Phylogenomic data reveal hard polytomies across the backbone of the large genus *Solanum* (Solanaceae)

Gagnon, Edeline[1,2]; Hilgenhof, Rebecca[1,2]; Orejuela, Andrés[1,2]; McDonnell, Angela[3]; Sablok, Gaurav[4,5] Aubriot, Xavier[6]; Giacomin, Leandro[7]; Gouvêa, Yuri[8]; Bohs, Lynn[9]; Dodsworth, Steven[10,11]; Martine, Christopher[12]; Poczai, Péter[4,13]; Knapp, Sandra[14]; Särkinen, Tiina[1].

[1]*Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK*

[2]*School of Biological Sciences, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, EH9 3JH, UK*

[3]*Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, 1000 Lake Cook Rd, Glencoe, IL 60022, USA*

[4]*Finnish Museum of Natural History (Botany Unit), University of Helsinki, PO Box 7 FI-00014 Helsinki, Finland*

[5]*Organismal and Evolutionary Biology Research Programme (OEB), Viikki Plant Science Centre (ViPS), PO Box 65, FI-00014 University of Helsinki, Finland*

[6]*Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91405, Orsay, France*

[7]*Instituto de Ciências e Tecnologia das Águas & Herbário HSTM, Universidade Federal do Oeste do Pará, Rua Vera Paz, sn, Santarém, CEP 68040-255, PA, Brazil*

[8]*Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais – UFMG, Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, CEP 31270-901, MG, Brazil*

[9]*Department of Biology, University of Utah, Salt Lake City, UT, USA*

[10]*School of Life Sciences, University of Bedfordshire, University Square, Luton LU1 3JU, UK*

[11]*Royal Botanic Gardens, Kew, Richmond TW9 3AE, Surrey, UK*

[12]*Department of Biology, Bucknell University, Lewisburg, PA 17837, USA*

[13]*Faculity of Environmental and Biological Sciences, University of Helsinki, FI-00014 Finland*

[14]*Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.*

Corresponding author: Edeline Gagnon, edeline.gagnon@gmail.com

ABSTRACT

Increased volumes of phylogenomic data have revealed incongruent topologies in gene trees, both between and within genomes across many organisms. Some of these incongruences indicate polytomies that may remain impossible to resolve. Here, widespread gene-tree discordance is uncovered along the backbone of *Solanum*, one of the largest flowering plant genera that includes the cultivated potato, tomato, and eggplant, as well as 24 minor crop plants. First, a densely sampled species-level phylogeny of *Solanum* is built using unpublished and publicly available Sanger sequences comprising 60% of all accepted species (742 spp.) and nine regions (ITS, *waxy*, and seven plastid markers). The robustness of the Sanger-based topology is tested by examining a plastome dataset with 140 species and a nuclear target-capture dataset with 39 species of *Solanum*. Clear incongruences between species trees generated from the supermatrix, plastome, and nuclear target-capture datasets are revealed. Discordance within the plastome and target-capture dataset are found at different evolutionary depths in three different areas along the backbone of these phylogenetic trees, with polytomy tests suggesting that most of these nodes have short branches and should be collapsed. We argue that incomplete lineage sorting due to rapid diversification is the most likely cause behind these polytomies, and that embracing the uncertainty that underlies them is crucial to depict the evolution of large and rapidly radiating lineages.

Keywords

Phylogenomic discordance, incongruence, hard polytomies, Solanaceae, *Solanum*, incomplete lineage sorting, supertree network

INTRODUCTION

Recent advances in high-throughput sequencing have provided larger molecular datasets, including entire genomes, for reconstructing evolutionary relationships (e.g., Ronco et al. 2021). In botany, considerable progress has been made since the publication of the first molecular-based classification of orders and families (APG 1998), with one of the most recent examples including a phylogenetic tree of the entire Viridiplantae based on transcriptome data from more than a thousand species (1KP 2019). Whilst large datasets have strengthened our understanding of evolutionary relationships and classifications across the Tree of Life, several of them have demonstrated repeated cases of persistent topological discordance across key nodes in birds (Suh et al. 2015; Suh 2016), mammals (Simion et al. 2017; Morgan et al. 2013; Romiguier et al. 2013), amphibians (Hime et al. 2021), plants (Wickett et al. 2014; 1KP 2019), and fungi (Kuramae et al. 2006).

Much debate surrounds whether persistent topological incongruences can be resolved with more data or whether they represent so-called hard polytomies that reflect complex biological realities due to non-bifurcating evolution (Jeffroy et al. 2006; Philippe et al. 2011). Discordance in phylogenetic signal can be due to three general classes of effects (Wendel and Doyle 1998): (1) technical causes such as gene choice, sequencing error, model selection or poor taxonomic sampling (Philippe et al. 2011; Philippe 2017); (2) organism-level processes such as rapid or convergent evolution, rapid diversification, incomplete lineage sorting (ILS), or horizontal gene transfer (Degnan and Rosenberg 2009) and (3) gene and genome-level processes such as interlocus interactions and concerted evolution, intragenic recombination, use of paralogous genes for analysis, and/or non-independence of sites used for analysis. Together, these biological and non-biological processes can lead to conflicting phylogenetic signals at

different loci in the genome and hinder the recovery of the evolutionary history of a group (Degnan and Rosenberg 2009). Consequently, careful assessment of phylogenetic discordance across mitochondrial, plastid, and nuclear datasets is critical for understanding realistic evolutionary patterns in a group, as traditional statistical branch support measures fail to reflect topological variation of the gene trees underlying a species tree (Liu et al. 2009; Kumar et al. 2012).

Here we explore the presence of topological incongruence in nuclear and plastome datasets of the large and economically important angiosperm genus *Solanum* L., which includes 1,228 accepted species and several major crops and their wild relatives, including potato, tomato and brinjal eggplant (aubergine), as well as at least 24 minor crop species. Building a robust species-level phylogeny for this genus has not only been challenging because of the sheer size of the genus, but also because of persistent poorly resolved nodes along the phylogenetic backbone. Bohs (2005) published the first plastid phylogenetic analysis for *Solanum* and established a set of 12 highly supported clades based on her strategic sampling of 112 (9%) species, spanning morphological and geographic variation. As new studies have emerged with increased taxonomic and genetic sampling (e.g., Levin et al. 2006; Weese and Bohs 2007; Stern et al. 2011; Särkinen et al. 2013; Tepe et al. 2016), the understanding of the overall phylogenetic relationships within *Solanum* has evolved to recognise three main clades: (1) Thelopodium Clade containing three species sister to the rest of the genus, (2) Clade I containing c. 350 mostly herbaceous and non-spiny species (including the Tomato, Petota, and Basarthrum clades that contain the cultivated tomato, potato, and pepino, respectively), and (3) Clade II consisting of c. 900 predominantly spiny and shrubby species, including the cultivated brinjal eggplant (Table 1). The two latter clades are further resolved into 10 major and 43 minor clades (Table 1).

Despite the establishment of the relatively robust major and minor clades, phylogenetic relationships between many of these clades of *Solanum* have remained poorly resolved, mainly due to limitations in taxon and molecular marker sampling. The most recent genus-wide phylogenetic study by Särkinen et al. (2013) was based on seven markers (two nuclear and 5 plastid) and included fewer than half (34%) of the species of *Solanum.* It failed to resolve the relationships between major and minor clades, especially within Clade II and the large component Leptostemonum Clade. A more robust, well-resolved, and densely sampled phylogeny of *Solanum* is needed to study the morphological, chemical, and genetic variation in this mega-diverse group, particularly for those interested in using wild species traits in plant breeding (Hardigan et al. 2016; Li et al. 2018; Smith et al. 2020).

To check whether significant increases in taxonomic and molecular sampling are key to phylogenetic improvements in *Solanum*, we explored phylogenetic relationships across this mega-diverse genus using plastid (PL) and nuclear target-capture (TC) phylogenomic datasets as well as a Sanger sequence dataset including 60% of species. We ask: (1) Does increased gene sampling of plastome and nuclear data resolve previously identified polytomies between major and minor clades?; (2) Is there evidence of discordance between genomic datasets?; (3) Within genomic datasets, do high branch support also show high gene concordance?; (4) Are areas of high discordance in the *Solanum* phylogeny better explained by polytomies rather than bifurcating nodes? Discordance analyses and comparison of branch support values across the 48 species trees built from the two phylogenomic datasets and the Sanger sequence supermatrix show phylogenetic discordance both within and between genomic datasets. Polytomy tests and filtered supertree networks indicate that the discordance is due to the presence of hard polytomies in at least three places. The hard polytomies are likely due to rapid speciation and

diversification coupled with ILS and we suggest that they represent an important aspect of the biology and evolution of the genus.

MATERIALS AND METHODS

Taxon Sampling

A Sanger sequence supermatrix included all available sequences from GenBank for nine regions, including the nuclear ribosomal internal transcribed spacer (ITS), low-copy nuclear region *waxy* (i.e., GBSSI), two protein-coding plastid genes *matK* and *ndhF*, and five non-coding plastid regions (*ndhF-rpl32*, *psbA-trnH*, *rpl32-trnL*, *trnS-G*, and *trnT-L*). GenBank results were blasted against target regions in USEARCH v.11 (Edgar 2010). Only vouchered and verified samples were kept. Taxon names were checked against SolanaceaeSource synonymy (solanaceaesource.org, Nov. 2020) and species that were duplicated were pruned out. A total of 817 Sanger sequences were generated and added to the matrix, adding 129 previously unsampled species and new data for 257 species (Table S1). Final species sampling across major and minor clades of *Solanum* varied from 13-100%, with 742 species of *Solanum* (60% of the 1,228 currently accepted species, Nov 2020; Table 1). Four taxa of *Jaltomata* Schltdl. were used as an outgroup (Table S1).

To assess discordance in the overall phylogeny of *Solanum*, a set of species was selected for the phylogenomic study to represent all 10 major and as many of the 43 minor clades of *Solanum* as possible (Table 1), as well as the outgroup *Jaltomata*. The final sampling included 151 samples for the PL dataset (140 *Solanum* species; Table 1, S2) and 40 samples for the TC dataset (39 *Solanum* species; Table 1, S3). For the PL dataset, 86 samples were sequenced using low-coverage genome skimming, and the remaining samples downloaded from GenBank (Nov

2019). For the TC dataset, 12 samples were sequenced as part of the Plant and Fungal Trees of Life project (PAFTOL; paftol.org) using the Angiosperms353 bait set (Johnson et al. 2019) and 17 sequences were added from an unpublished dataset provided by A. McDonnell and C. Martine. Sequences for the remaining 12 samples were extracted from the GenBank SRA archive using the SRA Toolkit 2.10.7 (https://github.com/ncbi/sra-tools; Table S3).

DNA Extraction, Library Preparation & Sequencing

DNA extraction, library preparation, and sequencing of both Sanger sequence and phylogenomic datasets followed established protocols. Full methods are available in Supplementary Material and Methods.

Phylogenetic Analyses

Testing the effect of methodological choices. – A total of 48 phylogenetic analyses were run using different methodological strategies to test their effect on tree topologies (Fig. 1a) because the choice of phylogenetic method, taxon sampling, missing data, data inclusion, and partitioning strategy can lead to topological discordance (e.g., Philippe et al. 2011, 2017; Saarela et al. 2018; Duvall et al. 2020; Gonçalves et al. 2020). Both Maximum Likelihood (ML) and Bayesian Inference (BI) were performed on Sanger sequence matrices (see Supplementary Material and Methods for details). Analyses were run on all nine loci individually, on the combined plastid dataset (seven loci), and on the final combined matrix. The ML and BI species trees from the final combined matrix, with their more complete taxonomic sampling, were used as a reference to compare results from the PL and TC species trees generated below.

