

RESEARCH ARTICLE

Congruent Deep Relationships in the Grape Family (Vitaceae) Based on Sequences of Chloroplast Genomes and Mitochondrial Genes via Genome Skimming

Ning Zhang, Jun Wen*, Elizabeth A. Zimmer*

Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington, DC, 20013–7012, United States of America

* wenj@si.edu (JW); zimmerl@si.edu (EAZ)



OPEN ACCESS

Citation: Zhang N, Wen J, Zimmer EA (2015) Congruent Deep Relationships in the Grape Family (Vitaceae) Based on Sequences of Chloroplast Genomes and Mitochondrial Genes via Genome Skimming. PLoS ONE 10(12): e0144701. doi:10.1371/journal.pone.0144701

Editor: Hector Candela, Universidad Miguel Hernández de Elche, SPAIN

Received: July 30, 2015

Accepted: November 22, 2015

Published: December 14, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All gene sequence alignments were deposited into Dryad with the DOI 10.5061/dryad.d15v3.

Funding: This study was supported by a Peter Buck Postdoctoral Fellowship from the National Museum of Natural History, Smithsonian Institution, awarded to Ning Zhang, the Laboratories of Analytical Biology of the Smithsonian National Museum of Natural History (NMNH), a grant from the National Science Foundation (DEB 0743474 to S.R. Manchester and J. Wen), a Smithsonian Endowment Grant, and the

Abstract

Vitaceae is well-known for having one of the most economically important fruits, i.e., the grape (*Vitis vinifera*). The deep phylogeny of the grape family was not resolved until a recent phylogenomic analysis of 417 nuclear genes from transcriptome data. However, it has been reported extensively that topologies based on nuclear and organellar genes may be incongruent due to differences in their evolutionary histories. Therefore, it is important to reconstruct a backbone phylogeny of the grape family using plastomes and mitochondrial genes. In this study, next-generation sequencing data sets of 27 species were obtained using genome skimming with total DNAs from silica-gel preserved tissue samples on an Illumina HiSeq 2500 instrument. Plastomes were assembled using the combination of *de novo* and reference genome (of *V. vinifera*) methods. Sixteen mitochondrial genes were also obtained via genome skimming using the reference genome of *V. vinifera*. Extensive phylogenetic analyses were performed using maximum likelihood and Bayesian methods. The topology based on either plastome data or mitochondrial genes is congruent with the one using hundreds of nuclear genes, indicating that the grape family did not exhibit significant reticulation at the deep level. The results showcase the power of genome skimming in capturing extensive phylogenetic data: especially from chloroplast and mitochondrial DNAs.

Introduction

Vitaceae is an economically important plant family, containing the fruit species, the grape (*Vitis vinifera*), that can be eaten as fresh fruit or made into juice, wine, jelly, and raisins. Some other Vitaceae species, such as *Parthenocissus tricuspidata* (Boston ivy) and *P. quinquefolia* (Virginia creeper), are commonly used as ornamentals. About 900 species from 15 genera are documented in the grape family [1]. Except for the temperate species in the genera *Ampelopsis*, *Causonis*, *Nekemias*, *Parthenocissus*, and *Vitis* (~ 135 temperate species in the five genera), most species in the family are distributed in tropical and subtropical regions as woody or herbaceous climbers or rarely shrubs [2–8].

Small Grants Program of the National Museum of Natural History.

Competing Interests: The authors have declared that no competing interests exist.

The molecular phylogenetic analyses of Vitaceae initially were studied using several plastid genes, such as *rbcl*, the *trnL-F* intron and spacer, the *atpB-rbcL* spacer, *rps16*, and the *trnC-petN* spacer [9–14]. The nuclear ribosomal ITS, and the nuclear gene *GAI1* (*GA INSENSITIVE 1*) that encodes one regulator of gibberellins were also used for resolving the Vitaceae phylogeny [10, 15]. Although not completely resolving the deep phylogeny of the Vitaceae, the phylogenetic study with the most comprehensive taxon sampling in the family by Ren et al. [14] identified five major clades in the family, i.e., the *Ampelocissus-Vitis-Nothocissus-Pterisanthes* clade, the *Parthenocissus-Yua* clade, the core *Cissus* clade, the *Cayratia-Cyphostemma-Tetrastigma* clade (hereafter called the CCT clade), and the *Ampelopsis-Rhoicissus-Clematicissus* clade. While taxa of three of the five major clades possess 5-merous flowers, taxa of the *Cissus* clade and the CCT clade have 4-merous flowers. However, because of the uncertain deep relationships, whether the 4-merous flower originated once or multiple times was difficult to determine, especially given the merosity of the outgroup Leeaceae species' being mostly 5, but with 4 also present in some taxa.

Wen et al. [5] reconstructed a deep phylogeny of the grape family using 417 single-copy nuclear genes retrieved from transcriptomes of 15 Vitaceae species and the published genome of the wine grape [16], *V. vinifera*. In this case, a highly supported deep phylogeny was recovered for Vitaceae. In this topology, the *Ampelopsis-Rhoicissus* clade is the earliest divergent lineage and is characterized by having 5-merous flowers. The *Vitis-Ampelocissus* clade and the *Parthenocissus-Yua* clade are sister groups, and they possess 5-merous flowers. The CCT and the *Cissus* clades are sister to each other and taxa of both clades possess 4-merous flowers, suggesting a single origin of 4-merous taxa in the grape family.

It has been commonly reported that the evolutionary history of nuclear genes and that of organellar genes may be different due to incomplete lineage sorting [17] and/or reticulate evolution (hybridization, introgression and allopolyploidy) [18, 19]. Moreover, only 15 species were sampled in the Wen et al. phylogenomic study [5], which might have produced systematic errors into the phylogenetic analysis due to sparse sampling [20]. Since using a few plastid genes had not resolved the deep phylogenetic relationships, we chose to produce a deep phylogeny for Vitaceae using complete chloroplast genomes as well as selected mitochondrial (mt) sequences from 28 species (27 species of Vitaceae plus an outgroup species from Leeaceae). Reference chloroplast and mitochondrial genomes of *V. vinifera* were included in our analyses [21, 22].

