Robustness of RADseq for evolutionary network reconstruction from gene trees

José Luis Blanco-Pastor[a b], Yann J.K. Bertrand[a c], Isabel María Liberal[d], Yanling Wei[e], E. Charles Brummer[e] and Bernard E. Pfeil[a]

a. Department of Biological and Environmental Sciences, University of Gothenburg, Box 461 40530, Göteborg, Sweden.
b. INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), 86600 Lusignan, France
c. Institute of Botany, Czech Academy of Sciences, Zámek 1, 25243 Průhonice, Czech Republic
d. Real Jardín Botánico de Madrid (RJB-CSIC), Madrid, Spain
e. Plant Breeding Center, Department of Plant Sciences, University of California, Davis, Davis, CA, USA.

Correspondance
José Luis Blanco-Pastor,
INRA, Centre Nouvelle-Aquitaine-Poitiers, UR4 (URP3F), 86600 Lusignan, France.
Email: jose-luis.blanco-pastor@inra.fr

Keywords: RADseq; gene trees; evolutionary networks; hybridization; *Medicago*; alfalfa

Competing interest statement

Authors have no competing interest to declare

Abstract

Although hybridization has played an important role in the evolution of many species, phylogenetic reconstructions that include hybridizing lineages have been historically constrained by the available models and data. Recently, the combined development of high-throughput sequencing and evolutionary network models offer new opportunities for phylogenetic inference under complex patterns of hybridization in the context of incomplete lineage sorting. Restriction site associated DNA sequencing (RADseq) has been a popular sequencing technique for evolutionary reconstructions of close relatives in the Next Generation Sequencing (NGS) era. However, the utility of RADseq data for the reconstruction of complex evolutionary networks has not been thoroughly discussed. Here, we used new molecular data collected from diploid perennial *Medicago* species using single-digest RADseq to reconstruct evolutionary networks from gene trees, an approach that is computationally tractable with datasets that include several species and complex patterns of hybridization. Our analyses revealed that complex network reconstructions from RADseq-derived gene trees were not robust under variations of the assembly parameters and filters. Filters to exclusively select loci with high phylogenetic information created datasets that retrieved the most anomalous topologies. Conversely, alternative clustering thresholds or filters on the number of samples *per* locus affected the level of missing data but had a lower impact on networks. When most anomalous networks were discarded, all remaining network analyses consistently supported a hybrid origin for *M. carstiensis* and *M. cretacea*.

1. Introduction

The reconstruction of a reticulate history in evolutionary close relatives has been considered from three different analytical perspectives: i) population genetic models including: approximate Bayesian Computation (Beaumont et al., 2002), full-likelihood genealogical samplers that make use of DNA sequences (Gronau et al., 2011; Hey, 2010; Sethuraman and Hey, 2016) and likelihood or pseudo-likelihood methods based on the joint allele frequency spectrum (Excoffier et al., 2013; Gutenkunst et al., 2009; Pickrell and Pritchard, 2012); ii) *D*-statistics (Durand et al., 2011; Eaton and Ree, 2013; Green et al., 2010; Meyer et al., 2012; Pease and Hahn, 2015); and iii) evolutionary network models (Solís-Lemus and Ané, 2016; Wen and Nakhleh, 2016; Yu et al., 2014, 2013; Yu and Nakhleh, 2015; Zhang et al., 2018; Zhu et al., 2017). The first two perspectives assume a previously known backbone phylogeny to formulate a hypothesis of hybridization. This backbone tree is usually constructed either using i) a total evidence approach with concatenation of full sequence information (Eaton and Ree, 2013; Escudero et al., 2014; Fernández-Mazuecos et al., 2017; Hipp et al., 2014; Wagner et al., 2013) or ii) coalescent based methods (Eaton and Ree, 2013; Fernández-Mazuecos et al., 2017; Rheindt et al., 2014) that reconcile individual gene trees. Despite being a standard approach, the construction of a backbone tree could be an incorrect representation of the main evolutionary history of the species under complex reticulate evolution (Clark and Messer, 2015; Huson et al., 2010; Yu et al., 2011), or molecular data can show a different "main" phylogeny when hybridization is first accounted for (Sousa et al., 2017).

Evolutionary networks provide an explicit model of evolutionary relationships that extends the tree model to allow for reticulations with internal nodes representing ancestral species. Recently developed phylogenetic network reconstruction methods are based on maximum parsimony (Yu et al., 2013), maximum likelihood (ML) (Yu et al., 2014), maximum pseudo-likelihood (Solís-Lemus and Ané, 2016; Yu and Nakhleh, 2015) and Bayesian inference (BI) methods (Wen et al., 2016; Wen and Nakhleh, 2016; Zhang et al., 2018; Zhu et al., 2017). Although ML and BI methods show promise, they are still limited to small datasets (usually fewer than 10 individuals and less than 3 reticulations, Yu and Nakhleh, 2015). In contrast, maximum pseudo-likelihood (summary) methods (Solís-Lemus and Ané, 2016; Yu and Nakhleh, 2015) are nowadays a convenient alternative for complex empirical datasets (Wen et al., 2017).

RADseq approaches (reviewed in Andrews et al., 2016) are widely used sequencing techniques for evolutionary reconstructions in the Next Generation Sequencing era. RADseq was first envisioned as a technique to find intraspecific genetic variation (Baird et al., 2008; Elshire et al., 2011; Hohenlohe et al., 2011). Later RADseq methods have been considered suitable for phylogenetic studies from shallow to deep timescales (Cariou et al., 2013; Eaton, 2014; Harvey et al., 2016; Rubin et al., 2012). RADseq are particularly appealing for systematics because they are easily applied to non-model organisms for which no reference genome or previous genomic information is available (Cariou et al., 2013; Fernández-Mazuecos et al., 2017; Rubin et al., 2012). For that reason RADseq has become a very popular technique for hybridization studies across a diversity of organisms and timescales (Escudero et al., 2014; Fernández-Mazuecos et al., 2017). Nevertheless, because RADseq datasets are limited in sequence length, contain relatively few variable sites, and do not generally yield resolved gene trees (Rubin et al., 2012), it is unknown if they are appropriate for maximum pseudo-likelihood phylogenetic network reconstruction methods. Their intrinsic

characteristics suggest that these datasets are limited for network inference from gene trees, but an in-depth evaluation of their utility is still lacking.

Medicago L. (Fabaceae) is a genus comprising 87 species (Small, 2011) and includes the economically important forage crop alfalfa (*M. sativa*, section *Medicago*) in addition to the model legume *M. truncatula* (Barker et al., 1990; Benedito et al., 2008; Branca et al., 2011; Cook, 1999; Young et al., 2011). The genus *Medicago* L. exhibits severe phylogenetic gene tree incongruence that has been mainly attributed to hybridization and ILS (Eriksson et al., 2018, 2017; Maureira-Butler et al., 2008; Sousa et al., 2017, 2014; Steele et al., 2010; Yoder et al., 2013). We collected new molecular data from diploid perennial *Medicago* species using single-digest RADseq (Genotyping-By-Sequencing, Elshire et al., 2011). We investigated the ability of RADseq data to unveil the evolutionary history of diploid species of *Medicago* section *Medicago* using a network reconstruction method that uses gene trees (Yu and Nakhleh, 2015). Specifically, we investigated the robustness of this method (i.e. the propensity to retrieve a set of optimal networks with similar topologies) under a variety of RADseq data assembly parameters and filters.