TABLE 1. Number of species and taxon sampling across major and minor clades of *Solanum*. Clades are based on groups identified in previous molecular phylogenetic studies (Bohs 2005; Weese and Bohs 2007; Stern et al. 2011; Stern and Bohs 2012; Särkinen et al. 2013; Tepe et al. 2016). Species number for each clade is based on current updated taxonomy in the SolanaceaeSource database. The 19 clades sampled in the pruned trees for the principal coordinate analysis in this study are in bold. New associated major clade names are given where applicable. Lines shaded in gray represent major and minor clades belonging to Clade II.

Minor clade	Associated major clade (Särkinen et al. 2013)	New associated major clade (This study)	Species	Sampled species (%)		
				Super-matrix	Plastome (PL)	Target Capture (TC)
Thelopodium	**Thelopodium**		**3**	**3 (100%)**	**1 (33%)**	**1 (33%)**
African non-spiny	M Clade	VANAns	14	5 (36%)	1 (7%)	-
Normania	**M Clade**	**VANAns**	**3**	**2 (67%)**	**1 (33%)**	**1 (33%)**
Archaesolanum	**M Clade**	**VANAns**	**8**	**8 (100%)**	**1 (13%)**	**1 (13%)**
Valdiviense	**M Clade**	**VANAns**	**1**	**1 (100%)**	**1 (100%)**	**1 (100%)**
Dulcamaroid	**M Clade**	**DulMo**	**45**	**25 (56%)**	**8 (18%)**	**1 (2%)**
Morelloid	**M Clade**	**DulMo**	**75**	**66 (88%)**	**15 (20%)**	**1 (1%)**
Regmandra	**Potato**	**Regmandra**	**12**	**6 (50%)**	**4 (33%)**	**1 (8%)**
Herpystichum	Potato		10	10 (100%)	-	-
Pteroidea	Potato		10	10 (100%)	1 (10%)	-
Oxycoccoides	Potato		1	1 (100%)	-	-
Articulatum	Potato		2	2 (100%)	-	-
Basarthrum	**Potato**		**16**	**10 (56%)**	**3 (19%)**	**3 (19%)**
Anarrhichomenum	Potato		12	8 (82%)		
Etuberosum	**Potato**		**3**	**2 (67%)**	**2 (67%)**	**1 (33%)**
Tomato	**Potato**		**7**	**14 (82%)**	**8 (47%)**	**3 (18%)**
Petota	**Potato**		**113**	**61 (54%)**	**38 (34%)**	**2 (2%)**
Clandestinum-Mapiriense	**Clandestinum-Mapiriense**		**3**	**3 (100%)**	**1 (33%)**	**1 (33%)**
Wendlandii-Allophyllum	**Wendlandii-Allophyllum**		**10**	**7 (70%)**	**1 (10%)**	**1 (10%)**
Nemorense	Nemorense		4	4 (100%)	1 (25%)	-

Pachyphylla	Cyphomandra	39	32 (82%)	1 (3%)	-
Cyphomandropsis	Cyphomandra	11	7 (64%)	1 (9%)	1 (9%)
Geminata	**Geminata**	**150**	**68 (45%)**	**5 (3%)**	**1 (1%)**
Reductum	Geminata	2	2 (100%)	1 (50%)	-
Brevantherum	Brevantherum	83	29 (35%)	3 (4%)	-
Gonatotrichum	Brevantherum	7	7 (100%)	1 (14%)	-
Inornatum	Brevantherum	5	2 (40%)	1 (20%)	-
Trachytrichium	Brevantherum	2	2 (100%)	-	-
Elaeagniifolium	**Leptostemonum**	**5**	**5 (100%)**	**1 (20%)**	**1 (20%)**
Micracantha	Leptostemonum	14	9 (64%)	1 (7%)	-
Torva	**Leptostemonum**	**54**	**34 (63%)**	**5 (9%)**	**1 (2%)**
Erythrotrichum	Leptostemonum	33	13 (39%)	1 (3%)	-
Thomasiifolium	Leptostemonum	9	4 (44%)	1 (11%)	-
Gardneri	Leptostemonum	10	8 (80%)	1 (10%)	-
Acanthophora	**Leptostemonum**	**22**	**13 (59%)**	**1 (5%)**	**-**
Lasiocarpa	Leptostemonum	12	12 (100%)	-	-
Sisymbriifolium	**Leptostemonum**	**4**	**4 (100%)**	**1 (25%)**	**1 (25%)**
Androceras	Leptostemonum	16	15 (94%)	-	-
Crinitum	Leptostemonum	23	10 (43%)	-	-
Bahamense	Leptostemonum	3	3 (100%)	-	-
Asterophorum	Leptostemonum	4	2 (50%)	-	-
Carolinense	Leptostemonum	11	8 (73%)	1 (9%)	-
Hieronymi	Leptostemonum	1	1 (100%)	1 (100%)	-
Old World	**Leptostemonum**	**332**	**197 (59%)**	**24 (7%)**	**16 (5%)**
Campechiense	Leptostemonum	1	1 (100%)	-	-
Crotonoides	Leptostemonum	3	2 (67%)	1 (33%)	-
Multispinum	Leptostemonum	1	1 (100%)	1 (100%)	-
Unplaced	Leptostemonum	9	1 (13%)	-	-
	TOTAL:	1,228	746 (60%)	140 (11%)	39 (3%)

Additional analyses were run on the two phylogenomic datasets; 40 on the PL and six on the TC datasets (Fig. 1a). The effect of phylogenetic method on tree topologies was tested by comparing topologies from ML (IQ-TREE 2; Minh et al. 2020a) and two coalescent methods (ASTRAL-III v.5.7.3, Zhang et al. 2018; and SVDquartets, Chifman and Kubatko 2014; Fig. 1a). The two coalescent methods have been shown to perform differently under different levels of ILS (Chou et al. 2015). In addition, the effect of missing data (i.e., data quality) was tested using complete (125 samples) versus partial plastid genomes (151 samples; Fig. 1a; Table S4). Each of the two matrices was analysed using different parts of the plastid genome (i.e., data type; Fig. 1a): (1) all regions; (2) exons only; (3) introns only; and (4) intergenic regions only. Finally, the effect of partitioning scheme amongst loci was tested with the PL dataset using three different methods (Fig. 1a): (1) no partition; (2) partition by loci; and (3) best-fit partition scheme. For the TC dataset, the effect of data quality (i.e., missing data) was tested using three different thresholds of the minimum number of recovered target genes (min 20, 10, and 4; Fig. 1a), using HybPiper (Johnson et al. 2016). ML (IQ-TREE2) and coalescent trees (ASTRAL-III) were generated for each of these datasets resulting in 6 species trees (Fig. 1a). Full methods of plastome and target capture assembly, and the phylogenetic methods used to generate the 48 species trees from all three datasets are indicated in Supplementary Material and Methods.

Discordance Analyses

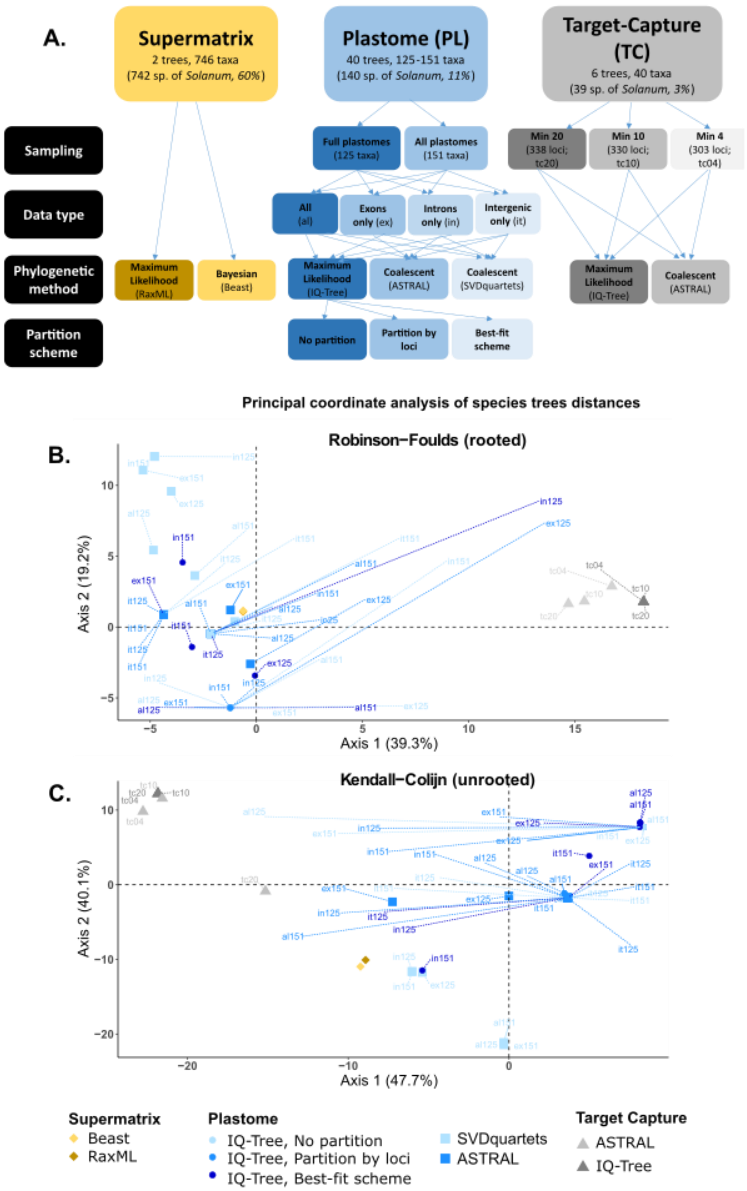
Principal Coordinate Analysis. – Phylogenomic discordance within and between datasets was assessed by comparing branch support values across the 48 species trees generated from the supermatrix, PL, and TC datasets. To compare topological differences, Principal Coordinates Analyses (PCoA) of species tree distances were carried out and visualised using the package "Treespace" (Jombart et al. 2017), adapting an R script from Gonçalves et al. (2019b). Distances

were measured between unrooted (Robinson and Foulds 1981) and rooted topologies (Kendall and Colijn 2016). To enable comparison, all 48 species trees were pruned to 27 taxa representing 19 minor clades of *Solanum* (Table 1). The species sampled in the PL and TC datasets were identical for all except three minor clades, in which different closely related species were sequenced (Acanthophora: *S. viarum* Dunal/*S. capsicoides* All.; Morelloid: *S. opacum* A.Braun & C.D.Bouché/*S. americanum* Mill.; Elaeagnifolium: *S. mortonii* Hunz./*S. elaeagnifolium* Cav.).

Concordance factors. – Phylogenomic discordance within the PL and TC datasets was measured using gene concordance factors (gCF) and site concordance factors (sCF) calculated in IQ-TREE 2 (Minh et al. 2020b). These metrics assess the proportion of gene trees that are concordant with different nodes along the phylogenetic tree and the number of informative sites supporting alternative topologies. Low gCF values can result from either limited information (i.e., short branches) or genuine conflicting signal; low sCF values ~30% indicate lack of phylogenetic information in loci (Minh et al. 2020b). The metrics were calculated using the ASTRAL-III topologies with 151 (PL) and 40 (TC) tips, where the PL topology was pruned to 43 tips to facilitate comparison, including the same 40 tips as the TC analysis with representatives of three minor clades not sampled in the TC phylogeny.