With the recent rapid development of next-generation sequencing (NGS) tools, it has become efficient and cost-effective to obtain genomic sequence data and mine markers for plant phylogenetic analyses [23–29]. The genome skimming approach [19, 23, 30, 31] is a rapid and cost-effective strategy for generating phylogenetically informative data via low- to high-density shotgun sequencing of total genomic DNA and was chosen for the study because it generates large phylogenetic data sets and is bioinformatically straight-forward. Genome skimming was developed initially to obtain phylogenetic informative sites from nuclear ribosomal, mitochondrial and chloroplast DNA markers because there are hundreds to thousands of copies in one plant cell and can be easily obtained with a low degree of sequencing coverage [32]. Genome skimming has been applied successfully in various plant groups, for example, in the Sonoran Desert clade of *Asclepias* (Apocynaceae) [30], in the tropical tree family Chrysobalanaceae [33], in the palm family (Arecaceae) [34], and in ferns [35]. In this study, we intend to test: (1) whether the deep phylogeny of the grape family reconstructed using plastomes, the one using mitochondrial genes, and the one using hundreds of nuclear genes reported previously are congruent; and (2) whether ancient hybridization occurred in the deep evolution of the grape family.

Table 1. Sampling design and information on the NGS and plastome data. All voucher specimens were deposited at the United States National Herbarium (US).

Species	Accession number	Data size (Gb)	Number of reads	Coverage of plastome	Plastome size (bp)
<i>Vitis flexuosa</i>	Wen 12461	3.52	23,496,856	1,689	160,964
<i>Vitis riparia</i>	Wen 12695	3.11	20,763,492	1,671	161,011
<i>Vitis vinifera</i>	N/A	N/A	NC_007957	N/A	160,928
<i>Vitis rotundifolia</i> var. <i>munsoniana</i>	Wen 12195	2.12	14,118,242	1,178	161,275
<i>Pterisanthes heterantha</i>	Wen 11820	2.95	19,648,540	773	155,700
<i>Ampelocissus ascendiflora</i>	Wen 11822	2.97	19,825,858	880	155,686
<i>Parthenocissus vitacea</i>	Wen 11976	2.70	17,988,788	679	161,640
<i>Parthenocissus heptaphylla</i>	Wen 11985	2.80	18,672,228	1,133	161,467
<i>Tetrastigma</i> sp. <i>nov.</i>	Wen 2011-fujian	3.12	20,780,600	1,064	156,880
<i>Tetrastigma rafflesiae</i>	Wen 11824	2.57	17,137,048	161	160,089*
<i>Tetrastigma lawsonii</i>	Wen 11680	2.67	17,809,412	602	160,232*
<i>Tetrastigma voinierianum</i>	Wen 12-012	3.01	20,078,184	142	159,986*
<i>Cayratia japonica</i>	Wen UNI68	2.71	18,063,906	829	158,240*
<i>Cyphostemma juttae</i>	Wen 2010-093	2.31	15,375,710	480	159,245
<i>Cyphostemma humile</i>	Wen 2010-090	2.15	14,353,720	136	158,800
<i>Cyphostemma sandersonii</i>	Wen 2010-094	2.17	14,450,758	724	159,196
<i>Cyphostemma adenopoda</i>	Wen UC Davis	2.47	16,487,240	400	159,495
<i>Cissus trifoliata</i>	Wen 11977	2.35	15,641,708	2,006	160,220
<i>Cissus tuberosa</i>	Wen 2010-091	1.93	12,848,572	995	159,928
<i>Cissus microcarpa</i>	Wen 11954	2.80	18,638,058	307	160,390
<i>Cissus discolor</i>	Wen 12-011	3.11	20,737,692	953	158,713
<i>Cissus quadrangularis</i>	2010-086	3.08	20,537,768	2,309	160,382
<i>Cissus antarctica</i>	Wen 2012-010	2.84	18,913,388	595	161,706*
<i>Ampelopsis aconitifolia</i>	Wen 2010-078	3.62	24,125,332	1,883	162,649*
<i>Ampelopsis cordata</i>	Wen 12620	2.93	19,539,876	879	161,955*
<i>Rhoicissus digitata</i>	Wen 2010-097	2.51	16,705,996	681	160,619*
<i>Nekemias arborea</i>	Wen 12005	3.23	21,529,740	713	162,625*
<i>Leea guineensis</i>	Wen 2010-095	2.45	16,356,482	254	160,555*

*: small gaps were not bridged for this plastid genome

doi:10.1371/journal.pone.0144701.t001

Materials and Methods

Ethics Statement

No specific permits were required for the collection of samples as they were all grown in the greenhouse, which complied with all relevant regulations. None of the samples represents endangered or protected species ([Table 1](#)).

Sampling and DNA extraction

To cover the major lineages of the grape family, 27 species were sampled in this study ([Table 1](#)). Total DNA was isolated from ~ 15 mg silica gel dried leaves with the NucleoSpin Plant II DNA extraction kit (Macherey-Nagel, REF 740770.250). Then those DNAs were sent to the Genomic Sequencing and Analysis Facility (GSAF), University of Texas, Austin, for library construction and sequencing. Paired-end reads of 2 x 150 bp for all 27 species were generated in a single lane on an Illumina HiSeq2500 instrument. All raw data have been deposited into the GenBank with the BioProject accession number being PRJNA298058.

Assembly of plastid genomes

The raw data obtained from the GSAF was filtered using Trimmomatic version 0.32 [36] with default settings. The plastome sequence of the wine grape (*V. vinifera*) was downloaded from GenBank (NC_007957) and was used as the reference plastid genome. For each species, the plastome was assembled using both the reference genome and *de novo* methods. The reference-guided assembly was performed using Bowtie2 [37] implemented in the Geneious program package (R8 version) with default settings [38]. The *de novo* assembly was conducted with Velvet [39] implemented in Geneious with the K-mer ranging from 69 to 99. The best K-mer was determined with the Velvet Optimiser implemented in Geneious with the K-mer choice and the coverage optimization being Lcon and Lbp, respectively. To correct errors or ambiguities resulting from either assembly method, the consensus sequence obtained using the reference genome was extracted and then used as the reference sequence for mapping of contigs obtained by *de novo* assembly. Then the sequence of each plastid genome was adjusted manually based on alignment of the consensus sequence and contigs mentioned above.

