42]. These hypotheses have been rejected by all molecular phylogenetic analyses [7, 17, 20, 21, 36–38, 43]. Both Ford and Gottlieb [38] and Johnson et al. [17] found support for a branch defining a sister relationship between *Hauya* and tribe Circaeae. However, we corroborate the result of Levin et al. [21] that *Hauya* is sister to all remaining members of subfamily Onagroideae (A = 100; ML = 100; PhyP = 106/11; Fig. 5b).

The monophyly of *Fuchsia* (A = 100; ML = 100; PhyP = 106/30) and its sister relationship to *Circaea* (PhyP = 153/21; Fig. 5c) is strongly supported, despite striking morphological differences between the two genera. *Fuchsia* (Fig. 1C) is a mainly tropical genus with 4-merous bird-pollinated flowers that are generally red in color, whereas *Circaea* (Fig. 1D) is restricted to northern latitudes and has 2-merous, autogamous, or insect-pollinated white flowers [1]. This sister relationship between the genera is overwhelmingly supported by the analysis of alternative quartets, and the node supporting this relationship shows very little influence of ILS (Fig. 5c). The monophyly of *Circaea* receives high support from both ASTRAL and ML trees (A = 100, ML = 100, Fig. 4). The low number of genes with phylogenetic signal at the node defining the monophyly of *Circaea* is due to the reduced number of genes recovered for sample *Circaea\_canadensis\_LOL\_668* (47 of 302 genes). Regardless, the majority of informative genes for this topology agree with the monophyly of *Circaea* (PhyP = 28/8), results that agree with a larger, more comprehensive study of the genus [44].

Johnson et al. [17] found weak support for the placement of *Lopezia* as sister to all members of Onagraceae except for *Ludwigia*. Somewhat similar is a clade of *Lopezia* + *Circaea* + *Fuchsia* obtained in the morphological family-wide phylogenetic study of Hoch et al. [41], but that relationship was based on the single character of integument histogenesis. Neither of these relationships has been recovered in any of the other studies except for Martin & Dowd [45]. Our analyses, however, strongly support the results of Levin et al. [20, 21] and Ford and Gottlieb [38] that tribe Lopezieae (including *Megacoras*) is sister to Gongylocarpeae + Onagreae + Epilobieae (A = 100; ML = 100; PhyP = 79/72; Fig. 5d). Further, we find strong support for a monophyletic *Epilobium* (A = 100; ML = 100; PhyP = 175/23; Fig. 4) and *Chamaenerion* (A = 100; ML = 100; PhyP = 170/24; Fig. 4) and their sister relationship composing Epilobieae (A = 100; ML = 100; PhyP = 151/35), which is consistent with a more detailed study by Baum et al. [46] and others [17, 20, 21]. Tribe Epilobieae is clearly sister to Onagreae, as has been previously reported [17, 20, 21].

#### Relationships within tribe Onagreae

The enigmatic Baja California endemic, and presumably hummingbird-pollinated, *Xylonagra arborea* (Fig. 1K) is strongly supported as sister to all other members of Onagreae in some of our analyses (A = 100; ML = 100); however, analysis with PhyParts reveals that only 22 trees support this topology (with 107 informative trees dissenting; Fig. 3). In addition, the DiscoVista relative frequency analysis supports *Xylonagra* as sister to the rest of Onagreae, but there is also moderate support for a sister relationship to the clade comprising *Camissoniopsis*, *Eremothera*, *Camissonia*, *Tetrapteron*, *Neoholmgrenia*, *Clarkia*, *Chylismiella*, and *Gayophytum* (Fig. 5e). The former relationship is consistent with Levin et al. [21] and potentially clarifies previous conflicting results about its placement within the tribe [17, 20]. Inside Onagreae, both the ASTRAL and RAxML analyses strongly support a sister relationship between a clade comprising *Camissoniopsis*, *Eremothera*, *Camissonia*, *Tetrapteron*, and *Neoholmgrenia* (Fig. 2d–f) and a clade comprising *Clarkia*, *Chylismiella*, and *Gayophytum* (Fig. 2a–c) with *Taraxia* as sister to these two clades together. A clade composed of *Clarkia*, *Chylismiella*, *Gayophytum*, and *Taraxia* was recovered with weak supported by Levin et al. [21]; however, Johnson et al. [17] recovered *Taraxia* as sister to all other members of Onagreae, and a relationship of *Clarkia*, *Chylismiella*, and *Gayophytum* as sister to all other remaining members of Onagreae except *Taraxia*. In Levin et al. [21], the relationships of the clade of *Camissoniopsis*, *Eremothera*, *Camissonia*, *Tetrapteron*, and *Neoholmgrenia*, and the clade of *Clarkia*, *Chylismiella*,

and *Gayophytum* within the tribe lacked resolution, and weak support was found for inclusion of *Taraxia* within *Clarkia*, *Chylismiella*, and *Gayophytum* as sister to *Clarkia* + *Gayophytum* + *Chylismiella*. The gene tree conflict surrounding the placement of *Taraxia* in the family reveals the source of previous confusion. Although both our ASTRAL and ML analyses give 100% support to the grouping *Taraxia* + (the clade of *Camissoniopsis*, *Eremothera*, *Camissonia*, *Tetrapteron*, and *Neoholmgrenia*, + the clade of *Clarkia*, *Chylismiella*, and *Gayophytum*), only 11 of 139 gene trees are in concordance with this relationship, whereas 11 gene trees also support a relationship of *Taraxia* + (*Eulobus* + (*Chylismia* + (*O. sect. Pachylophus* + *O. sect. Lauvaxia* + lineage B) + (*O. sect. Calylophus* + lineage A))) and 14 gene trees support a relationship of *Taraxia* + *Xylonagra*. This may be a case of an anomaly zone in our current dataset, where the true species tree is not represented by the majority of gene trees [33, 34].

