BIOLOGIJA. 2006. Nr. 1. P. 60–63

The phylogeny of woody Maloideae (Rosaceae) using chloroplast *trnL-trnF* sequence data

R. Verbylaitė,

B. Ford-Lloyd,

J. Newbury

School of Biosciences,
University of Birmingham, U. K.

In this study, the most suitable DNA extraction protocols for Maloideae subfamily species were determined. Also, it was shown that the most suitable method to analyse phylogenetic data, such as observed in this study is the maximum parsimony method.

The monophyletic origin of Maloideae subfamily including *Vauquelinia* and *Kageneckia* were confirmed. Close relationships between *Crataegus* and *Mespilus* were obtained. However, no intra-specific variation within the Maloideae genera according to *trnL-trnF* plastid region was observed, and the hypothesis of *Mespilus canescens* origin still needs more data to be confirmed or rejected.

Key words: phylogeny, *Maloideae*, *trnL-trnF*, sequencing

INTRODUCTION

The Rosaceae family is subdivided into four subfamilies. The subfamilies are: Spiraeoideae, Rosoideae, Amygdaloideae and Maloideae. To the family Rosaceae belong trees, shrubs and herbs. Leaves are usually deciduous; some members of the family are evergreen. Rosaceae family plants have hermaphrodite flowers and are mostly entomophilous, pollinated by flies. Flowers are solitary or aggregated in different types of inflorescence. Sepals are usually five, stamens 10–20 (rarely more or less). Fruits are varying: follicle, achene, drupaceous, baccate, capsule, berry, pome or a drupe [1]. The Rosaceae family displays a considerable diversity in anatomy, vegetative features, and fruit morphology but despite of this the family is considered monophyletic. Monophyly of Rosaceae is strongly supported by analyses of *rbcL* sequences [2]. The sister group of the Rosaceae according to *rbcL* sequences are Ulmaceae, Celtidaceae, Moraceae, Urticaceae and Rhamnaceae, although pre-molecular techniques suggest that the sister group of Rosaceae are Saxifragaceae, Fabaceae and Crassulaceae [2]. The base chromosome number, various chemical characters, distribution of rust parasites and DNA sequences from rDNA ITS1 and 2 and *rbcL* are data that delimit the current subdivision of *Rosaceae* into subfamilies [3].

The Maloideae subfamily members are characterised by fleshy pome, but phylogenetic data show that this subfamily includes also the *Kageneckia*, *Vauquelinia* and *Lindleya* genera which have drupaceous or follicle fruits [3]. A recent phylogenetic analysis in the subfamily Maloideae, based on ITS1, 5.8S rDNA and ITS2, shows that the genus *Mespilus* is nested within the *Crataegus* clade. This study also suggests that endemic to Arkansas *Mespilus canescens* could be of hybrid origin [4].

Though a huge amount of work has already been done in trying to solve Rosaceae phylogeny, a lot still needs to be done. In the main public databases there is a clear lack of Maloideae species *trnL-trnF* region sequences. This chloroplast DNA region was widely used for looking into the phylogeny of the other *Rosaceae* subfamilies [5–8]. This chloroplast DNA region should be interesting inferring also Maloideae subfamily evolutionary relationships, and new sequence data could be useful for other authors that will try to solve the phylogenetic problems of Rosaceae.

The aim of this study was to assess subfamily Maloideae phylogenetic relationships using *trnL-trnF* chloroplast sequence data and the genetic variation among the species studied.

MATERIALS AND METHODS

The plant species for molecular approach were chosen based on GenBank data (no sequence data were present for species with the loci previously planned to use in this study). For this study, 15 plant species were sampled (see Table 1). More than one individual collected in different locations represented some of plant species. In total, 30 plant samples were collected. Individuals were sampled in the Hilliers arboretum (UK) unless otherwise stated (see Table 1).