Assembly of genes of mitochondrial origin in the grape family

Goremykin *et al.* [22] sequenced the mitochondrial genome of the wine grape (GenBank accession number NC_012119). To obtain genes of mitochondrial origin in the other sampled taxa of Vitaceae, we mapped NGS data onto the published grape mitochondrial genome for each species and extracted the sequences of 16 regions (S2 Table). Complete mitochondrial genomes were not produced for the 27 taxa in this study, because they are significantly longer than the chloroplast genomes and evolve with frequent rearrangements and horizontal gene transfers [22] that make alignments and determination of homology difficult.

Genome annotation and phylogenetic analyses

In total, the plastomes of 28 species were aligned using MAFFT [40] implemented in Geneious. With the annotation of the grape plastome available, we could easily predict the coding and noncoding regions for the other 27 species. The alignments of coding and non-coding regions of each gene and those for intergenic regions were then obtained separately.

Because the sequences of plastid tRNA and rRNA are highly conserved, we excluded them from further analyses. In addition, sequences of only one copy of the inverted repeat (IR) region were included for the analyses because gene sequences of the two IR copies are completely or nearly identical. With the exclusions, we have a data set of 79 protein-coding genes (S1 Table), and we concatenated all these genes into one matrix. We also reconstructed the phylogeny of Vitaceae using intergenic regions and introns, excluding those in one of the two IR regions as well. In total, 95 non-coding regions were used for phylogenetic analyses, and they were concatenated into one matrix. We also concatenated sequences of the 79 protein-coding genes and the 95 non-coding regions into one matrix and conducted the phylogenetic analyses.

For genes of mitochondrial origin, we selected and concatenated 16 regions (S2 Table) with a significant number of variable sites across Vitaceae for phylogenetic analyses. All gene sequence alignments were deposited into the Dryad with the DOI being 10.5061/dryad.d15v3.

Maximum likelihood (ML) and Bayesian trees were inferred with RAXML [41] and MrBayes 3.2.1 [42, 43], respectively, with concatenated matrix being partitioned by genes. For ML analysis, the model was specified as GTR + CAT; 100 fast bootstrap ML reps were performed. For Bayesian analysis, two runs with 4 chains were run for 1,000,000 generations, and the model was specified as GTR + G. The MCMC convergence in Bayesian inference was monitored by AWTY (<http://ceb.csit.fsu.edu/awty>). Trees were sampled at every 100 generations. The first

25% of trees were discarded as burn-in with the remaining trees being used for generating the consensus tree.

Results and Discussion

Strategies for assembling plastomes

Overall, we obtained 74 Gb of data generated in a single Illumina lane for the 27 species. The minimum and the maximum size of the NGS data was 1.93 (12,848,572 reads) and 3.62 Gb (24,125,332 reads), for *Cissus tuberosa* and *Ampelopsis aconitifolia*, respectively. The average size of the data set for each of the 27 species was 2.75 Gb (Table 1).

For the three *Vitis* species, we obtained nearly complete plastid genomes with only two small gaps in AT-rich regions, using *V. vinifera* as the reference. We were able to bridge the two gaps in the *Vitis* species by mapping contigs from *de novo* assembly onto the consensus sequence from the reference genome method. However, for species from other genera, there were many gaps and ambiguous sites in the assembled plastomes from the reference genome method. For example, there were 23 gaps in the initial plastome of *Pterisanthes heterantha*, one species from the sister lineage of *Vitis* assembled with the reference method, in spite of 773X mapping coverage (981,188 [number of reads mapping on the grape plastome] * 147 bp [average length of reads] / 186789 bp [length of consensus sequence]). All gaps were bridged by using both reference and *de novo* assembly methods. Except for gaps, ambiguous sites generated using the reference genome assembly method were another source of errors. Nevertheless, in combination with the *de novo* assembly method, these ambiguous sites were all corrected in *Pterisanthes heterantha*. For plastome assembly at the family level, we found it necessary to interactively use both methods.

Another strategy we used for the plastome assembly was a successive reference approach. After obtaining a complete plastome with high quality from one species (for example, *Pterisanthes heterantha*), we used it as the reference genome for assembling the plastome of its most closely related relative based on our taxonomic understanding of the family (i.e., *Ampelocissus ascendiflora*), rather than using that of *V. vinifera*. There were many more poorly mapped regions (16 versus 3) and ambiguous sites (1345 versus 107) in *Ampelocissus ascendiflora*, if *V. vinifera* was used as the reference genome.

After extensive comparative assemblies, we obtained 17 complete plastomes with no gaps and ten plastomes with one to three gaps only (Table 1), all of which were in intergenic regions. These gaps were excluded from further phylogenetic analyses to reduce the impact of missing data.

Resolving the backbone of Vitaceae with chloroplast and mitochondrial data

Plastid genomes represent an important source for characters to be used in plant phylogenetic analyses. To test if the topology using plastomes was congruent with the one using the hundreds of nuclear genes reported by Wen *et al.* [5], we used various analytical strategies to explore our plastome data. We first performed a partitioned analysis of the concatenated data of 126,683 aligned sites of the 28-plastome matrix by excluding one inverted repeat region for phylogenetic reconstruction. The ML and Bayesian topologies from this analysis (Fig 1) were congruent with the one generated from 417 single-copy orthologous nuclear genes with high support values. The topologies differed only in one node concerning the relationships of *Nekemias arborea* and *Rhoicissus digitata*. In the nuclear tree, *Nekemias* was sister to *Rhoicissus* with 100% bootstrap support, whereas *Nekemias* was sister to the clade of *Rhoicissus* and *Ampelopsis*

