WILEY-BLACKWELL

Cladistics 34 (2018) 57–77

Cladistics

10.1111/cla.12191

Optimal data partitioning, multispecies coalescent and Bayesian concordance analyses resolve early divergences of the grape family (Vitaceae)

Limin Lu[a], Cymon J. Cox[b], Sarah Mathews[c], Wei Wang[a], Jun Wen[d,*] and Zhiduan Chen[a,*]

[a]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China;* [b]*Centro de Ciências do Mar, Universidade do Algarve, Gambelas, Faro 8005-319, Portugal;* [c]*CSIRO National Research Collections, Australian National Herbarium, Canberra, ACT 2601, Australia;* [d]*Department of Botany, National Museum of Natural History, MRC166, Smithsonian Institution, Washington, DC 20013-7012, USA*

Accepted 4 January 2017

Abstract

Evolutionary rate heterogeneity and rapid radiations are common phenomena in organismal evolution and represent major challenges for reconstructing deep-level phylogenies. Here we detected substantial conflicts in and among data sets as well as uncertainty concerning relationships among lineages of Vitaceae from individual gene trees, supernetworks and tree certainty values. Congruent deep-level relationships of Vitaceae were retrieved by comprehensive comparisons of results from optimal partitioning analyses, multispecies coalescent approaches and the Bayesian concordance method. We found that partitioning schemes selected by PartitionFinder were preferred over those by gene or by codon position, and the unpartitioned model usually performed the worst. For a data set with conflicting signals, however, the unpartitioned model outperformed models that included more partitions, demonstrating some limitations to the effectiveness of concatenation for these data. For a transcriptome data set, fast coalescent methods (STAR and MP-EST) and a Bayesian concordance approach yielded congruent topologies with trees from the concatenated analyses and previous studies. Our results highlight that well-resolved gene trees are critical for the effectiveness of coalescent-based methods. Future efforts to improve the accuracy of phylogenomic analyses should emphasize the development of new methods that can accommodate multiple biological processes and tolerate missing data while remaining computationally tractable.

Introduction

Reconstructing the tree of life of both extant and extinct organisms is one of the primary goals of evolutionary biology (Darwin, 1859; Rokas and Carroll, 2006). The application of molecular data and improvements in algorithmic approaches have provided unprecedented insights into organismal relationships at all taxonomic levels (APG III, 2009; Soltis et al., 2011; Jetz et al., 2012; Telford, 2013). Despite substantial progress in the reconstruction of many parts of the tree of life, accurately determining the relationships of some groups remains challenging and is particularly difficult when evaluating more ancient relationships (Jansen et al., 2006; Soltis et al., 2009; Telford and Copley, 2011; Sun et al., 2015; Zhang et al., 2016). Advances in high-throughput sequencing have enabled researchers to tackle this challenge with hundreds and even thousands of genes (Zou et al., 2008; Moore et al., 2010, 2011; Egan et al., 2012; Jiao et al., 2012; Zimmer and Wen, 2012, 2015; Weitemier et al., 2014). Indeed, many of these studies have focused on resolving seemingly intractable phylogenies among major

*Corresponding authors;
E-mail addresses: wenj@si.edu; zhiduan@ibcas.ac.cn

lineages of life using genomic-scale data (Delsuc et al., 2005; Jian et al., 2008; Zhong et al., 2010; Song et al., 2012; Xi et al., 2012; Liu et al., 2014; Misof et al., 2014; Zeng et al., 2014; Sun et al., 2015). Phylogenetic reconstructions at lower taxonomic levels (e.g. orders, families, genera and species) have also benefited greatly from phylogenomic investigation (e.g. Ma et al., 2014), especially at the species level, where finding enough variation requires sequencing many loci. Species of economic value are important targets for phylogenomic studies as their inter-relationships are significant not only for understanding the biogeography, adaptation and diversification of these groups, but also vital for conserving and exploiting valuable genetic resources (Rossetto et al., 2001; Wan et al., 2013; Wen et al., 2013b).

The use of concatenation methods has greatly advanced our understanding of phylogenetic relationships of many organisms, especially since the 1970s with the widespread analysis of genetic data obtained using the Sanger sequencing technology (Parfrey et al., 2010; Thomson and Shaffer, 2010; Soltis et al., 2013). Data concatenation methods typically assume that all characters track a single underlying tree topology (Kubatko and Degnan, 2007; Pirie, 2015). Increased application of multi-locus and large genome-scale data to phylogenetic reconstruction, however, has called to attention to the often ignored observation that genealogical histories of individual genes may differ considerably from the underlying organismal phylogeny (Maddison, 1997; Pollard et al., 2006; Szöllősi et al., 2015; Zimmer and Wen, 2015). It is often difficult in practice to determine whether systematic biases or biological processes have led to phylogenetic incompatibility within a specific group (e.g. Sanderson et al., 2000; Rokas et al., 2003; Philippe et al., 2005; Burleigh and Mathews, 2007; Lu et al., 2016a; Springer and Gatesy, 2016).

Erroneous phylogenetic reconstruction due to systematic biases may result from failure to adequately model the substitution process and can be exacerbated by insufficient phylogenetic signals and limited taxon sampling (Foster, 2004; Philippe et al., 2005; Rodríguez-Espeleta et al., 2007; Zhong et al., 2011; Cox et al., 2014; Liu et al., 2014). Systematists have attempted to resolve problematic phylogenies by expanding taxon and/or character sampling (Pollock et al., 2002; Zwickl and Hillis, 2002; Heath et al., 2008; Nabhan and Sarkar, 2012) as well as utilizing better-fitting models (Philippe et al., 2005; Telford and Copley, 2011). Schemes dividing the data into partitions that are modelled individually (Brandley et al., 2005; Brown and Lemmon, 2007; Lanfear et al., 2012; Xi et al., 2012) and computational strategies that use mixture models (Rambaut and Grass, 1997; Lartillot and Philippe, 2004; Jayaswal et al., 2014) have been introduced to accommodate variations in substitution rates across sites. These strategies, however, cannot address erroneous phylogenetic reconstructions that result from biological processes including gene duplication and loss, horizontal gene transfer, hybridization, selection and incomplete lineage sorting (ILS) (Maddison, 1997; Maddison and Knowles, 2006; Koonin et al., 2009; Kapralov et al., 2011). ILS is a population-level process that results from the failure of two allelic lineages to coalesce in a population, but rather one of the lineages coalesces with a more distantly related population (Degnan and Rosenberg, 2009). If these populations are separated through multiple speciation events, the "gene" tree of alleles will be incongruent with the species tree (Maddison, 1997; Nichols, 2001). A coalescent-based model (Rannala and Yang, 2003) has been developed to accommodate gene tree heterogeneity resulting from ILS (Carstens and Knowles, 2007; Song et al., 2012; Zhong et al., 2013; Xi et al., 2014). Moreover, hierarchical Bayesian models for the simultaneous computation of gene trees, coalescent trees and the implied species tree have been devised but they suffer from being highly parameterized and often too computationally intensive for even moderately sized data sets (Heled and Drummond, 2010; Zimmermann et al., 2014). As a consequence, several fast (or "short-cut") multispecies coalescent models have been implemented, which use pre-computed gene trees and are statistically consistent given sufficient numbers of input trees (Liu, 2008; Liu et al., 2009a,b, 2010, 2015a,b; Roch and Warnow, 2015; Mirarab et al., 2016). These methods, however, have been criticized not least because they assume the absence of recombination within an individual locus (Gatesy and Springer, 2013, 2014; Pyron et al., 2014; Springer and Gatesy, 2014, 2016). Consequently, the choice of using either coalescent-based species tree estimation methods or the concatenation approach is currently highly controversial especially when gene trees are probably influenced by ILS (Roch and Warnow, 2015). Approaches that make no particular assumption regarding the reason for gene tree discordance, such as the Bayesian concordance approach (BCA) (Ané et al., 2007; Larget et al., 2010), have been developed and merit further comparative studies.

The grape family Vitaceae includes 15 genera and ca. 900 species of perennial climbing plants that are distributed worldwide, primarily in tropical and subtropical regions (Wen, 2007b; Wen et al., 2015). Species of Vitaceae (except for a few species of *Cyphostemma*) are characterized by leaf-opposed tendrils that enable them to climb to the top of the canopy to optimize light interception (Zhang et al., 2015b). The family is well known for *Vitis vinifera*, which is a source of wine, fresh fruits, raisins, and juice, as well as species that are used as ornamentals

and in local medicines (Wen, 2007b; Ren et al., 2011; Gerrath et al., 2015; Torregrosa et al., 2015). Species of Vitaceae are also ecologically important lianas or vines in tropical and temperate forests (Gentry, 1991; Wen et al., 2013a; Wang et al., 2015), and can dominate tropical savannas and landscapes where limestone substrates are prevalent (Wen, 2007b; Lu et al., 2013, 2016b). Studies based on limited chloroplast and nuclear markers have generally supported five major clades in Vitaceae: the *Ampelopsis sensu lato* (*s.l.*) clade, the *Ampelocissus-Vitis* clade, the *Parthenocissus-Yua* clade, the core *Cissus* clade and the *Cayratia-Cyphostemma-Tetrastigma* (CCT) clade (labelled as clades I–V in Fig. 1a; Soejima and Wen, 2006; Wen et al., 2007; Ren et al., 2011; Trias-Blasi et al., 2012). Relationships among these major lineages, however, remain contentious due to incongruence among phylogenetic trees, poor branch resolution or a lack of consistent statistical support across different sampling schemes and phylogenetic methods. We have detected significant heterogeneity among lineages where most of the deep internal branches are extremely short and some terminal branches are very long (Fig. 1a; Ren et al., 2011); and this pattern might have resulted from rapid radiation during the early diversification of Vitaceae and subsequent extinctions along the stem of the major lineages of the family (Fig. 1b; Lu et al., 2013). Wen et al. (2013c) and Zhang et al. (2015a) recently investigated relationships among major clades of Vitaceae by concatenating hundreds of nuclear genes, chloroplast genomes and mitochondrial genes, which generally supported the topology inferred by Ren et al. (2011). Despite this progress, the early divergences of Vitaceae deserves further exploration with comparative strategies of the concatenation and coalescent-based approaches.

We herein conducted a systematic study of Vitaceae using both the concatenated data matrix and fast multispecies coalescent approaches with different taxon and character sampling schemes. The BCA was conducted on just a single data set due to its constraint that the data contain no missing entries. Together

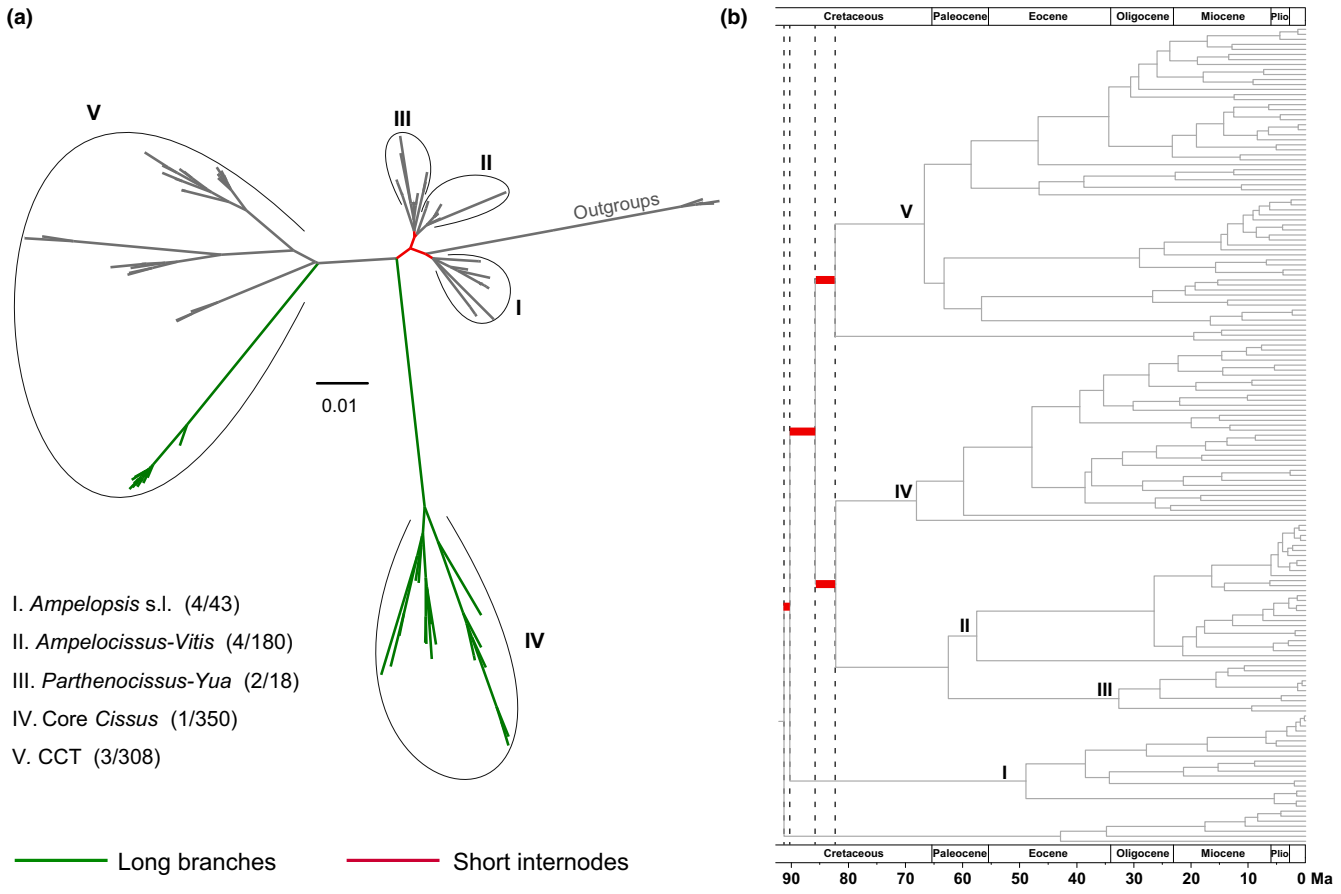

Fig. 1. (a) Phylogram based on Ren et al. (2011), showing significant rate heterogeneity among the five major clades of Vitaceae, with the internodes extremely short (red) and some terminal branches very long (green). (b) Chronogram of Bayesian divergence time estimates from Lu et al. (2013), suggesting rapid diversification of the major lineages in Vitaceae. Major clades are marked I–V and the approximate numbers of genera and species per clade are provided in parentheses.

these molecular data sets were used to conduct a detailed examination of conflicting signals among gene trees, and a comparison of the three phylogenetic methods. The objectives of this study were to: (i) test for phylogenetic incompatibility among Vitaceae gene trees, and (ii) reconstruct early divergences of Vitaceae using optimal partitioning, multispecies coalescent and Bayesian concordance methods.