Network analyses and polytomy tests.– The presence of polytomies was explored by generating filtered supertree networks in SplitsTree 4 (Huson & Bryant 2006). Networks were generated based on the 160 gene trees used for the PL dataset and the 303 gene trees from the TC filtered to a minimum of 20 taxa per loci, with the minimum number of trees set to 50% of each dataset (80 and 151 trees respectively). The PL network was pruned as described above to facilitate comparison with the TC networks. Polytomy tests were carried out in ASTRAL-III following Sayyari and Mirarab (2018). Gene trees were used to infer quartet frequencies for all

222 branches to determine the presence of polytomies while accounting for ILS. The analysis was
223 run using ASTRAL-III topologies of PL (151 tips) and three TC topologies (min 20, 10, and 4)
224 and were run twice to minimise error due to gene tree error (collapsing branches with <10% and
225 <75% local PP support).



226

227 **Figure 1**. Overview of all (N=48) phylogenetic analyses performed in this study. **a)** Schematic
228 representation of all the analyses run on the supermatrix, plastome (PL) and target capture (TC)

datasets, and distribution of all resulting 48 phylogenetic trees in Principal Coordinates Analyses (PcoA) space based on distances between **b)** unrooted (Robinson-Foulds) and **c)** rooted trees (Kendall-Colijn). Colours and symbols reflect three main datasets used in the study: Supermatrix (yellow-brown diamonds), PL (blue circles and squares), and TC (grey triangles). Different phylogenetic methods are shown in different shades for supermatrix and TC analysis. For PL analyses, different partition schemes analysed using Maximum Likelihood (IQ-TREE) are shown in blue circles (pale=no partition; middle=partition by loci; dark=best-fit scheme) and different coalescent methods in squares (pale=SVDquartets; dark=ASTRAL). Data quality and data source for PL and TC analyses is indicated using abbreviations: ex: exons; in: introns, it: intergenic, al: all plastome loci; 151: all plastomes, 125: full plastomes only; tc04: Min. 4 loci, tc10: Min. 10 loci, tc20: Min 20 loci.

RESULTS

Phylogenetic Analyses

Congruent recovery of clades. – All three datasets, including the supermatrix and the two phylogenomic datasets (PL and TC), recovered all previously recognized minor and major clades in *Solanum* (Fig. 2; Table S5); the only exception was the Mapiriense-Clandestinum minor Clade, which was found to be polyphyletic in the supermatrix phylogeny (only one species sampled in PL and TC phylogenies). The Potato Clade was strongly supported across all analyses (Fig. 2; Table S5), as was the Regmandra Clade in supermatrix and PL analyses (only one sample in TC phylogenies). Furthermore, all analyses recovered a clade here referred to as DulMo, that includes the Morelloid and the Dulcamaroid Clades as sister to each other (Fig. 2; Table S5). A new strongly supported clade, here referred to as VANAns Clade comprised of the

Valdiviense, Archaesolanum, Normania, and the African non-spiny Clades, was found across all analyses (Fig. 2; Table S5).

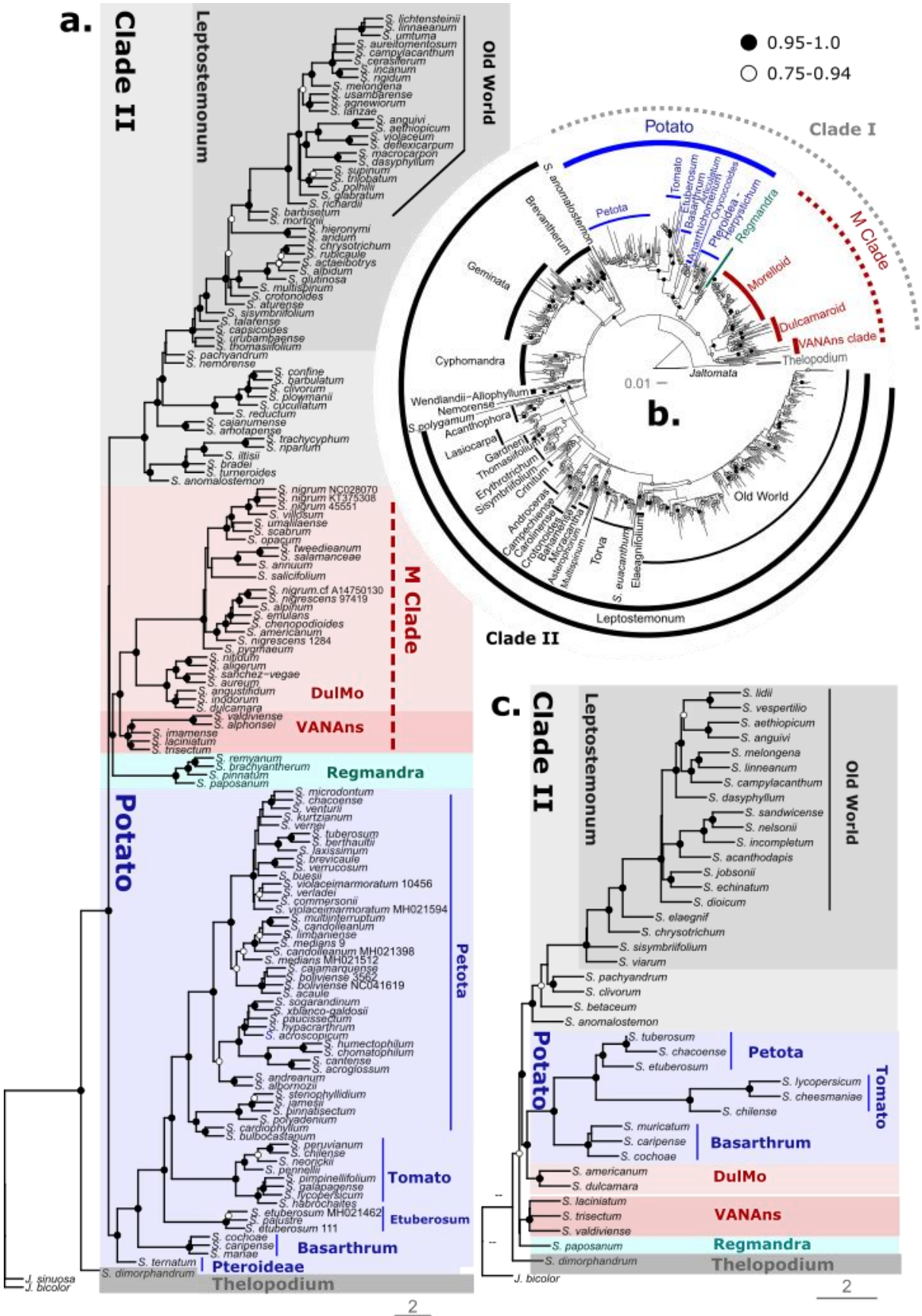
Clade II was supported as monophyletic across all topologies (Fig. 2), with maximum branch support in 42 out of 48 species trees (Table S5). As for the relationships of clades within Clade II, while differences in sampling hamper thorough comparisons, there was no deep incongruence detected amongst topologies obtained with the supermatrix, PL, and TC datasets (Fig 2; Fig. S1-S9). Within Clade II, the large Leptostemonum Clade (the spiny solanums) was strongly supported in 46 of 48 analyses, with strong branch support in all cases (Fig. 2; Fig S1-S9; Table S5).

Incongruent relationships amongst clades. – In contrast to the results above, major incongruence between species trees was observed with respect to the relationships of the main clades identified above (Fig. 2a,c; Table S5). All analyses recovered the Thelopodium Clade as sister to the rest of *Solanum*, except for three PL topologies based on coalescent analyses (SVDquartets) and all three ML analyses of the TC dataset (Fig. 2; Table S5). While the supermatrix phylogeny supported the monophyly of the previously recognised Clade I that includes most non-spiny *Solanum* clades (Fig. 2b; Fig. S1-S2), the PL and TC trees resolved these clades as a grade relative to a monophyletic Clade II (Fig. 2a,c; Fig. S1-S9). This was in large part due to the unstable position of the Regmandra Clade, which was subtended by a particularly short branch and resolved in different positions along the backbone in all three datasets (Fig. 2, Fig. S1-S9, Table S5). For example, the supermatrix ML analysis recovered the Regmandra Clade as sister to the Potato Clade with moderate branch support (Fig. 2b), whereas the PL analyses in 39 out of 40 species trees resolved Regmandra either as sister to the M Clade, Clade II, or both with varying levels of support (21-100% branch support; Table S5). The ML

TC species trees resolved Regmandra as sister to the Potato Clade, DulMo, and Clade II (100% branch support; Table S5). While two of the PL ASTRAL analyses also recovered this topology, the third analysis resolved Regmandra as sister to the VANAns Clade; in all cases, branch support was only weak to moderate (67 to 82%; Table S5).

The previously identified M Clade composed of the VANAns and DulMo Clades was not supported by all analyses (Fig. 2, Table S5). While all PL ML analyses recovered the M Clade with high branch support (> 89%), support for this relationship was more variable in the ASTRAL and SVDquartets analyses (Table S5, Fig S5-7). The TC analyses never recovered the M Clade, but rather resolved the DulMo Clade as sister to the Potato Clade (100% branch support; Table S5, FigS8-S9). Furthermore, the VANAns Clade was recovered as sister to the rest of *Solanum* (excluding the first diverging lineage Thelopodium Clade) with moderate support in the TC ML analyses, and with low or no support in the TC ASTRAL analyses (Fig. 2, Table S5, Fig. S8-S9).

In all analyses, the Potato Clade was a clearly congruent and strongly supported clade, but its position within *Solanum* remains incongruent between datasets (Fig. 2); it was resolved as sister to the remaining *Solanum* in PL, sister to M Clade in supermatrix, and sister to DulMo Clade in the TC analyses (Fig. 2), with many more strongly supported positions suggested in other analyses (Table S5). Interestingly, the phylogenomic datasets also showed incongruent positions for the larger Petota Clade, where TC analyses resolved it as sister to the Etuberosum Clade, with maximum support in the ASTRAL analyses, and moderate to low support in the ML analyses (72-86%). In contrast, PL analyses placed it as sister to the Tomato Clade, albeit with branch support values under 85% in 4/8 of the ASTRAL PL analyses (Table S5).



297 **FIGURE 2.** Comparison of *Solanum* clades recovered in the three different datasets. **a)** Plastome
298 (PL) phylogeny from coalescent analysis (ASTRAL) with 151 samples representing 140
299 *Solanum* species based on 160 loci representing exons, introns and intergenic regions. **b)**
300 Supermatrix phylogeny from Maximum Likelihood analysis (RaxML) of 742 *Solanum* species
301 based on two nuclear and seven plastid regions. **c)** Nuclear TC phylogeny with 40 species from
302 coalescent analysis (ASTRAL) where a minimum of 20 taxa per loci threshold was used (min20,
303 303 loci included). Clades are shown in the same colour in all three phylogenies to enable
304 comparison. Branch support values are colour coded (bootstrap values in B, local posterior
305 probability values in A and C): black = strong (0.95–1.0), white = moderate to weak support
306 (0.75–0.94). Scale bars = substitutions/site. *Jaltomata* taxa were used as the outgroup to root all
307 the three phylogenies. Dashed lines indicate clades from the supermatrix analysis that were not
308 recovered in both the TC and PL analyses. Collection or Genbank numbers are indicated in the
309 PL phylogeny for duplicate species sampled in the phylogenetic tree.