2. Materials and Methods

2.1 Sampling

Our choice of species (Table 1) was based on results of previous studies grouping diploid perennial *Medicago* taxa (Bena, 2001; Maureira-Butler et al., 2008; Sousa, 2015; Yoder et al., 2013). These includes: *M. marina, M. cretacea, M. rhodopea, M. prostrata, M. daghestanica and M. sativa* (section *Medicago* subsection *Medicago*), *M. hybrida and M. suffruticosa* (section *Medicago* subsection *Suffruticosae*); *M. carstiensis* (section *Carstiensae*)*; M. rugosa and M. scutellata* (section *Spirocarpos* subsection *Rotatae*). As outgroup we used the annual species *M. truncatula* (section *Spirocarpos* subsection *Pachyspirae*).

2.2 Sequence preparation

We extracted genomic DNA with a custom CTAB DNA Extraction Protocol and constructed a genotyping-by-sequencing (GBS) library following the library preparation protocol of Elshire et al. (2011) with minor modifications as described by Annicchiarico et al. (2017). In brief, GBS library was prepared using the frequent cutter ApeKI (R0643L; NEB) restriction enzyme. Sets of 8-bp barcoded adapters were ligated to restriction fragments for multiplex sequencing. The QIAquick PCR purification kit (28104; QIAGEN) was used to purify equal volumes of the pooled ligated products previous to the final PCR amplification step with the Kapa Library Amplification Readymix (Kapa Biosystems KK2611. Sequences were obtained at the Genomic Core Facility of the UT Southwestern Medical Center (Dallas, TX) with an Illumina HiSeq 2500 system that generated 100-bp single-end reads. This protocol was chosen based on comparisons made among a few protocols and different enzymes, including the two-enzyme protocol by Poland et al. (2012) and the 2b-RAD protocol by Wang et al. (2012). The decision was made based on the number of sites genotyped that were shared among representative individuals (Annicchiarico et al., 2017).

Raw single-end sequence reads were trimmed of adapter sequence and filtered with a minimum quality score of 20 using trimmomatic (Bolger et al., 2014). Assembly was then performed using ipyrad v. 0.7.19 (http://ipyrad.readthedocs.io/), a toolbox for reproducible assembly and analysis of RADseq type genomic data sets based on the pyRAD pipeline (Eaton, 2014). Assembly consisted of seven sequential steps, with parameters based on those recommended for single-end GBS data in the ipyrad documentation. We used the *de novo* + reference method, with the *M. truncatula* genome sequence (Mt4.0,

http://www.medicagohapmap.org) as a reference. Briefly, the steps of the ipyrad pipeline are described as follows: In step 1, sequences were demultiplexed according to barcode sequences. In step 2, low quality reads and Illumina adapters were filtered out. Step 3 removed amplification duplicates and then clustered reads within each sample according to a clustering threshold. This step tries to identify all the reads that map to the same locus within each sample. As we used the *de novo* + reference method, the *M. truncatula* reference was used to identify homology, and then the remaining unmatched sequences were clustered with the standard *de novo* ipyrad pipeline. Because phylogenetic results are known to be sensitive to the similarity threshold employed in step 3 for within-sample and step 6 (see below) for across-sample sequence clustering (Fernández-Mazuecos et al., 2017; Leaché et al., 2015; Shafer et al., 2017; Takahashi et al., 2014), five assemblies of GBS loci were generated using a range of clustering thresholds (clust parameter) from c=0.75 to c=0.95 (Table 2). Step 4 jointly estimated the error rate and heterozygosity to differentiate "good" reads from sequencing errors. Step 5 called the consensus of sequences within each cluster. Step 6 clustered consensus sequences across samples. Step 7 filtered the data and wrote output files. In step 7 we applied filters for the maximum number of indels *per* locus (8), max heterozygosity *per* locus (50% of samples) and max number of SNPs *per* locus (20). To evaluate the effect of missing data on network inference, for each assembly we generated datasets with two alternative values for the minimum number of samples *per* locus ("minimum taxon coverage" -min- parameter, 4 and 10). The effect of locus variation on networks was tested by generating datasets with two alternative values for the minimum number of parsimony-informative sites (PIS parameter, 4 and 10). We saved the data in the ipyrad format (*.loci) that was later on transformed in individual alignment files *per* locus in the phylip format using a custom R script. We obtained 20 RADseq datasets under different combinations of assembly parameters and filters described above (Table 2).

2.3 Network inference

We analyzed RADseq datasets alignments with PhyloNet (Than et al., 2008; Wen et al., 2017). Within PhyloNet we applied the method that infers species networks from gene trees using maximum pseudo-likelihood (InferNetwork_MPL command; Yu and Nakhleh, 2015) which is computationally fast. First, we analyzed separate sets of genes from each of the 20 RADseq datasets with RaxML v.7.2.8 (Stamatakis, 2006) using the GTRCAT substitution model and using *M. truncatula* as outgroup. We sampled several individuals/alleles for some species (see Table 1) that were mapped to single taxa with the -a parameter. Ten optimal networks were returned with the -n parameter. We chose 5 maximum allowed number of reticulation events. Remaining parameter values were set as default.

2.4 Network distances

To investigate dissimilarities between evolutionary networks computed with alternative RADseq datasets we used multidimensional scaling. We first calculated a matrix of distances among networks computed with the topological dissimilarity measure of Nakhleh (2010) (normalized to get values within [0, 1]), which is implemented in PhyloNet. Then we applied a Principal Coordinate Analysis (PCoA) to transform the distance matrix into a set of coordinates that were plotted to display network distances. We performed the PCoA using all pairwise distances between every network returned by the PhyloNet analyses.

3. Results

3.1 Sequence capture and RADseq data

Among the 20 RADseq datasets the number of loci ranked from 4 (clust95.min10.PIS10) to 3,405 (clust85.min4.PIS4), concatenated length (bp) ranged from 367 (clust95.min10.PIS10) to 303,272 (clust85.min4.PIS4) and missing data (%) ranged from 16.2 (clust95.min10.PIS10) to 56.6 (clust75.min4.PIS10).

3.2 Phylogenetic networks

Best networks (networks with highest likelihood scores) for each of the 20 RADseq datasets showed marked differences (Fig. 1). A hybrid origin was recovered for all species (excluding the outgroup species, *M. truncatula*) at least in one of the 20 best species networks (Table 3): *M. carstiensis* (observed as hybrid in 18 networks), *M. cretacea* (in 16 networks), *M. rhodopea* (in 10 networks), *M. marina* (in 6 networks), *M. rugosa* (in 4 networks), *M. scutellata* (in 4 networks), *M. daghestenica* (in 4 networks), *M. suffruticosa* (in 4 networks), *M. prostrata* (in 2 networks), *M. hybrida* (in 2 networks) and *M. sativa* (in 2 networks).

3.3 Network distances

The RADseq datasets that retrieved the highest distances from the "core" set of networks were those that were computed with datasets filtered to contain only the most variable loci (PIS10 filter, see Fig. 2). In general, these datasets contained a low number of loci and short concatenated sequence lengths. The PCoA did not show a marked effect of the filter on the minimum number of samples *per* locus (min filter) or the use alternative clustering thresholds (clust parameter).