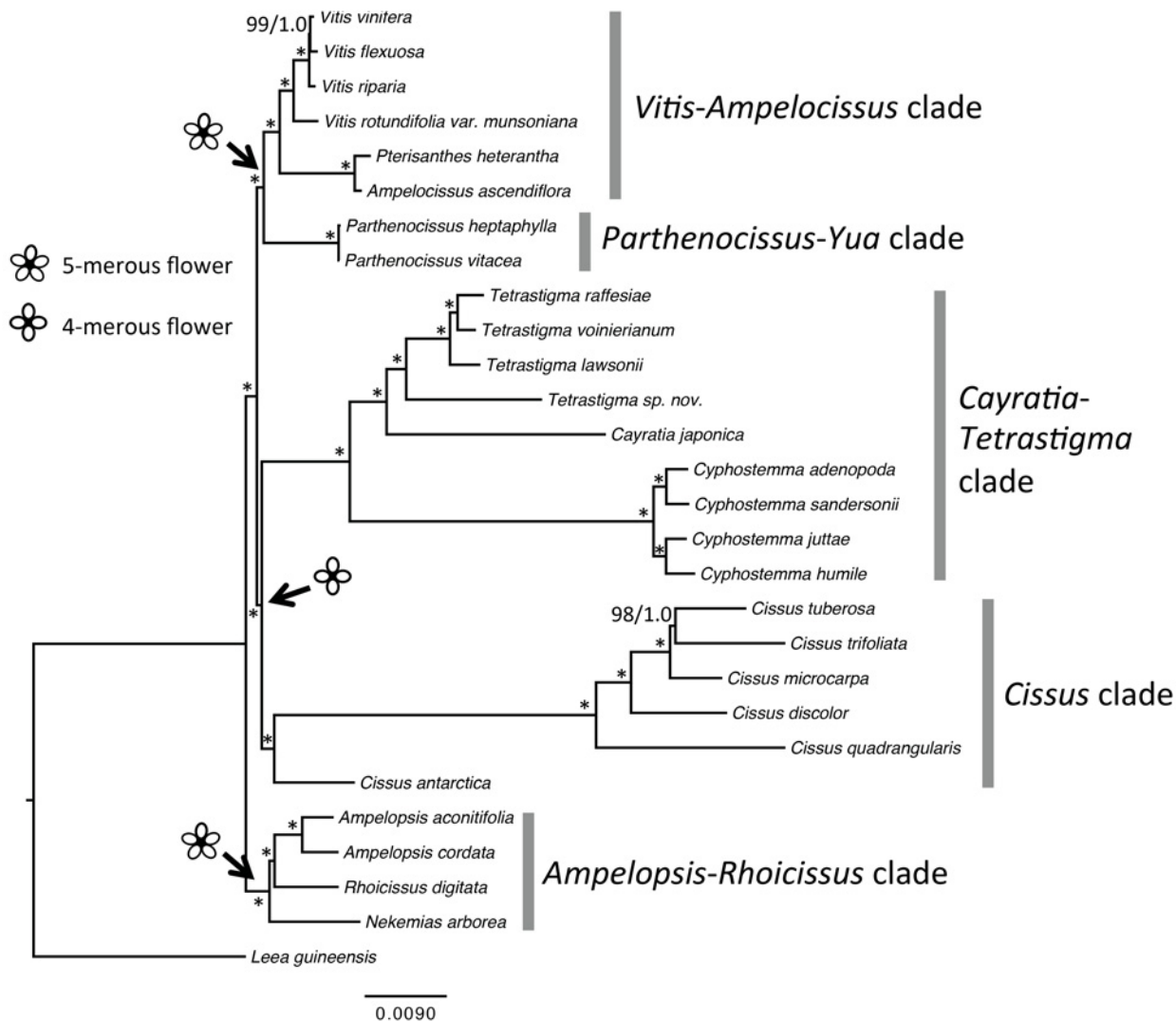


Fig 1. The backbone relationships of the grape family resolved by sequences of complete plastomes. The tree was reconstructed using RaxML and MrBayes with gene partitioning, which resulted in the same topology. Numbers associated with the branches are bootstrap value and posterior probabilities, with the asterisk indicating the node having a bootstrap value of 100% and a posterior probability of 1.0.

doi:10.1371/journal.pone.0144701.g001

s.s in plastome and mitochondrial topologies, with 100% and 98% bootstrap support, respectively (Figs 1 and 2). *Nekemias* was recently segregated from *Ampelopsis* s.l., because the latter was not monophyletic with *Rhoicissus* and the *Cissus striata* lineage nested within [6]. Either the nuclear or the organelle phylogenies still support the recognition of *Nekemias* as a distinct genus from *Ampelopsis*. With the consistent topology in the maternally inherited organelle phylogenies (Figs 1 and 2), the incongruence with the nuclear tree may be due to reticulate evolution, but we need to further explore the causes of such a topological difference.

We also tested the congruence of coding genes and noncoding regions in the plastomes, by analyzing the 79 protein-coding genes (S1 Fig) and the 95 non-coding regions separately (S2 Fig). Analyses from these two data sets generated the same topology with only slight differences in support values at some nodes.

In addition, we explored whether genes with different evolutionary rates may lead to congruent topologies. Evolutionary rates can be generally measured by sequence identity, i.e., the percentage of identical sites in all sites of an alignment, with faster evolving genes having lower identity scores. This strategy resulted in three data sets of 10 genes, 26 genes, 56 genes, corresponding with identity less than 80%, 85%, 90%, respectively (Table 2). Phylogenetic analyses were then performed using these three matrices. With the addition of more genes from low to high gene identity, the topology was almost identical except for the relationships among three *Cissus* species (S1 Fig). The 10-gene topology supported the sister relationship of *C. microcarpa* and *C. tuberosa* (bootstrap value 68), whereas the 26, 56 and 79-gene data sets all supported