310

311 Finally, while the BI and ML supermatrix phylogeny resolved the morphologically
312 unusual *S. anomalostemon* S.Knapp & M.Nee as sister to the rest of Clade II (BS 95%, PP 1.0),
313 PL analyses supported *S. anomalostemon*+Brevantherum Clade as the first branching lineage
314 within Clade II with high branch support (>90% for 27/40 PL trees; Fig. 2; Table S5). The
315 Brevantherum Clade was not included in the TC analyses preventing a strict comparison. Within
316 the Leptostemonum Clade, the Old World Clade is strongly supported with some exceptions:
317 while most analyses showed the Elaeagnifolium Clade as sister to the Old World Clade (Fig. 2),
318 ten PL analyses resolved the Elaeagnifolium Clade as nested within the Old World Clade (e.g.,
319 PL ASTRAL intron only dataset and PL SVDquartets analyses of all plastome data; Table S5).

There are also some differences between species closely related to the Eggplant clade and Anguivi grade, involving *S. campylacanthum* Hochst. ex A.Rich., *S. melongena* L., *S. linnaeanum* Hepper & P.-M.LJaeger, *S. dasyphyllum* Schum. & Thonn. and *S. aethiopicum* L. (Fig. 2., Fig. S1-9).

Discordance Analyses

Principal coordinate analysis. – The PCoA analysis (Fig. 1b-c) showed that all TC species trees were clearly separated from other species trees along the first and second axes in rooted and unrooted tree space (Fig. 1b-c). Trees derived from the supermatrix dataset were nested within the 40 PL species trees, the latter being spread more widely across the tree space (Fig. 1b-c).

Within datasets, species trees obtained with different phylogenetic methods are spread across the tree space, in line with the variety of different topologies observed between major clades and described above. Some weak clustering of PL topologies analysed using different methods was observed along the second axis (ML vs. coalescent; Fig. 1b-c). No clustering based on partitioning strategy or data sampling (151 vs. 125 taxa) was observed in PL dataset, or number of included loci in the TC dataset (min 4, 10 and 20; Fig. 1b-c).

Concordance factors. – Phylogenomic discordance was generally high across the PL and TC topologies, with gCF values >50% in only three nodes in the PL phylogeny (*Solanum* as a whole, *S. chilense* (Dunal) Reiche + *S. lycopersicum* L. or the Tomato Clade, and *S. hieronymi* Kuntze + *S. aridum* Morong in the Leptostemonum Clade; Fig. 3a). Elsewhere, gCF fell to 36% and below (16 nodes with gCF values 10% and below), with the lowest values found near branch nodes that varied the most amongst the different reconstructed species trees. This included the

node subtending Regmandra (gCF 1%, SCF 42%; Fig. 3a), and that positioning Regmandra + DulMo + VANans Clade as sister to Clade II (gCF 4%, SCF 33%). Similarly, low gCF and uninformative sCF values around 33% were found across Clade II, including the node placing *S. hieronymi* + *S. aridum* as sister to the Elaeagnifolium + Old World minor clades (gCF 6 %, sCF 34 %; Fig. 3a), as well as the placement of *S. capsicoides*, a representative of the Acanthophora lineage within Leptostemonum (gCF 7%, sCF 36%; Fig. 3a). Not all nodes with low gCF values <10% showed uninformative sCF values: for example, the node subtending the grouping of *S. chrysotrichum* Schltdl., *S. multispinum* N.E.Br., and *S. crotonoides* Lam. has a gCF value of 4%, but an sCF value of 51%. The close relationship amongst these three taxa, representing the Torva, Multispinum, and Crotonoides Clades respectively, is found in all species trees that included these taxa (Fig. S1-S7).

Across the TC phylogeny, gCF and SCF values were slightly higher on average, with six nodes presenting values >50% for both metrics: four within the Petota Clade, one at the base of the Leptostemonum Clade (gCF 64%, SCF 73%; Fig. 3a), and another at the base of the Old World Clade within Leptostemonum (gCF 58%, SCF 75%; Fig. 3a). Five nodes had low gCF values of >10% or less, with again some of the lowest values located near the base of the tree, such as the relationship of Regmandra as sister to the VANAns Clade (gCF 3%, sCF 39%; Fig. 3a), or placement of Potato as sister to the DulMo Clade (gCF 10%, sCF 41%; Fig. 3a), and the relationship of the Potato Clade + DulMo Clade as sister to Clade II (gCF 4%, sCF 41%; Fig. 3a).

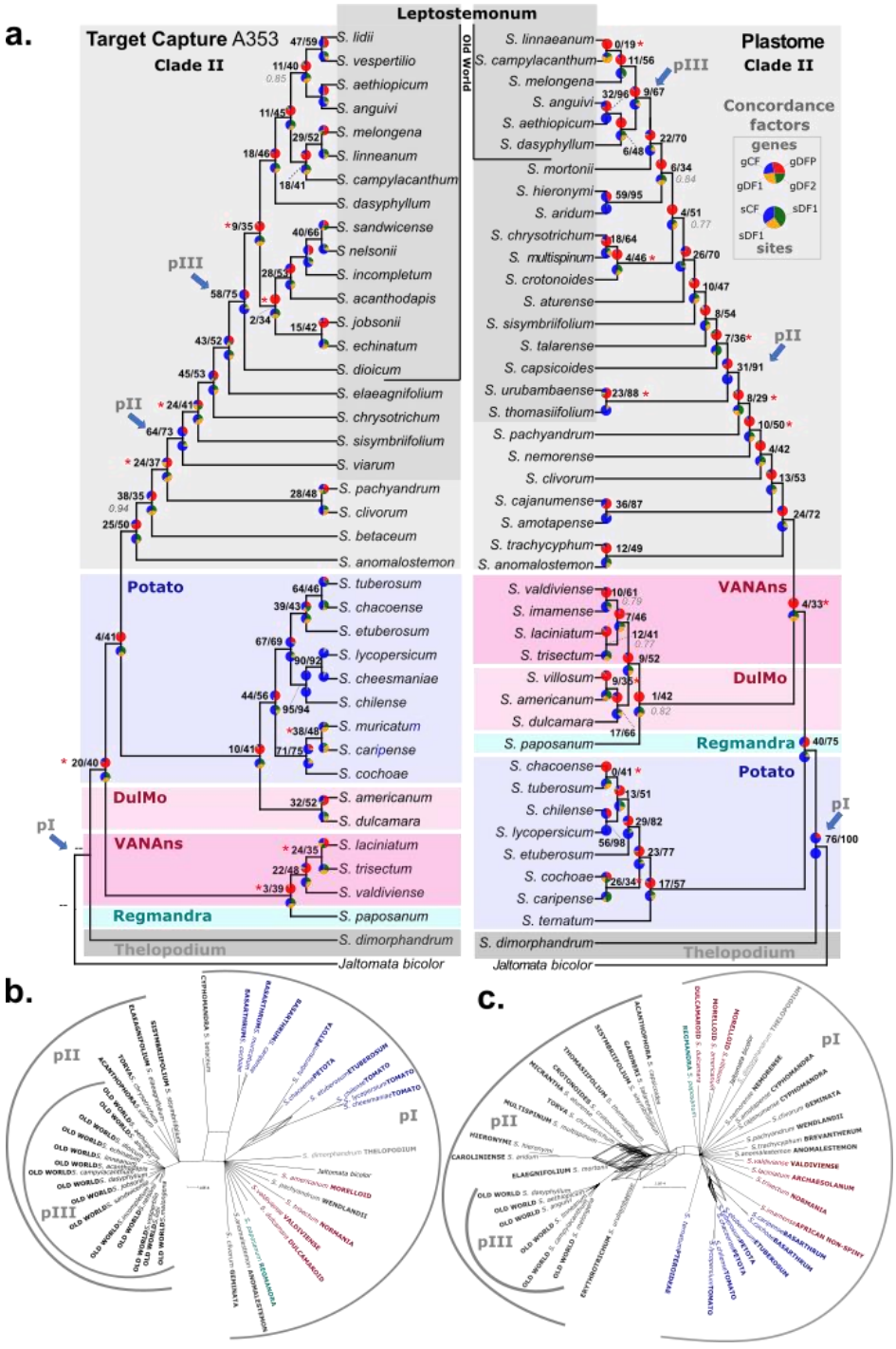
Network analyses and polytomy tests. – The filtered supertree network showed three large polytomies in the TC ASTRAL topology (Fig. 3b) corresponding to areas where incongruence was detected based on visual comparison (Fig. 1b-c) and low gCF and sCF values (Fig. 3a):

along the backbone (i) of *Solanum*, (ii) of the Leptostemonum Clade, and (iii) of the Old World Clade. These polytomies showed almost no boxed edges (i.e., reticulations) indicating lack of, rather than conflict in, signal (Fig. 3b). The same three polytomies were visible in the PL ASTRAL topology, with more boxed edges indicating conflict in signal (Fig. 3c).

The polytomy tests carried out for the three TC ASTRAL datasets resulted in 10 nodes each for which the null hypothesis of branch lengths equal to zero was accepted, suggesting they should be collapsed into polytomies (Fig. S11a,c,e); these nodes subtended the Regmandra Clade, and also suggested polytomies located within the VANAns Clade and Clade II, and at the crown nodes of the Leptostemonum Clade and Old World Clade. In two cases, a polytomy was detected at the base of the Tomato Clade (min 4 and min 10 datasets), and at the base of the Etuberosum + Petota +Tomato Clade (min 20 dataset). Repeating the analysis by collapsing nodes with <75% local PP support led to the collapse of 12 to 13 nodes across the analyses, most of them affecting the same clades as in the previous runs, but also leading to the collapse of the crown node of the *Solanum* clade. The effective number of gene trees was too low when nodes with <75% local PP support were collapsed to carry out the test for two nodes subtending *S. betaceum* Cav.and *S. anomalostemon*, most likely related to the low number of genes recovered for *S. betaceum* (Table S3).

For the PL dataset, only the results from collapsing branches with <10% in local PP support are reported (Fig. S10g) because collapsing branches with <75% local PP support resulted in the effective number of gene trees being too low to carry out the test across most of the nodes of the species trees (Fig. S10h). The polytomy tests for the PL ASTRAL topology resulted in a total number of 69 nodes across *Solanum* where the null hypothesis was accepted. With these nodes collapsed, the resulting topology included a polytomy between Regmandra,

388 Potato, M Clade, and Clade II. Polytomies were also detected at the base of Clade II, as well as
389 the Leptostemonum, Old World, Morelloid, Petota, and VANAns Clades (Fig. S10g).



390

Figure 3. Discordance analyses within and between the plastome (PL) and target capture (TC) phylogenomic datasets across *Solanum*. **a)** Rooted TC (left) and PL (right) ASTRAL phylogenies with gene concordance factor (gCF) and site concordance factor (sCF) values shown as pie charts, above and below each node respectively; plastome topology is based on analysis of all plastomes (151 samples pruned to 43 to facilitate comparison), including exons, introns and intergenic regions; and target capture topology is based on the analysis of 40 species with a minimum threshold of 20 taxa per loci. For gCF pie charts, blue represents proportion of gene trees concordant with that branch (gCF), green is proportion of gene trees concordant for 1st alternative quartet topology (gDF1), yellow support for 2nd alternative quartet topology (gDF2), and red is the gene discordance support due to polylphyly (gDFP). For the sCF pie charts: blue represents proportion of concordance across sites (sCF), green support for 1st alternative topology (quartet 1), and yellow support for 2nd alternative topology (quartet 2) as averaged over 100 sites. Percentages of gCF and sCF are given above branches, in bold. Branch support (local posterior probability) values ≥0.95 are not shown, between 0.75 and 0.94 are shown in italic grey, and nodes with support values below 0.75 are shown with asterisks (*); double-dash (--) indicates that the branch support was unavailable due to rooting of the phylogenetic tree. The nodes corresponding to the three main polytomies identified in the filtered supertree network (see b and c below) are identified in blue. **b)** Filtered supertree network of the TC dataset (40 taxa, min20) based on 303 gene trees with a 50% minimum tree threshold. **c)** Filtered supertree network of the PL dataset based on 160 genes trees (exon, intron, and intergenic regions) with a 50% minimum tree threshold. The three main polytomies discussed in the text are identified by arcs (pI, pII and pIII).