Materials and methods

Taxon sampling and data collection

Representative taxa of the five major clades of Vitaceae were selected based on the framework of Ren et al. (2011). A sixth major clade including ca. seven species of *Cissus* from Australia, Papua New Guinea and the Neotropics that is distinct from the core *Cissus* was proposed by Liu et al. (2013). In the present study, we designate this Australasian-Neotropical disjunct *Cissus* group as *Cissus* II. The number of species sampled to represent a genus was determined based on the species diversity in the genus. Where genera have been found previously to be non-monophyletic (e.g. *Cissus*, Liu et al., 2013; *Cayratia*, Lu et al., 2013; *Ampelocissus*, Liu et al., 2016), at least one species from each subclade of the genus was sampled. The monotypic *Acareosperma* Gagnep. from Laos is known only from the type specimen (Gagnepain, 1919) and was not available for this study. Species of Leeaceae, the only other family of the order Vitales to which Vitaceae belongs, were included as outgroups (Soejima and Wen, 2006; Wen, 2007a).

Seven data sets (42_11cp, 42_6nu, 42_17loci, 42_14loci, 42_5cp, 362_5cp and 16_229nu) with different taxon and character sampling schemes were used to infer the early divergences of Vitaceae (Table S1). The 42_11cp data set sampled 42 individuals representing 13–15 genera (circumscriptions of some genera have not been formally described yet) and all six major lineages of Vitaceae, and it includes 11 chloroplast regions (listed below). Of the 11 chloroplast regions, six loci (*atp*B-*rbc*L, *rbc*L, *rps*16, *trn*C-*pet*N, *trn*H-*psb*A and *trn*L-*trn*F) have been used in previous phylogenetic studies of Vitaceae (Ingrouille et al., 2002; Soejima and Wen, 2006; Chen et al., 2011; Liu et al., 2013, 2016; Lu et al., 2013, 2016b), while the other five loci (*atp*F-*atp*H, *mat*K, *psb*K-*psb*I, *rpl*16 and *rpo*C1) were used in this study to infer the phylogeny of Vitaceae for the first time. The 42_6nu data set represents the same 42 taxa included in the 42_11cp data set and it includes six nuclear regions (*aro*B, *at*103, *GAI*1, ITS, *phy*A and *sqd*1). Nuclear ribosomal internal transcribed spacer regions (ITS) and *GAI*1 have been used in previous phylogenetic analyses of Vitaceae (Rossetto et al., 2007; Wen et al., 2007; Nie et al., 2010; Liu et al., 2016). Primers for the nuclear *phy*A (305F and 820R) gene were newly designed for this study. The *at*103 and *sqd*1 protein-coding nuclear genes were amplified and sequenced with universal primers designed by Li et al. (2008). The 42_17loci data set was formed by combining the 42_11cp and 42_6nu data sets. The 42_14loci data set was constructed by excluding three conflicting nuclear markers (*aro*B, *at*103 and *GAI*1) from data set 42_17loci. A schematic graphical representation of locus structures and primers used for the 17 markers is provided in Fig. S1. Voucher information and NCBI GenBank accession numbers for the above data sets are listed in Table S2. The 362_5cp data set included 362 taxa for which five commonly used chloroplast markers (*atp*B-*rbc*L, *rps*16, *trn*C-*pet*N, *trn*H-*psb*A and *trn*L-*trn*F) were available. The 42_5cp data set represents a subset of the 42_11cp data set for these same five plastid markers. The data set with denser taxon sampling, 362_5cp, encompassed ca. 40% of the known Vitaceae species. To avoid identification errors in the public database (Chesters and Vogler, 2013), we selected only taxa with sequences from the same voucher. Voucher information and NCBI GenBank accession numbers for the 362_5cp data set are presented in Table S3. Finally, we re-analysed the data set of Wen et al. (2013c), which included 229 orthologous single-copy nuclear genes identified in the transcriptomes of 15 species of Vitaceae and one species of its sister family Leeaceae.

DNA extraction, PCR, sequencing and alignment

Total genomic DNA was extracted from silica gel-dried material or herbarium material using the DNeasy Plant Mini Kit protocol (Qiagen, Crawley, UK). All amplifications were performed in 25-μL reactions containing 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 mM of each primer, 1 U of *Taq* DNA polymerase (Qiagen) and approximately 10–50 ng of the DNA template. The amplification profiles of the plastid regions consisted of a 3 min initial denaturation at 95 °C, 37 cycles of 20 s denaturation at 94 °C, 30 s annealing at 48–50 °C and 40 s extension at 72 °C, followed by a final extension of 5 min at 72 °C. The PCR protocols for the nuclear regions were similar to those for the plastid regions except for the inclusion of a longer annealing time. The primers used to amplify each locus are provided in Table S4. The PCR products were purified using the polyethylene glycol (PEG) precipitation procedure following the protocol of Sambrook et al. (1989). The purified PCR products were sequenced in both directions by standard methods using BigDye 3.1 reagents with an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA) with the primers from the original

amplification. The forward and reverse sequences were assembled and base calls were checked using Geneious 6.1.2 (Biomatters, 2013). During sequence assembly, special attention was paid to those sites with overlapping peaks in the chromatograms, possibly indicating intra-individual variation (polymorphisms). If an obviously overlapping signal was detected in both the forward and the reverse chromatograms, the site was deemed to be putatively polymorphic between alleles or copies. Samples with polymorphic sites were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) following the supplied protocol. Bacterial cells picked from insert-containing colonies were directly used as a template for standard PCR with the M13 forward and reverse primers. Eight clones per individual were selected and sequenced.

Individual loci were aligned with MUSCLE v.3.8.31 (Edgar, 2004), followed by manual adjustment and annotation in Geneious v.6.1.2 (Biomatters, 2013). To maintain the correct reading frame, protein-coding genes were translated to amino acid sequences in Geneious and used to guide the alignment of the nucleotide sequences. Alignments of the 1st, 2nd and 3rd codon positions of the protein-coding genes were exported from SeaView v.4.4.2 (Gouy et al., 2010). Ambiguous alignment segments with many variable positions and/or gaps were excluded from data sets using Gblocks v.0.91b (Castresana, 2000) allowing gap positions = "with half" and leaving other settings as default.

Substitution saturation test

Sites that undergo high substitution rates and exhibit saturation can lead to violations of the assumption of homogeneity in substitution models if mutational biases vary among taxa (e.g. among-lineage compositional heterogeneity). The average of the uncorrected pairwise (p) distances between all taxa for each locus and the proportion of parsimony-informative (PI) characters were calculated in MEGA v.5 (Tamura et al., 2011) to assess the rate heterogeneity of the chloroplast and nuclear genes. Xia's information entropy-based index of substitution saturation for each locus and its subsets (intron, spacer, and the 1st, 2nd, 1st + 2nd and 3rd codon positions for each protein-coding region) was analysed in DAMBE v.5.3.78 (Xia and Xie, 2001; Xia, 2013). Only fully resolved sites were analysed and the proportion of invariant sites (P_{inv}) was estimated prior to the substitution saturation test. We conclude that the sequences have experienced little substitution saturation when the observed saturation index (I_{ss}) value is significantly smaller than the critical I_{ss} value ($I_{ss.cs}$ assuming a symmetrical topology and $I_{ss.ca}$ assuming an asymmetrical topology).

Incongruence analyses

The simplest way to identify incongruence is to compare trees visually (Wiens, 1998), so we performed optimal maximum-likelihood (ML) tree searches and bootstrap (BS) analyses of each locus of each of the combined data sets (42_11cp, 42_6nu, 42_17loci and 16_229nu) using RAxML v.8.0.24 (Stamatakis, 2006, 2014) with 1000 "rapid bootstrapping" (-x parameter) replicates (Stamatakis et al., 2008). To check for consistency in tree topologies, these analyses were repeated using GARLI v.2.1 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl, 2006) available at molecularevolution.org (Bazinet et al., 2014). An adaptive best tree search analysis and 1000 bootstrap replicates with the substitution model selected in MrModeltest v.2.3 (Nylander, 2004) were applied to the analysis of each locus. The optimal ML tree for each locus was visually compared and examined for strongly supported conflicts (BS > 70%; Mason-Gamer and Kellogg, 1996). The relative length of branches and extent of evolutionary rate variation among lineages in these trees were visualized using SplitsTree v.4.13.1 (Huson and Bryant, 2006). To further visualize conflicts among genes, we constructed a supernetwork for the same optimal ML gene trees of each multi-gene data set in SplitsTree. To reduce the risk of overestimating conflicts among single gene trees, only nodes with BS > 70% were preserved. The 70% majority-rule consensus of the 1000 bootstrap trees from each gene analysis was generated in PAUP* v.4.0 b10 (Swofford, 2003). To calculate the supernetwork, we used the Z-closure option and mean edge weights, set the splits transformation as equal angle, and left all other parameters as the default settings.

To quantify the incongruence among phylogenetic trees, we calculated the Internode Certainty (IC), Internode Certainty All (ICA), Tree Certainty (TC) and Tree Certainty All (TCA) (Salichos et al., 2014) in RAxML. The IC and ICA values quantify the specific degree of incongruence for a given branch, while the other two measures, TC (the sums of IC values) and TCA (the sums of ICA values), describe the global degree of incongruence among trees in the data set. The relative TC and TCA scores are normalized by the maximum possible TC and TCA values for a given phylogeny. As these metrics assume no missing gene sequence data entries for any taxon, these analyses could only be performed on the 16_229nu data set.

We measured the statistical significance of character disagreement between the chloroplast and nuclear data sets and among individual nuclear data sets by conducting the hierarchical likelihood ratio tests (hLRTs) as implemented in Concaterpillar v.1.4 (Leigh et al., 2008). Concaterpillar uses hierarchical clustering and LRTs of phylogenies calculated with RAxML v.7.2.8

(due to compatibility issues with newer versions) to detect incongruence among data partitions. All phylogenies were reconstructed in Concatepillar using a GTR + G substitution model. With the initial *P*-value defined as 0.05, Concaterpillar indicates which loci are most appropriately concatenated and those that should be analysed separately by assessing both topological and branch-length congruence among phylogenies reconstructed from combinations of the defined data partitions.