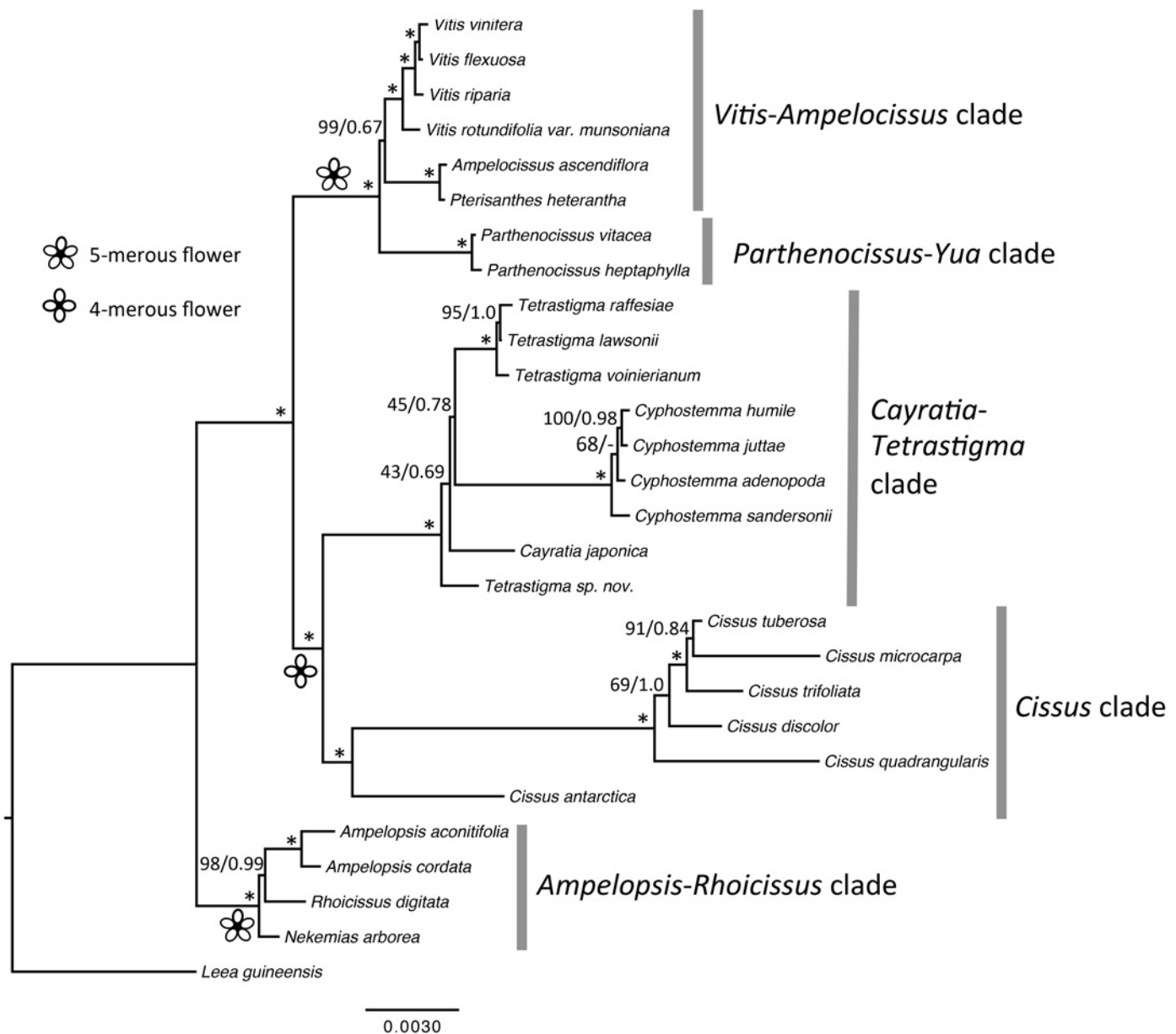


Fig 2. The backbone relationships of the grape family resolved using 16 mitochondrion origin regions. Numbers associated with the branches are bootstrap values and posterior probabilities obtained using RaxML and MrBayes, respectively. The asterisk indicates that the bootstrap value is 100% and the posterior probability is 1.0 at the node.

doi:10.1371/journal.pone.0144701.g002

Table 2. The four data matrices of the chloroplast protein coding genes based on gene identity.

Matrix based on gene identity thresholds	Number of genes	Size of alignment (bp)	Parsimony informative sites
Less than 80%	10	14,062	1,985
Less than 85%	26	25,883	2,880
Less than 90%	56	55,382	4,344
All 79 genes	79	69,800	4,832

doi:10.1371/journal.pone.0144701.t002

the sister relationship between *C. tuberosa* and *C. trifoliata* with bootstrap values of 41%, 23%, and 57%, respectively (S1 Fig).

We used mitochondrial genes as another source of phylogenetic data for Vitaceae. Since the mitochondrial genome of the wine grape is large, i.e., 773,279 bp [22], and 42.4% of chloroplast genome has been incorporated into the mitochondrial genome [22]. Therefore, it is impossible to assemble a complete mitochondrial genome using the genome skimming method because it is difficult to tease chloroplast-origin reads apart from mitochondrion-origin reads. Even though the Illumina sequencing generated 1.93–3.62 Gb data for each of the 27 species, and represented high-density coverage for a mt genome. We therefore mined the mitochondrion-specific genes conservatively to avoid potential error, using the grape mitochondrial genome as the reference. Because mitochondrial genes are relatively conserved, we selected 16 mt regions with a significant number of variable sites among taxa of the family for phylogenetic inference (S1 Table). The backbone topology reconstructed using these 16 mitochondrial regions (18,105 bp) is congruent with the one using the plastome data; however, the positions of *Tetrastigma* and *Cissus* can not be well resolved in the mt tree because only 3% of the total sites of 16 mitochondrial genes are parsimony-informative (Fig 2).

It has been reported extensively that phylogenies using genes from the three genomes (plastid, mitochondrial and nuclear) in plants may differ due to different evolutionary histories [17, 44–47]. Nevertheless, our study overall shows a congruent backbone phylogeny of the grape family using genes from the three compartments. Five major monophyletic lineages, i.e., the *Ampelopsis-Rhoicissus* clade, the *Cissus* clade, the *Cayratia-Cyphostemma-Tetrastigma* clade, the *Parthenocissus-Yua* clade, and the *Vitis-Ampelocissus* clade, were defined (Figs 1 and 2).