After excluding datasets with the PIS10 filter, a hybrid origin was recovered for eight species at least in one of the remaining 10 best species networks: *M. carstiensis* (observed as hybrid in all 10 networks), *M. cretacea* (in all 10 networks), *M. rugosa* (in 4 networks), *M. rhodopea* (in 3 networks), *M. marina* (in 2 networks), *M. suffruticosa* (in 2 network), *M. scutellata* (in 1 network) and *M. sativa* (in 1 network).

4. Discussion

Our empirical comparison among networks computed from the RADseq datasets reveal some general patterns in how assembly parameters and filters influence complex evolutionary network reconstructions from gene trees. Our study shows that RADseq datasets with a low number of loci retrieve the most atypical network topologies, regardless the high phylogenetic information contained in the loci. The RADseq networks that were the closest to the core set of networks were those that assembled the highest number of loci with very little impact on the clustering threshold or the minimum number of samples *per* locus and therefore with very little impact on the level of missing data. In general RADseq datasets showed low robustness (different best network topologies) under variation of the assembly parameters and filters. But, after excluding the most divergent networks, all remaining analyses supported a hybrid origin for two species: *M. carstiensis* and *M. cretacea*.

Recently Fernández-Mazuecos et al. (2017) showed a high robustness of coalescent approaches for RADseq-based species trees reconstructions. Contrastingly, here we observed variation among network topologies under variations of the assembly parameters and filters underlining the importance of RADseq data preparation on the final results. RADseq is a particularly appealing technique for systematics because of their potential for detecting both current and historical hybridization (Escudero et al., 2014; Twyford and Ennos, 2012) and

because they are easily applied with no previous genomic information and reduced lab costs. The pseudo-likelihood method of Yu and Nakhleh (2015) is also attractive because it does not require heavy computational resources. Nevertheless its use on RADseq data may produce misleading results without a proper evaluation of the optimal assembly parameters and filters. In phylogenetic analysis with RADseq, it is particularly challenging to establish general criteria for determining the assembly parameters that maximize the number of orthologous RAD sequences between samples and filtering parameters that retain loci with the optimal level of missing data or phylogenetic information. It has been suggested that low phylogenetic resolution of loci may constrain the identification of hybrids because poorly resolved gene trees, constructed from markers with limited sequence divergence between species, are likely to be uninformative in tracing the reticulate history of species (Linder and Rieseberg, 2004; Twyford and Ennos, 2012). In contrast our study suggests that high loci number increases the power for network inference from RADseq-gene trees despite the low phylogenetic information contained within each individual locus. Additionally, selectively choosing the most variable RADseq dataset may be detrimental as these loci may introduce potential biases typical of hypervariable regions of the genome. Indeed, the most variable regions could be those retaining ancestral polymorphisms or those representing regions of introgressed DNA (Eaton and Ree, 2013).

In recent years RADseq has been applied for the evolutionary reconstruction of complex taxonomic groups (Eaton and Ree, 2013; Escudero et al., 2014; Fernández-Mazuecos et al., 2017; Hipp et al., 2014; Wagner et al., 2013). Most previous studies using RADseq data relied on a "backbone tree" and placed a limited number of hybridization events upon it. Nevertheless it is known that this approach could provide an incorrect representation of the evolutionary history of the species under complex reticulate evolution with multiple hybridization events (Huson et al., 2010; Yu et al., 2011). New tools for evolutionary network reconstructions (Solís-Lemus and Ané, 2016; Wen et al., 2016; Wen and Nakhleh, 2016; Yu et al., 2014, 2013; Yu and Nakhleh, 2015; Zhang et al., 2018; Zhu et al., 2017) now offer the opportunity to study reticulate evolution including cases with multiple hybridization events and with no previous information on the "backbone tree" or where such main tree is potentially non-existent. Development of such evolutionary network models are now in full swing and should become standard methods for phylogenetic inference under incomplete lineage sorting (ILS) and hybridization. Despite these remarkable methodological advances, in the most complex cases computational limitations reduce the set of methods to those using maximum pseudo-likelihood inference of networks from gene trees (Solís-Lemus and Ané, 2016; Yu and Nakhleh, 2015). These methods have a great potential but there is no information in the literature about the adequacy of the commonly used RADseq datasets for the estimation of evolutionary networks using these type of analyses were a previous computation of gene trees is required. In general using RADseq for gene tree reconstruction poses a number of potential problems: the orthology relationships among sequences are unknown, mutations on restriction sites is expected to yield missing data that increases with evolutionary time and the genetic linkage relationships among loci are unknown (see Rubin et al., 2012). Additionally, given the short length of sequences, the phylogenetic information of each locus is very scarce and recombination detection is not straightforward.

Despite the varied result obtained with different RADseq datasets, general patterns emerged regarding the identification of hybrid species which were more evident when the PIS10 datasets were excluded. A hybrid origin was retrieved by all remaining PIS4 datasets for *M. carstiensis* and *M. cretacea*. This signal was clearly stronger than the hybridization signal detected for the remaining species (hybrid origin detected in ≤ 4 datasets for *M. rugosa*, *M.*

rhodopea, *M. suffruticosa*, *M. marina*, *M. scutellata* and *M. sativa*). *M. carstiensis* is the only *Medicago* species exclusively with rhizomes (which are found only sporadically in a few other species, especially *M. sativa*). Phylogenetic relationships around *M. carstiensis* has been enigmatic as it forms a monospecific section (Carstiensae, Small, 2011) and previous phylogenetic studies did not provide well-supported information on the relationships of this species (Maureira-Butler et al., 2008; Small, 2011). It has been speculated that *M. carstiensis* is a relic species that is ancestral to the much more widespread *M. orbicularis* (Bennett et al., 2006). But its particular characteristics could be also explained by speciation after disruptive selection on hybrids (Seehausen, 2004). Regarding *M. cretacea*, Urban (1873) (the first to prepare a comprehensive analysis of the genus *Medicago*) already considered that this species had controversial affinities. Lesins and Lesins (Lesins and Lesins, 1979), in the second comprehensive systematic analysis of *Medicago*, already included *M. cretacea* in the monotypic section *Cretaceae*. Later on, analyses by Bena (2001) and Maureira-Butler et al. (2008) showed alternative inconsistent phylogenetic relationships for *M. cretacea*. These contentious taxonomic and phylogenetic placement of *M. carstiensis* and *M. cretacea* observed in previous studies are consistent with hybridisation.

5. Conclusions

Here we inferred a hybrid origin for *M. carstiensis* and *M. cretacea* using RADseq data and a maximum pseudo-likelihood approach for network inference from gene trees. We observed that loci number had an important impact on network reconstruction from RADseq-gene trees, whereas the clustering threshold used in the data assembly or a filter on taxon coverage had a lower impact on network inference. Future research on methods that explore the parameter space for optimal assembly parameters and filters may be required to obtain a clear phylogenetic picture of all diploid perennial *Medicago* species and to consider these approaches sufficiently robust for their standard use in the phylogenetics community.

Tables

Table 1. Information on the *Medicago* samples used in the present study.