Data partitioning

Choosing an appropriate partitioning scheme and determining an appropriate substitution model for each partition are vital for inference of the correct phylogenetic tree because they can affect both the accuracy of the tree reconstruction and levels of node support (Brown and Lemmon, 2007; Xi et al., 2012). The following partitioning schemes based on gene identity or biochemical and evolutionary constraints were implemented: (i) by genome (two partitions: plastid sequences and nuclear sequences), (ii) by gene (each gene in a separate partition), (iii) by functional group (exon, intron and spacer in each region as a separate partition), (iv) by codon positions (the 1st, 2nd and 3rd codon positions in separate partitions; intron and spacer in each region also treated as individual partitions), and (v) an optimal partitioning scheme selected by PartitionFinder v.1.0.1 (Lanfear et al., 2012, 2014) based on the corrected Akaike information criterion (AICc). The AICc is recommended because it is theoretically more appropriate when the sample size n is small compared to K, the number of estimated parameters ($n/K < 40$; Burnham and Anderson, 2002). Best-fitting models of sequence evolution were chosen for each partition based on AIC as implemented in MrModeltest. For the PartitionFinder analyses, separate partitioning by codon position was defined as the default, and the most fine-grained partition scheme and the greedy heuristic search algorithm with linked branch lengths were used to search for the best-fit partitioning scheme. To avoid low precision from model bias and/or using partitions that were too short, sequences < 20 bp or regions containing no variable sites (no model could be fit to these partitions) were excluded from the partitioning analyses. Thus, the *rbc*L-*acc*D spacer region (< 20 bp) was excluded from the data sets for the partitioning analyses and coding regions were not partitioned by codon when the number of sites in each codon position was < 20, or when the region did not contain variable sites (e.g. the protein-coding region *psb*K was not partitioned by codon because each of the *psb*K 1st, *psb*K 2nd and *psb*K 3rd partitions had only 11 bp; Table S5). In analyses using the ML criterion, the optimal partitioning scheme was determined based on the AIC score, with the partition scheme with the smallest AIC considered the best (Brown and Lemmon, 2007). In the Bayesian framework, the Bayes factor (BF) was evaluated for selecting among competing partitioning strategies based on the ratio of two marginal likelihoods. The marginal likelihood was estimated using MrBayes v.3.2.1 (Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012) with both the harmonic mean (HM) and the stepping-stone sampling (SS) methods. The SS approach uses importance sampling to estimate each marginal likelihood ratio in a series bridging the posterior and prior distributions (Xie et al., 2011). Based on both simulated and empirical data it has been demonstrated that the SS approach is more precise than the commonly used HM method (Fan et al., 2011; Xie et al., 2011; Baele et al., 2012). BFs were compared for all partitioning schemes. If $2\ln BF > 10$, then the difference between the two models was considered to be significant (Kass and Raftery, 1995).

Concatenated phylogenetic reconstruction

Phylogenetic trees were reconstructed using the maximum-parsimony (MP) and ML optimality criteria, and Bayesian Markov chain Monte Carlo (MCMC) inference (BI). The MP analysis was performed with a heuristic search strategy followed by 100 random-stepwise-addition replicates with tree-bisection-reconnection (TBR) branch swapping and Mul-Trees in effect. Indels were treated as missing data. Bootstrap values (BS) were calculated from 1000 bootstrap replicates using heuristic searches as described above (Felsenstein, 1985). ML analysis using GARLI applied an adaptive best tree search analysis and 1000 bootstrap replicates, with the expert mode that allows users to specify their own character sets and model blocks for the partitioned analysis. The best-fitting model for each partition of data sets 42_5cp, 362_5cp, 42_11cp, 42_17loci and 42_14loci (Table 1) was determined either in MrModeltest or selected in PartitionFinder simultaneously with the optimal partitioning schemes. Bayesian inference was conducted with MrBayes with the best-fitting models applied to each data partition. Two independent MCMC chains were run in parallel for each data set for 10 000 000 generations, and sampling every 1000 generations. An average standard deviation of the split frequencies of < 0.01 was assumed to indicate that the two runs had converged to a stationary distribution. Tracer v.1.5 (Rambaut and Drummond, 2009) was used to confirm that the effective sample size (ESS) for all relevant parameters was > 200. After discarding the first 25% trees as burn-in, a 50% majority-rule consensus tree and posterior probabilities (PP) for node support were calculated using the remaining trees from both chains. The

Table 1
Model likelihood values, marginal likelihoods, Bayes factor comparisons and support values for major nodes of Vitaceae under different data partitioning schemes for data sets 42_5cp, 362_5cp, 42_11cp, 42_17 loci and 42_14loci

				GARLI		Marginal likelihood		Bayes factor		Node 1		Node 2		Node 3		Node 4		Node 5		Node 6		Node 7	
Data set	Missing characters	Partitioning strategy	No. of partitions	L	AIC	HM	SS	2ln (BF_{HM})	2ln (BF_{SS})	BP	PP	BP	PP	BP	PP	BP	PP	BP	PP	BP	PP	BP	PP
42_5cp	4847	I: Unpartitioned	1	−17 303.47	34 624.94	−17 380.66	−17 768.44	508	622	40	0.84	59	0.98	24	0.61	27	–	–	–	–	–	–	–
		II: Gene regions	5	−17 177.95	34 445.90	−17 273.68	−17 586.21	294	258	52	0.97	54	0.98	27	0.79	19	–	–	–	–	–	–	–
		III: Functional groups	9	−17 076.76	34 279.52	−17 186.47	−17 528.95	120	143	57	0.99	61	1.00	28	0.80	20	–	–	–	–	–	–	–
	0.50%	IV: Codon groups	11	−17 077.57	34 293.14	−17 204.68	−17 556.61	156	198	57	0.98	58	0.97	28	0.77	20	–	–	–	–	–	–	–
		VI: PF AICc	8	−17 059.64	**34 225.28**	−17 126.44	−17 457.37	**0**	**0**	55	0.99	62	1.00	30	0.80	23	–	–	–	–	–	–	–
362_5cp	5567	I: Unpartitioned	1	−45 428.12	90 874.24	−46 421.52	−49 902.20	7386	9191	68	1.00	54	0.99	–	–	39	–	–	–	–	–	35	0.93
		II: Gene regions	5	−44 990.17	90 070.34	−45 512.39	−48 103.79	5567	5594	74	1.00	55	1.00	–	–	38	0.60	–	–	–	–	39	0.82
		III: Functional groups	9	−44 824.12	89 772.24	−45 358.82	−47 968.01	5260	5323	71	1.00	59	1.00	–	–	36	–	–	–	–	–	43	0.90
	7.80%	IV: Codon groups	11	−44 857.96	89 853.92	−45 334.50	−47 965.56	5212	5318	75	1.00	56	1.00	–	–	35	0.52	–	–	–	–	40	0.90
		VI: PF AICc	8	−44 770.37	**89 672.74**	−42 728.70	−45 306.64	**0**	**0**	66	1.00	66	1.00	–	–	37	0.80	–	–	–	–	44	0.94
42_11cp	10 407	I: Unpartitioned	1	−37 045.49	74 108.98	−37 118.08	−37 555.25	2783	2681	87	1.00	76	1.00	–	–	–	–	–	–	–	–	–	–
		II: Gene regions	11	−36 452.94	73 043.88	−36 576.71	−37 030.50	1700	1631	88	1.00	79	1.00	42	0.92	35	0.73	–	–	–	–	–	–
		III: Functional groups	17	−36 320.31	72 870.62	−36 464.33	−36 951.18	1475	1472	89	1.00	78	1.00	45	0.91	38	0.76	–	–	–	–	–	–
	2.80%	IV: Codon groups	27	−36 039.84	72 415.68	−36 199.64	−36 765.80	946	1102	93	1.00	75	1.00	45	0.81	38	0.61	–	–	–	–	–	–
		VI: PF AICc	17	−36 022.45	**72 290.90**	−35 726.77	−36 214.94	**0**	**0**	90	1.00	73	1.00	38	0.68	37	–	–	–	–	–	–	–
42_17loci	14 690	I: Unpartitioned	1	−65 320.33	130 660.66	−62 091.72	−62 401.52	1460	734	99	1.00	50	–	64	–	98	1.00	–	0.98		0.95	–	–
		II: Genome regions	2	−64 441.35	128 920.70	−63 182.91	−63 520.32	3642	2972	99	1.00	52	0.89	60	0.89	97	1.00	–	–	–	–	–	–
		III: Gene regions	17	−63 137.67	126 541.34	−62 929.09	−63 497.92	3134	2927	100	1.00	–	–	54	–	99	1.00	55	0.94	–	0.91	–	–
	9.50%	IV: Functional groups	31	−62 832.9	126 041.80	−62 447.75	−63 110.84	2172	2153	100	1.00	–	–	48	–	99	1.00	57	1.00	–	1.00	–	–
		V: Codon groups	49	−62 316.55	125 165.10	−62 033.28	−62 832.66	1343	1597	100	1.00	–	–	–	–	98	1.00	62	1.00	56	1.00	–	–
		VII: PF AICc	27	−62 205.51	**124 769.02**	−61 361.86	−62 034.32	**0**	**0**	100	1.00	–	–	–	–	99	1.00	64	0.72	51	0.71	–	–
42_14loci	12 096	I: Unpartitioned	1	−50 346.87	100 711.74	−49 523.74	−50 703.99	4615	5832	92	1.00	89	1.00	46	0.97	30	–	–	–	–	–	–	–
		II: Genome regions	2	−49 294.45	98 626.90	−48 913.72	−49 226.32	3395	2877	97	1.00	93	1.00	–	–	–	–	–	–	–	–	–	–
		III: Gene regions	14	−48 387.51	97 003.02	−48 166.72	−48 671.62	1901	1768	98	1.00	94	1.00	42	0.64	33	0.51	–	–	–	–	–	–
	10.04%	IV: Functional groups	20	−48 255.68	96 777.36	−48 091.61	−48 641.93	1751	1708	99	1.00	93	1.00	–	0.62	–	0.53	–	–	–	–	–	–
		V: Codon groups	34	−47 809.10	96 040.20	−47 678.69	−48 354.28	925	1133	98	1.00	92	1.00	44	–	39	–	–	–	–	–	–	–
		VII: PF AICc	21	−47 803.77	**95 919.54**	−47 216.00	−47 787.77	**0**	**0**	98	1.00	92	1.00	44	–	37	–	–	–	–	–	–	–

PF AICc refers to the optimal partitioning scheme selected in PartitionFinder by the AICc metric. Genome regions: partition the data set into two partitions based on chloroplast and nuclear sequences; Gene regions: each region in a separate partition; Functional group: exon, intron and spacer in each region as a separate partition; Codon groups: the 1st, 2nd and 3rd codon positions in separate partitions; intron and spacer in each region are also treated as individual partitions. Values in bold indicate the optimal partitioning strategies for ML and BI analyses. Nodes 1–6 indicating major divergences correspond to those in Fig. 5 and node 7 can be found in Fig. S6b. "–" indicates nodes that are not supported by either ML or BI inferences.

Bayesian analyses and RAxML analyses were implemented on the CIPRES Science Gateway Portal (Miller et al., 2010) and the GARLI analyses were conducted through the gateway available at molecularevolution.org.

Long-branch attraction detection

Long-branch attraction (LBA) in phylogenetic trees is an artefact caused by the failure to model adequately large amounts of homoplasy. LBA artefacts may be common in phylogenetic reconstructions, but often remain unrecognized. Several methods have been proposed to detect LBA (e.g. Huelsenbeck, 1997; Bergsten, 2005). Because outgroup taxa often connect to ingroups by long branches, LBA caused by the attraction of relatively long-branched ingroup taxa to the outgroups has been recognized as a major source of tree artefacts (Bergsten, 2005). To test whether the outgroup affects the ingroup topology, we performed phylogenetic analyses that included or excluded the outgroups. LBA can also be detected by querying whether long ingroup branches are sufficiently long to be attracted in an MP analysis (Huelsenbeck, 1997, 1998). This test uses a Monte Carlo simulation method whereby 100 data sets are simulated on trees which differ in their placement of the long ingroup branch, using the program Seq-Gen v.1.3.2 (Rambaut and Grass, 1997). All simulations used the GTR + G + I model with fixed parameter values obtained from MrModeltest. The data were simulated on topologies from the conflicting optimal trees of the ML and the strict consensus tree of the MP analyses of the 16_229nu data, and two heuristic tree searches were performed under ML and MP on each replicate data set. Parsimony trees for these data sets were constructed with ten random taxon addition replicates and with one tree being held at each step. ML trees used the same GTR + G parameters as in the original analysis. If the parsimony analyses are significantly misled by LBA, then more than 5% of the MP trees should place the long branches together while the ML analyses should recover the topologies with the long branches separated, based on the data simulated on the ML topology.

Multispecies coalescent species tree reconstruction

Species trees of Vitaceae were inferred based on the data sets 42_17loci (assuming the 11 chloroplast regions to be one linkage group), 42_14loci (excluding *aro*B, *at*103 and *GAI*1 from data set 42_17loci) and 16_229nu. The maximum pseudo-likelihood (MP-EST; Liu et al., 2010) and the average ranks of coalescences (STAR; Liu et al., 2009b) methods were used for both data sets as implemented in the Species Tree Analysis Web (STRAW) server (Shaw et al., 2013). MP-EST and STAR are rapid methods for analysing data sets that involve a large number of genes and a moderate number of species compared with the Bayesian sampling methods such as BEST (Liu, 2008) or *BEAST (Heled and Drummond, 2010). MP-EST estimates species trees from a collection of gene trees by maximizing a pseudo-likelihood function of taxon triplets in the species tree and the results may be more robust to missing data (Liu et al., 2010). STAR uses average ranks of gene coalescence times to build a species tree from a set of gene trees, and thus is resistant to the effects of variable substitution rates across lineages (Liu et al., 2009b). Both methods require rooted input trees, which were obtained from the best-scoring ML trees in RAxML. Bootstrap support for both the MP-EST and the STAR methods was calculated based on 1000 ML bootstrap trees for each locus of each data set. To evaluate the effects of gene tree resolution on species tree inference, two subsets of 16_229nu were used to infer the species tree with MP-EST and STAR independently. One subset (16_75nu) included the 75 nuclear gene matrices that resolved at least one internode with > 70% BS support. The second subset (154nu_16) included the 154 nuclear gene matrices that provided no or poor support for the early-branching topology of the Vitaceae tree.