The position of *Cissus antarctica*, a species from Australia, has been difficult to resolve in previous studies [4, 10, 11]. It is resolved here using the complete plastomes as well as with the 16 mitochondrial genes. Both trees placed *Cissus antarctica* as the first diverged lineage of the *Cissus* clade. *Cissus antarctica* is closely related to four other *Cissus* species in Australia and New Guinea (*C. hypoglauca*, *C. oblonga*, *C. penninervis*, and *C. sterculiifolia*) and one species from the Neotropics (*C. trianae*) [4, 10, 48]. The position of this clade from our analyses thus supports treating the *Cissus antarctica* clade as part of *Cissus*, the largest genus of the family with about 300 species [1].

The *Cissus* clade and the CCT clade are sisters, and taxa of this large diverse clade is characterized by 4-merous flowers and well-developed thick floral discs [1]. This aspect of the topologies supports a single origin for 4-merous flowers from a 5-merous ancestor.

Genome skimming as a valuable resource for plant phylogenetic analyses

Although most studies using a genome skimming approach produced low-density coverage of the whole genome of a species, large data sets of chloroplast, mitochondrial, rDNA, or even other nuclear genes have been obtained [30, 35, 49, 50]. With the high-density coverage produced in this study, we obtained all chloroplast protein coding genes and even assembled the

complete plastomes as well as sequences of genes of mitochondrial origin using the genome skimming approach. As the sequencing costs using next-gen platforms have decreased rapidly in recent years, high-density genome skimming like this Vitaceae study is becoming affordable to colleagues in the systematics community. We thus advocate the high-density genome skimming approach, when possible, to potentially explore phylogenetic data from all three genomes, especially in cases where reticulation and horizontal gene transfer events may have been suspected. In the next phase of our study, we will continue to mine the nuclear data fraction from this Vitaceae Illumina run, especially for mining faster evolving regions/markers that our earlier transcriptome data largely did not cover, as the RNA-Seq approach produces sequences of only the coding regions of expressed genes [5].

Mitochondrial genes have been largely under-utilized in plant phylogenetic studies. Earlier plastome studies employed methods that isolated plastid DNA from fresh leaves via chloroplast enrichment using sucrose gradients [21], and mitochondrial genes were thus excluded in such studies. Another reason for the under-utilization of mt DNA comparisons is that plant mt DNA tends to evolve by structural rearrangements with slow rates of nucleotide substitution [51]. Furthermore, genome skimming can be tricky, because cp genes can be transferred into mt or vice versa [22, 52], so tests of whether a particular gene is most related to mitochondrial genes of reference species or to its chloroplast one may be useful. So far 295 mitochondrial genomes of 95 plant species have become available in GenBank. With the availability of a broad range of reference genomes, it is possible to mine mitochondrial-specific genes using the most closely related mitochondrial genome as the reference for many plant groups. For our analyses of Vitaceae, the mt genome of *V. vinifera* has been sequenced and assembled [22], which greatly facilitated the mining of genes of actual mitochondrial origin.

Supporting Information

S1 Fig. The phylogenetic relationships of Vitaceae reconstructed using four data sets. Numbers associated with the branches are bootstrap values obtained using 10, 26, 56 and 79 genes, corresponding to sequence identities lower than 80, 85, 90 and 100%, respectively. The asterisk indicates the bootstrap value as 100%. If each of the four bootstrap values from all matrices are 100%, one diamond is placed at the node. The dashes indicate incongruence of a relationship from the one using 79 protein-coding plastid genes (see text for details).

(EPS)

S2 Fig. The phylogenetic relationships of Vitaceae reconstructed using the 95 non-coding plastid regions. Numbers associated with the branches are bootstrap values. The asterisk indicates a bootstrap value of 100%.

(EPS)

S1 Table. The 79 protein-coding plastid genes used in the phylogenetic analyses of Vitaceae.

(DOCX)

S2 Table. The 16 regions of mitochondrial origin used in the phylogenetic analyses of Vitaceae.

(DOCX)

Acknowledgments

This study was supported by a Peter Buck Postdoctoral Fellowship from the National Museum of Natural History, Smithsonian Institution, awarded to Ning Zhang, the Laboratories of

Analytical Biology of the Smithsonian National Museum of Natural History (NMNH), a grant from the National Science Foundation (DEB 0743474 to J. Wen), a Smithsonian Endowment Grant, and the Small Grants Program of the National Museum of Natural History.

Author Contributions

Conceived and designed the experiments: NZ EAZ LW. Performed the experiments: NZ. Analyzed the data: NZ. Contributed reagents/materials/analysis tools: NZ JW. Wrote the paper: NZ EAZ JW.