DISCUSSION

Phylogenomic Discordance in Solanum

Our analyses, despite increasing taxon sampling to 60% of *Solanum* species in the supermatrix analyses and providing the first comprehensive PL and TC phylogenies for the genus, did not help resolve any of the polytomies detected in previous phylogenetic studies (Särkinen et al. 2013). While the same major and minor clades are resolved consistently within and between all datasets (except for the Mapiriense-Clandestinum Clade), clear discordance both within and between plastid and nuclear phylogenomic datasets is evident in *Solanum*. The PcoA (Fig. 1b-c) showed striking differences in the species trees generated using different phylogenomic datasets (TC versus the PL). The diversity of topologies within datasets was not strongly influenced by the phylogenetic reconstruction method employed but is likely linked to the repeated short branches along the phylogenetic backbone in *Solanum*, due to low levels of informative characters across our analyses for these nodes. Our analyses support high levels of phylogenomic discordance in both TC and PL topologies, with relatively low gCF and sCF values across both datasets. Large key polytomies were identified in the network analysis of the TC and PL datasets at different evolutionary depths, at the base of : (i) *Solanum*, in relation with the emergence of all the major clades, (ii) the largest *Solanum* lineage, the Leptostemonum Clade, and (iii) the species-rich Old World Clade (Fig. 3b,c). Polytomy tests confirmed that multiple nodes within these polytomies should be collapsed, lending weight to the idea that these are hard polytomies within the genus *Solanum*.

Phylogenomic discordance has been discovered in many plant lineages mainly at order or family level (e.g., Morales-Briones et al. 2021; Smith et al. 2015). Our study differs from previous studies in highlighting widespread discordance within a large genus both within and

between nuclear and plastome phylogenomic datasets at different evolutionary depths. The strong gene-tree discordance pattern we found along the backbone of *Solanum* contrasts sharply with results obtained from other mega-diverse angiosperm genera, such as *Carex* (Cyperaceae), where low gCF and sCF values were recovered in only two nodes along the backbone (Villaverde et al. 2020).

Interpreting the exact nature of the three main regions of discordance discovered here needs careful consideration, because distinguishing between processes such as gene flow, ILS, and hybridisation can be difficult at different evolutionary depths. The more recently diverged Leptostemonum Clade is by far the most species-rich clade in *Solanum* (approximately 566 currently recognised species) and contains the Old World Clade, the most rapidly diversifying lineage within *Solanum* (Echeverría-Londoño et al. 2020). The low gCF percentages observed in both the Leptostemonum and Old World Clade could be attributed to ILS due to rapid diversification; alternatively, the pattern could be explained by gene flow amongst populations and species. Similar patterns of strong discordance along the backbone of more recently diverged clades in Solanaceae such as the rapidly evolving Australian *Nicotiana* section *Suaveolentes* have been detected using the Angiosperms353 bait set (Dodsworth et al. 2020). Custom-designed baits that incorporate faster evolving and less universally conserved genes (Soto Gomez et al. 2018) or expansion of the number of loci sampled using a RAD-Seq approach show promise here (Chase et al., submitted). There are, however, examples where both custom-designed baits (Larridon et al. 2020) and whole-genome sequencing approaches do not provide more resolved phylogenetic trees; these cases are due to (i) ancient hybridization events (Morales-Briones et al. 2021) and/or (ii) rapid radiation and substantial gene-flow amongst species, such as has been shown in tomatoes (Pease et al. 2016), eggplants (Page et al. 2019), and cichlids (Malkinsky et

al. 2018; Matschiner et al. 2020). Given the discordance patterns observed here, both processes are likely in the case of *Solanum*.

The deeper polytomy along the backbone of *Solanum* is more difficult to interpret because the older age of the node means extinction has had more time to affect the observed pattern (Louca and Pennell 2020). The gene discordance observed along the backbone, combined with the extremely short branches subtending the early diverging lineages in *Solanum*, could be attributed to several phenomena, such as ILS caused by events of elevated diversification rates, gene flow, hybridisation, or polyploidy, as has been argued elsewhere to explain the phylogenetic incongruence patterns observed in some of the early-diverging lineages of mammals (Simion et al. 2017), Neoaves (Suh, 2016), Amaranthaceae s.l. (Morales-Briones et al. 2021) and the Leguminosae (Koenen et al. 2020a, b). High speciation rates in early diverging lineages of *Solanum* have not been detected in diversification-rate studies (Echeverría-Londoño et al. 2020), which may not be surprising given the debate surrounding the accuracy of current methods for estimating diversification rates (Louca and Pennell 2020). Studies of key morphological traits across the entire *Solanum* phylogeny have had difficulty in pinpointing distinct synapomorphies that would clearly establish the relationships amongst the clades (Bohs 2005). Whilst there is a three-fold increase in genome size between the distantly related potato (*S. tuberosum* L., Potato Clade) and eggplant (*S. melongena*, Leptostemonum Clade; Barchi et al. 2019), there is currently no conclusive evidence for genome duplication along the backbone of *Solanum*. This is supported by the almost complete absence of paralogs detected in the TC dataset, save for one locus. Species trees from the TC, PL, and supermatrix do not suggest any obvious events of chloroplast capture or introgression. Chromosome counts indicate that the ancestor of *Solanum* was diploid with a large majority of *Solanum* species reported to be diploid

(>97% of the 506 species for which chromosome counts are available; Chiarini et al. 2018). Mapping of polyploidy across the phylogeny indicates that most of the early branching lineages of *Solanum* are diploid (e.g., Regmandra, and most clades within VANAns and Potato clade [except the Petota Clade itself]) and that polyploids have risen independently within the Archaesolanum, Petota, and Morelloid Clades, and three minor clades within the larger Leptostemonum Clade (Chiriani et al. 2018). In contrast, little is known about genome size and chromosome content evolution, with only 62 *Solanum* species recorded in the plant DNA C-value database (Pellicer and Leitch 2019), and 86 species studied with chromosome banding and/or FISH techniques (Chiarini et al. 2018). A broader taxon and genomic sampling will be needed to confidently exclude major roles for introgression, hybridisation, and polyploidy on the observed gene-tree discordance, and to determine whether the processes causing discordance change at different evolutionary depths across the phylogenetic tree (Knowles et al. 2018).

Is There A Single Bifurcating "Truth"?

The idea that "well-supported and fully bifurcating" phylogenies are a requisite for evolutionary studies is built on the premise that such trees are the accurate way of representing evolution. Only a limited set of methods exist for inferring non-bifurcating evolutionary relationships and trait evolution (Than et al. 2008; Solís-Lemus et al. 2017; Wen et al. 2018), and they are usually limited to fewer than 30 taxa due to their computational intensity (Cao et al. 2019, BioRxv). The shift in systematics from "tree "- to "bush "-like thinking, where polytomies and non-reticulate patterns of evolution are considered as acceptable or real (Poczai 2013; Mallet et al. 2016; Edelman et al. 2019), comes from the accumulation of studies finding similar unresolvable phylogenetic nodes, despite using different large-scale genomic sampling strategies and various analytical methods (Suh 2016). Discordances have been shown in several groups of

Solanaceae, *Nicotiana* (Dodsworth et al. 2020), the Capsiceae (*Capsicum* and relatives, Spalink et al. 2018), subtribe Iochrominae (Gates et al. 2018), *Jaltomata* (Wu et al. 2019), tomato and its wild relatives (Strickler et al. 2015; Pease et al. 2016), and potatoes (Huang et al. 2019). Our study, which focuses on relationships across *Solanum*, suggests this pattern is widespread in the family and adds to evidence supporting the prevalence of phylogenomic discordance in plants at both shallow and deeper phylogenetic depths (see for instance Wickett et al. 2014; Collier-Zans 2015; Sun et al. 2015; Folk et al. 2017; Crowl et al. 2017; Lane et al. 2018, Morales-Briones et al. 2018; Walker et al. 2019; Charr et al. 2020; Dupin et al. 2020; Stull et al. 2020).

The discordance discovered here does not affect the previously identified infrageneric groupings of *Solanum* and our analyses demonstrate that the major and minor clades are stable (e.g., Weese and Bohs 2007; Särkinen et al. 2013). The new supermatrix phylogeny that includes 60% of the extant species in the genus is a significant advance for building a detailed understanding of trait evolution across the genus, but the prevalence of hard polytomies has important implications for researchers interested in trait evolution and historical biogeography, as it has been argued that standard methods of trait evolution may incorrectly infer how traits evolve (Hahn and Nakhleh 2016). The discordance between traits, gene trees, and species trees has been defined as hemiplasy (Avise and Robinson 2008), and studies have shown that depending on the level of ILS present in the data, it can lead to different interpretations of convergent evolution of traits across a phylogenetic tree (Mendes et al. 2016). Our understanding of trait change will likely be advanced through additional in-depth phylogenomic studies in the group, as we try to link population-level evolutionary processes to patterns generated at a macro-evolutionary scale (e.g., Pease et al. 2016; Page et al. 2019). Scaling up these approaches to understand evolution of traits across large clades such as *Solanum* will not only require more

genomic data, but also filling in some basic natural history knowledge gaps, such as genome size and chromosome structure, which are poorly known for lineages such as the Thelopodium and Regmandra Clade and for species that are not close relatives of major commercial crops such as eggplant, tomato or potato. To address these challenges, continued efforts and field work will be required, as seed collections are required to provide high quality material for chromosome studies and accurate measures of genome sizes (but see Viruel et al. 2019). Acknowledging and embracing the uncertainty that underlies hard polytomies is crucial if we are to design research programs aimed at understanding the biology of large and rapidly radiating lineages.

SUPPLEMENTARY MATERIALS

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.[NNNN].

Scripts used in the analyses and for the production of figures will be made available at https://github.com/edgagnon/Solanum_Phylogenomics

FUNDING

This work was supported by the Fonds de recherche du Québec en Nature et Technologies postdoctoral fellowship and a grant from the Department of Biological Sciences of the University of Moncton to EG, the Sibbald Trust fellowship to RH, the Ceiba Foundation to AO, CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico awards 479921/2010-54 and 427198/2016-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES/FAPESPA award 88881.159124/2017-01 to LLG, NSF through grant DEB-0316614 “PBI Solanum: a worldwide treatment” to SK and LB, the Calleva Foundation & Sackler Trust (Plant and Fungal Trees of Life Project at Kew), the LUOMUS Trigger Fund to PP, and the OECD CRP Grant. Field sampling was supported by the Northern Territory

Herbarium and the David Burpee Endowment at Bucknell University (Australia), and National Geographic Society Northern Europe Award GEFNE49-12 (Peru, TS). Peruvian specimens were collected and sequenced under the permission of Ministerio de Agricultura, Dirección General Forestal y de Fauna Silvestre (collection permits 084-2012-AG-DGFFSDGEFFS and 096-2017-SERFOR/DGGSPFFS, and genetic resource permit 008-2014-MINAGRI-DGFFS/DGEFFS).