Bayesian concordance analysis

A BCA was conducted for data set 16_229nu in BUCKy v.1.4.4 (Ané et al., 2007; Larget et al., 2010). For each of the 229 loci, we sampled trees from an MCMC analysis (MrBayes) using a single four-chain 2000 000 generations run, sampled every 1000 generations. After discarding 501 trees from each locus tree set as burn-in, the remaining trees from each locus were summarized by the program mbsum (distributed with BUCKy), and all trees were input into the program BUCKy and analysed using the default settings.

Results

Data characteristics

The five data sets constructed for this study (42_5cp, 362_5cp, 42_11cp, 42_6nu and 42_17loci) contained 515 newly generated sequences (Tables S2 and S3). The total aligned lengths, number of parsimony-informative characters and other characteristics of the data sets 42_17loci and 16_229nu are provided in Tables S5 and S6, respectively. No significant saturation of substitutions was detected for any locus, based on Xia's tests (Tables S5 and S6). The average pairwise distances ranged from 0.017 to 0.052 for data set

42_11cp, from 0.037 to 0.178 for 42_6nu, and from 0.044 to 0.186 for 16_229nu (Fig. S2).

Single gene and unpartitioned data incongruence

A summary tree of the ML analysis of the 11 chloroplast markers (42_11cp) is presented in Fig. 2a and shows strong support for the early-branching of *Ampelopsis s.l.* (89%/88%/1.0; RAxML BS, Garli BS and PP, respectively) and the monophyly of a group formed by *Ampelocissus-Vitis* and *Parthenocissus-Yua* (77%/76%/1.0). In contrast, the well-supported topology based on the nuclear *aro*B (Fig. 2b) was incongruent with those of data set 42_11cp (Fig. 2a), nuclear *GAI*1 (Fig. 2c) and nuclear ribosomal ITS (Fig. 2d) trees with respect to the positions of *Parthenocissus-Yua*, core *Cissus* and *Cissus* II. The ML trees based on *GAI*1 (Fig. 2c) and ITS (Fig. 2d), however, revealed only minor, poorly supported topological conflicts with each other and the chloroplast tree. Network visualization based on the 70% majority-rule consensus trees for each gene of data set 42_17loci also indicates only minor conflict among individual genes concerning the early divergences of Vitaceae (Fig. 2e). Phylogenetic relationships of the major lineages based on ML analysis of the individual nuclear genes *at*103, *phy*A and *sqd*1 were poorly supported (< 70% BS; data not shown). Concaterpillar analyses revealed two statistically distinct groups of nuclear loci in the 42_17loci data set based on topological congruence: (*aro*B, *at*103 and *GAI*1) and (ITS, *phy*A and *sqd*1). Moreover, the branch-length congruence assessment indicated that the loci from each of the two subgroups should not be concatenated. To explore this further, we analysed the data set with and without *aro*B, *at*103 and *GAI*1 (42_17loci and 42_14loci data sets, respectively) and assessed the impact of including these genes on tree topology and node support.

Of the 229 nuclear genes from the transcriptome data, 75 genes supported 28 different topologies with at least one major ingroup node having > 70% BS (Figs S3 and S4). The network visualization of these ML trees highlights well-supported topological conflicts among the major clades of Vitaceae (Fig. 3a). Of the 229 nuclear gene trees, the monophyly of each major clade of Vitaceae is well supported (> 70% BS) by a considerable number of genes (Fig. 3b). The optimal ML tree based on the unpartitioned 229_nu16 data set (Fig. 3c) is congruent with the analysis of chloroplast genes (Fig. 2a), but IC and ICA scores indicate considerable disagreement among genes (Fig. 3c). The TC and TCA values are 8.19 and 8.13, respectively. Moreover, a relative tree certainty (TCA) score of 0.625 suggests a moderate level of overall gene conflict in the tree. In the 75 well-resolved gene trees in Figs S3 and S4, support for the three major internal nodes of Vitaceae is more equivocal, using 70% ML BS as a cut-off value (Fig. 3d).

Phylogenetic analyses of concatenated data

Concatenated analyses. For the 42_5cp data set, all three phylogenetic methods yielded a poorly resolved backbone of Vitaceae, possibly due to insufficient phylogenetic signals (Fig. S5). The MP analysis revealed the CCT clade as the first diverged lineage of Vitaceae with low BS support (Fig. S5a). The ML and BI analyses weakly supported the following relationship: *Ampelopsis s.l.* was the first diverged lineage of Vitaceae; *Parthenocissus-Yua* and *Ampelocissus-Vitis* formed a clade as the second lineage; and core *Cissus*, *Cissus* II and CCT grouped together (Fig. S5b).

For the 362_5cp data set, all three methods supported *Ampelopsis s.l.* as the first diverged lineage and the close relationship between *Parthenocissus-Yua* and *Ampelocissus-Vitis* (Table 1; Fig. S6). The position of the clade of core *Cissus* and *Cissus* II, however, differed in the analyses. The MP analysis resolved core *Cissus* and *Cissus* II as a clade sister to CCT with BS values < 50% (Fig. S6a). The ML and BI analyses recognized the clade of core *Cissus* and *Cissus* II as the second diverged lineage but the support values were low (Table 1; Fig. S6).

The MP analysis of data set 42_11cp resulted in poor resolution of the major lineages of Vitaceae, only moderately supporting *Parthenocissus-Yua* as sister to the *Ampelocissus-Vitis* clade (BS = 76%; Fig. S7a). The ML and BI analyses yielded a fairly resolved backbone of Vitaceae, although support for (*Cissus* + CCT) and (core *Cissus* + *Cissus* II) was relatively low (Fig. S7b,c).

Long-branch attraction. The phylogenetic position of the core *Cissus* clade has varied in previous analyses and is characterized by having a long branch (cf. Wen et al., 2007, 2013c; Ren et al., 2011). For data set 16_229nu, the MP analysis resulted in a topology that strongly conflicted with the trees from the ML and the BI analyses (Fig. 4). Specifically, the model-based methods resolved *Ampelopsis s.l.* as sister to the rest of Vitaceae (Fig. 4a), while the MP analysis supported core *Cissus* diverging first and CCT diverging subsequently (Fig. 4b). This conflict appears to be due to LBA as the unrooted ingroup relationships in both Fig. 4a,b are identical with the incongruence being due to the attachment of the long-branched outgroup to the long core *Cissus* internal branch in the MP tree topology. We tested this using the method of Huelsenbeck (1997). We analysed data sets simulated on the ML topology, which unite core *Cissus* with the CCT clade (Fig. 4a) using both ML and MP. The ML

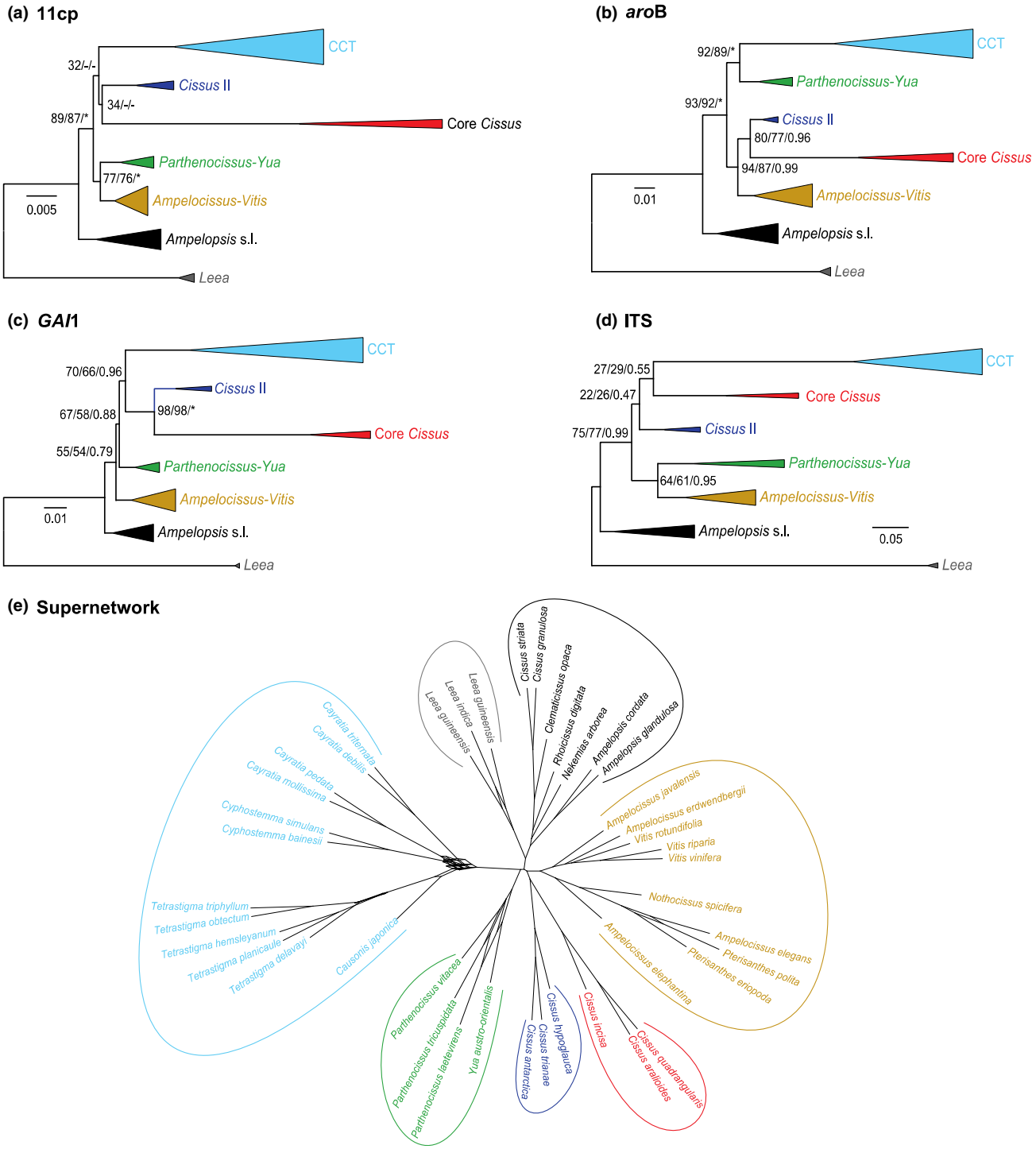

Fig. 2. Relationships among early-diverging lineages of Vitaceae based on the ML and BI analyses of data sets: (a) 42_11cp, (b) *aro*B, (c) *GAI*1 and (d) ITS. Nodal numbers are ML bootstrap values from RAxML and GARLI, and BI posterior probabilities, respectively. "*" represents BS = 100% or PP = 1.00. (e) A supernetwork based on data set 42_17loci (including 11 chloroplast and six nuclear regions of 42 taxa). The network was constructed with SplitsTree and the 70% majority-rule consensus of 1000 bootstrap trees was used as the input tree for each gene. Parallelograms indicate incongruence among gene trees.

analyses always recovered this topology, whereas MP always placed core *Cissus* as sister to the rest of the ingroups as in Fig. 4b. We also analysed data sets simulated on the MP tree (Fig. 4b) and found that both the ML and the MP analyses always recovered core *Cissus* as sister to the rest of the ingroups

