Molecular Phylogenetics and Evolution 107 (2017) 367–381

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A cryptic species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae)

J.S. Eriksson *, J.L. Blanco-Pastor, F. Sousa, Y.J.K. Bertrand, B.E. Pfeil

Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 40530 Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 30 April 2016
Revised 21 November 2016
Accepted 29 November 2016
Available online 2 December 2016

Keywords:
Hybridisation
Homoploid hybrid speciation
Gene capture
Next-generation sequencing
Coalescent simulations
Polyploid speciation

ABSTRACT

Although hybridisation through genome duplication is well known, hybridisation without genome duplication (homoploid hybrid speciation, HHS) is not. Few well-documented cases have been reported. A possible instance of HHS in *Medicago prostrata* Jacq. was suggested previously, based on only two genes and one individual. We tested whether this species was formed through HHS by sampling eight nuclear loci and 22 individuals, with additional individuals from related species, using gene capture and Illumina sequencing. Phylogenetic inference and coalescent simulations were performed to infer the causes of gene tree incongruence. We found no evidence that phylogenetic differences among *M. prostrata* individuals were the result of HHS. Instead, an autopolyploid origin of tetraploids with introgression from tetraploids of the *M. sativa* complex is likely. We argue that tetraploid *M. prostrata* individuals constitute a new species, characterised by a partially non-overlapping distribution and distinctive alleles (from the *M. sativa* complex). No gene flow from tetraploid to diploid *M. prostrata* is apparent, suggesting partial reproductive isolation. Thus, speciation via autopolyploidy appears to have been reinforced by introgression. This raises the intriguing possibility that introgressed alleles may be responsible for the increased range exploited by tetraploid *M. prostrata* with respect to that of the diploids.

1. Introduction

Our understanding of the scope of biological species diversity relies mainly on species primarily discovered through morphological and cytological investigations. There is, however, a barely explored continent of hidden diversity not easily diagnosed by either of these approaches, which is now beginning to emerge by DNA-based investigations (Hebert et al., 2004; Bucklin et al., 2011; Liu et al., 2011; Martinsson et al., 2014; Tedersoo et al., 2014). These methods are not without complications. One long-standing difficulty for traditional and DNA-based methods of species discovery is hybridisation. Hybridisation can shape existing diversity by generating new but fleeting genetic combinations that can confound morphological species assignment (e.g., in hybrid zones). Alternatively, hybridisation can transfer alleles characteristic of one species into another, sometimes with limited or no morphological change apparent in mosaic individuals (Rheindt et al., 2014). Hybridisation also has the potential to create entirely new species that may combine or exceed the traits of the parental species (Dittrich-Reed and Fitzpatrick, 2013), which in turn may produce a confusing mix of character state distributions among closely related species. The influence of hybridisation on existing species and newly created species is often not easy to detect, requiring large genetic data sets to do so in some cases (Green et al., 2010).

Hybridisation is also considered to be an evolutionary force that creates opportunities for adaptation (Rieseberg et al., 2003), and is especially prevalent among plants (Abbott, 1992). Hybrids can be formed by the interbreeding of individuals at the same ploidy level, either with or without subsequent genome doubling. These hybrid populations sometimes form new species isolated from both progenitor species, and are referred to as allopolyploid and homoploid hybrid species, respectively. Speciation via allopolyploidy is a common process, especially in plants (Wilson, 1994; Doyle et al., 2002; Lee et al., 2002; Wood et al., 2009; Marcussen et al., 2012), which contrasts with the few demonstrated cases of homoploid hybrid speciation (HHS), consistent with frequencies predicted from theory (Buerkle et al., 2000; Gross and Rieseberg, 2005). In order to understand the general patterns of HHS and the persistence of such species, it is important that more cases of HHS are uncovered and further investigated.

The main type of evidence for HHS, especially for events that are somewhat ancient (i.e., having taken place hundreds of thousands to millions of years ago) is phylogenetic incongruence among unlinked genomic regions, which can be caused by several pro-

* Corresponding author.
E-mail address: jonna.eriksson@bioenv.gu.se (J.S. Eriksson).

http://dx.doi.org/10.1016/j.ympev.2016.11.020
1055-7903/

cesses apart from hybridisation (e.g., Wendel and Doyle, 1998). One major challenge for the discovery of new cases resides in recognizing genuine hybridisation signals among incongruent gene trees. Another is the clear demonstration that a new species has in fact been formed. Fortunately, new analytical methods (Maureira-Butler et al., 2008; Joly et al., 2009; Meng and Kubatko, 2009; Yu et al., 2011; Jones et al., 2013, 2014; Grumer et al., 2014) and increasingly richer genetic data are allowing great progress on these fronts, but reports of putative hybrid species from earlier studies may lack sufficient data to be conclusive. The potential to collect incredible amounts of genetic data has dramatically changed in the last decade with next-generation sequencing technologies, removing a key bottleneck in hybrid species discovery.

With respect to the data required for discriminating between allopolyploidy and HHS, allopolyploidy is easier to infer, because each nuclear locus will (at least initially) be present in two copies – one for each parent. These copies can be maintained for many millions of years (Lynch and Connery, 2000; Blanc and Wolfe, 2004), potentially allowing detection of the origin of the polyploid far in time from the event that created it. With HHS, on the other hand, allelic fixation generally occurs more quickly, making the recovery of both parental contributions less likely at a single locus (Ungerer et al., 1998). Thus, more loci are typically required to infer HHS than for allopolyploid speciation. Another complicating factor for HHS detection proceeds from the uncertainty that often surrounds the exact ploidy level. Although many plant species have chromosome numbers reported and curated in the Index to Plant Chromosome Numbers (http://www.tropicos.org/Project/IPCN), this information is not always available for genetic variants. For example, the exact specimen counted may not be available to other researchers conducting DNA sequencing and phylogenetic inference. Moreover, a single count for a species does not adequately represent karyotypic diversity, if the ploidy level varies in a cryptic fashion with respect to the established taxonomy.

In the plant genus *Medicago* L. (Fabaceae), the phylogenetic work to date has resulted in several instances of gene tree incongruence, either within or among studies (Bena et al., 2005; Maureira-Butler et al., 2008; Steele et al., 2010; Havananda et al., 2011). *Medicago prostrata* Jacq., a perennial herb from eastern-central Europe, may have been formed by HSS as suggested in a previous study based on the observed phylogenetic incongruence between the nuclear genes CNGC5 and β-cop (Maureira-Butler et al., 2008). In CNGC5, *M. prostrata* is nested within the Sativa complex, whereas in β-cop *M. prostrata* forms a clade with *M. arborea* L. Complicating the picture, two chromosome numbers have been reported in *M. prostrata*: $2n = 16$ and $2n = 4x = 32$ (Small, 2011), although any correlation with geographic or morphological variation and these counts are as yet unknown. Unfortunately, no counts were made of the exact specimens used in the previous phylogenetic study that reported this initial case of incongruence (Maureira-Butler et al., 2008).

Thus, several questions arise with respect to the previous observations made on *M. prostrata*: Is this a species of hybrid origin derived from the *M. arborea* and *M. sativa* L. lineages or can incomplete lineage sorting (ILS) alone explain the observed incongruence? If hybridisation can be demonstrated, did it occur to form a new species at either tetraploid level (allopolyploidy or HHS via tetraploids) or at the diploid level (HHS via diploids), or was there just introgression between extant species at either ploidy level? Can we determine whether the tetraploids have an allo- or autopolyploid origin? We address these questions, in light of the challenges described earlier, by generating a large data set in terms of sampled loci and individuals that added to sequences produced in the previous work of (Maureira-Butler et al., 2008). We use gene capture and next-generation sequencing (NGS) to obtain alignments of eight loci from nearly 50 individuals. We investigate the causes of the observed gene tree incongruence with a test designed to distinguish between biological processes that are known to produce such patterns, namely incomplete lineage sorting (ILS) and hybridisation. We also determine the ploidy level of each individual in the study by assessing allele number in up to 10 loci, with a cross-check in some cases from chromosome counts.

2. Materials and methods

2.1. Taxon sample

A total of 49 *Medicago* individuals were successfully sampled: 22 belonging to *M. prostrata* (19 from herbarium sheets and three from fresh material) and 27 from related species (fresh material) (Table 1). Individuals not from herbarium sheets were grown from seed in growth chambers under uniform conditions.

2.2. Chromosome counts

The ploidy level for some of the living collections was determined by counting the chromosomes from root tips (2–3 cm) that were pre-treated in colchicine 0.1% for 2 h. After the pre-treatment the roots were fixated in Carnoy 1 solution (1:3 acetic acid and 2:3 alcohol) for >24 h, to minimise the shrinkage of the chromosomes. The fixed samples were hydrolysed with one normal HCl and cleaned with 70% alcohol for 10 min each, before being stained with aceto-orcein for 30 min. The roots were then attached to a slide and squashed gently to spread the cells. To make the samples permanent, liquid nitrogen was used to freeze the slides so that the coverslip could easily be removed. The slides were then treated with absolute alcohol, 1:1 absolute alcohol/histoclear, twice with histoclear, and then mounted with histomount for 24 h. Chromosomes were then counted under an optical microscope.

2.3. DNA Extraction, probe selection and library preparation

Genomic DNA was extracted from approximately 25–30 mg of dried leaf tissue using DNEasy Plant Mini Kit (Qiagen, Cat. No. 69104), following the manufacturer's protocol, with two exceptions: a pre-treatment with 390 µl AP2 and 10 µl proteinase K in 42 °C overnight for the herbarium specimens, and 30 min with AP3 solution for a higher concentration yield.

Genes for sequencing were selected from the reference genome sequence of *M. truncatula* in the *Medicago* HapMap project (http://www.medicagohapmap.org/) and a MyBaits kit constructed, as per (Sousa et al., 2014). Some genes were unlinked (on different chromosomes or chromosome arms), whereas others where physically tightly linked (within 30 kb of one another). All genes were blasted in NCBI (http://www.ncbi.nlm.nih.gov) in order to select only low copy nuclear genes present in genus *Medicago* with a threshold of 95% similarly along exons, the value recommended for successful gene capture at the standard conditions for these probes. The MyBaits RNA-based probe kit was made by the company (see below), with probes of 3× tiling density, to match the genes that we selected from the *M. truncatula* project for gene capture (see below).

Libraries of genomic DNA were prepared for each of the sampled individuals. A NanoDrop 2000c instrument (Thermo Fisher Scientific, Waltham, MA, USA) was used to measure extracted DNA concentrations and purity. The genomic DNA was then fragmented (sonicated) using a Covaris S220 instrument (Covaris, Woburn, Massachusetts, CA, USA), with fragments of 400–600 bp chosen. We prepared barcoded libraries with NEXTflex™ Rapid DNA-Seq Kit and NEXTflex™ Barcodes (BIOO Scientific, Austin,

Table 1
Species used in this study along with accession numbers. ENA stands for European Nucleotide Archive were all the raw sequence reads have been deposited. All the chromosome counts presented here refer to root tip squash counts made on fresh material grown at the University of Gothenburg. Sienna refers to the Sienna Botanical Garden. PI and W6 refers to USDA GRIN accession numbers. LD refers to University Lund herbarium. GB refers to the University of Gothenburg herbarium and 13-P3 refers to herbarium in Vienna. The labels from Vienna are temporary until added in the database.