Species	**Accession No**	**Germplasm Bank**	**Country of origin**	**Accession European Nucleotide Database**	**Chromosome number[d]**
M. carstiensis	PI641414	USDA	Russian Federation		16[a]
M. carstiensis	MED152/91	Ösnabruck Botanic Garden	NA		16[a]
M. cretacea	PI631721	USDA	Russian Federation		16[b]
M. daghestanica	Eoo337403; 1653*1	GB herbarium	Russian Federation		16[a]
M. daghestanica	Eoo337402; 1653*2	GB herbarium	Russian Federation		16[a]
M. hybrida	PI538998	USDA	Russian Federation		[d]
M. marina	PI516711	USDA	Morocco		16[a]
M. marina	PI419391	USDA	Greece		[d]
M. prostrata	PI577445	USDA	Italy		16[b]
M. rhodopea	SA43026	SARDI	NA		16[a]
M. rhodopea	W619154	USDA	Bulgaria		16[a]
M. rugosa	PI487386	USDA	Tunisia		[d]
M. sativa[c]	PI577567	USDA	Italy		16[a]
M. scutellata	PI505433	USDA	Spain		16[a]
M. suffruticosa	PI516913	USDA	Morocco		[d]
M. suffruticosa	W64952	USDA	Spain		16[a]
M. truncatula	W64996	USDA	Greece		[d]

[a], flow cytometry; [b], chromosome counts; [c], five individuals used; [d], samples confirmed as diploid by inspection of the phased samtools files (sam files) for all eight genes using Tablet (Milne et al., 2013, 2010).

Table 2. Characteristics of RADseq datasets generated in ipyrad and used for gene tree and network inference.

Dataset	Clustering threshold	Minimum taxon coverage	Minimum PIS *per* loci	Number of loci	Concatenated length (bp)	Missing data (%)
clust75.min4.PIS4	0.75	4	4	3194	283,753	49.7
clust75.min4.PIS10	0.75	4	10	346	29,974	56.6
clust75.min10.PIS4	0.75	10	4	1756	153,733	34.4
clust75.min10.PIS10	0.75	10	10	140	12,274	41.5
clust80.min4.PIS4	0.80	4	4	3364	299,468	49.5
clust80.min4.PIS10	0.80	4	10	317	28,744	55.3
clust80.min10.PIS4	0.80	10	4	1856	162,504	34.0
clust80.min10.PIS10	0.80	10	10	134	11,846	40.4
clust85.min4.PIS4	0.85	4	4	3405	303,272	48.8
clust85.min4.PIS10	0.85	4	10	205	18,789	51.8
clust85.min10.PIS4	0.85	10	4	1910	167,477	33.8
clust85.min10.PIS10	0.85	10	10	98	8,711	38.4
clust90.min4.PIS4	0.90	4	4	3036	271,409	48.0
clust90.min4.PIS10	0.90	4	10	81	7,451	45.5
clust90.min10.PIS4	0.90	10	4	1729	151,806	33.3
clust90.min10.PIS10	0.90	10	10	40	3,616	33.2
clust95.min4.PIS4	0.95	4	4	1506	135,056	46.1
clust95.min4.PIS10	0.95	4	10	5	455	24.2
clust95.min10.PIS4	0.95	10	4	912	80,449	32.7
clust95.min10.PIS10	0.95	10	10	4	367	16.2

Table 3. Positive hybridization signal detected for each taxon in each network. Only strict hybridization signal is considered, i.e. a taxa nested within a hybrid clade but represented with a single branch is not considered of hybrid origin.

Dataset	*M. carstiensis*	*M. cretacea*	*M. daghestanica*	*M. hybrida*	*M. marina*	*M. prostrata*	*M. rhodopea*	*M. rugosa*	*M. sativa*	*M. scutellata*	*M. suffruticosa*	*M. truncatula*
clust75.min4.PIS4	✓	✓						✓				
clust75.min4.PIS10	✓						✓			✓		
clust75.min10.PIS4	✓	✓					✓	✓				
clust75.min10.PIS10	✓	✓	✓		✓		✓					
clust80.min4.PIS4	✓	✓								✓		
clust80.min4.PIS10	✓						✓			✓		
clust80.min10.PIS4	✓	✓					✓					
clust80.min10.PIS10	✓	✓				✓	✓				✓	
clust85.min4.PIS4	✓	✓									✓	
clust85.min4.PIS10	✓		✓		✓	✓						
clust85.min10.PIS4	✓	✓										
clust85.min10.PIS10		✓		✓	✓				✓			
clust90.min4.PIS4	✓	✓									✓	
clust90.min4.PIS10	✓	✓	✓		✓							
clust90.min10.PIS4	✓	✓			✓		✓					
clust90.min10.PIS10	✓						✓				✓	
clust95.min4.PIS4	✓	✓			✓			✓				
clust95.min4.PIS10	✓	✓		✓			✓					
clust95.min10.PIS4	✓	✓						✓	✓			
clust95.min10.PIS10		✓	✓				✓			✓		

Figure Legends

Fig. 1 - Best networks (networks with highest likelihood scores) for each of the 20 RADseq datasets.

Fig. 2 - PCoA showing pairwise network distances calculated with the topological dissimilarity measure of Nakhleh (2010). The figure show pairwise distances between the 10 best networks of each of the 20 RADseq datasets. Transparency represents the filter on parsimony informative sites, shapes represent the filter on min. samples locus, and colors represent the clustering threshold used to generate the RADseq dataset.

Fig. 3 - Bar chart representing number of PIS4 RADseq datasets supporting a hybrid origin for the *Medicago* species analyzed in this study.

Acknowledgements

The authors thank Luay Nakhleh and Jiafan Zhu for their assistance with PhyloNet analyses and Filipe de Sousa for providing plant material.

Funding

This work was supported by a grant from the Swedish Research Council (grant reference 2009-5206) and by the Marie Curie Intra-European Fellowship "AlfalfaEvolution" (FP7-PEOPLE-2013-IEF, project reference 625308).

References

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92. https://doi.org/10.1038/nrg.2015.28

Annicchiarico, P., Nazzicari, N., Wei, Y., Pecetti, L., Brummer, E.C., 2017. Genotyping-by-Sequencing and Its Exploitation for Forage and Cool-Season Grain Legume Breeding. Front. Plant Sci. 8, 679. https://doi.org/10.3389/fpls.2017.00679

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376. https://doi.org/10.1371/journal.pone.0003376

Barker, D.G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S., Flament, P., Gallusci, P., Génier, G., Guy, P., Muel, X., Tourneur, J., Dénarié, J., Huguet, T., 1990. Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. Plant Mol. Biol. Report. 8, 40–49. https://doi.org/10.1007/BF02668879

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. Genetics 162.