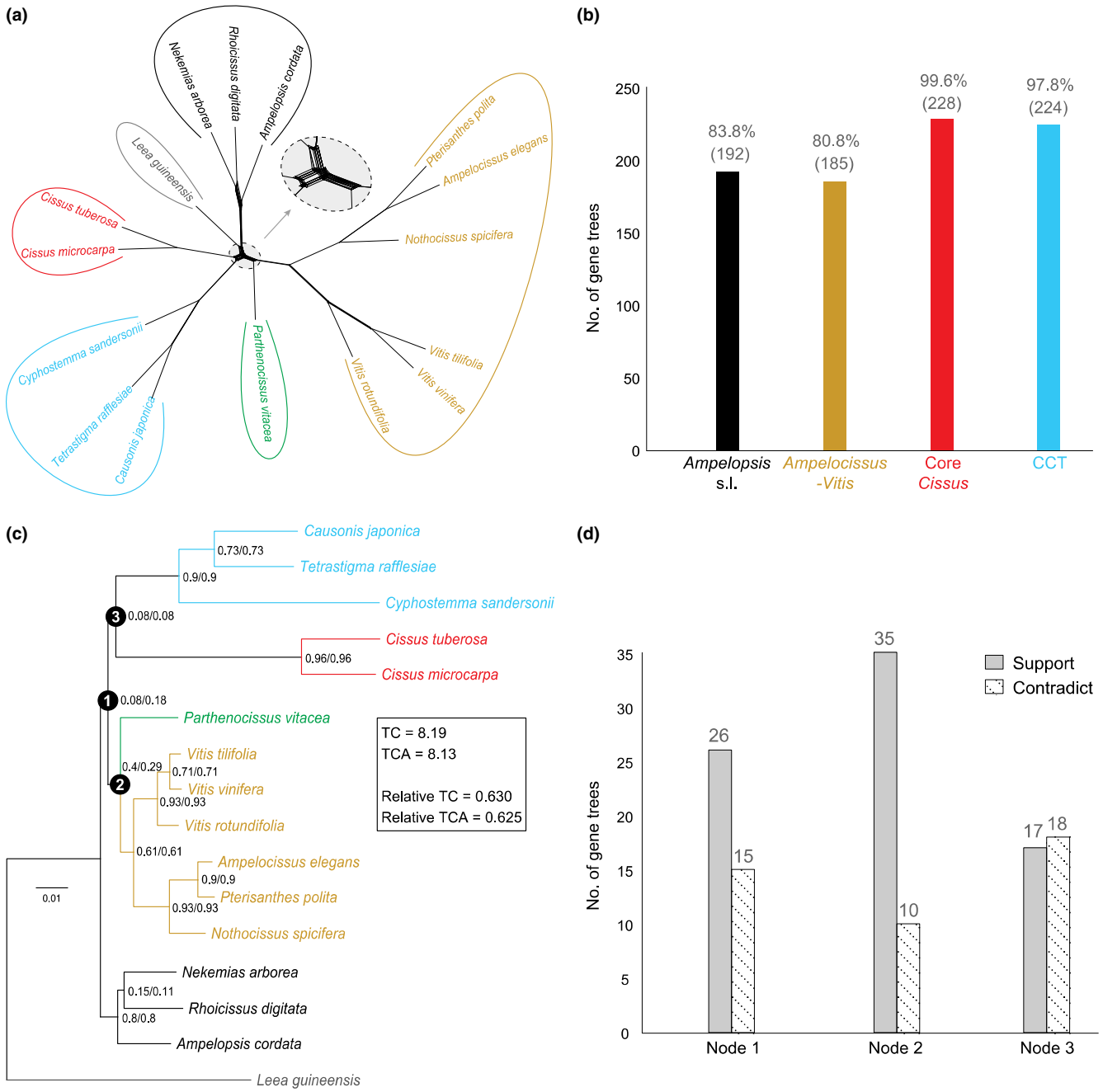

Fig. 3. (a) A supernetwork based on data set 16_229nu. The network was constructed with SplitsTree and the 70% majority-rule consensus of 1000 bootstrap trees was used as the input tree for each gene. Parallelograms indicate incongruence among gene trees. (b) Number of the 229 nuclear gene trees that support the monophyly of the major clades of Vitaceae with > 70% bootstrap support. (c) The optimal ML tree based on data set 16_229nu with the IC/ICA scores indicated for each internode to quantify incongruence among gene trees. TC and TCA values and their relative values are provided on the right-hand side. (d) Of the 229 gene trees of the 16_229nu data set, the number of trees that support and contradict all three major internodes of Vitaceae [as indicated in (c)] with 70% ML bootstrap as cut-off.

(Fig. 4b). Therefore, the discrepancy between the MP and ML trees can be attributed to LBA in the MP analyses. To examine the effect on tree topology and support values, we also excluded core *Cissus* from the 42_11cp data set and re-analysed the data with the MP, ML and BI methods. With core *Cissus* absent from the analysis, the MP analyses retrieved a topology identical to the Bayesian and ML trees, and the bootstrap values for the deep lineages improved considerably (Fig. S7d), again suggesting the MP topology was influenced by LBA between the outgroup and core *Cissus*.

Partitioning analyses. Site and model partitioning analyses were conducted for the three concatenated data sets consisting only of chloroplast data (42_5cp, 362_5cp and 42_11cp), the combined 11 chloroplast gene and six nuclear gene data set (42_17loci), and the combined 11 chloroplast gene and three nuclear gene data set (42_14loci). The optimal substitution models for each gene, codon position and spacer (where present) partition of the data sets 42_11cp and 42_6nu are listed in Table S5. The optimal partitioning schemes for each matrix and the best-fitting substitution model for each partition selected by PartitionFinder under AICc are provided in Table S7. Model likelihoods (using GARLI), marginal likelihoods, BF comparisons between optimal and suboptimal partitioning schemes of the 42_5cp, 362_5cp, 42_11cp, 42_17loci and 42_14loci data sets, and support values for major divergences in Vitaceae are summarized in Table 1. For the ML analyses, partitioning schemes selected by PartitionFinder under AICc were optimal for all the data sets examined (Table 1). Similarly, BFs based on both the HM and the SS methods demonstrated that partitioning schemes selected by PartitionFinder were optimal for all the five data sets (Table 1). In general, the partitioning analyses did not impact the phylogenetic results substantially (i.e. different partitioning schemes may retrieve distinct but not statistically supported topologies) except the Bayesian analyses of data set 42_17loci. The optimal partitioning schemes of three chloroplast data sets all supported *Ampelopsis s.l.* as sister to other Vitaceae and the close relationship between *Parthenocissus-Yua* and *Ampelocissus-Vitis* (Table 1). The phylogenetic positions of core *Cissus* and *Cissus* II were not well resolved by the chloroplast data (Figs 2a, S5b and S6b). Trees from ML analyses of the combined chloroplast and nuclear data (42_17loci) were mostly congruent with the chloroplast-only tree, but the trees from the combined data strongly supported core *Cissus* and *Cissus* II as sisters and weakly supported *Parthenocissus-Yua* as sister to CCT. Bayesian analyses of 42_17loci based on different partitioning schemes, however, resulted in conflicting topologies concerning the position of *Parthenocissus-Yua* (Fig. 5c; Table 1). When excluding the three conflicting nuclear genes from data set 42_17loci (i.e. 42_14loci), a topology congruent with that of the 42_11cp data set was obtained (Fig. S8b).

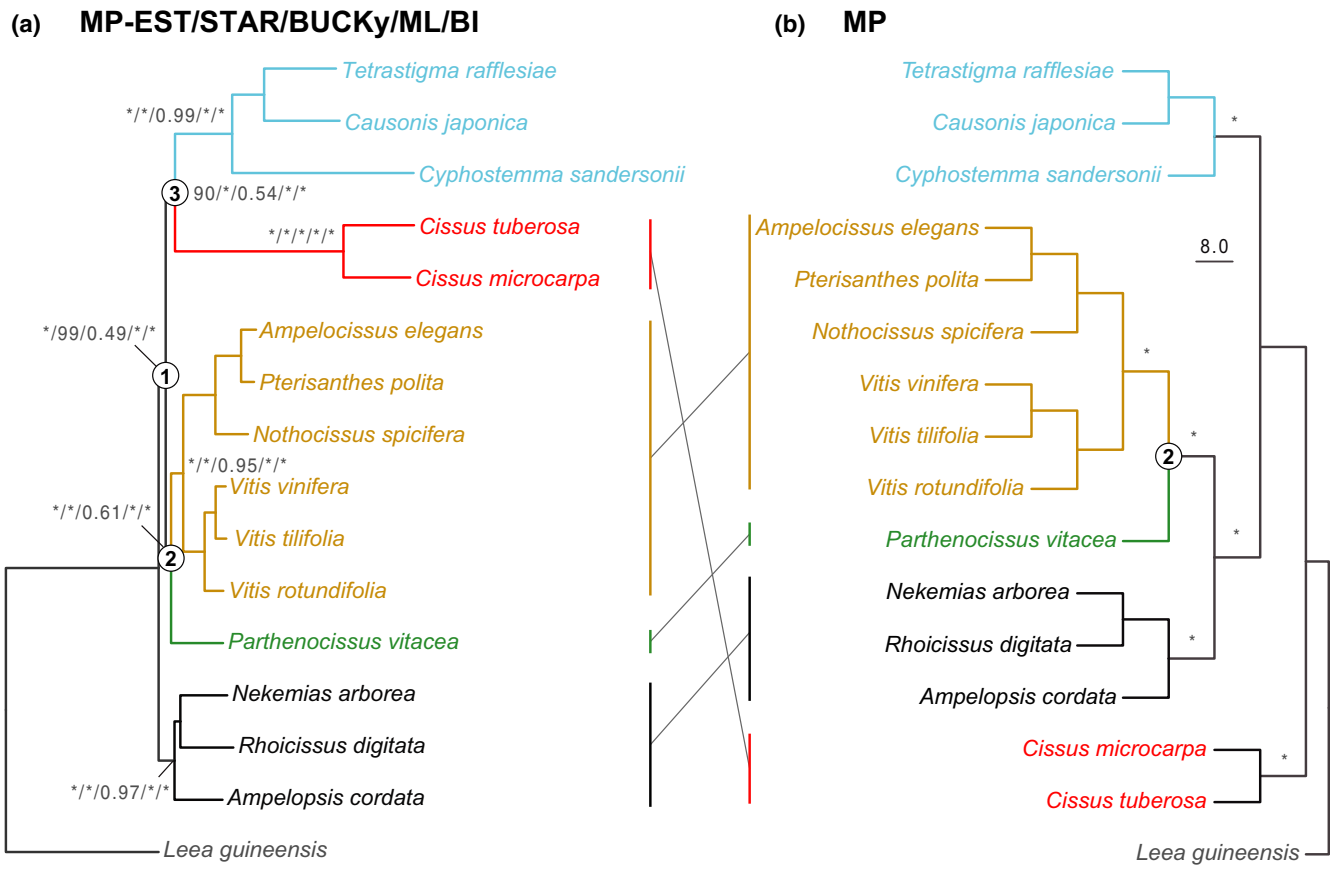

Fig. 4. (a) Topologies inferred from MP-EST, STAR, BUCKy, ML and BI analyses based on data set 16_229nu. Support values are bootstrap values (BP) of MP-EST and STAR, concordance factors of BUCKy, BP for RAxML and posterior probability (PP) of the BI analysis. (b) Topology inferred from the MP analysis with BP values indicated above branches. "*" indicates BS = 100% or PP = 1.00.

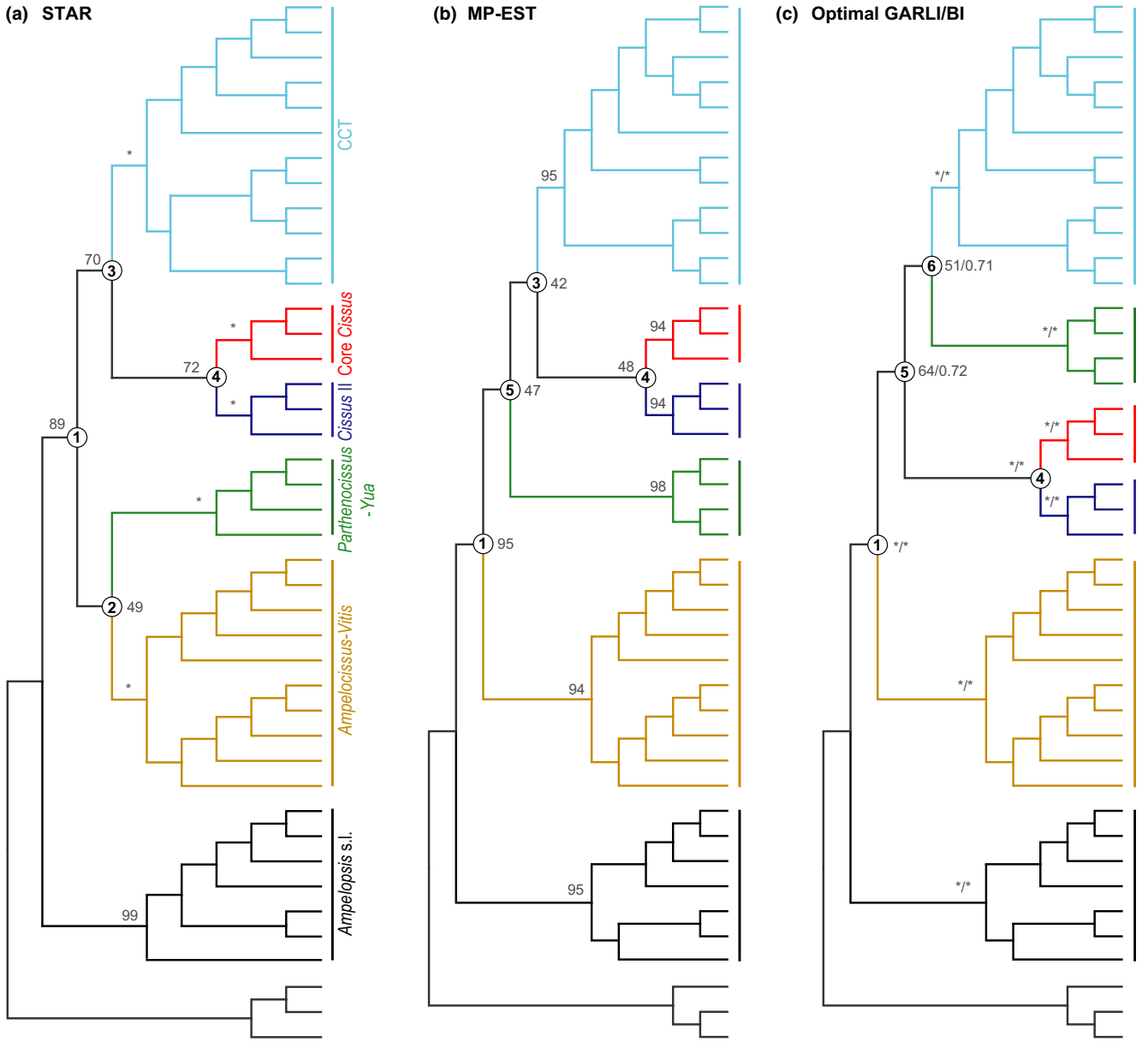

Fig. 5. Topologies inferred from the coalescent analyses using STAR (a) and MP-EST (b) and the optimal partitioning analyses using the ML and BI methods (c) of data set 42_17loci. ML bootstrap values and posterior probability for the major nodes are indicated. "*" represents BS = 100% or PP = 1.00.