ACKNOWLEDGEMENTS

We thank Elliot Gardner for sharing scripts and advice on phylogenomic analyses with HybPiper, Royce Steeves for providing advice on DNA extraction for genome skimming, Felix Forest and Olivier Maurin for providing technical support and providing feedback on the manuscript, and João R. Stehmann, Thais Almeida, Paul Gonzáles, and Maria Baden who greatly contributed to fieldwork and sample acquisition.

AUTHOR CONTRIBUTIONS

EG designed and performed the analyses of the paper, with guidance from PP, AO, SD and TS; EG produced all figures, and wrote the article, with major contributions from TS, and PP, SD, SK and XA. RH and TS helped in data gathering and analyses. All other authors contributed data to the main analyses. All authors read and contributed to the final version of the manuscript.

LITERATURE CITED

1KP (One Thousand Plant Transcriptomes Initiative). 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574: 679-685.

570 APG (Angiosperm Phylogeny Group). 1998. An ordinal classification for the families of
571 flowering plants. Ann. Mo. Bot. Gard. 85: 531–553.

572 Avise J.C., Robinson T.J. 2008. Hemiplasy: a new term in the lexicon of
573 phylogenetics. Syst. Biol. 57(3): 503–507.

574 Barchi L., Pietrella M., Venturini L., Minio A., Toppino L., et al. 2019. A chromosome-
575 anchored eggplant genome sequence reveals key events in Solanaceae evolution. Sci. Rep. 9:
576 11769.

577 Bohs L. 2005. Major clades in *Solanum* based on *ndhF* sequence data. In: Keating R.C.,
578 Hollowell V.C., Croat T.B., editors. A festschrift for William G. D'Arcy: the legacy of a
579 taxonomist. St. Louis: Missouri Botanical Garden Press. p. 27–49.

580 Cao Z., Liu X., Ogilvie H.A., Yan Z., Nakhleh L. 2019. Practical Aspects of
581 Phylogenetic Network Analysis Using PhyloNet. bioRxiv 746362; doi: /10.1101/746362

582 Charr J-C, Garavito A, Guyeux C, Crouzillat D, Descombes P, Fournier C, Ly SN,
583 Raharimalala EN, Rakotomalala J-J, Stoffelen P, et al. 2020. Complex evolutionary history of
584 coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular
585 emphasis on Coffea canephora (Robusta coffee). Mol Phylogenet Evol. 151:106906.

586 Chiarini F., Sazatornil F., Bernardello G. 2018. Data reassessment in a phylogenetic
587 context gives insight into chromosome evolution in the giant genus *Solanum* (Solanaceae).
588 System. Biodiver. 16(4): 397-416.

589 Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent
590 model. Bioinformatics 30: 3317–3324.

591 Chou J., Gupta A., Yaduvanshi S., Davidson R., Nute M., Mirarab S., Warnow T. 2015.
592 A comparative study of SVDquartets and other coalescent-based species tree estimation
593 methods. BMC Genomics 16: S2, doi: /10.1186/1471-2164-16-S10-S2

594 Collier-Zans E. 2015. Recombination in the chloroplasts of the florally diverse Andean
595 subtribe Iochrominae (Solanaceae). Undergraduate Honours Thesis, University of Colorado
596 Boulder, USA

597 Crowl A.A., Myers C., Cellinese N. 2017. Embracing discordance: Phylogenomic
598 analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean
599 *Campanula* (Campanulaceae) clade. Evolution 71(4): 913-922. doi: 10.1111/evo.13203.

600 Degnan J.H., Rosenberg N.A. 2009. Gene-tree discordance, phylogenetic inference and
601 the multispecies coalescent. Trends Ecol. Evol. 24(6): 332–340.

602 Dodsworth S., Christenhusz M.J.M., Conran J.G., Guignard M.S., Knapp S., Struebig M.,
603 Leitch A.R., Chase M.W. 2020. Extensive plastid-nuclear discordance in a recent radiation
604 of *Nicotiana* section *Suaveolentes* (Solanaceae). Bot. J. Linn. Soc. 193(4): 546–559.

605 Dupin J, Raimondeau P, Hong-Wa C, Manzi S, Gaudeul M, Besnard G. 2020. Resolving
606 the Phylogeny of the Olive Family (Oleaceae): Confronting information from organellar and
607 nuclear genomes. Genes. 11(12). doi:10.3390/genes11121508.
608 http://dx.doi.org/10.3390/genes11121508.

609 Duvall M.R., Burke S.V., Clark D.C. 2020. Plastome phylogenomics of Poaceae:
610 alternate topologies depend on alignment gaps. Bot. J. Linn. Soc. 192(1): 9–20.

Echeverría-Londoño S., Särkinen T., Fenton I.S., Purvis A., Knapp S. 2020. Dynamism and context-dependency in diversification of the megadiverse plant genus *Solanum* (Solanaceae). J. Syst. Evol 58(6): 767-782, https://doi.org/10.1111/jse.12638.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. Science. 366(6465):594–599.

Edgar R.C. 2010. USEARCH v.11. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19): 2460–2461.

Gates D.J., Pilson D., Smith S.D. 2018. Filtering of target sequence capture individuals facilitates species tree construction in the plant subtribe Iochrominae (Solanaceae). Mol. Phylogenet. Evol. 123: 26–34.

Gonçalves D.J.P., Jansen R.K., Ruhlman T.A., Mandel J.R. 2020. Under the rug: Abandoning persistent misconceptions that obfuscate organelle evolution. Mol. Phylogenet. Evol. 151:106903. doi: 10.1016/j.ympev.2020.106903.

Gonçalves D.J.P., Simpson B.B., Shimizu G.H., Jansen R.K., Ortiz E.M. 2019. Genome assembly and phylogenomic data analyses using plastid data: Contrasting species tree estimation methods. Data Brief 25: 104271. doi: 10.1016/j.dib.2019.104271

Hahn M.W., Nakhleh, L. 2016. Irrational exuberance for resolved species trees. Evolution 70: 7–17. doi: 10.1111/evo.12832.

Hardigan M.A., Laimbeer F.P.E., Newton L., Crisovan E., J.P. Hamilton, B. Vaillancourt, K. Wiegert-Rininger, Wood J.C., Douches D.S., Farré E.M., Veilleux R.E., Buell

C.R. 2017. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc. Natl. Acad. Sci. U.S.A. 114(46): E9999–E10008. doi: 10.1073/pnas.1714380114

Hime P.M., Lemmon A.R., Lemmon E.C.M., Prendini E., Brown J.M., Thomson R.C., Kratovil J.D., Noonan B.P., Pyron R.A., Peloso P.L.V., Kortyna M.L., Keogh J.S., Donnellan S.C., Mueller R.L., Raxworthy C.J., Kunte K., Ron S.R., Das S., Gaitonde N., Green D.M., Labisko J., Che J., Weisrock D.W. 2020 [2021]. Phylogenomics reveals ancient gene-tree discordance in the amphibian tree of life. Syst. Biol. 70(1): 49-66, doi: 10.1093/sysbio/syaa034

Huang B., H. Ruess H., Liang Q., Colleoni C., Spooner D.M. 2019. Analyses of 202 plastid genomes elucidate the phylogeny of *Solanum* section *Petota*. Sci. Rep. 9: 4454. doi: /10.1038/s41598-019-40790-5

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23(2): 254-267.

Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends. Genet. 22(4): 225–231.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant. Sci. 4: 1600016. doi: 10.3732/apps.1600016

Johnson M.G., Pokorny L., Dodsworth S., Botigué L.R., Cowan R.S., Devault A., Eiserhardt L.W., Epitawalage N., Forest F., Kim J.T., Leebens-Mack J.H., Leitch I.J., Maurin O.,

Soltis D.E., Soltis P.S., Wong G.K., Baker W.J., Wickett N.J. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-Medoids clustering. Syst. Biol. 68(4): 594–606. doi: 10.1093/sysbio/syy086

Jombart T., Kendall M., Almagro-Garcia J., Colijn C. 2017. Treespace: statistical exploration of landscapes of phylogenetic trees. Mol. Ecol. Resour. 17: 1385–1392. doi: 10.1111/1755-0998.12676

Kendall M., Colijn C. 2016. Mapping phylogenetic trees to reveal distinct patterns of evolution. Mol. Biol. Evol. 33(10): 2735–2743.

Knowles L.L., Huang H., Sukumaran J., Smith S.A. 2018. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. Am. J. Bot. 105: 376-384.

Koenen E.J.M., Ojeda D.I., Steeves R., Migliore J., Bakker F.T., Wieringa J.J., Kidner C., Hardy O.J., Pennington R.T., Bruneau A., Hughes C.E. 2020a. Large‐scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near‐simultaneous evolutionary origin of all six subfamilies. New Phytol. 225: 1355-1369. doi : 10.1111/nph.16290

Koenen E.J.M., Ojeda D.I., Bakker F.T., Wieringa J.J., Kidner C., Hardy O.J., Pennington R.T., Herendeen P.S., Bruneau A., Hughes C.E. 2020b. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. Syst. Biol. syaa041. doi: 10.1093/sysbio/syaa041

673 Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics
674 and truth in phylogenomics. Mol Biol Evol. 29(2):457–472.

675 Kuramae E.E., Robert V., Snel B., Weiß M., Boekhout T. 2006. Phylogenomics reveal a
676 robust fungal tree of life. FEMS Yeast Res. 6(8): 1213–1220. doi: 10.1111/j.1567-
677 1364.2006.00119.x

678 Lane AK, Augustin MM, Ayyampalayam S, Plant A, Gleissberg S, Di Stilio VS,
679 Depamphilis CW, Wong GK-S, Kutchan TM, Leebens-Mack JH. 2018. Phylogenomic analysis
680 of Ranunculales resolves branching events across the order. Bot J Linn Soc. 187(2):157–166.

681 Larridon I, Villaverde T, Zuntini AR, Pokorny L, Brewer GE, Epitawalage N, Fairlie I,
682 Hahn M, Kim J, Maguilla E, et al. 2019. Tackling rapid radiations with targeted squencing.
683 Front. Plant Sci. 10:1655.

684 Levin R.A., Myers N.R., Bohs L. 2006. Phylogenetic relationships among the 'spiny
685 solanums' (*Solanum* subgenus *Leptostemonum*, Solanaceae). Am. J. Bot. 93: 157–169.

686 Li Y., Colleoni C., Zhang J., Liang Q., Hu Y., Ruess H., Simon R., Liu Y., Liu H., Yu G.,
687 Schmitt E., Ponitzki C., Liu G., Huang H., Zhan F., Chen L., Huang Y., Spooner D., Huang B.
688 2018. Genomic analyses yield markers for identifying agronomically important genes in potato.
689 Mol. Plant 11(3): 473–484. doi: 10.1016/j.molp.2018.01.009

690 Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for
691 estimating phylogenetic trees. Mol. Phylogenet. Evol. 53(1): 320–328.

692 Louca S., Pennell MW. 2020. Extant timetrees are consistent with a myriad of
693 diversification histories. Nature 580: 502-505, doi: 10.1038/s41586-020-2176-1

694 Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018.
695 Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene
696 flow. Nat Ecol Evol. 2(12):1940–1955.

697 Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? Bioessays.
698 https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201500149.

699 Matschiner M, Böhne A, Ronco F, Salzburger W. 2020. The genomic timeline of cichlid
700 fish diversification across continents. Nat Commun. 11(1):5895.