The monophyly of *Neoholmgrenia*, *Camissoniopsis*, and *Tetrapteron* (Fig. 2f–h) is highly supported (A = 100, ML = 100, PhyP = 112/58). However, the node defining the sister relationship *Camissoniopsis* + *Tetrapteron* is highly supported in ML and ASTRAL analyses but received only moderate PhyParts support PhyP = (74/105), with the most common conflicting topology being *Tetrapteron* and *Neoholmgrenia* sister to *Camissoniopsis*. This sister relationship between *Camissoniopsis* and *Tetrapteron* was previously recovered albeit with weak support in Levin et al. [21] and with strong support in Johnson et al. [17]. The remaining two genera of lineage E, *Camissonia* and *Eremothera*, comprise a clade, a result previously suggested or strongly supported in previous analyses. This relationship is poorly supported in our analysis by both ASTRAL (A = 48) and PhyParts (28/127), but the short branch defining this relationship received 100% bootstrap support in our ML analysis. The clade of *Clarkia* + (*Chylismiella* + *Gayophytum*), which was also recovered in Levin et al. [21], receives 100% support in both ASTRAL and ML analyses but is poorly represented by gene trees overall (PhyP = 25/119). However, the sister relationship between *Gayophytum* and *Chylismiella* is supported by higher gene-tree concordance (79/65).

#### Relationships within Oenothera

As in previous studies [21, 47], we find strong support (A = 100, ML = 100, PhyP = 42/101) across all analyses for the relationship of *Eulobus* as sister to *Oenothera* + *Chylismia* (Fig. 2i–l). Within *Oenothera*, there is strong support among all analyses for the previously described lineages A and B [21]. These deep lineages within the genus were first detected through synapomorphic seed morphology, with lineage A

possessing radially enlarged endotestal cells, and lineage B either angled or winged capsules [48]. The monophyly of these lineages was subsequently phylogenetically confirmed, but the placement of the remaining sections (*Calylophus*, *Lavauxia*, *Pachylophus*) has been a mystery, with many conflicting topologies supported with regard to their relationships [17, 20, 21].

*Oenothera* sect. *Calylophus* is a group of 7 spp. and 13 taxa, with a suspected Pleistocene radiation centering around the southwestern U.S., and repeated evolution of both bee pollination (ancestrally hawkmoth-pollinated) and gypsum endemism [23, 49]. This section has previously garnered conflicting phylogenetic support for a sister relationship to lineage B, lineage A, or even to sect. *Pachylophus* [17, 21]. With complete taxon sampling of sect. *Calylophus*, both the ASTRAL and concatenation analyses strongly support a sister relationship of sect. *Calylophus* with lineage A, as does the DiscoVista gene tree analysis (Fig. 5f). Although no representative of sect. *Calylophus* was analyzed in the seed/capsule analysis of Tobe et al. [48], sect. *Calylophus* has since been predicted to be consistent with membership in lineage A, due to its cylindrical (non-angled) capsules [1]. The short branch defining this relationship in our ML analysis and the limited number of gene trees (17) in concordance vs. the 108 gene trees in conflict with this topology warrant caution with this result. Only 17 gene trees agree with our ASTRAL topology; however, there are no alternate topologies supported by more than three gene trees. There is strong support for the monophyly of the two subsections of sect. *Calylophus* (A=100, ML=100, PhyP=96/62) with the exception that *O. toumeyi* ([23, 49]), traditionally placed in subsect. *Salpingia*, is strongly supported as sister to all other members of subsect. *Calylophus* (A=100, ML=100, PhyP=57/52), corroborating the findings of Cooper et al. [23].

*Oenothera* sect. *Lavauxia*, which has also been historically difficult to place within *Oenothera*, is a widespread hawkmoth-pollinated group ranging from southern Canada to Mexico, with two South American species. The group exhibits striking floral variation: *O. flava*, which is restricted to sky islands in the southwestern U.S. and northern Mexico, exhibits possibly the longest floral tubes in North America [50, 51]; and references therein) despite the modest-sized flowers of geographically widespread conspecifics. We present almost complete taxon sampling of this Sect. (4 of 5 taxa) and find strong support in all analyses for sect. *Lavauxia* as sister to lineage B, corroborating a weakly supported result from Johnson et al. (2009). This relationship was predicted based on the distinctly winged capsules of species in sect. *Lavauxia* [1]. Twenty-five gene trees agree with this topology, and only five gene trees place this section in the lineage

containing *Calylophus*+lineage A instead. No other arrangements occurred in more than 3 gene trees.