Growth chamber samples (species, sample number)	Accession number	ENA accession number	Chromosome count	Herbarium samples (species, sample number)	Accession number	ENA accession number
Medicago arborea 1	PI368041	ERS1353866	2n = 32	*Medicago prostrata 1*	LD 1723202	ERS1353844
M. arborea 2	PI330677	ERS1353867		*M. prostrata 2*	GB-0144771	ERS1353845
M. arborea 3	PI368172	ERS1353868	2n = 32	*M. prostrata 3*	LD 1660145	ERS1353846
M. carstiensis 1	PI641414	ERS1353869		*M. prostrata 4*	LD 1723650	ERS1353847
M. carstiensis 2	MED152/91	ERS1353870		*M. prostrata 5*	LD 1738262	ERS1353848
M. ciliaris	PI498731	ERS719974		*M. prostrata 6*	LD 1723330	ERS1353849
M. coronata	PI498807	ERS1353871		*M. prostrata 7*	LD 1592206	ERS1353850
M. cretacea 1	PI631721	ERS1353872	2n = 16	*M. prostrata 8*	GB-0144775	ERS1353851
M. cretacea 2	W633709	ERS1353873	2n = 16	*M. prostrata 9*	GB-0144774	ERS1353852
M. intertexta	Sienna	ERS1353874		*M. prostrata 10*	GB-0144772	ERS1353853
M. italica	PI577295	ERS511665		*M. prostrata 11*	LD 1723522	ERS1353854
M. littoralis	PI537222	ERS719977		*M. prostrata 12*	GB-0144773	ERS1353855
M. marina	PI419391	ERS1353875		*M. prostrata 13*	13-P3/1	ERS1353856
M. medicaginoides	W624116	ERS511669		*M. prostrata 14*	13-P3/4	ERS1353857
M. papillosa 1	PI464699	ERS1353876		*M. prostrata 15*	13-P3/5	ERS1353858
M. papillosa 2	PI631778	ERS1353877		*M. prostrata 16*	13-P3/12	ERS1353889
M. pironae 1	PI577372	ERS1353878		*M. prostrata 17*	13-P3/9	ERS1353859
M. pironae 2	PI253450	ERS1353879		*M. prostrata 18*	13-P3/18	ERS1353860
M. platycarpa 2	PI258759	ERS1353880		*M. prostrata 19*	13-P3/38	ERS1353861
M. prostrata 21	PI577447	ERS1353863	2n = 16	*M. prostrata 20*	13-P3/26	ERS1353862
M. prostrata 22	PI577450	ERS1353864	2n = 32			
M. prostrata 23	PI577446	ERS1353865				
M. rhodopea 1	W619154	ERS1353881				
M. rhodopea 2	SA43026	ERS1353882				
M. rotata	PI495577	ERS1353883				
M. ruthenica	PI245002	ERS1353884				
M. sativa subsp sativa	PI220598	ERS511666				
M. secundiflora	PI537238	ERS1353885				
M. shepardii	PI459134	ERS1353886				
M. strasseri	G. Pediados, Iraclio	ERS1353887				
M. suffruticosa subsp leiocarpa	W64952	ERS1353888				
Melilotus neapolitanus	Sienna	ERS511668				
Melilotus sulcatus	Sienna	ERS511667				

Texas, U.S.A.), as per the manufacturer's instructions, except that half reactions were performed to decrease the cost. During library preparation, we increased the yield by a PCR amplification of 14 cycles using the PCR master and primer mix (provided by the kits), with the following program: 98 °C, 2′; 14×(98 °C, 30″; 65 °C, 30″; 72 °C, 60″); 72 °C, 4′. The amplified concentrations were measured by NanoDrop again to ensure that there was enough DNA for gene enrichment.

2.4. Gene capture and sequencing

Gene capture of the DNA libraries was done using the manufacturer's specifications, except that eight barcoded sample libraries were pooled for a single gene capture reaction. After gene capture, the samples were concentrated by another PCR amplification using Herculase II Fusion DNA Polymerase (Agilent, Waldbronn, Germany) with the following program: 98 °C, 30″; 14× (98 °C, 20″; 60 °C, 30″; 72 °C, 30–45″); 72 °C, 5′. A total of 48 samples were then pooled and sequenced on a single run of a MiSeq instrument (Illumina, San Diego, California, USA) at the Genomics Core Facility of the University of Gothenburg, generating 150 bp paired-end sequences. Raw sequence reads were deposited in the European Nucleotide Archive (Table 1; ERS1353844–ERS1353889).

2.5. Contig assembly, alignment

We selected eight genes for thorough analysis after the sequencing, using the following pipeline: The sequence reads were assembled into de novo contigs for each gene and individual using the CLC Assembly cell 4.0 software (CLC Bio, Aarhus, Denmark), which includes adapter and quality trimming for all samples using default parameters. These contigs were later used to produce a reference sequence, against which to map reads for each individual (Supplementary Fig. 1). Read mapping was done using CLC mapper. The resulting mapped files were then sorted and phased into two alleles by samtools (Li et al., 2009) (Supplementary Fig. 1). For samples where the phasing failed to produce two alleles (i.e., the locus was homozygous), a script was designed to retrieve the homozygous allele. All sequences derived from the same gene were aligned using MUSCLE (Edgar, 2004) and checked by hand using Geneious v.5.6.6 (Kearse et al., 2012).

However, if the read depth was low, which was especially the case in the intronic regions, full length alleles could not be obtained at the phasing step. Samtools phase instead produced fragmented alleles as several partially or non-overlapping contigs. These were manually merged in Geneious (Kearse et al., 2012) using the following criteria: (a) non-overlapping contigs were merged arbitrarily, (b) partially overlapping contigs were merged only if no differences existed in the overlapping segments, (c) overlapping contigs were kept separate if there was even a single reliable (>3 reads) polymorphism between them. This procedure could accidentally result in the merging of fragments derived from different alleles, giving a mixed (chimaeric) allele (Supplemental Fig. 1). In these cases, recombination testing (below) was expected to give a strong signal for these artificial alleles, especially for *M. prostrata*, where we sampled a high ratio of individuals/alleles. Chimaeric alleles identified through recombination testing were then manually corrected by swapping the two candidate fragments at the breakpoint suggested by RDP3 and re-running RDP3 to ensure that no recombination signal could be identified, if the recombination

was indicated to have taken place within the same individual. Otherwise, the smallest recombined piece was deleted and the remaining contig was used in downstream analyses (Supplemental Fig. 1).

2.6. Polyploid allele phasing

Tetraploid individuals were expected to harbour two copies of each locus (homoeologues), together with any allelic variants. These homoeologues contained either only a single allele each for homozygous loci, or two alleles per locus for heterozygous loci. Since samtools phase assumes that only one or two alleles are present at a locus (i.e., that the individuals are diploid), retrieving additional alleles from tetraploid individuals needs to be done in another way. We identified polyploid *M. prostrata* individuals by inspecting the phased samtools files (sam files) for all eight genes using Tablet (Milne et al., 2013). Polyploid individuals were identified as such when several loci contained more than two alleles.

Phased sam files often contained a small proportion of reads with shared polymorphisms. Evidence for the presence of more than two alleles per locus could be found when two polymorphic sites were close enough together to be spanned by numerous reads (i.e., within the 150 bp read length). Four possible alleles with states 1 or 2 at the first and second site would be: 1–1, 1–2, 2–1 and 2–2. If we observed either three or four of these combinations, then we took this as confirmation that the locus contained three or four alleles. Two of these combinations were usually represented by the consensus sequences of each of the phased sam files (i.e., the original products of the phasing step). Combinations of polymorphisms were considered to represent real alleles (and not the product of PCR-mediated recombination during earlier steps) when supported by more than three independent sequence reads per combination. Reads were considered to be independent if they had different lengths, thus deriving from differently sheared genomic DNA pieces rather than PCR duplicates. This stringency appears to also rule out sequence error (either due to polymerase artefacts or in the Illumina step), because singleton bases (that are presumably erroneous) could be seen to reside (by and large) only on single independent reads among our sequences.

Although more than two alleles were present at several loci in tetraploid individuals in this sample, we did not attempt to reconstruct the additional alleles, but used the phased results directly. This was because numerous single polymorphic sites were often separated by invariant regions longer than the read length, and thus the reads could not unambiguously reveal the phase of all sites. Furthermore, the original phasing products preferentially separated the most different alleles present, and were thus sampling homoeologues preferentially, which was most useful to understand the origin of tetraploid individuals. Even with the original phasing products, chimaeric alleles could be present in our merged contigs, if invariant regions longer than the read length were present, or if the read depth was shallow at one part of the locus. We identified chimaeric alleles using recombination tests (Supplemental Fig. 1).

2.7. Recombination testing

A recombination test with RDP 3.44 (Martin et al., 2010) was performed on each alignment. We used a detection *p*-value of 0.1 and five methods (RDP (Martin and Rybicki, 2000), MaxChi (Maynard Smith, 1992), Bootscan/Recscan (Martin et al., 2005), SiScan (Gibbs et al., 2000) and Chimaera (Posada and Crandall, 2001)) to initially screen for recombination events. Any putative recombination event was then re-checked with all methods in the program, initially accepting any event with *p*-value $\leqslant$0.05. To check if putative recombination events were consistent with phylogenetic expectations, phylogenies of each recombining and non-recombining part of the locus were inferred using the built-in neighbor-joining (NJ) tool. If an event was significantly detected in any two methods at *p*-value $\leqslant$0.05 and had a phylogenetic pattern consistent with recombination, we trimmed the smallest parts of each affected allele (Supplemental Fig. 1) or eliminated the entire sequence before phylogenetic analysis.

2.8. Phylogenetic analysis

Individual gene phylogenies were inferred by Bayesian Inference (BI) and Maximum Likelihood (ML) analyses. To determine the best fitting sequence evolution model, we analysed each alignment with jModeltest2 (Darriba et al., 2012) and selected the optimal model using the Bayesian information criterion. BI was conducted in MrBayes 3.2.3 (Ronquist and Huelsenbeck, 2003) via the CIPRES Science Gateway (http://www.phylo.org). For each alignment, we performed two MrBayes analyses, running each with four Markov chains for 5 million generations, sampled every 1000 generations. The analyses were automatically stopped when the average standard deviation across runs dropped below 0.01. Posterior distributions of both runs converged on the same stable distribution as seen in TRACER 1.5 (Rambaut and Drummond, 2009). Trees for the first 20% generations were discarded as burn-in. ML analyses were conducted using PhyML 3.0 (Guindon and Gascuel, 2003; Guindon et al., 2010) via the ATGC web server (http://www.atgc-montpellier.fr/phyml/). Node supports for ML analysis were estimated using 1 000 non-parametric bootstrap (BS) replicates. Alignments and trees were deposited in the TreeBase repository (http://purl.org/phylo/treebase/phylows/study/TB2:S19925).