References

1. Wen J. Vitaceae. In: Kubitzki K, editor. *The Families and Genera of Vascular Plants*. 9. Berlin: Springer; 2007. p. 466–78.
2. Nie Z-L, Sun H, Chen Z-D, Meng Y, Manchester SR, Wen J. (2010) Molecular phylogeny and biogeographic diversification of *Parthenocissus* (Vitaceae) disjunct between Asia and North America. *Am J Bot* 97(8): 1342–53. doi: [10.3732/ajb.1000085](https://doi.org/10.3732/ajb.1000085) PMID: [21616887](https://pubmed.ncbi.nlm.nih.gov/21616887/).
3. Nie Z-L, Sun H, Manchester SR, Meng Y, Luke Q, Wen J. (2012) Evolution of the intercontinental disjunctions in six continents in the *Ampelopsis* clade of the grape family (Vitaceae). *BMC Evol Biol* 12(1): 17. doi: [10.1186/1471-2148-12-17](https://doi.org/10.1186/1471-2148-12-17)
4. Liu X-Q, Ickert-Bond SM, Chen L-Q, Wen J (2013) Molecular phylogeny of *Cissus* L. of Vitaceae (the grape family) and evolution of its pantropical intercontinental disjunctions. *Mol Phylogenet Evol* 66(1): 43–53. doi: [10.1016/j.ympev.2012.09.003](https://doi.org/10.1016/j.ympev.2012.09.003) PMID: [23000818](https://pubmed.ncbi.nlm.nih.gov/23000818/).
5. Wen J, Xiong ZQ, Nie ZL, Mao LK, Zhu YB, Kan XZ, et al. (2013) Transcriptome sequences resolve deep relationships of the grape family. *PLoS One* 8(9): e74394. doi: [10.1371/journal.pone.0074394](https://doi.org/10.1371/journal.pone.0074394) PMID: [24069307](https://pubmed.ncbi.nlm.nih.gov/24069307/).
6. Wen J, Boggan J, Nie Z-L (2014) Synopsis of *Nekemias* Raf., a segregate genus from *Ampelopsis* Michx. (Vitaceae) disjunct between eastern/southeastern Asia and eastern North America, with ten new combinations. *PhytoKeys* (42): 11. doi: [10.3897/phytokeys.42.7704](https://doi.org/10.3897/phytokeys.42.7704) PMID: [25383008](https://pubmed.ncbi.nlm.nih.gov/25383008/).
7. Lu L, Wang W, Chen Z, Wen J (2013) Phylogeny of the non-monophyletic *Cayratia* Juss. (Vitaceae) and implications for character evolution and biogeography. *Mol Phylogenet Evol* 68(3): 502–15. doi: [10.1016/j.ympev.2013.04.023](https://doi.org/10.1016/j.ympev.2013.04.023) PMID: [23669013](https://pubmed.ncbi.nlm.nih.gov/23669013/).
8. Wen J, Lu L, Boggan JK (2013) Diversity and evolution of Vitaceae in the Philippines. *Philipp J Sci* 142: 223–44.
9. Rossetto M, Jackes BR, Scott KD, Henry RJ (2001) Intergeneric relationships in the Australian Vitaceae: new evidence from cpDNA analysis. *Genet Resour Crop Evol* 48(3): 307–14. doi: [10.1023/A:1011225319360](https://doi.org/10.1023/A:1011225319360)
10. Rossetto M, Jackes BR, Scott KD, Henry RJ (2002) Is the genus *Cissus* (Vitaceae) monophyletic? Evidence from plastid and nuclear ribosomal DNA. *Syst Bot* 27(3): 522–33.
11. Rossetto M, Crayn DM, Jackes BR, Porter C (2007) An updated estimate of intergeneric phylogenetic relationships in the Australian Vitaceae. *Can J Bot* 85(8): 722–30. doi: [10.1139/B07-022](https://doi.org/10.1139/B07-022)
12. Ingrouille MJ, Chase MW, Fay MF, Bowman D, Bank MVD, Buijtin AD. (2002) Systematics of Vitaceae from the viewpoint of plastid *rbcL* DNA sequence data. *Bot J Linn Soc* 138(4): 421–32. doi: [10.1046/j.1095-8339.2002.00028.x](https://doi.org/10.1046/j.1095-8339.2002.00028.x)
13. Soejima A, Wen J (2006) Phylogenetic analysis of the grape family (Vitaceae) based on three chloroplast markers. *Am J Bot* 93(2): 278–87. doi: [10.3732/ajb.93.2.278](https://doi.org/10.3732/ajb.93.2.278) PMID: [21646189](https://pubmed.ncbi.nlm.nih.gov/21646189/).
14. Ren H, Lu L-M, Soejima A, Luke Q, Zhang D-X, Chen ZD, et al. (2011) Phylogenetic analysis of the grape family (Vitaceae) based on the noncoding plastid *trnC-petN*, *trnH-psbA*, and *trnL-F* sequences. *Taxon* 60(3): 629–37.
15. Wen J, Nie Z-L, Soejima A, Meng Y (2007) Phylogeny of Vitaceae based on the nuclear *GAI1* gene sequences. *Can J Bot* 85(8): 731–45. doi: [10.1139/B07-071](https://doi.org/10.1139/B07-071)
16. Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463–7. doi: [10.1038/nature06148](https://doi.org/10.1038/nature06148) PMID: [17721507](https://pubmed.ncbi.nlm.nih.gov/17721507/).
17. Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Syst Biol* 55(1): 21–30. doi: [10.1080/10635150500354928](https://doi.org/10.1080/10635150500354928) PMID: [16507521](https://pubmed.ncbi.nlm.nih.gov/16507521/).