*Oenothera* sect. *Pachylophus* is a group of five species and nine taxa with conspicuous, hawkmoth-pollinated flowers that ranges from Canada through the western U.S. to Mexico. Its seeds possess a synapomorphic “collar”, a large, hollow chamber that dramatically imbibes water and has been attributed to its colonization of an impressive habitat range including deserts, dune systems, grasslands, pinyon-juniper woodlands, and coniferous forests [1]. Cladistic analysis of seed coat anatomy suggested an affinity with members of lineage A [48] and previous molecular phylogenetic analyses have left the placement of sect. *Pachylophus* within *Oenothera* either unresolved [21] or weakly supported as sister to lineage A [17]. With complete taxon sampling for sect. *Pachylophus*, we find moderate support in our ASTRAL analysis (A=85) for a sister relationship between sect. *Pachylophus* and lineage B. However, our ML analysis could not resolve the placement of sect. *Pachylophus*. A deeper exploration of this node reveals nine gene trees in concordance with the ASTRAL topology (i.e., a sister relationship with lineage B), whereas 5 gene trees support a sister relationship of sect. *Pachylophus* to lineage A + sect. *Calylophus*. In addition, the DiscoVista analysis showed relatively equal gene tree frequencies supporting the ASTRAL topology, as well as sect. *Pachylophus* as sister to lineage B, and lineage A + sect. *Calylophus* (Fig. 5f). The patterns of gene tree conflict we observe relative to the summary topology suggest that these relationships fall within the anomaly zone [33, 34], where short times between speciation events and high levels of ILS result in a majority of gene histories that are inconsistent with the species history. The increased taxon and gene sampling of these analyses has revealed underlying conflict in phylogenetic signal among gene trees and confirmed previous difficulties in resolving the phylogenetic affinities of sect. *Pachylophus*. Within sect. *Pachylophus*, the widespread and morphologically diverse species *Oenothera cespitosa* appears paraphyletic as currently defined due to its exclusion of *O. psammophila* and *O. harringtonii*; a result previously suspected based on morphological data [52], and shown recently [22] to be the result of budding speciation arising from edaphic specialization. This more detailed investigation into the taxon relationships in this group has revealed complex relationships among taxa, including potential hybridization [22].

#### Relationships within *Oenothera* lineage A

All analyses recover strong support for a sister relationship between *O. xylocarpa* (sect. *Contortae*) and *O. primiveris* (sect. *Eremia*), with high gene concordance (A=100, ML=100, PhyP=111/47). Additionally,

ASTRAL analysis supports the inclusion of *O. tubifera* (sect. *Ravenia*) within this clade ( $A=97$ ). However, ML analysis recovers a conflicting relationship with *O. tubifera* as sister to all remaining members of lineage A (Fig. 4). Sections *Oenothera* and *Anogra* receive 100% support in both analyses, albeit with only 32 of 113 gene trees in concordance. Subsections within the large section *Oenothera* are generally supported as monophyletic, but with short branch lengths in our ML analysis and high gene conflict reported by PhyParts. Subsections *Oenothera*, *Raimannia*, *Munzia*, and *Candela* all receive 100% support from ASTRAL and ML analyses.

In the only case in our analyses where a topology disagrees with the monophyly of a section, we find evidence that sect. *Anogra* is not monophyletic. In both ASTRAL and ML analyses, sect. *Anogra* forms a strongly supported clade ( $A=95$ ,  $ML=100$ ) that includes *O. albicaulis* from sect. *Kleinia*, a result previously recovered in phylogenetic analyses focusing on lineage A [53]. The two white-flowered species of sect. *Kleinia* (only *O. albicaulis* is present in our analysis) share several seed morphology characters with subsect. *Raimannia* that are unlike anything in sect. *Anogra* [1, 48]. However, phylogenetic analyses including representatives from both sections have consistently placed *Kleinia* within sect. *Anogra* [53, 54], and have even found evidence that *Kleinia* itself is not monophyletic within sect. *Anogra* [17, 54]. These phylogenetic findings suggest that the aforementioned seed characters, which are both external seed coat features and internal anatomy, are not synapomorphies of the species pair currently circumscribed under sect. *Kleinia* and have potentially been gained or lost more than once within lineage A. The uncertainty with the topology of sects. *Kleinia* and *Anogra* is potentially due to gene tree discordance, or ancient hybrid events that have made the species tree reconstruction challenging. Including *O. coronopifolia*, the other member of sect. *Keinia*, and diploid and tetraploid cytotypes of *O. nuttallii* in future work may help with the reconstruction of this recalcitrant group.

#### Relationships within *Oenothera* lineage B

Strong support among analyses was recovered for *O. brachycarpa* (sect. *Megapterium*) as sister to the remainder of lineage B ( $A=100$ ,  $ML=100$ ,  $PhyP=31/101$ ). The monophyly of the large sect. *Gaura* within lineage B is 100% supported by ASTRAL and ML analyses with 44 vs. 98 gene trees in concordance. A sister relationship between sect. *Gaura* and a clade containing sections *Kneiffia*, *Paradoxus*, and *Peniophyllum* received high support in ASTRAL but was defined by a short, weakly supported branch in our ML analysis that is supported by only 4 gene trees ( $A=97$ ,  $ML=52$ ,  $PhyP=4/154$ ). The clade containing sections *Hartmannia* (Fig. 1K),

*Leucocoryne*, and *Xanthocoryne* is also strongly supported by ASTRAL and ML analyses (100%), but with only 9 genes of 126 agreeing with this topology.

#### Conclusions

Here we present the first phylogenomic analysis of relationships in Onagraceae using 303 nuclear, putatively single-copy genes. Depending on the relationships in question, these increased data resolved relationships that were previously unclear or revealed significant levels of gene tree conflict. Both ASTRAL and RAxML produced virtually identical topologies with regard to the relationships among tribes and genera with high bootstrap support throughout and in many cases high gene tree concordance. However, relationships among genera and sections, especially within the more heavily taxon-sampled Onagreae, reveal high conflict among gene trees for lower-level relationships. These cases, where increased gene number has still failed to confidently resolve relationships within the family reveal deep conflict among gene trees which could be due to rapid radiation, ILS, hybridization (ancient and recent), selection, as well as lack of information content [55–57]. Future analyses must explore these relationships, potentially using increased genomic sampling and lower-level taxonomic case studies.