The proportion of alleles found in tetraploid *M. prostrata* individuals (see Results), in either the "sativa group" clade or the *M. prostrata* core clade was examined. For each tetraploid individual, we summed the number of "sativa group" alleles (scoring 0.5 for recombined alleles and adding alleles for which there was evidence although not a full sequence – see Polyploid Allele Phasing, above) and divided this by the total number of alleles present.

We also performed preliminary BEAST v. 1.8 (Drummond et al., 2012) analyses with all alleles. In this set of species we had three different calibrations, derived from a reanalysis of *mat*K data (Steele et al., 2010) and a secondary calibration of the *Trifolium-Medicago* split at 24 Mya (mean = 24.7 Mya, SD = 2.3) from (Lavin et al., 2005), depending on the clades retrieved from the gene trees in MrBayes. The three calibrations were as follows: the divergence between *Trigonella-Medicago*, between the *M. truncatula* and *M. sativa* clades and finally the crown age of *Medicago*. These divergence priors were modelled with normal distributions with a mean of 15 Mya and a SD of 2.7, mean 6.14 Mya and SD = 1.2 and a mean of 11 Mya and SD = 2.35, respectively. From an earlier study (Sousa et al., 2014) that included the eight genes we used here, we set a prior on the substitution rate (ucld.mean) using a normal distribution with mean = 3.6E−9 and SD = 1.4E−9 substitution/site/year.

For the purpose of the hybridisation test (see below) and dating the nodes in our gene trees, further BEAST runs were performed on a subset of the alleles for three genes only (109, 148, 150). This was because additional incongruence among the diploid samples was seen among the trees, which could also contribute to incongruence detectable by the test. We wanted to limit the test to only a single case of incongruence (the position of tetraploid *M. prostrata* alleles). This would make it easier to pinpoint the cause of any rejection of the null hypothesis of ILS alone to explain gene tree incongruence. The Yule prior was used for the birth of new lineages, with an uncorrelated relaxed lognormal clock for the clock prior and UPGMA starting trees. This second set of BEAST analyses included a calibration point for the split between *M. littoralis*

(*M. truncatula* clade) and *M. sativa*, with mean = 6.14 Mya and SD = 1.2 as per the preliminary analysis. These analyses were run for 50 million generations, sampled every 1000 generations.

2.9. Hybridisation test

To test whether the observed incongruence between gene trees could be explained by ILS alone, without having to invoke hybridisation, we used a pipeline introduced in (Sousa et al., in press) that extends the method of (Maureira-Butler et al., 2008), with subsequent modifications (Blanco-Pastor et al., 2012; Ramadugu et al., 2013).

Under ILS, observed gene trees that have evolved under a single species tree are only expected to differ to a limited degree, governed by the effective population size and species tree branch lengths in generations (Degnan and Rosenberg, 2009). Hybridisation, on the other hand, makes it possible for gene trees to differ more than expected by ILS alone, especially when hybridisation occurs among distantly related species, effective population sizes are low and there are many generations between speciation events. When hybridisation is effective (i.e., introgressed alleles persist), more than one tree is required to describe species' relationships (more than one principal tree: see (Holland et al., 2008)).

Our test procedure compares the observed differences between gene trees to null expectations generated under simulation, to determine whether the observed differences are large enough to reject the null hypothesis of ILS alone as the cause of these differences (Maureira-Butler et al., 2008). The null expectations are derived by assuming that each gene tree is a good surrogate for the species tree containing it. Simulations are conducted under the multispecies coalescent, on each gene tree, using previously derived effective population size estimates applied across the tree, and branch lengths from an ultrametric time-calibrated posterior distribution of gene trees.

We subsampled to one allele per species (to match the test's current implementation), with the following taxa: *M. arborea, M. strasseri, M. littoralis, M. marina, M. cretacea, M. rhodopea, M. pironae, M. papillosa, M. sativa, M. prostrata_2n* (individual 21) and *M. prostrata_4n* (individual 22). Subsampling was done to maximise incongruence among allele positions among genes in the tetraploid *M. prostrata* individuals but to minimise incongruence in the other species – i.e., to test the position of the tetraploid alleles maximally affecting incongruence, which are also those that are least plausible under ILS alone. Effective population size (N_e) was set to 240 000 gene copies as estimated by (Maureira-Butler et al., 2008). We performed the test using a set of 100 randomly selected input trees drawn from the posterior set of gene trees from the BEAST analyses wherein the branch length was reported in years (to match the calibration, also in years), therefore we did not have to take into account the generation time. The critical value used was 65%, determined through simulations (Sousa et al., in press), unlike the overly conservative 95% used in previous studies (Blanco-Pastor et al., 2012; Ramadugu et al., 2013) (although results were identical with a 95% critical value).

We did not perform any species tree analyses because we suspected that hybridisation among several diploid species in the sample has also occurred (manuscript in preparation). This would either violate the assumptions of all current species tree and summary methods, or exceed the number of homoploid hybridisations allowable (e.g., in STEM; Meng and Kubatko, 2009).

2.10. Morphological observations and distribution

We checked whether diploid and tetraploid *M. prostrata* individuals could be distinguished from one another using morphological characters. We examined leaflet shape and the number of teeth on the leaflet apex margin of the central leaflet, both characters that varied among our specimens. Although fruit often varies between closely related species in *Medicago*, not all of our specimens had fruit, so we were unable to compare features in this organ among specimens of different ploidy. However, the fruit has been noted as lacking "conspicuous differences" between ploidy levels in an earlier study (Lesins and Lesins, 1960). We also mapped the locations of diploid and tetraploid individuals we examined as well as those of earlier observations (Lesins and Lesins, 1960; Havananda et al., 2011) to see if there were any differences in their distributions.

3. Results

3.1. Chromosome number variation

We examined chromosome number variation (root tip squash counts) in six individuals from *M. arborea*, *M. cretacea* and *M. prostrata* (Table 1) in order to confirm the counts reported in the literature (e.g., (Small, 2011) and references therein). We were able to confirm that specimens conforming to the description of *M. prostrata* were either diploid ($2n = 16$) or tetraploid ($2n = 32$).

3.2. Sequence recovery from gene capture

From the gene capture procedure, we managed to recover sequences from over 96% of the individuals from the eight genes. The only individual that we failed to recover was *M. prostrata* 13. A generally lower read depth between the exons could be seen for most of the individuals. The mean length after assembly and alignment was 2760 bp (Supplementary Table 1).

3.3. Allele assessment using Tablet

Manual inspection of the phased alleles with Tablet (Milne et al., 2013) revealed evidence for a third and fourth allele in some individuals for at least one of the eight loci (Table 2). This group included the known tetraploid, *M. prostrata* 22 but also the individuals 1–3, 5–8, 10, 14, 15 and 20. We interpret the individuals with up to four alleles per locus as tetraploids. The remaining individuals with up to three alleles per locus could be either triploids or tetraploids. However, no triploids have ever been reported for this species (Small, 2011). In fact, only a single count of a *Medicago* with 24 chromosomes was reported by Small in his monograph (Small, 2011), and this was of a species where the majority of counts are 28 (*M. polyceratia*). This presumably makes the 24 count either a mistake, or an aneuploid reduction from a tetraploid (based on $2n = 14$), rather than a triploid. Apart from artificial crosses between diploids and tetraploids (e.g., Ledingham, 1940; Gillies, 1972; Lesins, 1972; Smith, 1984; Veronesi et al., 1986), it appears that natural triploidy is unknown in the genus, so it is simpler to assume that all the individuals with additional alleles in our sample are in fact tetraploids.

3.4. Recombination events and elimination of paralogous (non-homoeologous) copies

Several likely cases of recombination were detected using the thresholds described in the Methods section. We found between zero and three supported events per gene, which suggests that recombination is common, but not overwhelming (Supplementary Table 2). For all alleles with evidence for recombination, we either deleted the recombinant region (where the region was short and near one end of the alignment) or removed the allele completely from the phylogenetic analysis.

Table 2
Numbers of alleles found in the different *M. prostata* individuals and genes used in the present study, with the inferred ploidy level.

Species + individual (ploidy)	Number of alleles per locus								
	Gene 103	Gene 105	Gene 107	Gene 109	Gene 119	Gene 131	Gene 147	Gene 148	Gene 150
Counted diploid									
M. prostrata 21 (2n = 2x = 16)	–[a]	–	2	2	2	2	2	2	–
Presumed diploid									
M. prostrata 4	–	–	2	*[b]	2	2	2	2	–
M. prostrata 9	–	–	2	2	2	2	2	2	–
M. prostrata 11	–	–	2	2	2	2	–	*	–
M. prostrata 12	–	–	2	2	2	2	2	2	–
M. prostrata 17	–	–	2	*	2	*	2	2	–
M. prostrata 18	–	–	2	2	2	2	2	2	–
M. prostrata 19	–	–	2	2	2	2	2	2	–
M. prostrata 23	–	–	2	2	2	2	2	2	–
Counted tetraploid									
M. prostrata 22 (2n = 4x = 32)	–	–	3	3	3	3	3	4	–
Presumed tetraploid (up to 4 alleles)									
M. prostrata 2	–	–	3	3	3	3	4	2	–
M. prostrata 7	–	–	4	2	3	3	*	3	–
M. prostrata 8	–	–	3	3	2	2	3	4	–
M. prostrata 10	–	–	3	3	3	4	3	4	–
M. prostrata 20	–	–	3	2	4	3	3	3	–
Polyploid (either triploid or tetraploid)									
M. prostrata 1	2	2	2	2	3	2	2	2	3
M. prostrata 3	–	–	3	2	3	3	3	3	–
M. prostrata 5	–	–	3	2	3	3	3	2	–
M. prostrata 6	–	–	3	3	3	3	3	2	–
M. prostrata 14	–	–	3	2	3	3	2	3	–
M. prostrata 15	2	2	3	2	3	2	2	2	3
Unknown									
M. prostrata 13	*	*	*	*	*	*	*	*	*
M. prostrata 16	*	*	*	*	*	*	*	*	*

[a] This locus not checked.
[b] Not of sufficient quality to score.

Some alleles displayed site patterns consistent with recombination, but did not satisfy the significance thresholds within RDP3. In gene 119, *M. prostrata*10_0 was one such case. In phylogenetic analyses (NJ trees: not shown) we found that one part of the alignment (aligned positions 1930–2752) placed this allele in a clade containing other diploid and tetraploid *M. prostrata* alleles (85% BS), whereas the remainder of the alignment placed this allele with the *M. sativa* clade (97% BS). In gene 109, *M. prostrata*10_1 was placed in the *M. sativa*_1 clade (100% BS) for part of the alignment (first 1600 bases), but with the *M. prostrata* core clade (plus some non-*M. sativa* species) for the remainder of the alignment (75% BS). We also excluded these alleles from the data set.