Multispecies coalescent and Bayesian concordance analyses

Species trees with bootstrap values estimated by the fast coalescent methods MP-EST and STAR for data sets 16_229nu, 42_17loci and 42_14loci are presented in Figs 4, 5 and S8a, respectively. For data set 16_229nu, the Bayesian concordance tree estimated with BUCKy is congruent with topologies of the fast coalescence methods and the concatenated ML and BI analyses (Fig. 4a). For data set 42_17loci, the fast coalescent analyses by STAR and MP-EST are topologically inconsistent, but the major nodes were not well supported (Fig. 5a,b). Species trees based on STAR and MP-EST for data set 42_14loci were generally consistent except that relationships among the four-petalled taxa (taxa with four petals, including the core *Cissus* clade and the CCT clade) were not well resolved (Fig. S8a). Species trees estimated using MP-EST and STAR for two subdata sets of 16_229nu (16_75nu and 16_154nu) had identical topologies compared with that of the 16_229nu (Fig. 4a). The species

tree based on analyses of 75 nuclear gene trees with at least one supported node yielded similar support values (Fig. S9a) to that of the 16_229nu data set, whereas analyses based on 154 nuclear gene trees that were unresolved resulted in a species tree that was not well supported (see internodes 1 and 3 in Fig. S9b).

Discussion

Insights into early divergences of Vitaceae

Divergence time estimations inferred that Vitaceae have a crown age of ca. 85–95 Ma and the major lineages may have diversified rapidly within a short period of time during the Late Cretaceous (Liu et al., 2013; Lu et al., 2013; Fig. 1b). The difficulty for reconstructing the early divergences of the family might have been caused by an ancient rapid radiation, perhaps coupled with extinction of lineages that existed early in the history of Vitaceae. By comprehensively comparing the results from optimal partitioning, multispecies coalescent and Bayesian concordance analyses, we retrieved early-branching relationships of Vitaceae that are congruent with recent investigations (Wen et al., 2013c; Zhang et al., 2015a) as summarized in Fig. 6: the *Ampelopsis s.l.* clade as sister to the rest of Vitaceae; *Parthenocissus-Yua* and *Ampelocissus-Vitis* as the second diverged lineage; and the four-petalled taxa forming a clade with *Cissus* II and core *Cissus* as a clade sister to CCT. The *Ampelopsis s.l.* clade contains ca. 43 species with an intercontinental disjunction in six continents, including *Ampelopsis*, *Nekemias*, *Rhoicissus*, *Clematicissus* and the *Cissus striata* complex (Nie et al., 2012; Wen et al., 2014). Some previous studies have recognized *Ampelopsis s.l.* as sister to other extant Vitaceae, albeit with only weak support (Ingrouille et al., 2002; Ren et al., 2011). As the second diverged lineage of Vitaceae, *Parthenocissus-Yua* and *Ampelocissus-Vitis* share characteristics of five-petalled flowers, circular to oval seed chalaza and anomocytic stomatal apparatuses (Lu et al., 2012; Manchester et al., 2013), although the synapomorphies of this major clade still need to be rigorously explored morphologically. The four-petalled clade was first recognized by Ren et al. (2011) with moderate support. In addition to sharing four-petalled flowers, members of this clade possess a thick floral disc. *Cissus* is the largest genus of Vitaceae and has been confirmed to be non-monophyletic. Most species of *Cissus* belong to core *Cissus*; the South American *Cissus striata* complex including *C. striata*, *C. simsiana* and *C. tweediana* is nested within *Ampelopsis s.l.* The position of *Cissus* II was uncertain in previous studies (Ingrouille et al., 2002; Rossetto et al., 2002, 2007; Liu et al., 2013; Rodrigues et al., 2014). The present study resolved

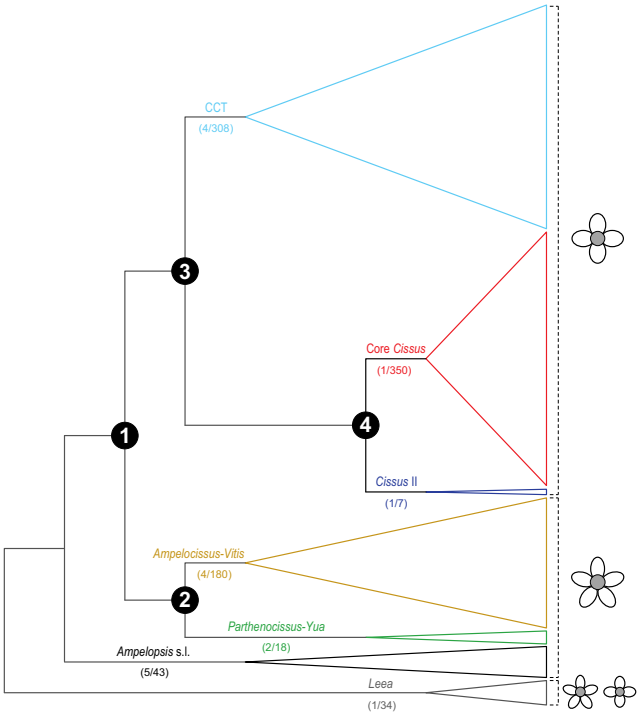

Fig. 6. Phylogenetic backbone of Vitaceae with names of major lineages displayed above branches and the approximate numbers of genera and species per clade indicated below branches. The four- and five-petalled groups are indicated with flower symbols on the right-hand side.

core *Cissus* and *Cissus* II forming a clade sister to CCT, even though the branch lengths indicate that the two clades of *Cissus* are quite divergent (Fig. 6). The gynoecial structure also supports the current phylogeny; highly reduced septa characterize the more derived clades including CCT and core *Cissus*, whereas the septa nearly approach the gynoecial centre in the more basal groups such as *Ampelopsis s.l.* (Ickert-Bond et al., 2014). The only unsampled monotypic genus *Acareosperma* is endemic to Laos and has only been collected once without flowering materials. This genus was established largely based on its distinct seeds with whorled spiny rugae (Gagnepain, 1919; Chen, 2009). *Acareosperma* may be most closely related to *Cayratia* based on their shared pedate leaf architecture, axillary inflorescence and unique seed characteristics (narrow chalaza, polygonal endotestal sclereids, multiple layers of endotestal sclereids and sarcotestal stomata; Wen, 2007b; Chen, 2009).

Previous phylogenomic studies have emphasized the importance of exploring the data for conflicting and potentially misleading phylogenetic signals and applying appropriate models to accommodate the underlying processes attributed to the discordance (e.g. Burleigh and Mathews, 2004; Nabhan and Sarkar, 2012; Townsend et al., 2012; Salichos and Rokas, 2013; Smith et al., 2015; Simmons et al., 2016). Although congruent

topologies for the early-diverged lineages of Vitaceae have been found by concatenating many loci from transcriptome and plastome data, well-supported conflicts were observed between trees from nuclear genes *aro*B and *GAI*1 (Fig. 2b,c), and among the 229 gene trees in the 16_229nu data set (Figs S3 and S4). The supernetwork of nuclear gene trees (Fig. 3a), the IC and ICA scores (Fig. 3c), and the low concordance factors of the BCA (nodes 1–3 in Fig. 4a) provide further evidence that there are topological conflicts concerning the placements of the early-branching lineages of Vitaceae. For instance, nuclear marker *aro*B strongly supports *Parthenocissus-Yua* as sister to the CCT clade, despite no known morphological characters that could be synapomorphies supporting such a relationship. Substitution saturation does not appear to be an explanation for the conflicts because no significant saturation was detected based on Xia's saturation tests (Tables S5). From the distribution of topologies among the 229 nuclear gene trees, the relationships shown in Fig. 6 were widely supported (Fig. S3a) although 27 alternative topologies were also recognized (Figs S3 and S4). We speculate that incongruence among nuclear genes might be caused by ILS or gene duplication and/or loss. These incongruences prompted us to explore strategies to resolve the phylogeny of Vitaceae using models with different assumptions.

Strategies to resolve early divergences of Vitaceae

Various studies have argued that simple data concatenation cannot accommodate the biologically meaningful processes causing phylogenetic incongruence among gene trees (e.g. Edwards et al., 2007; Pirie, 2015). Partitioned analyses are one of the most common approaches to account for substitution heterogeneity among sites without discarding data (DeBry, 2003; Blair and Murphy, 2011; Petkovits et al., 2011; Liu et al., 2012; Xi et al., 2012; Powell et al., 2013). Tools such as the Bayesian mixture model (Pagel and Meade, 2004) and PartitionFinder (Lanfear et al., 2012) have been developed and found to be effective in identifying the optimal partitioning schemes (Xi et al., 2012), but they are often computationally intractable for large phylogenomic data. Another common approach is to use multispecies coalescent models that have been developed for the estimation of accurate species trees in the presence of ILS from phylogenomic and smaller multi-locus data sets (Liu et al., 2009a). A third approach relies on BCAs, which make no particular assumption regarding the underlying causes of gene tree discordance. Considering the potential limitations of each of these approaches, we reconstructed the early divergences of Vitaceae using partitioning analyses of the concatenated data, multispecies coalescent approaches and BCAs.

Our comparative partitioning analyses, using both ML and BI, of five data sets (42_5cp, 362_5cp, 42_11cp, 42_17loci and 42_14loci) indicated that the partitioning schemes selected by PartitionFinder were preferred and that unpartitioned models performed worse for most data sets (Table 1). Phylogenomic studies conducted by Xi et al. (2012) and Liu et al. (2014) using either chloroplast or mitochondrial genes also revealed that the unpartitioned model performed poorly compared with the partitioned ones. For data set 42_17loci, however, which includes three nuclear genes that harbour conflicting signals concerning the placement of the *Parthenocissus-Yua* clade, the unpartitioned model was preferred for the Bayesian inference over the commonly used partitioning schemes based on gene identity and general biochemical or evolutionary constraints (Table 1). Similar results were retrieved by Liu et al. (2012), who found the most partitioned model as the best-fit scheme under the likelihood criterion whereas the unpartitioned model was preferred over more partitioned ones under the Bayesian criterion when there were conflicts among partitions. It remains to be seen if these observations have greater generality. Additionally, it is unclear to what extent the degree of incongruence among gene trees might impact the selection of the best partitioning scheme. When the three conflicting nuclear genes were excluded from the 42_17loci data set, partitioning analyses of the 42_14loci data set retrieved a topology congruent with that of the 42_11cp data set (Fig. S8b). Therefore, the 42_17loci data set may possess incongruence that could not be accommodated by partitioning models.

Biological processes that violate the model assumptions of concatenated methods can also lead to inaccurate reconstructions, even in the face of seemingly sufficient taxon and character sampling. Simulations revealed that concatenated data analysis can produce misleading trees with strong support, especially when there is an "anomaly zone" of nodes and internodes resulting from consecutive rapid speciation events in the species tree (Kubatko and Degnan, 2007; Edwards, 2009; Edwards et al., 2016). Previous studies have indicated that in regions of parameter space where analyses of concatenated data perform poorly, multispecies coalescent models may perform well (Liu et al., 2015a; Xi et al., 2015; Edwards et al., 2016). In the present study, the fast coalescent methods (STAR and MP-EST) for the 229 nuclear gene trees resulted in the same topology as the ML and BI analyses of the concatenated data, and the major nodes of Vitaceae were well supported (Fig. 4a). Furthermore, the species tree based on two subsets of the 16_229nu (16_75nu and 16_154nu) data set retrieved identical topologies (Fig. S9). The 75 resolved nuclear gene trees yielded a species tree with higher support values (node 1: 100/

100, node 3: 90/100; Fig. S9a) than that based on the 154 unresolved nuclear gene trees (node 1: 92/83, node 3: 77/97; Fig. S9b), although the latter was inferred from more than twice the number of loci than the former. This is consistent with the observation of Xi et al. (2015), who emphasized the significance of phylogenetic information in individual genes for coalescent analyses. For data set 42_17loci, the fast coalescent-based analyses using STAR (Fig. 5a) generated a species tree congruent with the 16_229nu data set, whereas the species tree based on MP-EST weakly support *Parthenocissus-Yua* as sister to CCT (node 5 in Fig. 5b). By excluding three conflicting nuclear genes from data set 42_17loci, STAR and MP-EST analyses based on data set 42_14loci retrieved a species tree congruent with that of the 16_229nu data set. These findings highlight concerns surrounding the use of coalescent approaches when gene trees conflict (Simmons et al., 2016).