Genomic sequencing approaches such as HybSeq provide a cost-effective way to gather hundreds of nuclear genes for phylogenetic analysis at multiple phylogenetic scales [58, 59]. With these large multi-gene datasets, however, gene tree conflict presents a significant challenge to species tree inference. Large datasets can help detect patterns of incomplete lineage sorting or hybridization [60, 61], but more data does not always help with estimating a strongly supported bifurcating species tree [62]. In some cases, only a small portion of the genome is unaffected by inter-taxon gene flow or ILS, making it more difficult to determine which loci or genes are the most appropriate for tree inference [31, 35, 62] or whether a bifurcating tree is an accurate representation of the evolutionary history of the group. Gene tree conflict, and the evolutionary mechanisms behind it, are likely causing some of the difficulties in reconstructing relationships within Onagreae [47]. For example, *Xylonagra arborea* is generally supported as sister to the rest of the tribe, but alternative topologies are supported by some analyses [17, 20]. The same is true for *Oenothera* sect. *Pachylophus*. This section has been difficult to place [17, 20, 21], and our DiscoVista analysis indicates the relationships with lineages A and B plus sect. *Calylophus* may fall into the anomaly zone for our current dataset. An important consideration is that speciation may have happened rapidly in these groups, leaving little to no trace of the true

evolutionary history [33, 34], or reflecting the fact that a bifurcating tree may not be an accurate representation of a rapid radiation in the presence of gene flow and ILS. We emphasize that the bifurcating species tree presented here represents a hypothesis of relationships rather than the true history, and that the measures of gene tree conflict alongside the tree suggest that evolution was not consistently tree-like throughout the history of the family. Analyzing varying subsets of gene trees under different evolutionary scenarios [62] and using population-level sampling and analysis may be necessary to better elucidate the true evolutionary history of lower-level clades within Onagraceae. Understanding the evolutionary history of such groups, whether it be a polytomy or bifurcating tree, is important to provide a contextual building block for further research in ecology and evolution.

## Methods

### Taxon and tissue sampling

Individuals from across Onagraceae were chosen to represent as many lineages as possible, with sampling focused most extensively on tribe Onagreae and *Oenothera*. Leaf material was sampled for 148 individuals from either field-collected (wild specimens), silica-dried tissue, or herbarium vouchers (with a maximum age of 49 years from collection date). Specimen collection details including voucher information, determination, and NCBI SRA accession numbers can be found in Table S1 (Additional file 1). DNA extractions were performed using a modified CTAB protocol [23] involving purification with silica, except in a few cases where repeated attempts resulted in insufficient DNA after the silica cleaning stage, in which case this stage was omitted.

### Library construction, bait capture, sequencing

Genomic libraries with an insert size of 550 bp were prepared using the TruSeq Nano HT DNA Library Preparation Kit (Illumina San Diego, CA, USA) following manufacturer's instructions, except that all reagent volumes (except PCR reagents) were cut in half beginning with the second addition of AMPure (SPRI) beads (Beckman Coulter, Beverly, MA). Successful library preparation was confirmed with the Qubit 2.0 fluorometer (Invitrogen Carlsbad, CA, USA) using the dsDNA HS Assay Kit, as well as BioAnalyzer 2100 traces (Agilent Technologies, Santa Clara, CA, USA) on a subset of samples. Target enrichment with liquid hybridization was performed using a MYbaits custom target enrichment kit (Mycroarray, Ann Arbor, MI, USA) designed for use in *Oenothera* [22, 23]. Libraries were multiplexed into pools containing 6–18 samples, roughly organized by taxonomic affiliation (e.g., *Oenothera* samples were hybridized together), with 100 ng of total starting library per sample in each

pool. In the few cases where less than 100 ng was present, we used the total amount available (lowest successfully attempted ~25 ng). In all cases, we did not exceed 1.2 µg of total DNA per pool as recommended by the manufacturer. Hybridization was performed at 65 °C for ~18 h and enriched library pools were amplified with 14–18 PCR cycles as needed. No correlation was observed between PCR cycle number and ultimate target recovery by sample, suggesting that 18 cycles (or possibly more) results in little target loss through library bias under our multiplexing and sequencing parameters. In many cases, using higher PCR-cycle numbers was crucial for gaining sufficient product concentration for sequencing, especially for samples from older collections and herbarium vouchers. Each resulting PCR-amplified pool was then cleaned with a QiaQuick PCR Purification Kit (Qiagen, Hilden, Germany). Excess adapter, as revealed through the BioAnalyzer, was removed pre-sequencing with a 0.7 to 1 volume ratio of Ampure beads to product. Sequencing of enriched pools of libraries containing 60–80 individual samples was carried out on the Illumina MiSeq System (600 cycle, v3 chemistry) with a final loading concentration of 16.5 pM (estimated from Qubit and BioAnalyzer output) and a 1% molar ratio of PhiX Control (Illumina). Individual sequencing runs resulted in approximately 28 million read-pairs passing Illumina quality filtering with an average of approximately 1–1.5% of reads assigned to each individual. Raw reads are deposited at the NCBI Sequence Read Archive (BioProject ID PRJNA544074); gene alignments, gene trees and species trees, and other related files and codes are deposited at Dryad (<https://datadryad.org/stash/share/Um2cZ0ubGDzGAhdXjuXPxzTGy8iL8ceewfVO5yRoLSc>).