We found that recombination involving *M. prostrata* alleles was only detected for tetraploid individuals. The putative parents included the *M. sativa* clade in several genes (109, 119, 150) and was the most common pattern we detected. We re-checked the mapped reads from the putative recombinants in Tablet to confirm that no mistake had been made in the samtools allele phasing step. In each case, we could confirm that several reads containing SNPs from either side of the breakpoints were present in the putative recombinants, thereby ruling out an assembly or allele phasing error. It is clear, then, that several *M. prostrata* tetraploid individuals displaying recombined sequences harbour genetic material derived from these two clades.

We first performed a preliminary phylogenetic analysis with MrBayes using the parameters described below. After inspecting the gene tree topologies, suspected paralogous sequences were removed from the matrices used in the final analyses. These sequences formed additional clades of species already seen in other parts of the trees. In some cases the assessment of paralogy was complemented by observing sequences from outgroup species being more closely related to one putative copy than to the additional clades from the other copy (details not shown).

3.5. Phylogenetic analysis

Phylogenetic analyses were performed using either one or two alleles (or partial alleles) per individual recovered after the assembly-mapping-phasing-recombination testing procedure. The Bayesian analysis of the individual genes (MrBayes) had high (>200) effective sample sizes for all parameters in all genes. The substitution models were HKY (Hasegawa et al., 1985) for all genes except 105, which was GTR (Rodríguez et al., 1990). In the MrBayes tree, *M. prostrata* alleles usually formed a core clade, a monophyletic group containing all alleles from diploids (and some alleles from tetraploids) that was often sister to, or closely related to, the "sativa group" taxa (*M. sativa* + *M. pironae*1 + either of the two *M. papillosa* samples). We also observed in the trees from genes 109, 119, 148 and 150 (Figs. 1–4) a highly supported clade with only a few alleles from some *M. prostrata* tetraploid individuals that were more closely related to the *M. sativa* clade than to diploid *M. prostrata* individuals (see also Table 3). Apart from this, alleles from tetraploid *M. prostrata* individuals found in the *M. prostrata* core clade could not be distinguished from diploid *M. prostrata* alleles. They did not form a separate clade to the diploid alleles in any gene tree (Figs. 1–4, Supp. Figs. 2–5). We placed four of the gene trees into the Supplementary Material in order to focus on those genes that were most decisive (i.e., had highest support for the relevant nodes).

Given that the "sativa group" associated alleles from *M. prostrata* individuals were only sequenced from tetraploids, we interpret these patterns as evidence for hybridisation (further

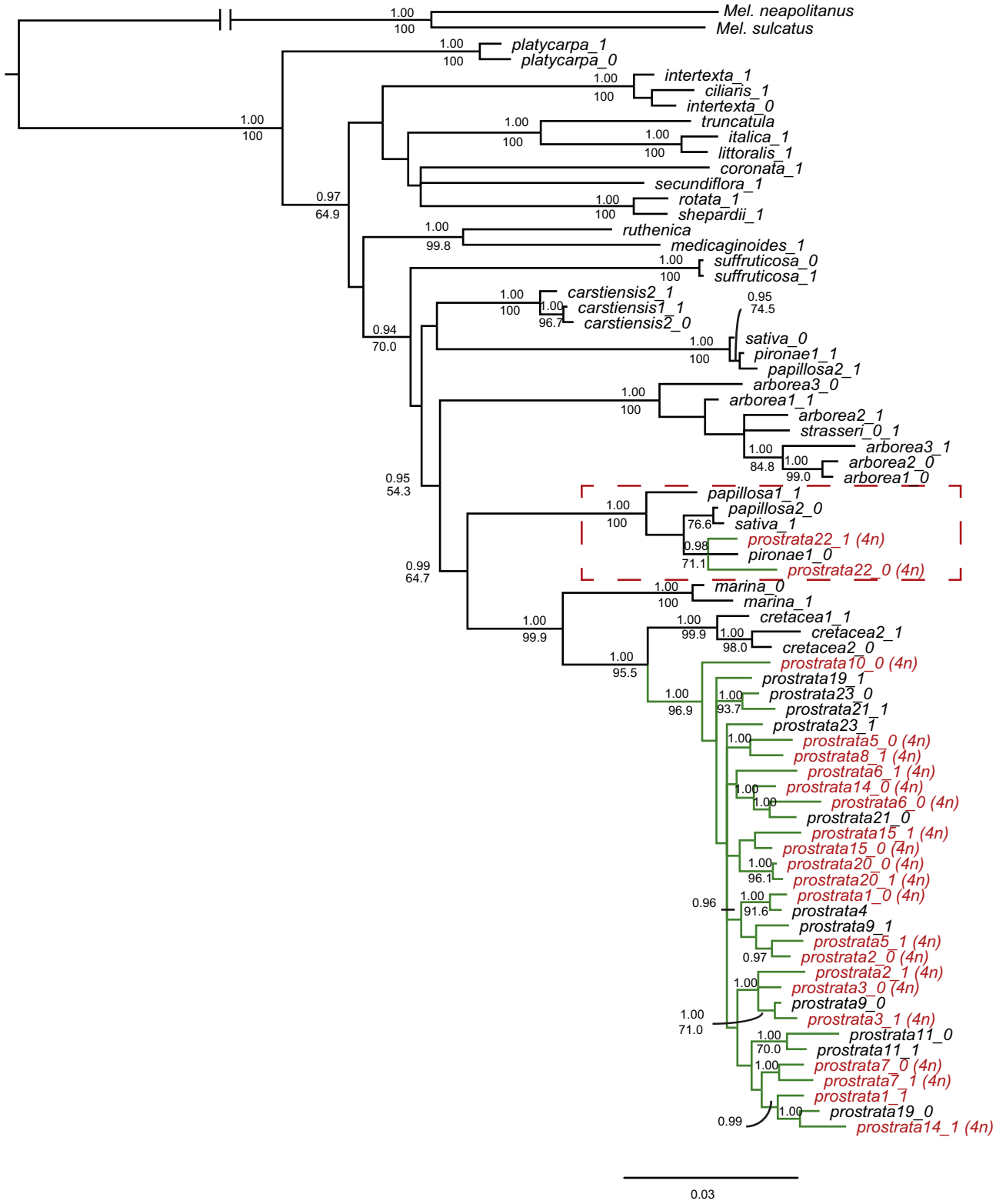

Fig. 1. Phylogenetic relationship of *Medicago* alleles based on gene 109. A 50% majority-rule consensus tree of the BI analysis. Numbers above branches are BI posterior probability values, number below branches are ML bootstrap support values. Tetraploid individuals represented with red labels (online version). *Medicago sativa* complex clade, the genome donor lineage of the tetraploid *M. prostrata,* is indicated with a red dotted square. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tested below) acting to either form an allopolyploid (composed of *M. prostrata* tetraploid individuals) or hybridisation between previously formed autopolyploid species from each respective clade. However, alleles from specific tetraploid individuals were found in both the core clade and the "sativa group" clade in only three of the gene trees (119, 148 150). When we include the

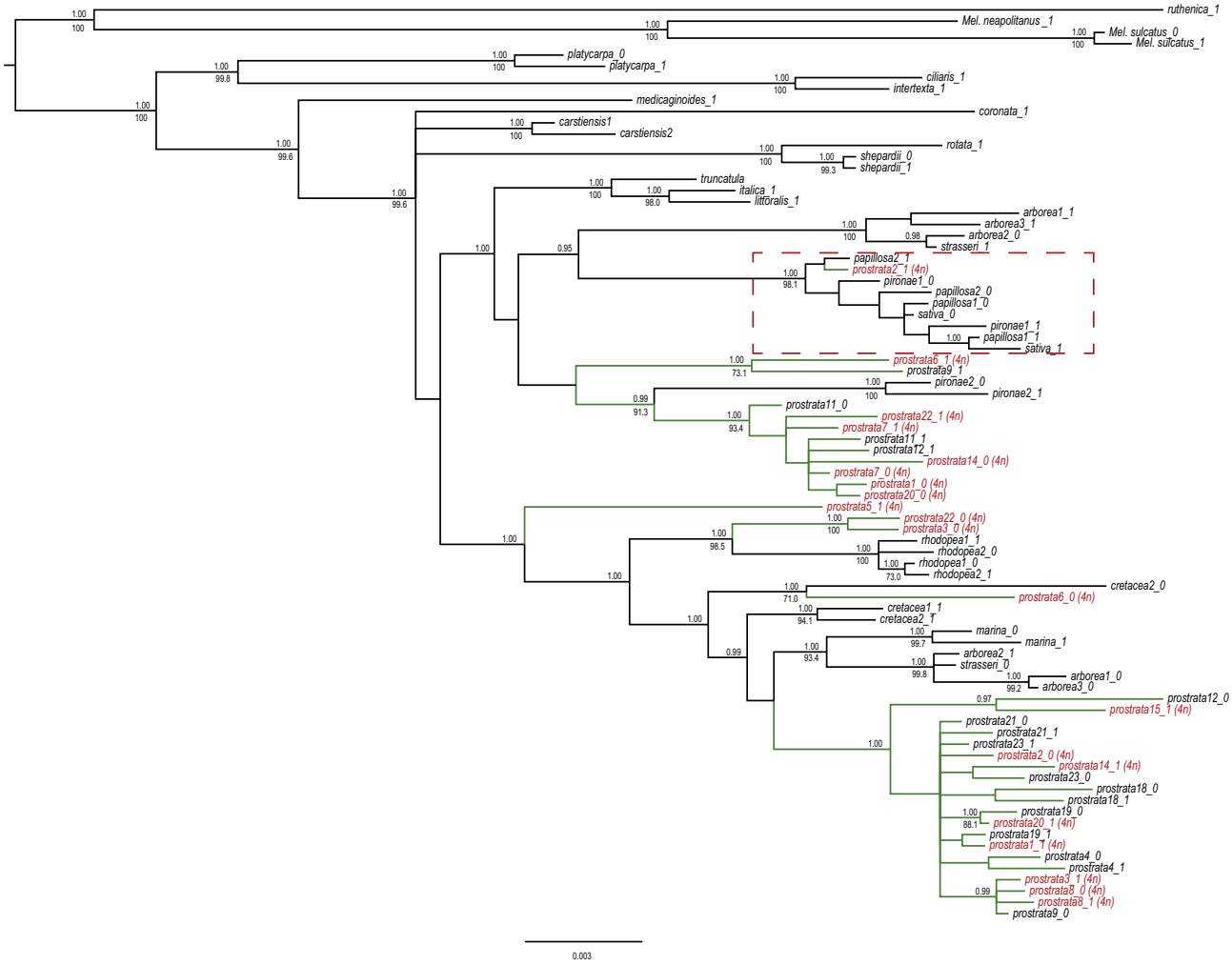

Fig. 2. Phylogenetic relationships of *Medicago* alleles from gene 119. Other details as per Fig. 1.

recombination evidence, gene 109 also shows signs that a *M. prostrata* individual contains sequences from both *M. sativa* group and *M. prostrata* core clades.