Bena, G., 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" Trigonella species to the genus Medicago L. Plant Syst. Evol. 229, 217–236. https://doi.org/10.1007/s006060170012

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., Udvardi, M.K., 2008. A gene expression atlas of the model legume Medicago truncatula. Plant J. 55, 504–513. https://doi.org/10.1111/j.1365-313X.2008.03519.x

Bennett, S.J., Broughton, D.A., Maxted, N., 2006. Ecogeographical analysis of the perennial Medicago. CRC for Plant-Based Management of Dryland Salinity.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Branca, A., Paape, T.D., Zhou, P., Briskine, R., Farmer, A.D., Mudge, J., Bharti, A.K., Woodward, J.E., May, G.D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M.J., Ronfort, J., Bataillon, T., Young, N.D., Tiffin, P., 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. Proc. Natl. Acad. Sci. 108, E864–E870. https://doi.org/10.1073/pnas.1104032108

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852.

https://doi.org/10.1002/ece3.512
Clark, A.G., Messer, P.W., 2015. Conundrum of jumbled mosquito genomes. Science. 347, 27–28. https://doi.org/10.1126/science.aaa3600
Cook, D.R., 1999. Medicago truncatula - A model in the making! Curr. Opin. Plant Biol. https://doi.org/10.1016/S1369-5266(99)80053-3
Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28, 2239–2252. https://doi.org/10.1093/molbev/msr048
Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121
Eaton, D.A.R., Ree, R.H., 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Syst. Biol. 62, 689–706. https://doi.org/10.5061/dryad.bn281
Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379. https://doi.org/10.1371/journal.pone.0019379
Eriksson, J.S., De Sousa, F., Bertrand, Y.J.K., Antonelli, A., Oxelman, B., Pfeil, B.E., 2018. Allele phasing is critical to revealing a shared allopolyploid origin of Medicago arborea and M. strasseri (Fabaceae). BMC Evol. Biol. 18, 9. https://doi.org/10.1186/s12862-018-1127-z
Eriksson, J.S.S., Blanco-Pastor, J.L.L., Sousa, F., Bertrand, Y.J.K.J.K., Pfeil, B.E.E., 2017. A cryptic species produced by autopolyploidy and subsequent introgression involving Medicago prostrata (Fabaceae). Mol. Phylogenet. Evol. 107, 367–381. https://doi.org/10.1016/j.ympev.2016.11.020
Escudero, M., Eaton, D.A.R., Hahn, M., Hipp, A.L., 2014. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in Carex (Cyperaceae). Mol. Phylogenet. Evol. 79, 359–367. https://doi.org/10.1016/j.ympev.2014.06.026
Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M., 2013. Robust Demographic Inference from Genomic and SNP Data. PLoS Genet. 9, e1003905. https://doi.org/10.1371/journal.pgen.1003905
Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., Glover, B.J., 2017. Resolving Recent Plant Radiations: Power and Robustness of Genotyping-by-Sequencing. Syst. Biol. https://doi.org/10.1093/sysbio/syx062
Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y.Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V.B., Golovanova, L. V, Lalueza-Fox, C., De La Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A draft sequence of the neandertal genome. Science. 328, 710–722. https://doi.org/10.1126/science.1188021
Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., Siepel, A., 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 43, 1031–1035. https://doi.org/10.1038/ng.937
Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695. https://doi.org/10.1371/journal.pgen.1000695
Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T., 2016. Sequence

Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. Syst. Biol. 65, 910–924. https://doi.org/10.1093/sysbio/syw036

Hey, J., 2010. Isolation with Migration Models for More Than Two Populations. Mol. Biol. Evol. 27, 905–920. https://doi.org/10.1093/molbev/msp296

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS One 9, e93975. https://doi.org/10.1371/journal.pone.0093975

Hohenlohe, P.A., Amish, S.J., Catchen, J.M., Allendorf, F.W., Luikart, G., 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. 11, 117–122. https://doi.org/10.1111/j.1755-0998.2010.02967.x

Huson, D.H., Rupp, R., Scornavacca, C., 2010. Phylogenetic networks: concepts, algorithms and applications. Cambridge University Press.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W., 2015. Phylogenomics of phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol. Evol. 7, 706–719. https://doi.org/10.1093/gbe/evv026

Lesins, K.A., Lesins, I., 1979. Genus Medicago (Leguminosae), Dr. W. Junk Publishers, The Hague. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-009-9634-2

Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91, 1700–1708. https://doi.org/10.3732/ajb.91.10.1700

Maureira-Butler, I.J., Pfeil, B.E., Muangprom, A., Osborn, T.C., Doyle, J.J., 2008. The reticulate history of Medicago (Fabaceae). Syst. Biol. 57, 466–482. https://doi.org/10.1080/10635150802172168

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M. V, Derevianko, A.P., Patterson, N., Andrés, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222–6. https://doi.org/10.1126/science.1224344

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., Marshall, D., 2010. Tablet—next generation sequence assembly visualization. Bioinforma. Appl. NOTE 26, 401–402. https://doi.org/10.1093/bioinformatics/btp666

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., Marshall, D., 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief. Bioinform. 14, 193–202. https://doi.org/10.1093/bib/bbs012

Nakhleh, L., 2010. A metric on the space of reduced phylogenetic networks. IEEE/ACM Trans. Comput. Biol. Bioinforma. 7, 218–222. https://doi.org/10.1109/TCBB.2009.2

Pease, J.B., Hahn, M.W., 2015. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. Syst. Biol. 64, 651–662. https://doi.org/10.1093/sysbio/syv023

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.-L., 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7, e32253.

Rheindt, F.E., Fujita, M.K., Wilton, P.R., Edwards, S. V., 2014. Introgression and phenotypic assimilation in zimmerius flycatchers (Tyrannidae): Population genetic and phylogenetic inferences from genome-wide SNPs. Syst. Biol. 63, 134–152.

https://doi.org/10.1093/sysbio/syt070
Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7, 1–12. https://doi.org/10.1371/journal.pone.0033394
Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207. https://doi.org/10.1016/j.tree.2004.01.003
Sethuraman, A., Hey, J., 2016. IMa2p - parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. Mol. Ecol. Resour. 16, 206–215. https://doi.org/10.1111/1755-0998.12437
Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. 8, 907–917. https://doi.org/10.1111/2041-210X.12700
Small, E., 2011. Alfalfa and Relatives: Evolution and Classification of Medicago. NRC Research Press, Ottawa, Ontario, Canada. https://doi.org/doi:10.1139/9780660199795
Solís-Lemus, C., Ané, C., 2016. Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. PLoS Genet. 12, e1005896. https://doi.org/10.1371/journal.pgen.1005896
Sousa, F., 2015. Next-generation Molecular Systematics and Evolution: Insights into Medicago. University of Gothenburg, Gothenburg (Sweden).
Sousa, F., Bertrand, Y.J.K., Doyle, J.J., Oxelman, B., Pfeil, B.E., 2017. Using Genomic Location and Coalescent Simulation to Investigate Gene Tree Discordance in Medicago L. Syst. Biol. 66, 934–949. https://doi.org/10.1093/sysbio/syx035
Sousa, F., Bertrand, Y.J.K., Nylinder, S., Oxelman, B., Eriksson, J.S., Pfeil, B.E., 2014. Phylogenetic Properties of 50 Nuclear Loci in Medicago (Leguminosae) Generated Using Multiplexed Sequence Capture and Next-Generation Sequencing. PLoS One 9, e109704. https://doi.org/10.1371/journal.pone.0109704
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446
Steele, K.P., Ickert-Bond, S.M., Zarre, S., Wojciechowski, M.F., 2010. Phylogeny and character evolution in Medicago (Leguminosae): Evidence from analyses of plastid trnK/matK and nuclear GA3ox1 sequences. Am. J. Bot. 97, 1142–1155. https://doi.org/10.3732/ajb.1000009
Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 77–81. https://doi.org/10.1016/j.ympev.2014.07.016
Than, C., Ruths, D., Nakhleh, L., Bioinformatics, B., Than, C., Ruths, D., Nakhleh, L., 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9, 322. https://doi.org/10.1186/1471-2105-9-322
Twyford, A.D., Ennos, R.A., 2012. Next-generation hybridization and introgression. Heredity (Edinb). https://doi.org/10.1038/hdy.2011.68
Urban, I., 1873. Prodomus einer Monographie der Gattung Medicago L. Rudolph Gaertner.
Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation, in: Molecular Ecology. Wiley/Blackwell (10.1111), pp. 787–798. https://doi.org/10.1111/mec.12023
Wang, S., Meyer, E., McKay, J.K., Matz, M. V, 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat. Methods 9, 808–810.
Wen, D., Nakhleh, L., 2016. Co-estimating Reticulate Phylogenies and Gene Trees from