### Quality filtering, assembly and alignment

Raw, demultiplexed reads from the MiSeq platform were downloaded and quality filtered as paired reads with Trimmomatic [63] using the following settings: ILLUMINA-CLIP: illumina\_adapters.fasta:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:20 2. All reads retaining a mated pair were saved for downstream analysis. These data were combined with an additional 21 transcriptomes from the 1KP project ([www.onekp.com](http://www.onekp.com)) for which orthologous genes were assembled from reads using HybPiper, for a final dataset comprising 169 accessions representing 152 taxa and 129 species (including four outgroups in the Lythraceae; Table S1, Additional file 1). To extract exon sequences from raw reads, we used the HybPiper v1.2 pipeline [59]. Briefly, HybPiper searches reads against a file of target gene sequences, assembles reads into contigs with SPAdes [64], aligns contigs to reference targets, and then scaffolds and translates them. For a given target ortholog, both *O.*

*capillifolia* subsp. *capillifolia* and *O. serrulata* sequences were available from the IKP project, but for many genes only partial exon coverage was available for either of the two species. To avoid issues from samples mapping to only one or the other partial reference sequence, where necessary we created a chimeric sequence representing both species in our HybPiper target file. HybPiper was run with default settings except for specifying `-bwa`, which uses nucleotide-level data when raw reads are matched to target genes. Due to concerns about sequence divergence affecting gene recovery, for 31 samples outside the tribe Onagreae where less than 275 genes were recovered, HybPiper was rerun using the default BLASTX method that matches reads to targets and aligns SPAdes contigs with targets at the protein level. In five cases, gene recovery improved and for these samples the data based on protein alignment was used instead for downstream analyses. Two python scripts, `short_seqs.py` and `remove_seqs.py` ([github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)) were then used to remove gene files in the HybPiper output that represented sequences with <25% of target sequence length. After these short sequences were removed, samples retained between 47–308 genes (Fig. S1, Additional file 2). For CDS alignments, sequences for a given gene were gathered from all samples into a single FASTA file, with independent files for nucleic and amino acid. Protein coding sequences were searched for stop codons, which were replaced with the letter “X” and these sequences were aligned with MAFFT v7.130b [65] using the following settings: `-localpair -maxiterate 1000`. Nucleic acid sequences were then mapped to amino acid alignments using `pal2nal v14` [66] with default settings. Empty gene files were subsequently removed and positions in alignments which were represented by less than 50% of samples were removed with the alignment trimmer `trimAl v1.4.rev.15` (Capella-Gutiérrez et al., 2009).

#### Gene tree estimation

Unrooted gene trees with 100 bootstraps each were produced with RAXML-HPC v8.2.0 (Stamatakis, 2014) using partitioning based on codon position, the rapid bootstrap method, the GTRCAT model of nucleotide substitution, and all other parameters on default. In some cases, poor alignments of individual sequences resulted from poor sequence recovery and/or misidentified orthology between HybSeq and transcriptome sequences. To identify poorly aligned sequences, gene trees were searched for branches of unreasonable lengths, defined as branches with lengths exceeding a percentage of the total gene tree depth: 25% for terminal branches, 50% for internal branches, and 75% for outgroup branches. A total of 88 gene trees were flagged by the script `brlen_outliers.py` ([github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)). After

manually investigating each tree, the offending sequence was removed and the corresponding alignments and gene trees were again generated with MAFFT, `pal2nal`, and RAXML. This manual pruning resulted in 228 sequences that were removed from 137 gene alignments. The distribution of manually removed sequences was such that only 13 samples had their sequences removed from greater than five gene alignments and the maximum number of genes removed for a single sample was 16 genes. After manual pruning, six genes were removed entirely from all analyses due to poor sample representation (contained <15 sequences) resulting in a final target gene list of 303 genes. Prior to downstream analyses, gene tree branches with <33% support were collapsed across all gene trees using `DendroPy v4.2.0` [67] and `sumtrees.py v4.2.0` [68].

To explore an alternative gene tree-building method, we also constructed gene trees for each of the 303 loci using IQ-TREE 1.6.9 and [69], performed model selection with ModelFinder for each locus [70]. To test branch support we used ultrafast bootstrap approximations [71], as well as single branch tests with the approximate likelihood ratio test, both with 1000 replicates [72].

#### Species tree estimation

Species tree estimation was carried out in two ways: using ASTRAL [73] to conduct a summary gene tree/species tree analysis and using a concatenated supermatrix. We used the ASTRAL-II implementation (for large datasets) of ASTRAL v4.10.2 with 303 gene trees. Support was evaluated using 100 multilocus bootstrap replicates (which samples from the gene tree bootstraps and accounts for gene tree uncertainty) and the local posterior probability (which evaluates quartet support at each node). This was done separately using the RAXML and IQ-Tree gene trees; the resulting trees were essentially identical. For our concatenated analysis, the 303 genes that passed quality filtering were concatenated into a single FASTA file with 169 samples and a resulting matrix length of 260,466 bases containing 175,265 variable sites. A partition file for this combined alignment was generated using a script distributed with HybPiper ([https://github.com/mossmatters/HybPiper/blob/master/hybpi\\_per/fast\\_merge.py](https://github.com/mossmatters/HybPiper/blob/master/hybpi_per/fast_merge.py)), which partitioned the alignment by both gene and codon position. A maximum likelihood tree with 100 bootstraps was generated with the same settings specified above for individual gene trees using RAXML HPC v8 [26] on XSEDE using Cipres Science Gateway [74].

#### Examining gene tree conflict

Gene tree conflict was examined with PhyParts [29] using the script `reroot_trees.py` to root gene trees that

contained outgroup taxa. This left a remaining subset of 206 gene trees where at least one of our four outgroups were present. An ASTRAL tree was generated with these rooted gene trees using the same parameters as previously mentioned for ASTRAL analysis. This resulting ASTRAL tree was rooted to the outgroups and analyzed with PhyParts in combination with the rooted gene trees. The output from PhyParts was then displayed on the ASTRAL topology (Fig. 3) with phypartspiecharts.py. Note that the remainder of the 206 gene trees not in the first two groups are those with low gene tree support values < 33. The scripts for rerooting gene trees and visualizing the pie charts are available at [github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts).