701 Mendes F.K., Hahn Y., Hahn M.W. 2016. Gene-tree discordance can generate patterns of
702 diminishing convergence over time. Mol. Biol. Evol. 33(12): 3299–3307, doi:
703 10.1093/molbev/msw197

704 Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler
705 A., Lanfear R. 2020a. IQ-TREE 2: new models and efficient methods for phylogenetic inference
706 in the genomic era. Mol. Biol. Evol. 37(5): 1530–1534. doi: 10.1093/molbev/msaa015

707 Minh B.Q., Hahn M.W., Lanfear R. 2020b. New methods to calculate concordance
708 factors for phylogenomic datasets. Mol. Biol. Evol. 37(9): 2727–2733. doi:
709 10.1093/molbev/msaa106.

710 Morgan C.C., Foster P.G., Webb A.E., Pisani D., McInerney J.O., O'Connell MJ. 2013.
711 Heterogeneous models place the root of the placental mammal phylogeny. Mol. Biol. Evol. 30:
712 2145–2156.

713 Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF,
714 Timoneda A, Yim WC, Cushman JC, Yang Y. 2021. Disentangling sources of gene tree

715 discordance in phylogenomic sata sets: Testing ancient hybridizations in Amaranthaceae s.l. Syst
716 Biol. 70(2):219–235.

717 Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep
718 history of hybridization and polyploidy in the Neotropical genus Lachemilla (Rosaceae). New
719 Phytol. 218(4):1668–1684.

720 Page A, Gibson J, Meyer RS, Chapman MA. 2019. Eggplant domestication: Pervasive
721 gene flow, feralization, and transcriptomic divergence. Mol Biol Evol. 36(7):1359–1372.

722 Pease J.B., Haak D.C., Hahn M.W., Moyle L.C. 2016. Phylogenomics reveals three
723 sources of adaptive variation during a rapid radiation. PLoS Biol. 14(2): e1002379. doi:
724 10.1371/journal.pbio.1002379

725 Pellicer J., Leitch I.J. 2020. The Plant DNA C‐values database (release 7.1): an updated
726 online repository of plant genome size data for comparative studies. New Phytol. 226: 301–305.
727 doi : 10.1111/nph.16261

728 Philippe H., Vienne D.M. de, Ranwez V., Roure B., Baurain D., Delsuc F. 2017. Pitfalls
729 in supermatrix phylogenomics. Eur. J. Taxon. 283. doi : 10.5852/ejt.2017.283

730 Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G.,
731 Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not
732 enough. PLoS Biol. 9(3): e1000602. doi: 10.1371/journal.pbio.1000602

733 Poczai P. 2013. To network or not to network that is the question. J. Genet. 92: 703–705.
734 doi: 10.1007/s12041-013-0293-4

735 Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53(1–
736 2): 131-147.

737 Romiguier J., Ranwez V., Delsuc F., Galtier N., Douzery E.J. 2013. Less is more in
738 mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of
739 placental mammals. Mol. Biol. Evol. 30: 2134–2144.

740 Ronco F., Matschiner M., Böhne A., Boila A., Büscher H.H., El Taher A., Indermaur A.,
741 Malinsky M., Ricci V., Kahmen A., Jentoft S., Salzburger W. 2021. Drivers and dynamics of a
742 massive adaptive radiation in cichlid fishes. Nature 589: 76-81, doi: 10.1038/s41586-020-2930-
743 4

744 Saarela J.M., Burke S.V., Wysocki W.P., Barrett M.D., Clark L.G., Craine J.M., Peterson
745 P.M., Soreng R.J., Vorontsova M.S., Duvall M.R. 2018. A 250 plastome phylogeny of the grass
746 family (Poaceae): topological support under different data partitions. PeerJ 6: e4299. doi:
747 /10.7717/peerj.4299

748 Särkinen T., Bohs L., Olmstead R.G., Knapp S. 2013. A phylogenetic framework for
749 evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. 13**:**
750 214. doi: 10.1186/1471-2148-13-214

751 Sayyari E., Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using
752 quartet frequencies. Genes 9(3): 132. doi: 10.3390/genes9030132.

753 Simion P., Philippe H., Baurain D., Jager M., Richter D.J., Di Franco A., Roure B., Satoh
754 N., Queinnec E., Ereskovsky A., Lapebie P. 2017. A large and consistent phylogenomic dataset
755 supports sponges as the sister group to all other animals. Curr. Biol. 27: 958–967.

Smith S.D., Pennell M.W., Dunn C.W., Edwards S.V. 2020. Phylogenetics is the New Genetics (for Most of Biodiversity). Trends Ecol. Evol. 35(5): 415-425. doi: 10.1016/j.tree.2020.01.005.

Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol Biol. 15:150.

Solís-Lemus C., P. Bastide, C. Ané. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34(12): 3292–3298. doi: 10.1093/molbev/msx235.

Soto Gomez M, Pokorny L, Kantar MB, Forest F, Leitch IJ, Gravendeel B, Wilkin P, Graham SW, Viruel J. 2019. A customized nuclear target enrichment approach for developing a phylogenomic baseline for Dioscorea yams (Dioscoreaceae). Appl Plant Sci. 7(6):e11254.

Spalink D., Stoffel K., Walden G.K., Hulse-Kemp A.M., Hill T.A., Van Deynze A., Bohs L. 2018. Comparative transcriptomics and genomic patterns of discordance in Capsiceae (Solanaceae). Mol. Phylogenet. Evol. 126: 293–302. doi: 10.1016/j.ympev.2018.04.030.

Stern S., Agra M.F. de, Bohs L. 2011. Molecular delimitation of clades within New World species of "spiny solanums" (*Solanum* subg. *Leptostemonum*). Taxon 60: 1429–1441.

Strickler S.R., Bombarely A., Munkvold J.D., York T., Menda N., Martin G.B., Mueller L.A. 2015. Comparative genomics and phylogenetic discordance of cultivated tomato and close wild relatives. PeerJ 3: e793. doi: 10.7717/peerj.793

775 Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. 2020. Nuclear
776 phylogenomic analyses of asterids conflict with plastome trees and support novel relationships
777 among major lineages. Am J Bot. 107(5):790–805.

778 Suh A., Smeds L., Ellegren H. 2015. The dynamics of incomplete lineage sorting across
779 the ancient adaptive radiation of neoavian birds. PLoS Biol. 13: e1002224.

780 Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root
781 of Neoaves. Zool. Scr. 45: 50–62.

782 Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and
783 reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9: 322. doi:
784 10.1186/1471-2105-9-322

785 Tepe E.J., Anderson G.J., Spooner D.M., Bohs L. 2016. Relationships among wild
786 relatives of the tomato, potato, and pepino. Taxon 65: 262–276.

787 Villaverde T, Jiménez-Mejías P, Luceño M, Waterway MJ, Kim S, Lee B, Rincón-
788 Barrado M, Hahn M, Maguilla E, Roalson EH, Hipp AL, THE GLOBAL CAREX GROUP.
789 2020. A new classification of *Carex* (Cyperaceae) subgenera supported by a HybSeq backbone
790 phylogenetic tree. Bot. J. Linn. Soc. 194(2): 141–163.

791 Viruel J, Conejero M, Hidalgo O, Pokorny L, Powell RF, Forest F, Kantarm MB, et al.
792 2019. A target capture-based method to estimate ploidy from herbarium specimens. Front. Plant
793 Sci. 10: 937.

794 Walker J.F., Walker-Hale N., Vargas O.M., Larson D.A., Stull G.W. 2019.
795 Characterizing gene tree conflict in plastome-inferred phylogenies. PeerJ 7: e7747. doi:
796 10.7717/peerj.7747

797 Weese T.L., Bohs L. 2007. A three-gene phylogeny of the genus *Solanum* (Solanaceae).
798 Syst. Bot. 32: 445–463.

799 Wen D., Yu Y., Zhu J., Nakhleh L. 2018. Inferring phylogenetic networks using
800 PhyloNet. Syst. Biol. 67(4): 735–740. doi: 10.1093/sysbio/syy015

801 Wendel J.F., Doyle J.J. 1998. Phylogenetic incongruence: window into genome history
802 and molecular evolution. In: Soltis D.E., Soltis, P.S. Doyle J.J. (eds.) Molecular systematics of
803 plants II: DNA sequencing. Boston: Sringer, 265-296.

804 Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N.,
805 Ayyampalayam S., Barker M.S., Burleigh J.G., Gitzendanner M.A., Ruhfel B.R., Wafula E., Der
806 J.P., Graham S.W., Mathews S., Melkonian M., Soltis D.E., Soltis P.S., Miles N.W., Rothfels
807 C.J., Pokorny L., Shaw A.J., DeGironimo L., Stevenson D.W., Surek B., Villarreal J.C., Roure
808 B., Philippe H., Pamphilis C.W. de, Chen T., x Deyholos T., Baucom R.S., Kutchan T.M.,
809 Augustin M.M., Wang J., Zhang Y., Tian Z., Yan Z., Wu X., Sun X., Wong G.K., Leebens-Mack
810 J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc.
811 Natl. Acad. Sci. U.S.A. 111 (45): E4859-E4868. doi: 10.1073/pnas.1323926111

812 Wu M., Kostyun J.L., Moyle L.C. 2019. Genome sequence of *Jaltomata* addresses rapid
813 reproductive trait evolution and enhances comparative genomics in the hyper-diverse
814 Solanaceae. Genome Biol. Evol. 11(2): 335–349.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19(S6): 153. doi: 10.1186/s12859-018-2129-y.

SUPPLEMENTARY INFORMATION

Supplementary Material and Methods

Table S1. Supermatrix sample information, including voucher details and Genbank numbers for sequences used.

*** (All sequences submitted, save for three).

Table S2. Plastome (PL) sample information, including voucher details and plastome assemblies results. Total length, as well as length for the long-single copy region (LSC), the short-single copy region (SSC) and the two inverted repeat regions (IR1 and IR2) is shown; statistics of mean coverage per base pair and standard deviation are also provided.

*** (Submission in progress in Genbank).

Table S3. Target capture (TC) sample information, including voucher details and sequence recovery statistics. The number of reads (NumReads), the number of reads mapped to the targets (ReadsMapped), the percentage of reads on target (PctOnTarget), the number of genes with reads (GenesMapped), the number of genes with contigs (GenesWithContigs), (GenesWithSeqs_, GenesAt25pct, GenesAt50pct, GenesAt75pct, GenesAt150pct, and the number of genes with paralog warnings (ParalogWarnings) is shown.

Table S4. Statistics for plastome alignments. Data shows number of sequences, trimming mode, the number of loci retained for coalescent analysis after checking for excessive gene tree branch lengths, alignment length, number of informative and constant sites, pairwise identity, average GC content, percentage of gaps, and average locus length for the exon, intron and intergenic regions.