Multi-locus Sequence Data. bioRxiv 26, 1–13. https://doi.org/10.1101/095539

Wen, D., Yu, Y., Nakhleh, L., 2016. Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent. PLoS Genet. 12, 1–17. https://doi.org/10.1371/journal.pgen.1006006

Wen, D., Yu, Y., Zhu, J., Nakhleh, L., 2017. Inferring Phylogenetic Networks Using PhyloNet. Syst. Biol. 00, 197–204. https://doi.org/10.1093/sysbio/syy015

Yoder, J.B., Briskine, R., Mudge, J., Farmer, A., Paape, T., Steele, K., Weiblen, G.D., Bharti, A.K., Zhou, P., May, G.D., Young, N.D., Tiffin, P., 2013. Phylogenetic signal variation in the genomes of medicago (Fabaceae). Syst. Biol. 62, 424–438. https://doi.org/10.1093/sysbio/syt009

Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F.X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A.T., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.-M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.-H., Jing, Y., Jöcker, A., Kenton, S.M., Kim, D.-J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S.L., Magdelenat, G., Matthews, L., Mccorrison, J., Monaghan, E.L., Mun, J.-H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.-B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., Roe, B.A., O 'bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.-B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., Bruce, &, 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480, 520–524. https://doi.org/10.1038/nature10625

Yu, Y., Barnett, R.M., Nakhleh, L., 2013. Parsimonious Inference of Hybridization in the Presence of Incomplete Lineage Sorting. Syst. Biol. 62, 738–751. https://doi.org/10.1093/sysbio/syt037

Yu, Y., Cuong, T., Degnan, J.H., Nakhleh, L., 2011. Coalescent Histories on Phylogenetic Networks and Detection of Hybridization Despite Incomplete Lineage Sorting. Syst. Biol. 60, 138–149. https://doi.org/10.1093/sysbio/syq084

Yu, Y., Dong, J., Liu, K.J., Nakhleh, L., 2014. Maximum likelihood inference of reticulate evolutionary histories. Proc. Natl. Acad. Sci. U. S. A. 111, 16448–16453. https://doi.org/10.1073/pnas.1407950111

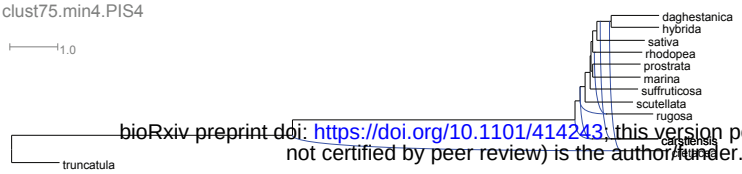
Yu, Y., Nakhleh, L., 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics 16, S10. https://doi.org/10.1186/1471-2164-16-S10-S10

Zhang, C., Ogilvie, H.A., Drummond, A.J., Stadler, T., 2018. Bayesian inference of species networks from multilocus sequence data. Mol. Biol. Evol. 35, 504–517.

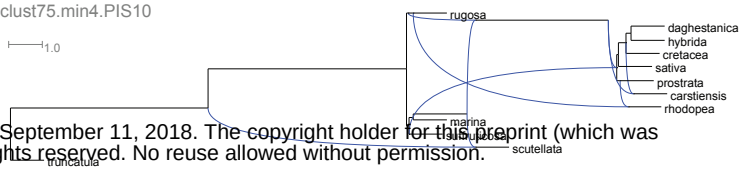
https://doi.org/10.1093/molbev/msx307

Zhu, J., Wen, D., Yu, Y., Meudt, H.M., Nakhleh, L., 2017. Bayesian Inference Of Phylogenetic Networks From Bi-allelic Genetic Markers. PLoS Comput Biol 14, e1005932. https://doi.org/10.1371/journal.pcbi.1005932