DiscoVista was used to investigate and visualize phylogenomic discordance. Specifically, the gene tree compatibility and branch quartet frequencies tools were used to look at the monophyly of genera and historically difficult clade topologies, respectively. The monophyly of genera in the ASTRAL species tree was compared to the differing topologies of the 303 gene trees (Fig. 4). The positions of clades within Onagraceae that have been historically difficult to resolve were examined using quartet trees (Fig. 5). DiscoVista analyzes the relative frequency of gene tree topologies that match the species tree topology for a given subset of clades. The ASTRAL tree was compared to alternative topologies for certain clades of interest; if a tree was present in 1/3 or more of gene trees, it was considered the most likely topology. If different topologies were present in roughly equal frequencies, this indicates a hard polytomy in the phylogeny.

#### Abbreviations

PTH	Permanent translocation heterozygosity
ML	Maximum likelihood
ILS	Incomplete lineage sorting

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-023-02151-9>.

**Additional file 1: Table S1.** Table of all samples in the analysis, collecting locations, and where deposited.

**Additional file 2: Figure S1.** (a) Matrix gene recovery per sample for genes with >50% target recovery. Samples are ordered by section of Onagraceae. (b) Matrix of gene recovery ordered by age of sample. **Figure S2.** Tanglegram of ASTRAL species tree (left) and concatenated ML tree (right).

#### Acknowledgements

We are extremely grateful to the following collectors who provided material for this project: R. Benkendorf, E. Brodie, D. Bruzesse, A. Cisternas, B. Cooper, M. Eshelmen, N. Fraga, E. Gardner, H. Goodmuth, A. Gruver, L. Hendrickson, E. Hilpman, L. Hintz, T. Jogesh, K. Kay, N. Khan, C. Klase, K. Krakos, L. Landis, E. Lewis, E. Levy, T. Miller, R. O'Dell, H. Ochoterena, H. Olvera, W. K. Ostler, M. Rhodes, A. Rork, S. Sianta, L. Steger, S. Steele, S. Still, N. Talkington, S. Todd, and S. Weller. We thank the Chicago Botanic Garden Herbarium (CHIC) and the United

States National Herbarium (US) for providing leaf material and the following land owners/agencies for providing permission to collect: Anza-Borrego State Park, California Botanic Garden, Chicago Botanic Garden, Chugach State Park, Bureau of Land Management (AZ, CA, CO, ID, NM, and UT), Department of Defense (Nevada National Security Site and White Sands Missile Range), Los Angeles County Arboretum, National Park Service (Carlsbad Caverns National Park, Death Valley National Park, Guadalupe Mountains National Park, Mojave National Preserve, and Pinnacles National Park), San Francisco Botanic Garden, Tejon Ranch, and U.S.D.A. Forest Service (Regions 2, 3, 4, and 5). Material for *O. serrulata* and *O. capillifolia* subsp. *berlandieri* that were used for designing probes for target enrichment (1KP accession codes SJAN and EQYT, respectively) was originally collected by RAR and prepared for RNA sequencing in the lab of Marc Johnson at the University of Toronto, Mississauga. Several additional 1KP transcriptomes were included in this study that were published in [75]; we gratefully acknowledge the work carried out by Marc Johnson and collaborators, and the 1KP Project, that went into generating these data. Sequencing of target-enriched genomics libraries was carried out at Pritzker Laboratory for Molecular Systematics and Evolution located at the Field Museum, Chicago, IL.

#### Authors' contributions

RPO, MGJ, LLB, SPK, NAD, JBF, PCH, RAL, MJM, RAR, WLW, KAS, and NJW designed the study, interpreted the data, and wrote and edited the manuscript. RPO, LLB, JBF, RAL, RAR, WLW, and KAS coordinated and/or participated in field work and the collection of plant material. RPO and LLB collected the DNA sequence data. RPO, MGJ, LLB, SPK, and NJW analyzed the data. All authors read and approved the final manuscript.

#### Funding

This work was funded by grants from the National Science Foundation (USA) (DEB-1342873 to KAS, JBF, and NJW; DEB-1239992 to NJW; DEB-1342792 to RAR; DEB-1342805 to RAL; DBI-1461007 to JBF; DEB-1054539 and DEB-1352907 to MJM). Additional funding was provided by Amherst College, the Negaunee Institute for Plant Conservation Science and Action at the Chicago Botanic Garden, the National Geographic Society, and Oberlin College.

#### Availability of data and materials

All DNA sequence data generated for this project can be accessed at the NCBI Sequence Read Archive (SRA), BioProject PRJNA544074. Individual BioSample accession number can be found in Table S1 (Additional file 1). Sequence alignments, tree files, and analysis of discordance file can be accessed via the Dryad Digital Repository: <https://datadryad.org/stash/share/Um2cZ0ubGDzGAhdXJuXPxzTGy8iL8ccefVfO5yRoLSc>. A supplementary table is presented in Additional file 1 and supplementary figures in Additional file 2.

#### Declarations

##### Ethics approval and consent to participate

Plant collections and land access complied with all relevant institutional, regional, and national guidelines, with appropriate permissions from the Chicago Botanic Garden Herbarium (CHIC), the United States National Herbarium (US) Anza-Borrego State Park, California Botanic Garden, Chugach State Park, Bureau of Land Management (AZ, CA, CO, ID, NM, and UT), Department of Defense (Nevada National Security Site and White Sands Missile Range), Los Angeles County Arboretum, National Park Service (Carlsbad Caverns National Park, Death Valley National Park, Guadalupe Mountains National Park, Mojave National Preserve, and Pinnacles National Park), San Francisco Botanic Garden, Tejon Ranch, and U.S.D.A. Forest Service (Regions 2, 3, 4, and 5).