Given that the allele phasing method preferentially recovers alleles that are more distinct from one another, the finding that several tetraploid individuals do not have alleles in both clades in several genes is unexpected. One possible explanation is that following polyploidisation some homoeologues are no longer recoverable, through either pseudogenisation (they have changed beyond the gene capture threshold), or that some loci have been removed from the genome. Another possibility is that the loci remain, but retain only alleles from one lineage in several cases (e.g., by the joint action of tetrasomic inheritance and genetic drift). In favour of the latter explanation is that evidence for additional alleles (and therefore the retention of homoeologous loci) is seen (Table 2) for several cases where the phased alleles of tetraploid *M. prostrata* individuals are found in only one of either the *M. prostrata* core clade or the *M. sativa* clade. An example is gene 148, where *M. prostrata* 20_0 and 20_1 are in the core clade, but evidence for a third allele was found (Table 2). Another example is in gene 109, where *M. prostrata* 22_0 and 22_1 are in the *M. sativa* clade, but a third allele is present (Table 2). Another case was observed in gene 119. Thus, it seems simpler to assume that the homoeologous loci are usually retained, but that the allelic variation is not.

The proportion of "sativa group" alleles compared to core *M. prostrata* alleles found in the tetraploid *M. prostrata* individuals was assessed in six loci (109, 119, 148, 150, 103 and 107) where we inferred the phylogeny and found it to be biased with respect to allelic contribution from each source. The other two loci (105 and 147) did not contain sufficient resolution in key parts of the tree to clearly demonstrate "sativa group" parentage for any alleles. Less than half of the alleles came from the "sativa group" in all cases. On average, the tetraploids contained only 0.13 "sativa group" alleles, compared to total alleles (range = 0–0.39; standard deviation 0.14; two outlier values were 0.38 and 0.39; the remaining 10 values range = 0–0.25).

3.6. Hybridisation test

In order to simplify the hybridisation tests we derived a reduced dataset from the genes 109, 148 and 150. Results of the hybridisation test indicated a rejection of the ILS null hypothesis to explain the incongruence between alleles in tetraploid *M. prostrata* individuals from the following pairs of genes: (i) 148 (allele1) – 148 (allele2); (ii) 148 (allele2) – 150; (iii) 148 (allele1) – 109 and (iv)

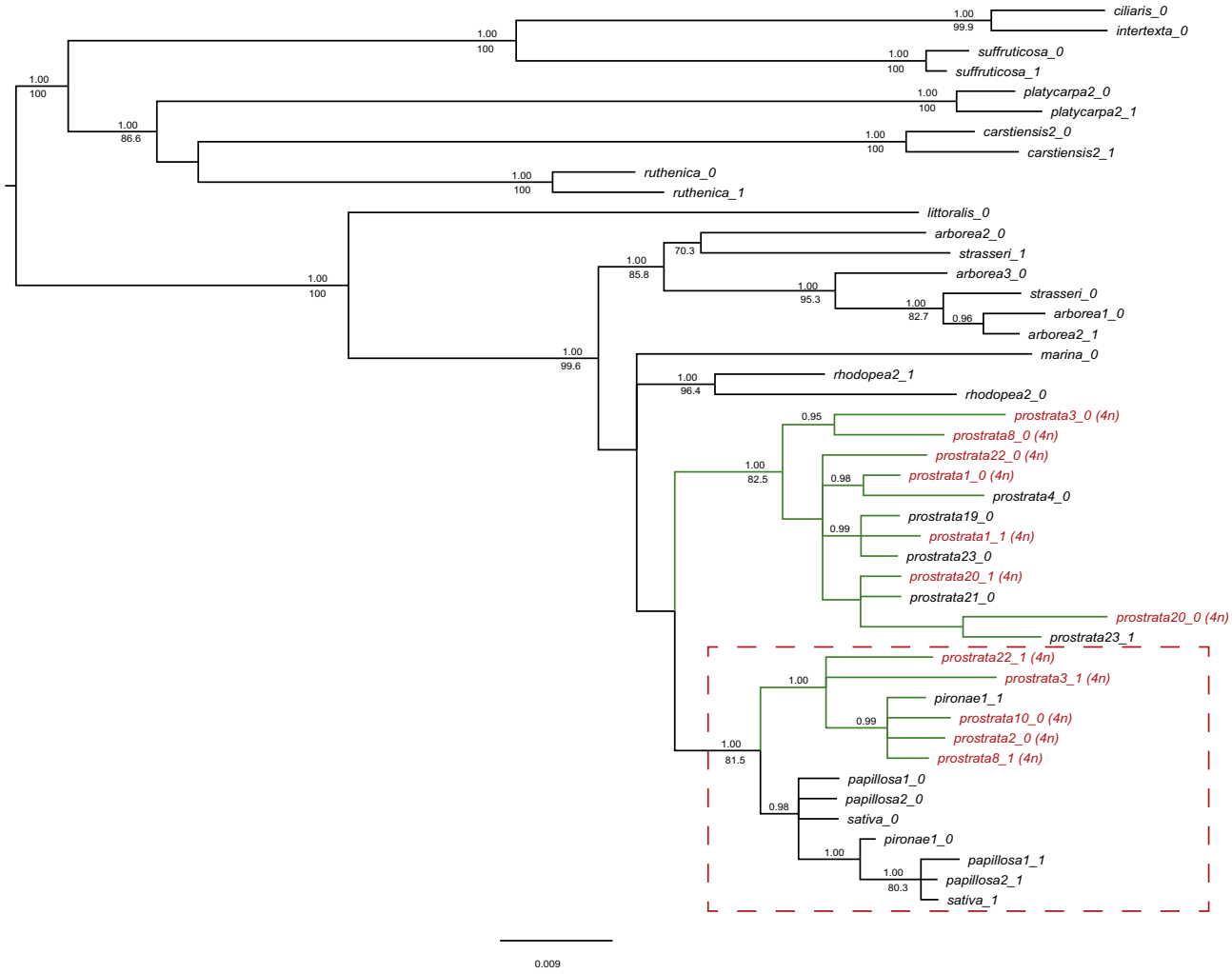

Fig. 3. Phylogenetic relationships of *Medicago* alleles from gene 148. Other details as per Fig. 1.

109–150 (Fig. 5). In each of these comparisons, the tetraploid allele was inferred to be with the diploid in one gene tree, but with the *M. sativa* group in the other gene tree. The differences between the gene trees among these positions (blue[1] distributions; Fig. 5) was significantly greater than the differences expected under the null expectations (red and green distributions; Fig. 5).

The remaining two gene tree comparisons (109–148 [allele2], 148[allele1] – 150) are between topologically similar placements with respect to the tetraploid allele (i.e., within the *M. sativa* clade or sister to the diploid, respectively), and neither of these comparisons rejected the ILS null (Fig. 5). In cases where the hybridisation test rejected the ILS null hypothesis, the removal of the tetraploid allele from the dataset resulted in the test no longer rejecting the ILS null (not shown). Therefore, the hybridisation test results clearly show that ILS alone cannot explain the differences among the gene trees, and that the different positions of the tetraploid alleles are the cause of this result. We accept hybridisation as the best explanation for these observations.

3.7. Morphological observations and distribution

Leaflets of our *M. prostrata* specimens ranged from narrowly obovate to more typically narrowly cuneate. The variation overlapped between diploid and tetraploid specimens. The terminal leaflet apex margin is weakly serrate, with typically three to five teeth, although sometimes more. Wider leaflets tend to have more teeth than narrower leaflets, but the range of tooth number overlapped both within specimens and between diploid and tetraploid specimens. Even by restricting observations to leaflets subtending open flowers (to control for developmental effects), we did not discover distinct patterns of variation in these features between ploidy levels.

The distributions of diploids and tetraploids examined here and previously reported, however, only partially overlapped (Supplementary Fig. 6). The tetraploids scarcely reach north-eastern Italy, whereas diploids have been found in several locations in that country. The distributions may extend beyond the locations we included here, but the ranges are consistent with that reported by Small for the diploids and tetraploids treated as a single group (Small, 2011). The type material apparently hails from Italy (Small, 2011: p. 502) and is therefore most likely diploid.

4. Discussion

4.1. Medicago prostrata is not the result of homoploid hybrid speciation

We found no evidence to support the hypothesis that *Medicago prostrata* is the result of homoploid hybrid speciation. Instead, the observations are consistent with one of several closely allied possi-

[1] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.