	Exons (ex)		**Introns (in)**		**Intergenic (it)**		**Combined plastome (al)**	
Number of taxa	151	125	151	125	151	125	151	125
Trimming mode	visual	visual	visual	visual	strict (trimAl)	strict (TrimAl)	strict & visual	strict & visual
Total sites	64,117	64,117	15,881	15,881	51,358	51,358	131,356	131,356
Parsimony informative sites	3,561 (5.6%)	3,197 (4.9%)	1,496 (9.4%)	1,364 (8.6%)	4,309 (8.4%)	3,984 (7.7%)	9,359 (7.1%)	8,545 (6.5%)
Constant sites	56,604 (88%)	57,665 (90%)	12,864 (81%)	13,187 (83%)	42,024 (82%)	42,898 (84%)	111,523 (85%)	113,731 (87%)
Pairwise identity (%)	94	99	90	98	90	97	91	97
Average GC content (%)	38.8	38.8	34.7	34.6	35.6	35.4	37.1	36.8
Gaps (%)	4	2	14	10	6	2	6.5	10
Average locus length	823	823	1,058	1,058	740	740	816	816
Loci retained for ASTRAL analysis	76	76	15	15	69	69	160	160

Table S5. Comparison of branch support values for monophyly and sister relationships of key clades and species in *Solanum,* across the 48 species trees. Branch support for maximum likelihood (RaxML and IQ-TREE) and SVDQuartet analyses are bootstrap support values in percentages, and for ASTRAL analyses are in local posterior probability values. Focus is given on monophyly and relationships between main clades in Solanum mentioned in the text. Abbreviations used in the phylogenetic analyses follow figure 1 in the main text: ex=exon, in=intron, it=intergenic region. Branch support ≥50 bootstrap (BS) and ≥0.5 posterior probability values are shown and colour coded: black = ≥95 BS/0.95 PP; dark grey = 75-94 BS/0.75-0.94 PP; light grey = 50-74 BS/0.5-0.74 PP. NA indicates nodes that were not tested due to taxon sampling. * excluding Thelopodium Clade; ** excluding S. anomalostemon; *** includes Pteroidea Clade in ASTRAL intron trees with partial and full plastomes (151 species); **** excludes *S. virginianum* in some analyses of TC dataset with low min. loci threshold.

Table S6. Supermatrix alignment details, with details about the nine regions selected for this study. Number of species sampled per region, accumulative percentage of species sampled per region, aligned length, proportions of parsimony informative characters (PI), and variable sites (VS) per region in the dataset are indicated. Values are calculated with outgroups, and with ambiguous regions and repeats excluded. bp=base pairs.

Region	**Genome**	**Species (% of total)**	**% of total**	**Aligned length before exclusion (bp)**	**Excluded bp**	**Aligned length after exclusion (bp)**	**PI**	**PI of total (%)**	**VS**	**VS of total (%)**
trnT-L	Plastid	622	84	2,077	164	1,913	593	31	928	45
ITS	Nuclear ribosomal	609	82	881	355	526	352	67	465	88
waxy	Nuclear	532	72	1,792	81	,711	923	54	1,195	70
trnS-G	Plastid	334	45	932	141	791	173	22	330	42
ndhF	Plastid	261	35	2,088	0	2,088	344	16	581	28
ndhF-rpl32	Plastid	261	35	899	16	883	232	26	404	46
matK	Plastid	242	32	1,148	0	1,148	220	19	409	36
rpL32-trnL	Plastid	206	27	1,377	124	1,253	282	23	494	39
psbA-trnH	Plastid	163	22	650	55	595	145	24	238	40
Total		746	--	11,844	936	10,908	3,263	30	5,044	46
Plastid only		678	91	9,171	500	8,671	1,989	23	3,384	39

Table S7. List of polyploid taxa in *Solanum.*

Table S8. Target capture (TC) alignment statistics. Loci excluded refer to the number of excluded loci based on excessively long branch lengths, and loci retained is the final number of loci retained for both ML and coalescent analyses. Empty sequences inserted refers to amount of missing data. Min = minimum; Bp = base pairs.

	TC Alignment 1 (min4)	**TC Alignment 2 (min10)**	**TC Alignment 3 (min20)**
Species	40	40	40
Mininum number of species per locus	4	10	20
Loci retained for ML analysis	348	337	310
Loci excluded	10	7	7
Loci retained for Astral analysis	338	330	303
Length of concatenated sequences (bp)	261,975	257,519	244,272
Empty sequences inserted	3,147	2,781	2,092

Table S9. Optimal substitution model used in ML analyses for the PL and TC datasets, determined using ModelFinder in IQ-TREE2. For each loci, the number of taxa, sites, informative sites, and invariable sites are indicated, as well as the model selected and the AICc score. Worksheet titles correspond to the following: PLUnpartitioned = Models selected for PL unpartitioned datasets (exons, introns, intergenic regions and combined, for 151 taxa and 125 taxa); PLByLociEx125 = Models selected for PL exons only, 125 taxa, partitioned-by-loci; PLByLociIn125 = Models selected for PL introns only, 125 taxa, partitioned-by-loci; PLByLociIt125 = Models selected for PL intergenic regions only, 125 taxa, partitioned-by-loci; PLByLociAll125 = Models selected for PL combined regions, 125 taxa, partitioned-by-loci; PLByLociEx151 = Models selected for PL exons only, 151 taxa, partitioned-by-loci; PLByLociIn151 = Models selected for PL introns only, 151 taxa, partitioned-by-loci; PLByLociIt151 = Models selected for PL intergenic regions only, 151 taxa, partitioned-by-loci; PLByLociAll151 = Models selected for PL combined regions, 151 taxa, partitioned-by-loci; PLBestPartScheme = Models selected for PL datasets analysed according to the best-partition scheme; TCPartitioned_Min04= Models selected for loci of the TC dataset, with minimum 4 taxa per loci; TCPartitioned_Min10= Models selected for loci of the TC dataset, with minimum 20 taxa per loci; TCPartitioned_Min20= Models selected for loci of the TC dataset, with minimum 20 taxa per loci.

Figure S1. Detailed RaxML of supermatrix phylogenetic tree with 746 taxa. Nodes with bootstrap support equal and above 95% are in cyan, and with branch support between 75% and 94% in red. Bootstrap support values for each node indicated in italic. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1.

Figure S2. Detailed Bayesian inference (Beast) supermatrix phylogenetic tree with 746 taxa. Nodes with posterior probability equal and above 0.95 are in cyan, and nodes with posterior probabilities between 0.75 and 0.95 are in red. Posterior probability values for each indicated in italic. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1.

Figure S3. ML results for each of the nine individual loci and combined plastid loci: (a) ITS ; (b) *matK*; (c) *ndhF*; (d) *ndhF-rpL32*; (e) *psbA-trnH;* (f) *rpL32-trnL*; (g) *trnL-trnT*; (h) *trnS-trnG*; (i) *waxy*; (j) seven plastid loci. Nodes with bootstrap support equal and above 95% are in cyan, and with branch support between 75% and 94% in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1.

Figure S4. BI results for each of the nine individual loci and combined plastid loci: (A) ITS *; (B) matK; (C) ndhF, (D) ndhF-rpL32, (E) psbA-trnH (F) rpL32-trnL; (G) trnL-trnT; (H) trnS-trnG; (I) waxy*; (J) seven plastid loci. Nodes with posterior probability equal and above 0.95 are in cyan, and nodes with posterior probabilities between 0.75 and 0.95 are in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1.

Figure S5. ML phylogenetic trees of plastome datasets. Nodes with bootstrap support equal and above 95% are in cyan, and with branch support between 75% and 94% in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1. a) 151 taxa, exons only, unpartitioned; b) 125 taxa, exons only, unpartitioned; c) 151 taxa, exons only, partitioned by loci; d) 125 taxa, exons only, partitioned by loci; e) 151 taxa, exons only, best partition scheme; f) 125 taxa, exons only, best partition scheme; g) 151 taxa, Introns only, unpartitioned; h) 125 taxa, Introns only, unpartitioned; i) 151 taxa, Introns only, partitioned by loci; j) 125 taxa, Introns only, partitioned by loci; k) 151 taxa, Introns only, best partition scheme; l) 125 taxa, Introns only, best partition scheme; m) 151 taxa, intergenic regions, unpartitioned; n) 125 taxa, intergenic regions, unpartitioned; o) 151 taxa, intergenic regions, partitioned by loci; p) 125 taxa, intergenic regions, partitioned by loci; q) 151 taxa, intergenic regions, best partition scheme; r) 125 taxa, intergenic regions, best partition scheme; s) 151 taxa, all data, unpartitioned; t) 125 taxa, all data, unpartitioned; u) 151 taxa, all data, partitioned by loci; v) 125 taxa, all data, partitioned by loci; w) 151 taxa, all data, best partition scheme; x) 125 taxa, all data, best partition scheme,

Figure S6. ASTRAL-III phylogenetic trees of plastome datasets. Nodes with multi-locus local posterior probability support equal and above 0.95 are in cyan, and with support between 0.75 and 0.94 in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1. a) 151 taxa, exons only; b) 125 taxa, exons only; c) 151 taxa, introns only; d) 125 taxa, introns only; e) 151 taxa, intergenic regions; f) 125 taxa, intergenic regions; g) 151 taxa, all data; h) 125 taxa, all data.

Figure S7. SVDquartets phylogenetic trees of plastome datasets. a) 151 taxa, exons only; b) 125 taxa, exons only; c) 151 taxa, introns only; d) 125 taxa, introns only; e) 151 taxa, intergenic regions; f) 125 taxa, intergenic regions; g) 151 taxa, all data; h) 125 taxa, all data.

Figure S8. ML phylogenetic trees of A353 target capture datasets (IQ-TREE2). Nodes with bootstrap support equal and above 95% are in cyan, and with branch support between 75% and 94% in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1. a) filtering threshold of a minimum of 4 taxa per loci; b) IQ-TREE2, filtering threshold of a minimum of 10 taxa per loci; c) filtering threshold of a minimum of 20 taxa per loci.

Figure S9. Coalescent phylogenetic trees of A353 target-capture datasets (ASTRAL-III). Nodes with multi-locus local posterior probability support equal and above 0.95 are in cyan, and with support between 0.75 and 0.94 in red. Tips indicate species names, followed by major and/or minor clade, as indicated in Table 1. a) filtering threshold of minimum of 4 taxa per loci; b) filtering threshold of minimum of 10 taxa per loci; c) filtering threshold of minimum of 20 taxa per loci;

Figure S10. Polytomy test results with ASTRAL-III. a) Target Capture A353 species tree ASTRAL-III, filtering threshold of minimum 4 taxa per loci, branches in gene trees with 10% or less branch support collapsed; b) Target Capture A353, ASTRAL-III, filtering threshold of minimum 4 taxa per loci, branches in gene trees with 75% or less branch support collapsed; c) Target Capture A353, ASTRAL-III, filtering threshold of minimum 10 taxa per loci, branches in gene trees with 10% or less branch support collapsed; d) Target Capture A353, ASTRAL-III, filtering threshold of minimum 10 taxa per loci, branches in gene trees with 75% or less branch support collapsed; e) Target Capture A353, ASTRAL-III, filtering threshold of minimum 20 taxa per loci, branches in gene trees with 10% or less branch support collapsed; f) Plastome, All Data, ASTRAL-III, 151 taxa, branches in gene trees with 10% or less branch support collapsed; g) Plastome, All Data, ASTRAL-III, 151 taxa, branches in gene trees with 75% or less branch support collapsed;

Datasets

File1: Newick treefile, with all 48 species trees (TreeBase)

File2**:** Newick treefile, with all 48 species trees pruned to 27 taxa (TreeBase).

File3: Alignment file of concatenated Sanger supermatrix, with partition between genes (746 taxa).

File4: Plastome Exons, Introns only, Intergenic regions used in phylogenetic analyses, 151 taxon alignment;

File5. Target Capture A353 alignments of 338 loci, filtering threshold of minimum 4 taxa per loci;

File6. Target Capture A353 alignments of 330 loci, filtering threshold of minimum 10 taxa per loci;

File7. Target Capture A353 alignments of 303 loci, filtering threshold of minimum 20 taxa per loci;

SCRIPTS

(Are being deposited in Github, in progress)

Rscripts for producing figures

Beast xml file

Scripts for Astral analyses

Scripts for IQtree analyses

Scripts for SVDQuartet analyses

Scripts for gCF and sCF calculations

Scripts for Polytomy tests