ILS is considered to be one of the major causes of gene tree conflict (Edwards, 2009), but other biological processes, such as hybridization, gene duplication and/or loss, and selection, may also lead to phylogenetic discordance (Szöllősi et al., 2015). When selection acts independently in non-sister lineages for the same phenotypic traits, it has the potential to cause phylogenetic incongruence at the gene level, although few cases have been documented (e.g. Kapralov and Filatov, 2007; Kapralov et al., 2011). Indeed, because the chloroplast genome acts as a single linkage group, selection acting on the chloroplast could lead to significant incongruence between the species tree and the chloroplast gene trees. Nevertheless, we believe that invoking selection as a general cause for incongruence among genes in the case of Vitaceae is *ad hoc* when contrasting with other (in our view, more likely) mechanisms (such as ILS and hybridization) and requires further specific hypothesis testing. BCA integrates over gene tree uncertainty and makes no assumptions about the underlying biological processes causing gene tree incongruence (Ané et al., 2007; Larget et al., 2010). There is, however, limited application of BCA in phylogenomic studies compared with the multispecies coalescent approaches, possibly due to its strictness on missing data. Our BCA for data set 16_229nu yielded an identical topology to that from the fast coalescent-based approaches (Fig. 4a). The same analyses, however, failed to find much support for the major divergences of Vitaceae, namely *Ampelopsis s.l.* as sister to other Vitaceae (0.49; node 1, Fig. 4a), *Parthenocissus-Yua* plus *Ampelocissus-Vitis* (0.61; node 2 in Fig. 4a) and core *Cissus* plus CCT (0.54; node 3 in Fig. 4a), suggesting that there were strong conflicts among genes which may be masked under high support values of the concatenated analyses (Fig. 4 with 100% support values for all nodes by the concatenated analyses).

Conflicts among and within genomes are ever more easily detected with next-generation sequencing technology efficiently generating large transcriptome and complete plastid or nuclear genomic data for phylogenetic analyses (Szöllősi et al., 2015; Zimmer and Wen, 2015). Additional work is needed to improve models of nucleotide and genome evolution for analyses of genome-scale data sets. Nevertheless, recent advances in complementary strategies for phylogenetic analyses of these data sets have improved our capacity to obtain the most likely species tree of a particular group (Wielstra et al., 2014; Simmons et al., 2016).

Acknowledgements

We thank Fernando Chiang, Yunfeng Deng, Michael Dillon, Jean Gerrath, Deden Girmansyah, Nguyen Hiep, Pete Lowry, Quentin Luke, Esteben Martinez, Michael Nee, Zelong Nie, Leng-Guan Saw, Yumin Shui, Elizabeth Widjaja and Tingshuang Yi for field assistance and/or sample collection. We are grateful to Steven R. Manchester, Yang Liu, Mark Fishbein, Frank E. Anderson and two anonymous reviewers for their constructive comments on earlier versions of the manuscript. This work was supported by grants from the National Natural Science Foundation of China (NNSF 31500179, 31590822, 31270268), the National Basic Research Program of China (2014CB954101), the National Science Foundation (DEB0743474). The Smithsonian Scholarly Studies Grant Program and the Endowment Grant Program, and the John D. and Catherine T. MacArthur Foundation to J.W.; and the CAS/SAFEA International Partnership Program for Creative Research Teams. Laboratory work was supported by the Laboratory of Analytical Biology of the National Museum of Natural History, Smithsonian Institution. Fieldwork was partially supported by the Sino-African Joint Research Center, the Chinese Academy of Sciences and Science and Technology Basic Work (2013FY112100).

References

Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426.

APG III, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161, 105–121.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167.

Bazinet, A.L., Zwickl, D.J., Cummings, M.P., 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63, 812–818.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.
Biomatters, 2013. Geneious 6.1.2 created by Biomatters. Available at: http://www.geneious.com/.
Blair, C., Murphy, R.W., 2011. Recent trends in molecular phylogenetic analysis: where to next? J. Hered. 102, 130–138.
Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.
Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.
Burleigh, J.G., Mathews, S., 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am. J. Bot. 91, 1599–1613.
Burleigh, J.G., Mathews, S., 2007. Assessing among-locus variation in the inference of seed plant phylogeny. Int. J. Plant Sci. 168, 111–124.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Inference: A Practical Information-Theoretical Approach. Springer, New York.
Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst. Biol. 56, 400–411.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Chen, I., 2009. History of Vitaceae inferred from morphology-based phylogeny and the fossil record of seeds. PhD thesis, University of Florida, Gainesville.
Chen, P.T., Chen, L.Q., Wen, J., 2011. The first phylogenetic analysis of *Tetrastigma* (Miq.) Planch., the host of Rafflesiaceae. Taxon 60, 499–512.
Chesters, D., Vogler, A.P., 2013. Resolving ambiguity of species limits and concatenation in multilocus sequence data for the construction of phylogenetic supermatrices. Syst. Biol. 62, 456–466.
Cox, C.J., Li, B., Foster, P.G., Embley, T.M., Civáň, P., 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63, 272–279.
Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.
DeBry, R.W., 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). Syst. Biol. 52, 604–617.
Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.
Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.
Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl Acad. Sci. USA 104, 5936–5941.
Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol., 94(Part A), 447–462.
Egan, A.N., Schlueter, J., Spooner, D.M., 2012. Applications of next-generation sequencing in plant biology. Am. J. Bot. 99, 175–185.
Fan, Y., Wu, R., Chen, M.-H., Kuo, L., Lewis, P.O., 2011. Choosing among partition models in Bayesian phylogenetics. Mol. Biol. Evol. 28, 523–532.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Foster, P.G., 2004. Modeling composition heterogeneity. Syst. Biol. 53, 485–495.
Gagnepain, F., 1919. *Acareosperma,* un genre nouveau d'Ampélidacées. Bull. Mus. Natl. Hist. Nat. 25, 131–132.
Gatesy, J., Springer, M., 2013. Concatenation versus coalescence versus "concatalescence". Proc. Natl Acad. Sci. USA 110, E1179.
Gatesy, J., Springer, M., 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80, 231–266.
Gentry, A.H., 1991. The distribution and evolution of climbing plants. In: Putz, F.E., Mooney, H.A. (Eds.), The Biology of Vines. Cambridge University Press, Cambridge, pp. 3–49.
Gerrath, J., Posluszny, U., Melville, L., 2015. Taming the Wild Grape: Botany and Horticulture in the Vitaceae. Springer, Heidelberg.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46, 239–257.
Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.
Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69–74.
Huelsenbeck, J.P., 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? Syst. Biol. 47, 519–537.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Ickert-Bond, S.M., Gerrath, J., Wen, J., 2014. Gynoecial structure of Vitales and implications for the evolution of placentation in the rosids. Int. J. Plant Sci. 175, 998–1032.
Ingrouille, M.J., Chase, M.W., Fay, M.F., Bowman, D., van der Bank, M., Bruijn, A.D.E., 2002. Systematics of Vitaceae from the viewpoint of plastid *rbcL* DNA sequence data. Bot. J. Linn. Soc. 138, 421–432.
Jansen, R., Kaittanis, C., Saski, C., Lee, S.-B., Tomkins, J., Alverson, A., Daniell, H., 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, 32.
Jayaswal, V., Wong, T.K.F., Robinson, J., Poladian, L., Jermiin, L.S., 2014. Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. Syst. Biol. 63, 726–742.
Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. Nature 491, 444–448.
Jian, S., Soltis, P.S., Gitzendanner, M.A., Moore, M.J., Li, R.Q., Hendry, T.A., Qiu, Y.L., Dhingra, A., Bell, C.D., Soltis, D.E., 2008. Resolving an ancient, rapid radiation in Saxifragales. Syst. Biol. 57, 38–57.
Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J., McKain, M., McNeal, J., Rolf, M., Ruzicka, D., Wafula, E., Wickett, N., Wu, X., Zhang, Y., Wang, J., Zhang, Y., Carpenter, E., Deyholos, M., Kutchan, T., Chanderbali, A., Soltis, P., Stevenson, D., McCombie, R., Pires, J., Wong, G., Soltis, D., dePamphilis, C., 2012. A genome triplication associated with early diversification of the core eudicots. Genome Biol. 13, R3.
Kapralov, M.V., Filatov, D.A., 2007. Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evol. Biol. 7, 73.
Kapralov, M.V., Kubien, D.S., Andersson, I., Filatov, D.A., 2011. Changes in Rubisco kinetics during the evolution of C4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. Mol. Biol. Evol. 28, 1491–1503.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Koonin, E.V., Wolf, Y.I., Puigbò, P., 2009. The phylogenetic forest and the quest for the elusive tree of life. Cold Spring Harb. Symp. Quant. Biol. 74, 205–213.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.
Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82.
Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 22, 2910–2911.
Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.
Leigh, J.W., Susko, E., Baumgartner, M., Roger, A.J., 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57, 104–115.
Li, M., Wunder, J., Bissoli, G., Scarponi, E., Gazzani, S., Barbaro, E., Saedler, H., Varotto, C., 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. Cladistics 24, 727–745.
Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.
Liu, L., Yu, L., Kubatko, L., Pearl, D.K., Edwards, S.V., 2009a. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53, 320–328.
Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009b. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.
Liu, L., Yu, L., Edwards, S., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10, 302.
Liu, Y., Budke, J.M., Goffinet, B., 2012. Phylogenetic inference rejects sporophyte based classification of the Funariaceae (Bryophyta): rapid radiation suggests rampant homoplasy in sporophyte evolution. Mol. Phylogenet. Evol. 62, 130–145.
Liu, X.Q., Ickert-Bond, S.M., 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, 43–53.
Liu, Y., Cox, C.J., Wang, W., Goffinet, B., 2014. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. Syst. Biol. 63, 862–878.
Liu, L., Wu, S., Yu, L., 2015a. Coalescent methods for estimating species trees from phylogenomic data. J. Syst. Evol. 53, 380–390.
Liu, L., Xi, Z., Davis, C.C., 2015b. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32, 791–805.
Liu, X.Q., Ickert-Bond, S.M., Nie, Z.L., Zhou, Z., Chen, L.Q., Wen, J., 2016. Phylogeny of the *Ampelocissus-Vitis* clade in Vitaceae supports the New World origin of the grape genus. Mol. Phylogenet. Evol. 95, 217–228.
Lu, L.M., Wen, J., Chen, Z.D., 2012. A combined morphological and molecular phylogenetic analysis of *Parthenocissus* (Vitaceae) and taxonomic implications. Bot. J. Linn. Soc. 168, 43–63.
Lu, L.M., Wang, W., Chen, Z.D., Wen, J., 2013. Phylogeny of the non-monophyletic *Cayratia* Juss. (Vitaceae) and implications for character evolution and biogeography. Mol. Phylogenet. Evol. 68, 502–515.
Lu, L.M., Chen, Z.D., Lu, A.M., 2016a. Will there ever be a tree of life that systematists can agree on? Chin. Sci. Bull. 61, 958–963.
Lu, L.M., Wen, J., Chen, Z.D., 2016b. *Cayratia cheniana* (Vitaceae): an endangered new species endemic to the limestone mountains of Ninh Thuan province, Vietnam. Syst. Bot. 41, 49–55.
Ma, P.F., Zhang, Y.X., Zeng, C.X., Guo, Z.H., Li, D.Z., 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). Syst. Biol. 63, 933–950.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.
Manchester, S.R., Kapgate, D.K., Wen, J., 2013. Oldest fruits of the grape family (Vitaceae) from the Late Cretaceous Deccan Cherts of India. Am. J. Bot. 100, 1849–1859.
Mason-Gamer, R.J., Kellogg, E.A., 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. 45, 524–545.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orlens, LA, pp. 1–8.
Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380.
Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.
Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl Acad. Sci. USA 107, 4623–4628.
Moore, M.J., Hassan, N., Gitzendanner, M.A., Bruenn, R.A., Croley, M., Vandeventer, A., Horn, J.W., Dhingra, A., Brockington, S.F., Latvis, M., Ramdial, J., Alexandre, R., Piedrahita, A., Xi, Z., Davis, C.C., Soltis, P.S., Soltis, D.E., 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. Int. J. Plant Sci. 172, 541–558.
Nabhan, A.R., Sarkar, I.N., 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Brief. Bioinform. 13, 122–134.
Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358–364.
Nie, Z.L., Sun, H., Chen, Z.D., Meng, Y., Manchester, S.R., Wen, J., 2010. Molecular phylogeny and biogeographic diversification of *Parthenocissus* (Vitaceae) disjunct between Asia and North America. Am. J. Bot. 97, 1342–1353.
Nie, Z.L., Sun, H., Manchester, S.R., 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.
Nylander, J.A.A., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University, Uppsala.
Pagel, M., Meade, A., 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 571–581.

Parfrey, L.W., Grant, J., Tekle, Y.I., Lasek-Nesselquist, E., Morrision, H.G., Sogin, M.L., Patterson, D.J., Katz, L.A., 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Syst. Biol. 59, 518–533.

Petkovits, T., Nagy, L.G., Hoffmann, K., Wagner, L., Nyilasi, I., Griebel, T., Schnabelrauch, D., Vogel, H., Voigt, K., Vágvölgyi, C., Papp, T., 2011. Data partitions, Bayesian analysis and phylogeny of the zygomycetous fungal family Mortierellaceae, inferred from nuclear ribosomal DNA sequences. PLoS ONE 6, e27507.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., Delsuc, F., 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol. Biol. 5, 50.