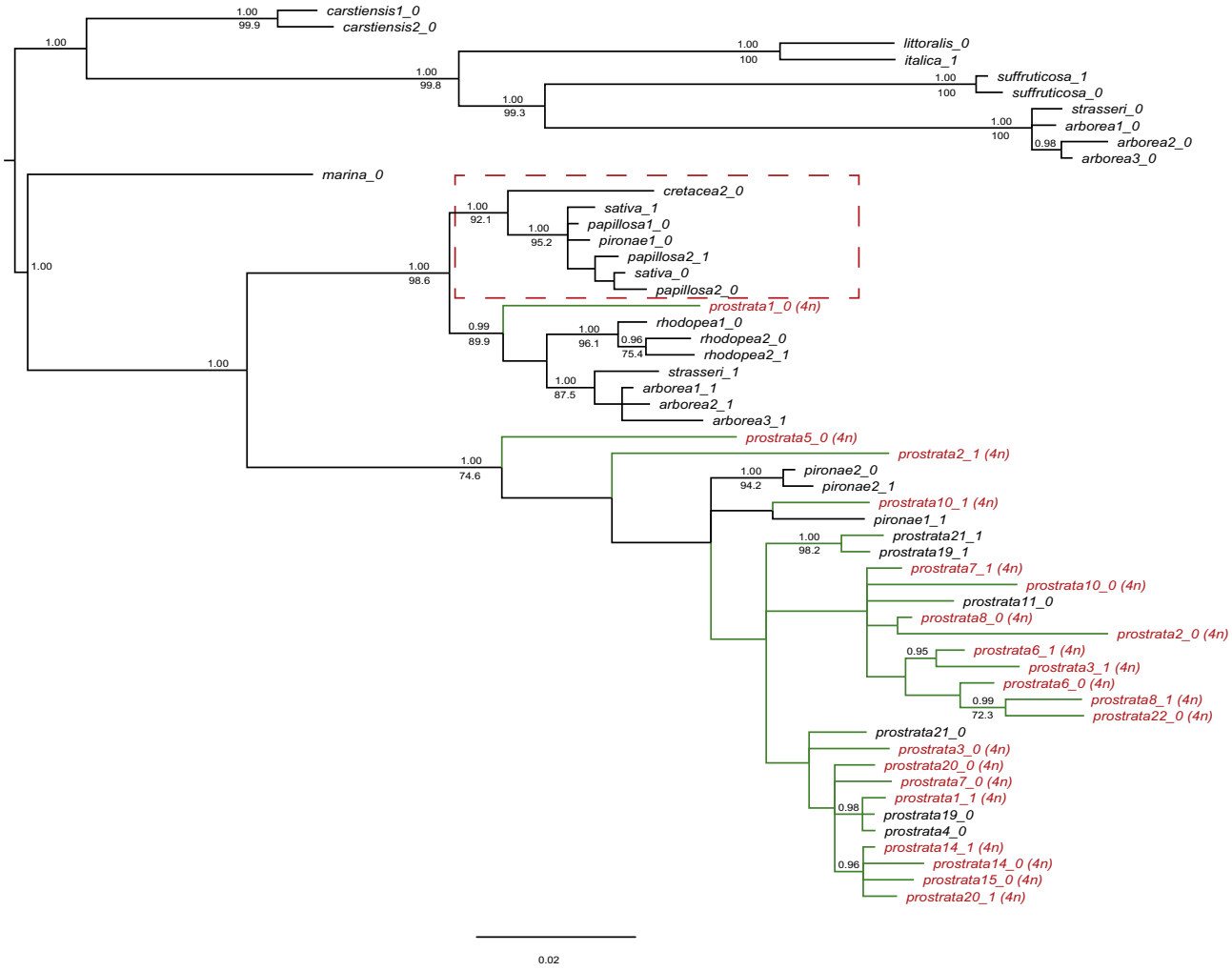

Fig. 4. Phylogenetic relationships of *Medicago* alleles from gene 150. Other details as per Fig. 1.

Table 3
M. prostrata core clade is a clade that includes the majority of *M. prostrata* alleles, including all alleles from diploid individuals and excluding any "sativa group" samples (*M. sativa* + *M. pironae*1 + either *M. papillosa* sample). The second pattern is a clade containing tetraploid *M. prostrata* alleles in addition to some or all of the "sativa group" alleles. The posterior probabilities for each clade are given. Left, Bayesian posterior probabilities of the MrBayes tree; Right, Bootstrap support obtained in the PhyML Maximum Likelihood trees.

Gene tree #	Diploid and tetraploid *M. prostrata* core clade	Tetraploid only *M. prostrata* + "sativa group" clade
103	0.95/29	0.95/18
105	0.51/0[a]	Not distinct
107	Two core clades	1.0/69
109	1.0/99	0.98/77
119	Several core clades	1.0/89
147	0.54/35	Polytomy
148	1.0/78	1.0/77
150	1.0/70[a]	0.99/92

[‡] Includes alleles from *M. pironae*, usually associated with *M. sativa*.

bilities: (1) an allopolyploid origin that gave rise to tetraploid individuals, contributing those alleles that are closer to the *M. sativa* clade from one progenitor, with diploid *M. prostrata* as the other genome donor species. (2) An autopolyploid origin, followed by hybridisation between tetraploid *M. prostrata* and a tetraploid *M. sativa*-clade species. (3) A segmental allopolyploid origin – essentially option (1) coupled with tetrasomic inheritance of some or all of the genome (Stebbins, 1947; Gaut and Doebley, 1997; Parisod et al., 2010). We favour the second explanation and present the lines of reasoning that support this, below.

4.2. *Medicago prostrata* tetraploids are of autotetraploid origin, introgressed with *M. sativa*

We found evidence in favour of hybridisation in the observation that alleles from *M. prostrata* individuals were found in a core clade and also in a clade with *M. sativa*. The differences in these topological positions were not consistent with ILS alone, as this hypothesis was rejected by the coalescent-based hybridisation test. Hybridisation, however, appears only to have introduced *M. sativa* clade alleles into tetraploid individuals of *M. prostrata*. Diploid *M. prostrata* individuals never contained *M. sativa*-clade alleles in the eight genes we examined. Tetraploid individuals were also the only ones containing recombinants between *M. sativa* and *M. prostrata*-clade alleles, consistent with hybridisation only affecting tetraploid *M. prostrata*.

Eight of twelve tetraploid *M. prostrata* individual contained *M. sativa*-clade alleles, but not at each locus we sampled, suggesting that tetrasomic inheritance may be occurring. If the genome were fully disomic (as expected for a classic allopolyploid), then we would expect that both parental alleles would be retained at all independently segregating loci across the genome (i.e., fixed hybridity). Tetrasomic inheritance, on the other hand, can allow

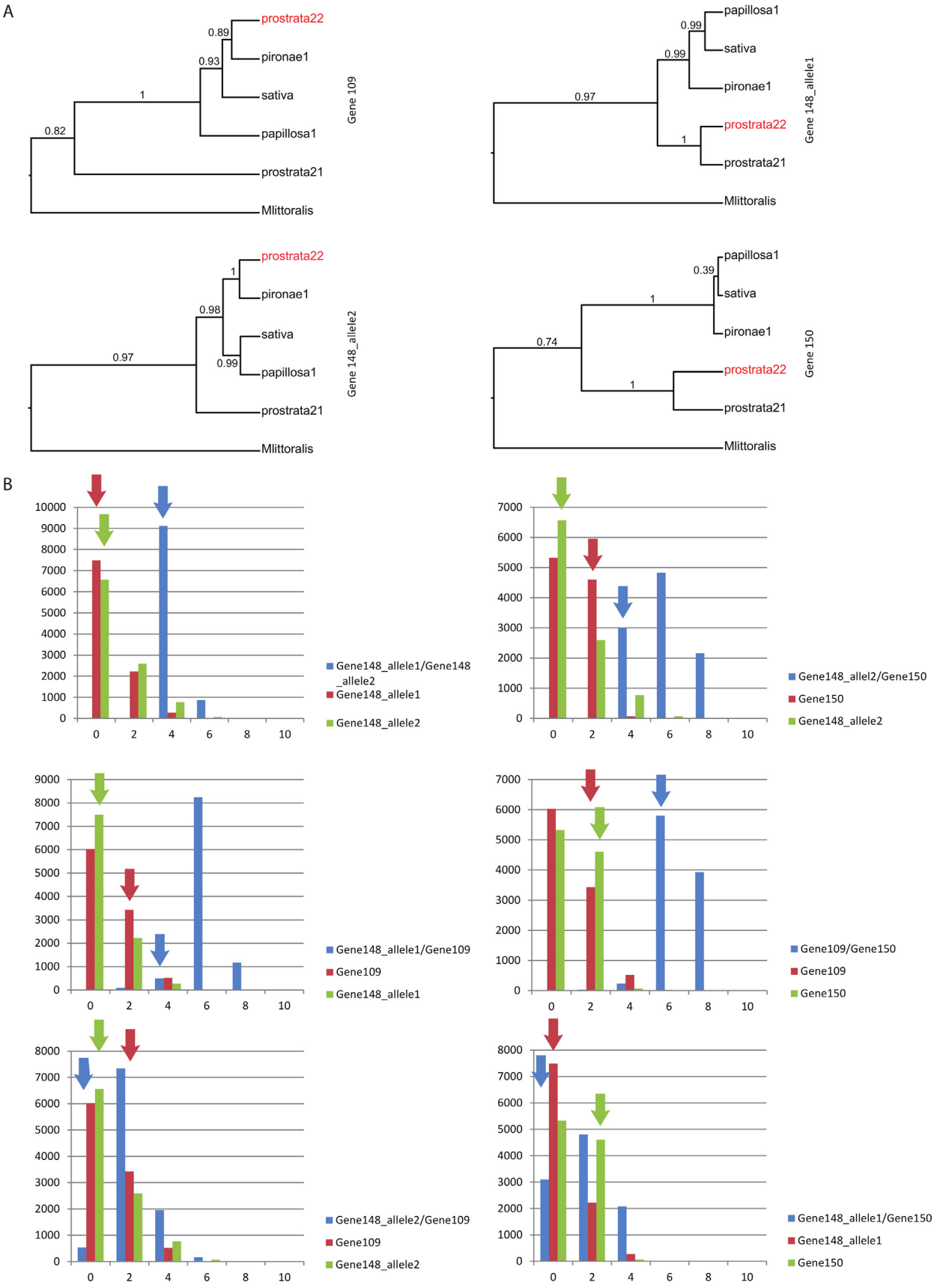

Fig. 5. Coalescent simulation hybridisation test results. (A) Summary trees showing the set of alleles used in the hybridisation test (trees include PP values). (B) Tree-to-tree distances. Right hand side: distances between 100 representative trees drawn from the Bayesian posterior distribution of each observed gene tree from one another (10000 combinations per pair-wise comparison). Left hand side: distances from 100 representative trees to 100 simulated gene trees (from each representative tree) that make up the null distribution for each observed gene tree (10000 comparisons in total).

the loss of alleles from one or the other parent from the population at each locus in the absence of, e.g., balancing selection. Furthermore, the low representation of alleles from the *M. sativa* clade in tetraploid *M. prostrata* is also consistent with the former alleles being at low frequency in the original population. This is not consistent with classic allopolyploidy (where each individual would initially contain alleles from both parents for every gene), but rather with the introgression of a limited number of *M. sativa* alleles into a background of *M. prostrata* (where *M. prostrata* was of autotetraploid origin).

A segmental allopolyploid origin of *M. prostrata* tetraploids (where part of the hybrid-origin genome tetrasomically segregates) is also not supported, because we would expect an initially even sampling of alleles at each locus, followed by a stochastic shift to either parent's alleles for loci undergoing tetrasomic inheritance (in the absence of selection). However, we observed a low frequency of *M. sativa* alleles only, at all loci among the tetraploid individuals. Consistent with an autopolyploid origin is the lack of morphological differentiation between diploid and tetraploid *M. prostrata* individuals seen here and previously (Lesins and Lesins, 1960). This is also what we might expect if the introgression of *M. sativa* alleles was limited to begin with, thus not allowing much additional variation to be explored by selection (and possible fixation to promote differentiation between ploidy levels).

The origin of tetraploid *M. prostrata* appears to be very recent, given that there is no distinct clade of tetraploid alleles among alleles from diploid *M. prostrata* individuals. It is also possible that repeated autotetraploidy continues to introduce diploid allelic variation into the tetraploids (as occurs in the *M. sativa* complex (Small, 2011)), thus obscuring how recently the tetraploids have formed. This would be rather typical of recently formed autotetraploids, but testing this will require more detailed sampling and paternity analyses that have not been considered yet. If gene flow is occurring between the ploidy levels, we expect that it is one directional. Both theory and our observations support this: the former because of the known mechanism of unreduced diploid gametes produced by diploids but the inability of tetraploids to form haploid gametes, and the latter by the lack of *M. sativa* alleles in any of the 72 combinations (eight loci sampled from nine diploid *M. prostrata* individuals) where they might have been found.