##### Consent for publication

Not applicable.

##### Competing Interests

The authors declare no competing interests.

Received: 11 March 2022 Accepted: 16 August 2023  
Published online: 17 November 2023

## References

- Wagner WH, Hoch PC, Raven PH. Revised classification of the Onagraceae. *Syst Bot Monogr*. 2007;83:1–240.
- Katinas L, Crisci JV, Wagner WL, Hoch PC. Geographical diversification of tribes Epilobieae, Gongylocarpeae, and Onagreae (Onagraceae) in North America, based on parsimony analysis of endemicity and track compatibility analysis. *Ann Mo Bot Gard*. 2004;91(1):159–85.
- Hoch PC, Raven PH. *Boisduvalia*, a coma-less *Epilobium* (Onagraceae). *Phytologia*. 1992;73:456–9.
- Raven PH. A survey of reproductive biology in Onagraceae. *NJ J Bot*. 1979;17:575–93.
- Berry PE. The systematics and evolution of *Fuchsia* sect. *Fuchsia* (Onagraceae). *Ann Mo Bot Gard*. 1982;69:1–198.
- Sytsma KJ, Smith JF. DNA and morphology: Comparisons in the Onagraceae. *Ann Mo Bot Gard*. 1988;75:1217–37.
- Sytsma KJ, Smith JF, Berry PE. The use of chloroplast DNA to assess biogeography and evolution of morphology, breeding systems, and flavonoids in *Fuchsia* sect. *Skinnera* (Onagraceae). *Syst Bot*. 1991;16:257–69.
- Berry PE, Skvarla JJ, Partridge AD, Macphail MK. *Fuchsia* pollen from the tertiary of Australia. *Aust Syst Bot*. 1990;3:739–44.
- Raven PH. Onagraceae as a model of plant evolution. In: Gottlieb LD, Jain SK, editors. *Plant Evolutionary Biology*. Dordrecht: Springer, Netherlands; 1988. p. 85–107.
- Holsinger KE, Ellstrand NC. The evolution and ecology of permanent translocation heterozygotes. *Am Nat*. 1984;124:48–71.
- Maron JL, Johnson MTJ, Hastings AP, Agrawal AA. Fitness consequences of occasional outcrossing in a functionally asexual plant (*Oenothera biennis*). *Ecology*. 2018;99:464–73.
- Allen GE. Hugo de Vries and the reception of the “mutation theory.” *J Hist Biol*. 1969;2:55–87.
- Golczyk H, Massouh A, Greiner S. Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. *Plant Cell*. 2014;26:1280–93.
- Heiser DA, Shaw RG. The fitness effects of outcrossing in *Calylophus serrulatus*, a permanent translocation heterozygote. *Evolution*. 2006;60:64–76.
- Levy M, Levin DA. Genic heterozygosity and variation in permanent translocation heterozygotes of the *Oenothera biennis* complex. *Genetics*. 1975;79:493–512.
- Raguso RA, Kelber A, Pfaff M, Levin RA, McDade LA. Floral biology of North American *Oenothera* sect. *Lavauxia* (Onagraceae): Advertisements, rewards, and extreme variation in floral depth. *Ann Mo Bot*. 2007;94(1):236–57.
- Johnson MTJ, Smith SD, Rausher MD. Plant sex and the evolution of plant defenses against herbivores. *Proc Natl Acad Sci U S A*. 2009;106:18079–84.
- Dudareva N, Cseke L, Blanc VM, Pichersky E. Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell*. 1996;8:1137–48.
- Jogesh T, Overson RP, Raguso RA, Skogen KA. Herbivory as an important selective force in the evolution of floral traits and pollinator shifts. *AoB Plants*. 2016. <https://doi.org/10.1093/aobpla/plw088>.
- Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Chris Pires J, Zimmer EA, et al. Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. *Am J Bot*. 2003;90:107–15.
- 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;29:147–64.
- Patsis A, Overson RP, Skogen KA, Wickett NJ, Johnson MG, Wagner WL, et al. Elucidating the Evolutionary History of *Oenothera* Sect. *Pachylophus* (Onagraceae): A Phylogenomic Approach. *Syst Bot*. 2021;46:799–811.
- Cooper BJ, Moore MJ, Douglas NA, Wagner WL, Johnson MG, Overson RP, et al. Target enrichment and extensive population sampling help untangle the recent, rapid radiation of *Oenothera* sect. *Calylophus*. *Syst Bot*. 2022. <https://doi.org/10.1093/sysbio/syac032>.
- Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, et al. Data access for the 1,000 Plants (1KP) project. *Gigascience*. 2014;3:17.
- Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018;19(Suppl 6):153.
- Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol*. 2016;33:1654–68.
- Smith SA, Moore MJ, Brown JW, Yang Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol*. 2015;15:150.
- Sayyari E, Whitfield JB, Mirarab S. DiscoVista: Interpretable visualizations of gene tree discordance. *Mol Phylogenet Evol*. 2018;122:110–5.
- Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*. 2009;24:332–40.
- Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol*. 2007;56:17–24.
- Liu L, Edwards SV. Phylogenetic analysis in the anomaly zone. *Syst Biol*. 2009;58:452–60.
- Huang H, Knowles LL. What is the danger of the anomaly zone for empirical phylogenetics? *Syst Biol*. 2009;58:527–36.
- Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genet*. 2006;2:e68.
- Bult CJ, Zimmer EA. Nuclear ribosomal RNA sequences for inferring tribal relationships within Onagraceae. *Syst Bot*. 1993;18:48–63.
- Conti E, Fischbach A, Sytsma KJ. Tribal relationships in Onagraceae: Implications from rbcL sequence data. *Ann Mo Bot Gard*. 1993;80:672–85.
- Ford VS, Gottlieb LD. Tribal relationships within Onagraceae inferred from pgiC sequences. *Syst Bot*. 2007;32:348–56.
- Eyde RH. Reproductive structures and evolution in *Ludwigia* (Onagraceae). III. Vasculature, nectaries, conclusions. *Ann Mo Bot Gard*. 1981;68:379–412.
- Eyde RH. Evolution and systematics of the Onagraceae: Floral anatomy. *Ann Mo Bot Gard*. 1982;69:735–47.
- Hoch PC, Crisci JV, Tobe H, Berry PE. A cladistic analysis of the plant family Onagraceae. *Syst Bot*. 1993;18:31–47.
- Raven PH. The generic subdivision of Onagraceae, tribe Onagreae. *Brittonia*. 1964;16:276–88.
- Crisci JV, Zimmer EA, Hoch PC, Johnson GB, Mudd C, Pan NS. Phylogenetic implications of ribosomal DNA restriction site variation in the plant family. *Ann Mo Bot Gard*. 1990;77:523–38.
- Xie L, Wagner WL, Ree RH, Berry PE, Wen J. Molecular phylogeny, divergence time estimates, and historical biogeography of *Circaea* (Onagraceae) in the Northern Hemisphere. *Mol Phylogenet Evol*. 2009;53:995–1009.
- Martin PG, Dowd JM. Phylogenetic studies using protein sequences within the order Myrtales. *Ann Mo Bot Gard*. 1986;73:442–8.
- Baum DA, Sytsma KJ, Hoch PC. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. *Syst Bot*. 1994;19:363–88.
- Maurin O, Anest A, Bellot S, Biffin E, Brewer G, Charles-Dominique T, et al. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *Am J Bot*. 2021;108:1087–111.
- Tobe H, Wagner WL, Chin H-C. A systematic and evolutionary study of *Oenothera* (Onagraceae): Seed coat anatomy. *Bot Gaz*. 1987;148:235–57.
- Towner HF. The biosystematics of *Calylophus* (Onagraceae). *Ann Mo Bot Gard*. 1977;64:48–120.
- Raguso RA, Pichersky E. Floral volatiles from *Clarkia breweri* and *C. concinna* (Onagraceae): Recent evolution of floral scent and moth pollination. *Osterr bot Z*. 1995;194:55–67.
- Summers HE, Hartwick SM, Raguso RA. Geographic variation in floral allometry suggests repeated transitions between selfing and outcrossing in a mixed mating plant. *Am J Bot*. 2015;102:745–57.
- Wagner WL, Stockhouse RE, Klein WM. The systematics and evolution of the *Oenothera caespitosa* species complex (Onagraceae). in *systematic botany from the ...*. 1985;12.
- Evans MEK, Hearn DJ, Hahn WJ, Spangle JM, Venable DL. Climate and life-history evolution in evening primroses (*Oenothera*, Onagraceae): A phylogenetic comparative analysis. *Evolution*. 2005;59:1914–27.
- Evans MEK, Smith SA, Flynn RS, Donoghue MJ. Climate, niche evolution, and diversification of the “bird-cage” evening primroses (*Oenothera*, sections *Anogra* and *Kleinia*). *Am Nat*. 2009;173:225–40.