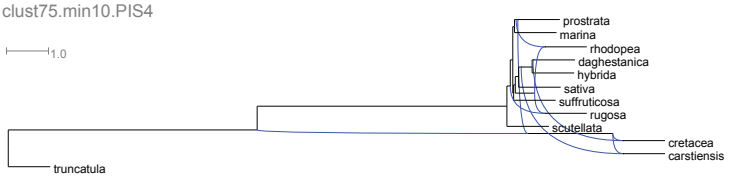
clust75.min4.PIS4



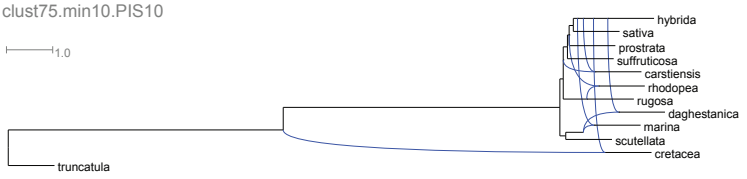
clust75.min4.PIS10



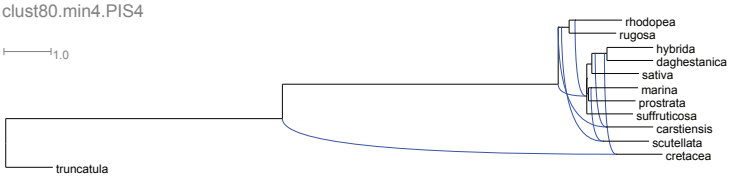
clust75.min10.PIS4



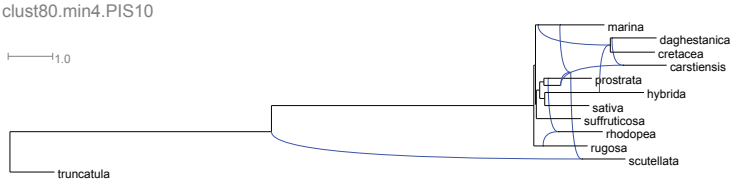
clust75.min10.PIS10



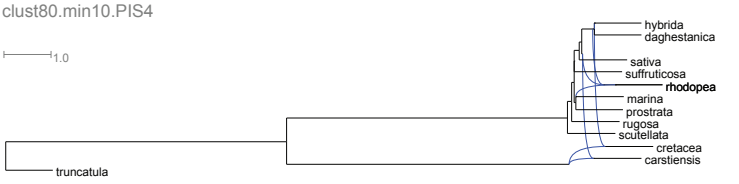
clust80.min4.PIS4



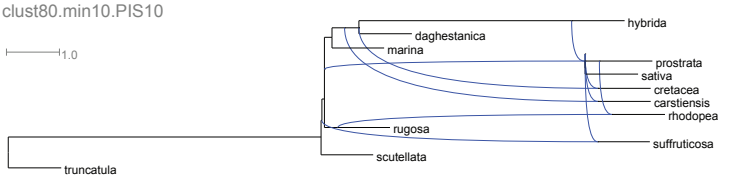
clust80.min4.PIS10



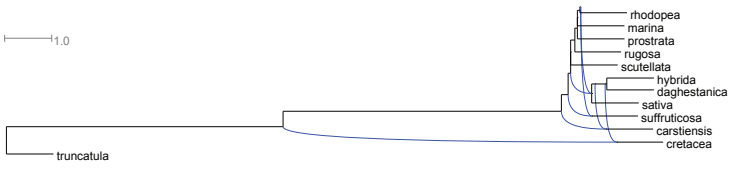
clust80.min10.PIS4



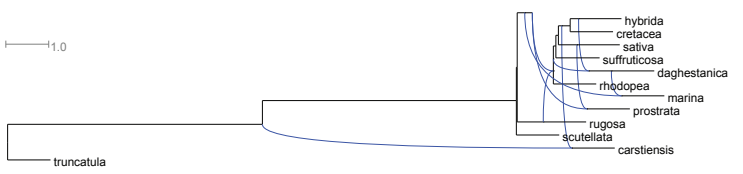
clust80.min10.PIS10



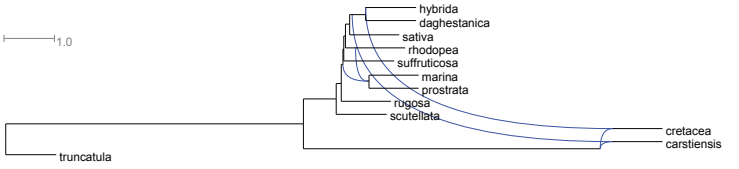
clust85.min4.PIS4



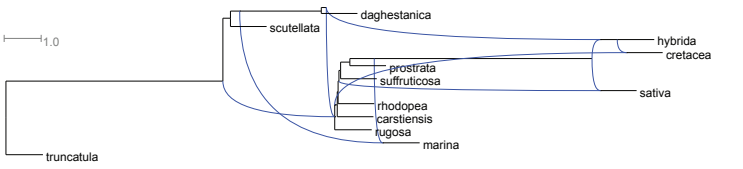
clust85.min4.PIS10



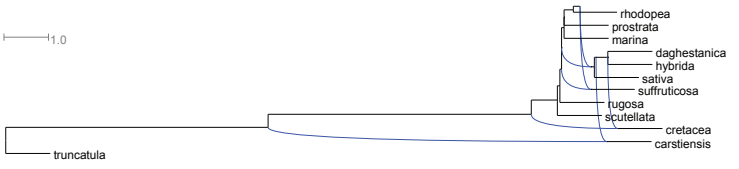
clust85.min10.PIS4



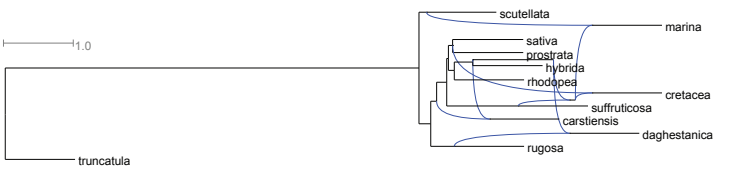
clust85.min10.PIS10



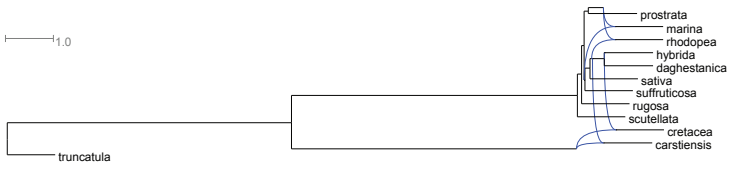
clust90.min4.PIS4