Although paralogy is, in general, another possible explanation for gene tree incongruence, the results from most genes are not consistent with this explanation. We found a strongly supported pattern of some tetraploid alleles drawn from the "sativa group" clade in four genes (and weakly supported in two more), a result that is extremely unlikely due to stochastic gene duplication and loss. Paralogous alleles maintained through several speciation events, with loss of copies to return them to single copy in all individuals, would typically result in gene trees where the paralogous sequences attached more deeply than orthologous ones would. This was clearly not the case in the test set of alleles we used for rejection of the ILS null (Fig. 5).

4.3. M. Prostrata tetraploids are a different species

As argued previously, autopolyploids that differ cryptically from their progenitors can in some cases be considered to be different species (Soltis et al., 2007). The criteria by which complexes containing more than one ploidy can be judged to contain more than one species should be no different to any other cases (Soltis et al., 2007). However, in many cases single morphologically-delimited species have been described that include multiple ploidy levels (Soltis et al., 2007 and references therein). In *Medicago* this has also been the case for some species, the well-studied alfalfa and relatives notwithstanding. *Medicago cretacea*, *M. intertexta*, *M. lupulina*, *M. papillosa* and *M. scutellata*, along with *M. prostrata* have been attributed with chromosome counts of $2n = 16$ and $2n = 32$ (Small, 2011). Without commenting on the complexities in each case, it is sufficient to say that *Medicago* taxonomy has not escaped the possibility that autotetraploids and diploids may be presently included under single names.

We propose that *M. prostrata* tetraploids should be regarded as a different species to the diploids, despite no known morphological distinctiveness from the diploids. If a group of individual organisms on an independent evolutionary trajectory to all other groups can be regarded as a separate species (i.e., they are separately evolving metapopulation lineages, sensu Queiroz, 2007), then *M. prostrata* tetraploids qualify. The diploids and tetraploids have different distributions (Supplementary Fig. 6), although partly overlapping, that suggests that they also have differing ecological preferences. Given that autopolyploid speciation must occur in sympatry with the diploid progenitor, the shift in distribution since polyploidy indicates that the tetraploids have changed their ecological preferences, rather than dispersing to a new suitable area that the diploids simply failed to colonise by chance. Gene flow does not appear to occur, at least from tetraploids to diploids. Gene flow in the other direction has not been tested. The tetraploids contain alleles from the *M. sativa* group, and are thus forming a distinct genetic cluster to the exclusion of the diploids. Using the unified species concept (Queiroz, 2007), we suggest that there is some evidence that ecological, interbreeding and genetic cohesion criteria have been fulfilled, providing grounds to recognise the tetraploids as a new species that is partway along the road to complete separation from the diploids.

Although our sampling in the *M. sativa* complex was not extensive, a previous result (Havananda et al., 2011) suggests that the source of the alleles that have introgressed into *M. prostrata* tetraploids may be *M. falcata* ssp. *falcata* (a tetraploid taxon). Havananda et al. (2011) found that two short non-coding chloroplast sequences were identical between some *M. prostrata* diploids and *M. falcata* ssp. *falcata*. Assuming that *M. prostrata* diploids and tetraploids share these sequences, this result could have occurred by introgression from *M. prostrata* tetraploids *into M. falcata* ssp. *falcata*, with the former acting as the male progenitor. A barrier to hybridisation has been observed between *M. sativa* and *M. prostrata* (at diploid and tetraploid levels) when the latter was the female progenitor (Lesins, 1962), but it should be noted that plastid inheritance in *M. sativa* has been shown to be partly paternal to a greater or lesser degree (Schumann and Hancock, 1989; Forsthoefel et al., 1992).

Medicago falcata ssp. *falcata* has a distribution that broadly overlaps with that of tetraploid *M. prostrata* in north-eastern Italy, Slovenia, Croatia, Hungary, Slovakia and eastern Austria, so there appears to be ample opportunity for hybridisation between these taxa to have occurred. It is also reasonable to speculate that the increased range of tetraploid *M. prostrata* may in fact be due to the influence of *M. falcata* ssp. *falcata* alleles. The latter taxon also has a broader distribution, notably to the north and east, presumably indicating a greater tolerance to cold winters than diploid *M. prostrata*, which is confined to a region much closer to the Mediterranean Sea.

4.4. Previous results

Our conclusion of an autopolyploid origin for tetraploid *M. prostrata* individuals, with no clear evidence for hybridisation at the diploid level for this species, also needs to be reconciled with the previously published observations in Maureira-Butler et al. (2008). That study reported discord between two gene trees within a smaller sample of *M. prostrata* alleles, a result that could be obtained from our data set if we select only one *M. prostrata* allele in each gene, which initially suggested HHS. In gene 150, if we had

only sampled *M. prostrata1* allele1 we would have observed it as sister to *M. arborea* + *M. rhodopea,* whereas in gene 147 *M. prostrata1* (either allele) would be sister to the *M. sativa* clade with high support – results largely similar to the previous ones (Maureira-Butler et al., 2008). This is a pattern we could see in all our trees, if selected alleles were to be sampled. We conclude, therefore, that it is extremely important to sample many individuals, loci and alleles, to be able to determine what processes are causing phylogenetic discordance.

4.5. Implications for the study of autopolyploidy

Speciation is not necessarily accompanied by morphological change. The true number of species (separately evolving metapopulation lineages *sensu* (Queiroz, 2007)) is likely to be much greater than the number of morphologically circumscribed species. In recent years, advances in DNA sequencing have revolutionised the ability of scientists to describe and delimit biological diversity (Hebert et al., 2004). With new sequencing technologies we can go far beyond the charting of species diversity, into the description of the evolutionary processes that generate this diversity. Although the speciation process described here is that of autopolyploidy, hybridisation has also played an important role. The introgression of alleles only into the tetraploid is a key indicator of the boundaries of successful gene flow between the *M. prostrata* diploid and tetraploid species. Thus, we have uncovered evidence from gene flow that demonstrates that the true species boundaries in this species pair do not match that suggested by morphology alone.

In a more general way, our results suggest a mechanism by which autopolyploidy may not necessarily be an evolutionary dead end. When more than one closely related autopolyploid species co-occurs, the potential to exchange chromosomes via hybridisation, but maintain fertility (because the chromosomes come with pairing partners via diploid gametes), may allow some of the benefits of an allopolyploid to be realised. For example, chromosome pairing may be more regular (i.e., forming bivalents, as in classic allopolyploids, rather than multivalents), as pairs are most likely to form between chromosomes derived only from each parent, thus increasing fertility (Stebbins, 1950). The action of natural selection on specific alleles derived from one parent in the genetic background of the other may allow for novel genetic combinations that improve fitness. Although the total number of alleles that can be maintained at each locus does not change, each locus can maintain alleles that are more distinct than without hybridisation (both by the introgression of more distinct alleles, but also the expected increase in diploidisation at these chromosomal segments), which may promote fitness where genetic diversity is beneficial, e.g., among disease resistance loci.

How much a particular autopolyploid can benefit from hybridisation with other polyploids will be very dependent on the particular situation – the number and proximity of other species at the same ploidy level, how closely related they are to one another and the presence of particular alleles from one species that are beneficial when introgressed into the other species. Nonetheless, the apparently common occurrence of autopolyploids (Soltis et al., 2007; Parisod et al., 2010 and references therein; Barker et al., 2015) may be a way for plants to share genes across species boundaries in new combinations without disrupting adaptive gene complexes in diploid species. If the new combinations are fitter than their parents, then they may end up outcompeting and replacing them. If not, the diploids will not have been 'polluted' by maladaptive alleles or combinations from the autopolyploid species, thus remaining suited to their current environments.

Autopolyploidy could be an ongoing experiment in testing new genotypes, rather than only of short-term evolutionary significance (Stebbins, 1950). If so, we might expect that ploidy and chromosome number gradually increase over time, as the occasional autopolyploid experiment is successful, with aneuploid reductions simultaneously operating to reverse 'genomic bloat'. There are certainly many known examples of allopolyploid series in plant taxa where diploid species are few in number compared to polyploids, or where diploids are not known (and perhaps extinct), matching the expectation of 'genomic bloat' as outlined above [*Hibiscus*: (Wilson, 1994), *Viola*: (Marcussen et al., 2012), *Fumaria*: (Bertrand et al., 2015), *Glycine*: (Kumar and Hymowitz, 1989)], as well as examples where the mode of polyploidy is incompletely known but probably also mostly allopolyploid [*Pavonia*: (Fryxell, 1999), Poaceae: (Levy and Feldman, 2002), Brassicaceae: (Kagale et al., 2014)]. The grass family Poaceae, as a case in point, contains more recent polyploids than diploids (including auto-, segmental- and allopolyploids) and was probably derived from a polyploid ancestor in any case (Cui et al., 2006; Soltis et al., 2009).

There are, however, far fewer direct examples of the predicted 'genomic bloat' in groups of species comprising only or mostly autopolyploids. This may be because information regarding autopolyploidy is harder to find, because autopolyploids are rarely described as distinct species. In *Helianthus decapetalus*, three chromosome numbers have been reported. If each was assigned to different species, then this group would have two polyploids (4*n* and 6*n*) and one diploid species (Church and Spaulding, 2009). In *Medicago sativa*, two or three diploid subspecies are recognised in some classifications, with perhaps the same number of autopolyploids known, which are also known to hybridise (Small, 2011; Kaljund and Leht, 2013). If hybridisation is an important mechanism for autopolyploids to escape from their diploid progenitors and possibly outcompete them, as we speculate, very few studies of autopolyploids are detailed enough to show this. Long term polyploid success and repeated increases in ploidy over time has certainly been a viable strategy in seed plant evolution (Cui et al., 2006; Van de Peer et al., 2009; Jiao et al., 2011) and may not necessarily be restricted to classic allopolyploids, if some of the benefits can be achieved via autopolyploidy followed by hybridisation among polyploids.

5. Taxonomy

Medicago tetraprostrata J.S. Eriksson & B.E. Pfeil, sp. nov.

Most similar to *M. prostrata* Jacq., but tetraploid (2*n* = 4*x* = 32). No distinguishing morphological features are known.

Type: SWEDEN: Västra Götaland. Species grown and collected in phytotron at the University of Gothenburg, Dept. Biological and Environmental Sciences. Date of collection: 7th May 2013, *Jonna S. Eriksson 51* (GB 0151165, holotype). Original provenance: Seed grown from USDA National Plant Germplasm System (http://www.ars-grin.gov/npgs/) collection PI577450. Original collection: No further details available, *K.A. Lesins and I. Lesins 2286*. (Listed as *M. prostrata 22* in Table 1).

Description: We refer readers to previously published descriptions of *M. prostrata* (e.g. in Small (2011)).