18. Yi T-S, Jin G-H, Wen J (2015) Chloroplast capture and intra-and inter-continental biogeographic diversification in the Asian–New World disjunct plant genus *Osmorhiza* (Apiaceae). *Mol Phylogen Evol* 85: 10–21. doi: [10.1016/j.ympev.2014.09.028](https://doi.org/10.1016/j.ympev.2014.09.028) PMID: [25585153](https://pubmed.ncbi.nlm.nih.gov/25585153/).
19. Govindarajulu R, Parks M, Tennesen JA, Liston A, Ashman T-L (2015) Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *Am J Bot* 102(4): 544–54. doi: [10.3732/ajb.1500026](https://doi.org/10.3732/ajb.1500026) PMID: [25878088](https://pubmed.ncbi.nlm.nih.gov/25878088/).
20. Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* 52(1): 124. doi: [10.1080/10635150390132911](https://doi.org/10.1080/10635150390132911) PMID: [12554446](https://pubmed.ncbi.nlm.nih.gov/12554446/).
21. Jansen RK, Kaitanis C, Saski C, Lee S-B, Tomkins J, Alverson AJ, et al. (2006) Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evol Biol* 6(1): 32. doi: [10.1186/1471-2148-6-32](https://doi.org/10.1186/1471-2148-6-32) PMID: [16603088](https://pubmed.ncbi.nlm.nih.gov/16603088/).
22. Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol Biol Evol* 26(1): 99–110. doi: [10.1093/molbev/msn226](https://doi.org/10.1093/molbev/msn226) PMID: [18922764](https://pubmed.ncbi.nlm.nih.gov/18922764/).
23. Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res* 36(19): e122–e. doi: [10.1093/nar/gkn502](https://doi.org/10.1093/nar/gkn502) PMID: [18753151](https://pubmed.ncbi.nlm.nih.gov/18753151/).
24. Zimmer EA, Wen J (2013) Using nuclear gene data for plant phylogenetics: Progress and prospects. *Mol Phylogen Evol* 66(2): 539–50. doi: [10.1016/j.ympev.2013.01.005](https://doi.org/10.1016/j.ympev.2013.01.005) PMID: [23375140](https://pubmed.ncbi.nlm.nih.gov/23375140/).
25. Rothfels CJ, Larsson A, Li F-W, Sigel EM, Huiet L, Burge DO, et al. (2013) Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS One* 8(10): e76957. doi: [10.1371/journal.pone.0076957](https://doi.org/10.1371/journal.pone.0076957) PMID: [24116189](https://pubmed.ncbi.nlm.nih.gov/24116189/).
26. Mandel JR, Dikow RB, Ehrendorfer F, Funk VA (2015) Using phylogenomics to resolve mega-families: an example from Compositae. *J Syst Evol*. doi: [10.1111/jse.12167](https://doi.org/10.1111/jse.12167)
27. Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, Kozik A, et al. (2014) A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Appl Plant Sci* 2(2). doi: [10.3732/apps.1300085](https://doi.org/10.3732/apps.1300085) PMID: [25202605](https://pubmed.ncbi.nlm.nih.gov/25202605/).
28. Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. (2014) Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun* 5.
29. Weitemier K, Straub SC, Cronn RC, Fishbein M, Schmickl R, McDonnell A, et al. (2014) Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl Plant Sci* 2(9). doi: [10.3732/apps.1400042](https://doi.org/10.3732/apps.1400042) PMID: [25225629](https://pubmed.ncbi.nlm.nih.gov/25225629/).
30. Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am J Bot* 99(2): 349–64. doi: [10.3732/ajb.1100335](https://doi.org/10.3732/ajb.1100335) PMID: [22174336](https://pubmed.ncbi.nlm.nih.gov/22174336/).
31. McPherson H, van der Merwe M, Delaney SK, Edwards MA, Henry RJ, McIntosh E, et al. (2013) Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. *BMC Ecol* 13(1): 8. doi: [10.1186/1472-6785-13-8](https://doi.org/10.1186/1472-6785-13-8) PMID: [23497206](https://pubmed.ncbi.nlm.nih.gov/23497206/).
32. Zimmer EA, Wen J (2015) Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *J Syst Evol* 53(5): 371–9. doi: [10.1111/jse.12174](https://doi.org/10.1111/jse.12174)
33. Malé PJG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, et al. (2014) Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol Ecol Resour* 14(5): 966–75. doi: [10.1111/1755-0998.12246](https://doi.org/10.1111/1755-0998.12246) PMID: [24606032](https://pubmed.ncbi.nlm.nih.gov/24606032/).
34. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. (2015) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol*.
35. Wolf PG, Sessa EB, Marchant DB, Li F-W, Rothfels CJ, Sigel EM, et al. (2015) An exploration into fern genome space. *Genome Biol Evol*.
36. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–20. doi: [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170) PMID: [24695404](https://pubmed.ncbi.nlm.nih.gov/24695404/).
37. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3): R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25) PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/).
38. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647–9. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/).

39. Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18(5): 821–9. doi: [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107) PMID: [18349386](https://pubmed.ncbi.nlm.nih.gov/18349386/).
40. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4): 772–80. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/).
41. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688–90. PMID: [16928733](https://pubmed.ncbi.nlm.nih.gov/16928733/)
42. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12): 1572–4. doi: [10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180) PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/).
43. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3): 539–42. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/).
44. Zhang N, Zeng L, Shan H, Ma H (2012) Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol* 195(4): 923–37. doi: [10.1111/j.1469-8137.2012.04212.x](https://doi.org/10.1111/j.1469-8137.2012.04212.x) PMID: [22783877](https://pubmed.ncbi.nlm.nih.gov/22783877/).
45. Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst Bot*: 144–63. doi: [10.1093/sysbio/syu048](https://doi.org/10.1093/sysbio/syu048)
46. Maddison WP (1997) Gene trees in species trees. *Syst Biol* 46(3): 523–36. doi: [10.1093/sysbio/46.3.523](https://doi.org/10.1093/sysbio/46.3.523)
47. Fehrer J, Gemeinholzer B, Chrtek J, Bräutigam S (2007) Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (Hieracium, Cichorieae, Asterales). *Mol Phylog Evol* 42(2): 347–61. doi: [10.1016/j.ympev.2006.07.004](https://doi.org/10.1016/j.ympev.2006.07.004) PMID: [16949310](https://pubmed.ncbi.nlm.nih.gov/16949310/).
48. Wen J, Lu L-M, Nie Z-L, Manchester SR, Ickert-Bond SM, Chen ZD. (2015) Phylogenetics and a revised classified of Vitaceae. *Botany 2015*. Available: <http://www.botanyconference.org/engine/search/index.php?func=detail&aid=669>.
49. Blischak PD, Wenzel AJ, Wolfe AD (2014) Gene prediction and annotation in *Penstemon* (Plantaginaceae): A workflow for marker development from extremely low-coverage genome sequencing. *Appl Plant Sci* 2(12). doi: [10.3732/apps.1400044](https://doi.org/10.3732/apps.1400044) PMID: [25506519](https://pubmed.ncbi.nlm.nih.gov/25506519/).
50. Bock DG, Kane NC, Ebert DP, Rieseberg LH (2014) Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytol* 201(3): 1021–30. doi: [10.1111/nph.12560](https://doi.org/10.1111/nph.12560) PMID: [24245977](https://pubmed.ncbi.nlm.nih.gov/24245977/).
51. Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J Mol Evol* 28(1–2): 87–97. PMID: [3148746](https://pubmed.ncbi.nlm.nih.gov/3148746/).
52. Richardson AO, Palmer JD (2007) Horizontal gene transfer in plants. *J Exp Bot* 58(1): 1–9. doi: [10.1093/jxb/erl148](https://doi.org/10.1093/jxb/erl148) PMID: [17030541](https://pubmed.ncbi.nlm.nih.gov/17030541/).