55. Lagomarsino LP, Frankel L, Uribe-Convers S, Antonelli A, Muchhala N. Increased resolution in the face of conflict: phylogenomics of the Neotropical bellflowers (Campanulaceae: Lobelioideae), a rapid plant radiation. *Ann Bot*. 2022;129:723–36. <https://doi.org/10.1093/aob/mcac046>.
56. Morales-Briónes DF, Gehrke B, Huang C-H, Liston A, Ma H, Marx HE, et al. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *Alchemilla* s.l. (Rosaceae). *Syst Biol*. 2021;71:190–207.
57. Liu Y-Y, Jin W-T, Wei X-X, Wang X-Q. Phylotranscriptomics reveals the evolutionary history of subtropical East Asian white pines: further insights into gymnosperm diversification. *Mol Phylogenet Evol*. 2022;168:107403.
58. Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, et al. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl Plant Sci*. 2014;2:apps.1400042.
59. Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl Plant Sci*. 2016;4:apps.1600016.
60. Nelson TC, Stathos AM, Vanderpool DD, Finseth FR, Yuan Y-W, Fishman L. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet*. 2021;17:e1009095.
61. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Mol Ecol*. 2016;25:2337–60.
62. Zhang D, Rheindt FE, She H, Cheng Y, Song G, Jia C, et al. Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. *Syst Biol*. 2021. <https://doi.org/10.1093/sysbio/syab024>.
63. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
64. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
65. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
66. Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34:W609–12. <https://doi.org/10.1093/nar/gkl315>.
67. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 2010;26:1569–71.
68. Sukumaran J, Holder MT. SumTrees: Summarization of split support on phylogenetic trees. Version 1.0. 2 Part of the: DendroPy Phylogenetic Computation Library Version 2.0. 3. 2008.
69. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
70. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
71. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35:518–22.
72. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
73. Mirarab S, Warnow T. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*. 2015;31:i44–52.
74. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). New Orleans, LA, USA: IEEE; 2010. p. 2.
75. 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.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