Distribution: Known tetraploid specimens have been found in the southern and western Balkans (Greece, Bosnia and Herzegovina, Croatia and Slovenia) to Eastern Italy, Western Romania, Hungary, Eastern Austria and Slovakia. It is presumably also found in regions in between (i.e., Albania, Macedonia, Kosovo, Montenegro and Serbia). Diploids overlap with the tetraploids in the Western Balkans (Greece, Croatia and Slovenia) and Eastern Italy. They extend beyond the known tetraploid range to Central Italy.

Selected specimens: GREECE: Grevenon, Mt. Vourinos. Lat.: 40 09′N Long.: 21 39′E, altitude 800–1000 m. Collected 25th May 1999, *Strid & al. 48434* (LD 1592206) (Listed as *M. prostrata 7* in Table 1). HUNGARY: Budapest. Collected June 1920, *Wagner s.n.*

(GB 0144771) (Listed as *M. prostrata 2* in Table 1). HUNGARY: Pest. Steinige Abhänge des Adlersberges b. Ofen. – Dolomit, 200 m. Collected 4th July 1873, *J. Freyn, s.n.* (GB 0144775) (Listed as *M. prostrata 8* in Table 1). ITALY: Trieste. F. Schultz, herbarium normale, Cent. 11, exsiccata 1038. Lieux herbeux, secs et rocailleux de la montagne calcaire Monte Spiccato. 1852, *M. de Tommasini s.n.* (GB 0144772) (Listed as *M. prostrata 10* in Table 1).

Acknowledgments

We thank two anonymous reviewers and especially Christoph Oberprieler for comments that improved the paper. This work was supported by grants from the Swedish Research Council, the Royal Swedish Academy of Sciences (Grant 2009-5206), Lars Hiertas Minne Fund, The Royal Physiographic Society in Lund, Helge Ax:son Johnsons Fund and the Lundgrenska Fund to BEP; from the P. A. Larssons Fund and Lars Hiertas Minne Fund to FS; from the European Research Council (Marie Curie postdoctoral fellowship) to JLBP.

JSE, JLBP, YJKB and BEP wrote the paper; JSE, FS and BEP designed the probes; JSE and FS performed the lab work; FS provided materials; FS and YJKB designed the bioinformatics scripts; JSE, JLBP, FS and BEP performed analyses; BEP designed the study.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.11.020.

References

Abbott, R.J., 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. Trends Ecol. Evol. 7, 401–405.

Barker, M.S., Arrigo, N., Baniaga, A.E., Li, Z., Levin, D.A., 2015. On the relative abundance of autopolyploids and allopolyploids. New Phytol. http://dx.doi.org/10.1111/nph.13698.

Bena, G., Lyet, A., Huguet, T., Olivieri, I., 2005. *Medicago - Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. J. Evol. Biol. 18, 1547–1558.

Bertrand, Y.J.K., Scheen, A.-C., Marcussen, T., Pfeil, B.E., Sousa, F., Oxelman, B., 2015. Assignment of homoeologues to parental genomes in allopolyploids for species tree inference, with an example from *Fumaria* (Papaveraceae). Syst. Biol. 64, 448–471.

Blanc, G., Wolfe, K.H., 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667–1678.

Blanco-Pastor, J.L., Vargas, P., Pfeil, B.E., 2012. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. PLoS One 7, e39089.

Bucklin, A., Steinke, D., Blanco-Bercial, L., 2011. DNA barcoding of marine metazoa. Ann. Rev. Mar. Sci. 3, 471–508.

Buerkle, C.A., Morris, R.J., Asmussen, M.A., Rieseberg, L.H., 2000. The likelihood of homoploid hybrid speciation. Heredity 84, 441–451.

Church, S.A., Spaulding, E.J., 2009. Gene expression in a wild autopolyploid sunflower series. J. Hered. 100, 491–495.

Cui, L., Wall, P.K., Leebens-Mack, J.H., Lindsay, B.G., Soltis, D.E., Doyle, J.J., Soltis, P.S., Carlson, J.E., Arumuganathan, K., Barakat, A., Albert, V.A., Ma, H., dePamphilis, C. W., 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16, 738–749.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9. 772-772.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

Dittrich-Reed, D.R., Fitzpatrick, B.M., 2013. Transgressive hybrids as hopeful monsters. Evol. Biol. 40, 310–315.

Doyle, J.J., Doyle, J.A., Brown, A.H.D., Palmer, R.G., 2002. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: Histone H3-D gene sequences. Evolution 56, 1388–1402.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Forsthoefel, N.R., Bohnert, H.J., Smith, S.E., 1992. Discordant inheritance of mitochondrial and plastid DNA in diverse alfalfa genotypes. J. Hered. 83, 342–345.

Fryxell, P.A., 1999. Pavonia Cavanilles (Malvaceae). Flora Neotropica Monogr. 76.

Gaut, B.S., Doebley, J.F., 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA 94, 6809–6814.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Gillies, C.B., 1972. Pachytene chromosomes of perennial *Medicago* species. Hereditas 72, 289–302.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W.W., Fritz, M.H.Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, 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, Z., Gusic, 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.

Gross, B.L., Rieseberg, L.H., 2005. The ecological genetics of homoploid hybrid speciation. J. Hered. 96, 241–252.

Grumer, J.A., Bryson, R.W., Reeder, T.W., 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst. Biol. 63, 119–133.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21, 160–174.

Havananda, T., Brummer, E.C., Doyle, J.J., 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). Am. J. Bot. 98, 1633–1646.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101, 14812–14817.

Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V., Huber, K.T., 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evol. Biol. 8, 202.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S. E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100.

Joly, S., McLenachan, P.A., Lockhart, P.J., 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174, E54–E70.

Jones, G., Aydin, Z., Oxelman, B., 2014. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31, 991–998.

Jones, G., Sagitov, S., Oxelman, B., 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. Syst. Biol. 62, 467–478.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., Sharpe, A.G., Parkin, I.A.P., 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell 26, 2777–2791.

Kaljund, K., Leht, M., 2013. Extensive introgressive hybridization between cultivated lucerne and the native sickle medic (*Medicago sativa* ssp. *falcata*) in Estonia. Annales Botanici Fennici 50, 23–31.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A.J., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kumar, P.S., Hymowitz, T., 1989. Where are the diploid (2n = 2x = 20) genome donors of *Glycine* Willd. (Leguminosae, Papilionoideae)? Euphytica 40, 221–226.

Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst. Biol. 54, 575–594.

Ledingham, G.F., 1940. Cytological and developmental studies of hybrids between *Medicago sativa* and a diploid form of *M. falcata*. Genetics 25, 1–15.

Lee, J.-Y., Mummenhoff, K., Bowman, J.L., 2002. Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). Proc. Natl. Acad. Sci. USA 99, 16835–16840.

Lesins, K., 1962. Interspecific crosses involving alfalfa. III. *Medicago sativa* L. × *M. prostrata* Jacq. Can. J. Genet. Cytol. 4, 14–23.

Lesins, K., Lesins, I., 1960. Sibling species in *Medicago prostrata* Jacq. Can. J. Genet. Cytol. 2, 416–417.

Lesins, K.A., 1972. Interspecific crosses involving alfalfa. VII. *Medicago sativa* × *M. rhodopea*. Can. J. Genet. Cytol. 14, 221–226.

Levy, A.A., Feldman, M., 2002. The impact of polyploidy on grass genome evolution. Plant Physiol. 130, 1587–1593.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, J., Möller, M., Gao, L.-M., Zhang, D.-Q., Li, D.-Z., 2011. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. Mol. Ecol. Resour. 11, 89–100.

Lynch, M., Connery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.
Marcussen, T., Jakobsen, K.S., Danihelka, J., Ballard, H.E., Blaxland, K., Brysting, A.K., Oxelman, B., 2012. Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). Syst. Biol. 61, 107–126.
Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102.
Martin, D.P., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martinsson, S., Rota, E., Erséus, C., 2014. Revision of *Cognettia* (Clitellata, Enchytraeidae): reestablishment of *Chamaedrilus* and description of cryptic species in the sphagnetorum complex. Syst. Biodivers. 13, 257–277.
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.
Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
Meng, C., Kubatko, L.S., 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor. Popul. Biol. 75, 35–45.
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. Bioinf. 14, 193–202.
Parisod, C., Holderegger, R., Brochmann, C., 2010. Evolutionary consequences of autopolyploidy. New Phytol. 186, 5–17.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762.
Queiroz, K.d., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.
Ramadugu, C., Pfeil, B.E., Manjunath, K.L., Lee, R.F., Maureira-Butler, I.J., Roose, M.L., 2013. Coalescence simulation testing of hybridization versus lineage sorting in *Citrus* (Rutaceae) using six nuclear genes. PLoS One 8, e68410.
Rambaut, A., Drummond, A.J., 2009. Tracer.
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.
Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J.L., Schwarzbach, A.E., Donovan, L.A., Lexer, C., 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301, 1211–1216.
Rodríguez, F., Oliver, J.L., Marín, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Schumann, C.M., Hancock, J.F., 1989. Paternal inheritance of plastids in *Medicago sativa*. Theor. Appl. Genet. 78, 863–866.
Small, E., 2011. Alfalfa and Relatives: Evolution and Classification of *Medicago*. NRC Research Press, Ottawa.
Smith, S.E., 1984. Use of meiotic analysis to describe genomic affinities in *Medicago*. Can. J. Genet. Cytol. 26, 679–681.
Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Kerr Wall, P., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96, 336–348.
Soltis, D.E., Soltis, P.S., Schemske, D.W., Hancock, J.F., Thompson, J.N., Husband, B.C., Judd, W.S., 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56, 13–30.
Sousa, F., Bertrand, Y.J.K., Doyle, J.J., Oxelman, B., Pfeil, B.E., in press. Using genomic location and coalescent simulation to identify paralogy, hybridisation and incomplete lineage sorting in nuclear phylogenies. Syst. Biol. (in press)
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.
Stebbins, G.L., 1947. Types of polyploids: their classification and significance. Adv. Genet. 1, 403–429.
Stebbins, G.L., 1950. Variation and Evolution in Plants. Columbia University Press, New York, NY.
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.
Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L.-D., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and geography of soil fungi. Science 346, 6213.
Ungerer, M.C., Baird, S.J.E., Pan, J., Rieseberg, L.H., 1998. Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95, 11757–11762.
Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., Vandepoele, K., 2009. The flowering world: a tale of duplications. Trends Plant Sci. 14, 680–688.
Veronesi, F., Mariani, A., Bingham, E.T., 1986. Unreduced gametes in diploid *Medicago* and their importance in alfalfa breeding. Theor. Appl. Genet. 72, 37–41.
Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II: Dna Sequencing. Kluwer Academic Publishers, Norwell, pp. 265–296.
Wilson, F.D., 1994. The genome biogeography of *Hibiscus* L. section *Furcaria* DC. Genet. Resour. Crop Evol. 41, 13–25.
Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L. H., 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA 106, 13875–13879.
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.