Pirie, M.D., 2015. Phylogenies from concatenated data: is the end nigh? Taxon 64, 421–423.

Pollard, D.A., Iyer, V.N., Moses, A.M., Eisen, M.B., 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2, e173.

Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51, 664–671.

Powell, A.F.L.A., Barker, F.K., Lanyon, S.M., 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. Mol. Phylogenet. Evol. 66, 69–79.

Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R., Burbrink, F.T., 2014. Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). Mol. Phylogenet. Evol. 81, 221–231.

Rambaut, A., Drummond, A.J., 2009. Tracer v1.5. Available at: http://beast.bio.ed.ac.uk/Tracer.

Rambaut, A., Grass, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235–238.

Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.

Ren, H., Lu, L.M., Soejima, A., Luke, Q., Zhang, D.X., Chen, Z.D., Wen, J., 2011. Phylogenetic analysis of the grape family (Vitaceae) based on the noncoding plastid *trnC-petN*, *trnH-psbA*, and *trnL-F* sequences. Taxon 60, 629–637.

Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol. 64, 663–676.

Rodrigues, J.G., Lombardi, J.A., Lovato, M.B., 2014. Phylogeny of *Cissus* (Vitaceae) focusing on South American species. Taxon 63, 287–298.

Rodríguez-Espeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56, 389–399.

Rokas, A., Carroll, S.B., 2006. Bushes in the tree of life. PLoS Biol. 4, e352.

Rokas, A., Williams, B.L., King, N., Carrol, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rossetto, M., Jackes, B.R., Scott, K.D., Henry, R.J., 2001. Intergeneric relationships in the Australian Vitaceae: new evidence from cpDNA analysis. Genet. Resour. Crop Evol. 48, 307–314.

Rossetto, M., Jackes, B.R., Scott, K.D., Henry, R.J., 2002. Is the genus *Cissus* (Vitaceae) monophyletic? Evidence from plastid and nuclear ribosomal DNA Syst. Bot. 27, 522–533.

Rossetto, M., Crayn, D.M., Jackes, B.R., Porter, C., 2007. An updated estimate of intergeneric phylogenetic relationships in the Australian Vitaceae. Can. J. Bot. 85, 722–730.

Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–333.

Salichos, L., Stamatakis, A., Rokas, A., 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. 31, 1261–1271.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sanderson, M.J., Wojciechowski, M.F., Hu, J.-M., Khan, T.S., Brady, S.G., 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Mol. Biol. Evol. 17, 782–797.

Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: species tree analysis web server. Nucleic Acids Res. 41, W238–W241.

Simmons, M.P., Sloan, D.B., Gatesy, J., 2016. The effects of subsampling gene trees on coalescent methods applied to ancient divergences. Mol. Phylogenet. Evol. 97, 76–89.

Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 1–15.

Soejima, A., Wen, J., 2006. Phylogenetic analysis of the grape family (Vitaceae) based on three chloroplast markers. Am. J. Bot. 93, 278–287.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P.K., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96, 336–348.

Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlsward, B.S., Bell, C.D., Latvis, M., Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth, C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.L., Hilu, K.W., Davis, C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue, M.J., Soltis, P.S., 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98, 704–730.

Soltis, D.E., Mort, M.E., Latvis, M., Mavrodiev, E.V., O'Meara, B.C., Soltis, P.S., Burleigh, J.G., de Casas, R.R., 2013. Phylogenetic relationships and character evolution analysis of Saxifragales using a supermatrix approach. Am. J. Bot. 100, 916–929.

Song, S., Liu, L., Edwards, S.V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl Acad. Sci. USA 109, 14942–14947.

Springer, M., Gatesy, J., 2014. Land plant origins and coalescence confusion. Trends Plant Sci. 19, 267–269.

Springer, M., Gatesy, J., 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94, 1–33.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771.

Sun, M., Soltis, D.E., Soltis, P.S., Zhu, X.Y., Burleign, J.G., Chen, Z.D., 2015. Deep phylogenetic incongruence in the angiosperm clade *Rosidae*. Mol. Phylogenet. Evol. 83, 156–166.

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0b10. Sinauer Associates, Sunderland, MA.

Szöllősi, G.J., Tannier, E., Daubin, V., Boussau, B., 2015. The inference of gene trees with species trees. Syst. Biol. 64, e42–e62.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Telford, M.J., 2013. The animal tree of life. Science 339, 764–766.
Telford, M.J., Copley, R.R., 2011. Improving animal phylogenies with genomic data. Trends Genet. 27, 186–195.
Thomson, R.C., Shaffer, H.B., 2010. Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. Syst. Biol. 59, 42–58.
Torregrosa, L., Vialet, S., Adiveze, A., Iocco-Corena, P., Thomas, M.R., 2015. Grapevine (*Vitis vinifera* L.). In: Wang, K. (Ed.), Agrobacterium Protocols, 3rd edn. Humana Press, New York, vol. 2, pp. 177–194.
Townsend, J.P., Su, Z., Tekle, Y.I., 2012. Phylogenetic signal and noise: predicting the power of a dataset to resolve phylogeny. Syst. Biol. 61, 835–849.
Trias-Blasi, A., Parnell, J.A.N., Hodkinson, T.R., 2012. Multi-gene region phylogenetic analysis of the grape family (Vitaceae). Syst. Bot. 37, 941–950.
Wan, Y., Schwaninger, H., Baldo, A., Labate, J., Zhong, G.-Y., Simon, C., 2013. A phylogenetic analysis of the grape genus (*Viti*s L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. BMC Evol. Biol. 13, 141.
Wang, Y.H., Jiang, W.M., Comes, H.P., Hu, F.S., Qiu, Y.X., Fu, C.X., 2015. Molecular phylogeography and ecological niche modelling of a widespread herbaceous climber, *Tetrastigma hemsleyanum* (Vitaceae): insights into Plio–Pleistocene range dynamics of evergreen forest in subtropical China. New Phytol. 206, 852–867.
Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A., 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2, apps.1400042.
Wen, J., 2007a. Leeaceae. In: Kubitzki, K. (Ed.), The Families and Genera of Vascular Plants. Springer-Verlag, Berlin, vol. 9, pp. 221–225.
Wen, J., 2007b. Vitaceae. In: Kubitzki, K. (Ed.), The Families and Genera of Vascular Plants. Springer-Verlag, Berlin, vol. 9, pp. 467–479.
Wen, J., Nie, Z.L., Soejima, A., Meng, Y., 2007. Phylogeny of Vitaceae based on the nuclear *GAI1* gene sequences. Can. J. Bot. 85, 731–745.
Wen, J., Lu, L.M., Boggan, J.K., 2013a. Diversity and evolution of Vitaceae in the Philippines. Philipp. J. Sci. 142, 223–244.
Wen, J., Ree, R.H., Ickert-Bond, S., Nie, Z.L., Funk, V., 2013b. Biogeography: where do we go from here? Taxon 62, 912–927.
Wen, J., Xiong, Z., Nie, Z.L., Mao, L., Zhu, Y., Kan, X.Z., Ickert-Bond, S.M., Gerrath, J., Zimmer, E.A., Fang, X.D., 2013c. Transcriptome sequences resolve deep relationships of the grape family. PLoS ONE 8, e74394.
Wen, J., Boggan, J.K., 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, 11–19.
Wen, J., Lu, L.M., Nie, Z.L., Manchester, S.R., Ickert-Bond, S.M., Chen, Z.D., 2015. Phylogenetics and a revised classification of Vitaceae. Presented at Botany 2015, Edmonton, Alberta.
Wielstra, B., Arntzen, J.W., van der Gaag, K.J., Pabijan, M., Babik, W., 2014. Data concatenation, Bayesian concordance and coalescent-based analyses of the species tree for the rapid radiation of *Triturus* Newts. PLoS ONE 9, e111011.
Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.
Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc. Natl Acad. Sci. USA 109, 17519–17524.
Xi, Z., Liu, L., Rest, J.S., Davis, C.C., 2014. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. Syst. Biol. 63, 919–932.
Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92, 63–71.
Xia, X., 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30, 1720–1728.
Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.
Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.H., 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60, 150–160.
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, 4956.
Zhang, N., Wen, J., Zimmer, E.A., 2015a. Congruent deep relationships in the grape family (Vitaceae) based on sequences of chloroplast genomes and mitochondrial genes via genome skimming. PLoS ONE 10, e0144701.
Zhang, N., Wen, J., Zimmer, E.A., 2015b. Expression patterns of *AP1*, *FUL*, *FT* and *LEAFY* orthologs in Vitaceae support the homology of tendrils and inflorescences throughout the grape family. J. Syst. Evol. 53, 469–476.
Zhang, N., Wen, J., Zimmer, E.A., 2016. Another look at the phylogenetic position of the grape order Vitales: chloroplast phylogenomics with an expanded sampling of key lineages. Mol. Phylogenet. Evol. 101, 216–223.
Zhong, B., Yonezawa, T., Zhong, Y., Hasegawa, M., 2010. The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. Mol. Biol. Evol. 27, 2855–2863.
Zhong, B., Deusch, O., Goremykin, V.V., Penny, D., Biggs, P.J., Atherton, R.A., Nikiforova, S.V., Lockhart, P.J., 2011. Systematic error in seed plant phylogenomics. Genome Biol. Evol. 3, 1340–1348.
Zhong, B., Liu, L., Yan, Z., Penny, D., 2013. Origin of land plants using the multispecies coalescent model. Trends Plant Sci. 18, 492–495.
Zimmer, E.A., Wen, J., 2012. Using nuclear gene data for plant phylogenetics: progress and prospects. Mol. Phylogenet. Evol. 65, 774–785.
Zimmer, E.A., Wen, J., 2015. Using nuclear gene data for plant phylogenetics: progress and prospects II. Next-gen approaches. J. Syst. Evol. 53, 371–379.
Zimmermann, T., Mirarab, S., Warnow, T., 2014. BBCA: improving the scalability of *BEAST using random binning. BMC Genom. 15, S11.
Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T., Ge, S., 2008. Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol. 9, R49.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, The University of Texas at Austin.
Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Scaled map of 11 chloroplast regions (based on *Vitis vinifera* chloroplast genome) and six nuclear

regions (based on *Vitis vinifera* genome) surveyed in this study.

Fig. S2. Average pairwise distances of six nuclear, 11 chloroplast and 229 nuclear loci (in the order of highest to lowest average distance).

Fig. S3. Topologies representing relationships among major clades of Vitaceae supported by two or more of the 229 nuclear gene trees (with at least one major clade > 70% BS supported).

Fig. S4. Topologies representing relationships among major clades of Vitaceae supported by only one of the 229 nuclear gene trees (with at least one major clade > 70% BS supported).

Fig. S5. (a) Phylogenetic relationships among Vitaceae based on the MP analysis of data set 42_5cp. (b) Topology inferred from the optimal partitioning analyses of the ML and BI methods of data set 42_5cp, indicating bootstrap values and posterior probability for the major nodes.

Fig. S6. (a) Phylogenetic relationships among Vitaceae based on the MP analysis of data set 362_5cp. (b) Topology inferred from the optimal partitioning analyses of the ML and BI methods of data set 362_5cp, indicating bootstrap values and posterior probability for the major nodes.

Fig. S7. Phylogenetic backbone of Vitaceae based on (a) maximum parsimony (MP), (b) maximum likelihood (ML) and (c) Bayesian inference (BI) analyses of data set 42_11cp. (d) A congruent backbone topology of the MP, ML and BI analyses based on the 42_11cp data set excluding the core *Cissus* clade.

Fig. S8. Topologies inferred from the coalescent analyses using MP-EST and STAR (a) and the optimal partitioning analyses using the ML and BI methods (b) of data set 42_14loci (excluding *aro*B and *GAI*1, which conflict strongly with other markers from data set 42_17loci).

Fig. S9. Species trees inferred from MP-EST and STAR based on data sets 16_75nu (a) and 16_154nu (b).

Table S1. Data sets used and analyses conducted for this study. "cp" and "nu" are used to represent chloroplast and nuclear genome, respectively.

Table S2. Taxon sampling and GenBank accession numbers of DNA sequences for the 17-marker data set (including 11 chloroplast and six nuclear regions of 42 taxa).

Table S3. Taxon sampling and GenBank accession numbers of DNA sequences for five chloroplast regions of 362 taxa.

Table S4. Primers used for PCR and sequencing in this study.

Table S5. Sequence characteristics, best substitution model, average pairwise distance and Xia's index of substitution saturation for each data partition based on data sets of 11 chloroplast and six nuclear regions for 42 taxa.

Table S6. Gene length, number of parsimony characters, best-fit substitution model, average pairwise distance and the substitution saturation level as implemented in DAMBE for the 229 nuclear genes from the transcriptome data.

Table S7. Optimal partitioning schemes of data sets 42_5cp, 362_5cp, 42_11cp, 42_17loci and 42_14loci and best-fitting substitution model for each partition selected by PartitionFinder based on the corrected Akaike Information Criterion (AICc).