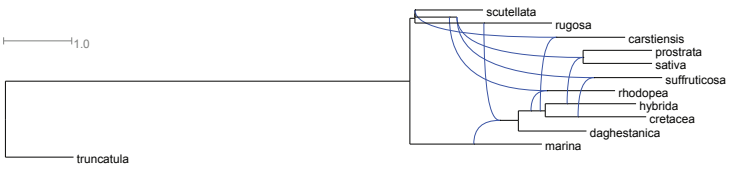
clust90.min4.PIS10



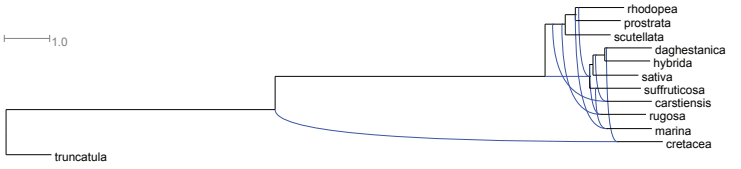
clust90.min10.PIS4



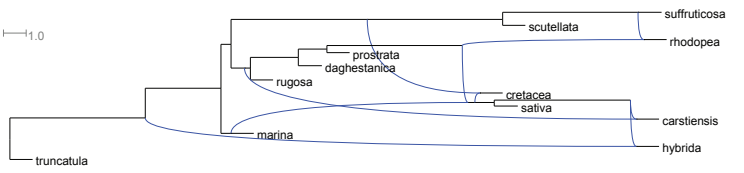
clust90.min10.PIS10



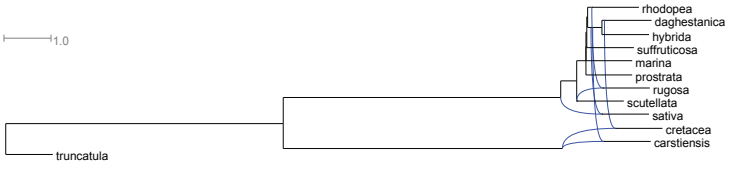
clust95.min4.PIS4



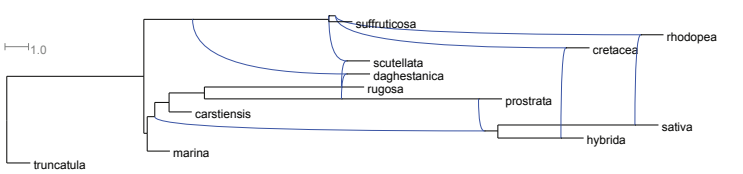
clust95.min4.PIS10

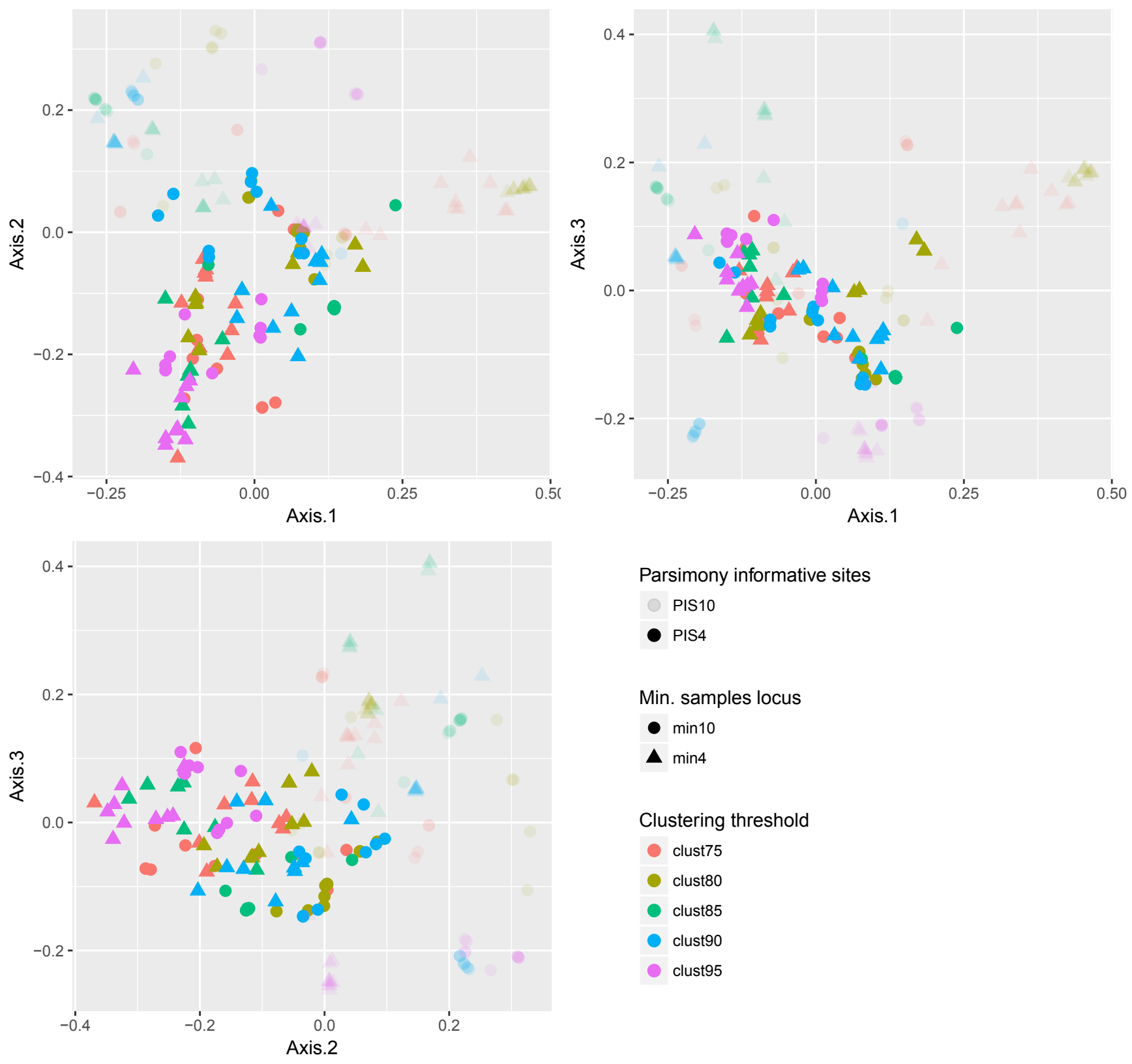


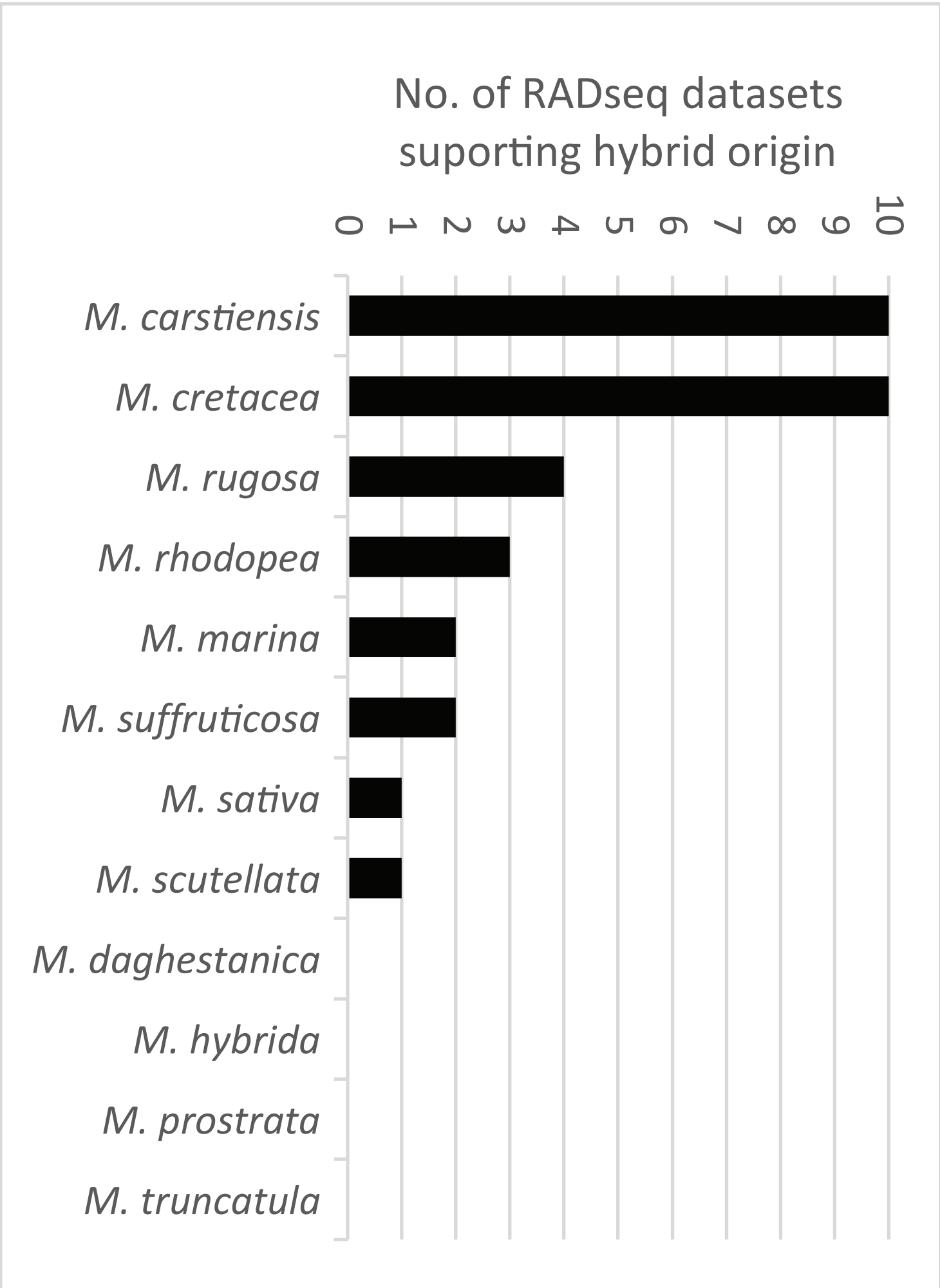
clust95.min10.PIS4



clust95.min10.PIS10




Axis.2
Axis.1
Axis.3
Axis.1
Axis.3
Axis.2
Parsimony informative sites
PIS10
PIS4
Min. samples locus
min10
min4
Clustering threshold
clust75
clust80
clust85
clust90
clust95


No. of RADseq datasets suporting hybrid origin
0
1
2
3
4
5
6
7
8
9
10
M. carstiensis
M. cretacea
M. rugosa
M. rhodopea
M. marina
M. suffruticosa
M. sativa
M. scutellata
M. daghestanica
M. hybrida
M. prostrata
M. truncatula