Corresponding author: Rita Verbylaitė, Liepų 15-2, Girionys, Kauno raj., LT-53103 Lithuania, e-mail: rita_verbylaite@yahoo.com

Five different DNA extraction protocols were used to find the most suitable to extract DNA from different *Rosaceae* family plants. Three of the methods used were based on the CTAB extraction protocol first described by Doyle and Doyle [9]. This method was employed in many other studies [6, 10–12]. However, the most useful proved to be the DNeasy Plant Mini Kit (Qiagen) also used by many other researchers [7, 8, 13].

Primers suitable for phylogenetic analysis among Maloideae tree species were selected using literature data. A primer set (*trnL5'^{UAA} F - trnF^{GAA}*) originally used by Shaw et al. [10] was chosen. The primer pair selected for this study was designed to amplify the non-coding chloroplast DNA region situated in a large single copy region of the chloroplast genome. At lower taxonomic levels, such primers are known to be very useful [14]. Primers used in this phylogenetic study amplify the *trnL-trnF* chloroplast DNA region. This region comprises two tRNA genes, *trnL* and *trnF*. The non-coding portion of this region is an inter-generic spacer between these two genes. Primer sequences suitable for this region amplification were first described by Taberlet et al. [15], and because of the near-universal nature of these primers, the regions became the most widely used non-coding cpDNA sequences in plant systematics. Although this region is not the most informative one according to Shaw et al. [10], the possibility to obtain some more sequence data from the GenBank of Maloideae subfamily determined the selection. PCR conditions were as given in Shaw et al. [10]. The PCR samples were purified using a QIAquick PCR purification Kit (Qiagen). The purified PCR products were sequenced using the Functional Genomics and Proteomics Lab facilities at the School of Biosciences, University of Birmingham (http://www.genomics.bham.ac.uk/sequencing.htm). The forward and reverse DNA sequence information was aligned and consensus DNA sequences for each sequenced individual were obtained. The consensus sequences were aligned using ClustalX computer software [16]. The phylogenetic trees were inferred using two different methods: maximum likelihood [17] and maximum parsimony [18, 19]. DNA sequences obtained during this molecular study, as well as the Maloideae subfamily species sequences of the *trnL – trnF* locus that were available in GenBank were used in the phylogenetic analysis. Phylogenetic trees were rooted using randomly chosen species from other Rosaceae subfamilies. From each subfamily, five different species belonging to different genera were selected. The plant species whose sequences were obtained from GeneBank and used in this study are listed in Table 1.

The phylogenetic trees were inferred using two different methods: maximum likelihood and maximum parsimony. The phylogenetic trees were generated using 500 replications. Matching sequences obtained from the same species but from different individuals were excluded from the analysis, as the computing time increases exponentially with every operational taxonomic unit (OTU) added. The generated phylogenetic trees were compared, and the maximum parsimony tree proved to be more statistically reliable. A difference between these two methods might arise, because the maximum likelihood method treats indels as multiple evolutionary events [20].

RESULTS AND DISCUSSION

The obtained different species sequence information was aligned, and the multiple alignment revealed that the *Crataegus* species collected in Birmingham Uni-

Table 1. **Species used in this study and trnL-trnF region sequence accession numbers**

Species	No. of individuals sampled	TrnL-trnF sequence length in bp	GenBank accession number
Cydonia oblonga Miller.	2	885	AM157398
Pyrus pyraster Burgsd.	2	771	AM157399
Pyrus communis L.	6 (4*)	933	AM157400
Malus sylvestris (L.) Mill	1*	830	AM157404
Malus domestica Poir. non Borkh	1	883	AM157405
Crataegus submollis Sarg.	1	891	AM157406
Crataegus azarolus L.	1	881	AM157407
Crataegus coccinea L.	1	836	AM157408
Crataegus monogyna Jacq.	5 (2**)	881	AM157397
Crataegus laevigata (Poir.) DC	3	924	AM157401
Crataegus crus-galli L.	2	906	AM157402
Crataegus persimilis Sarg.	2 (1**)	857	AM157403
Mespilus germanica L.	1	875	AM157409
Mespilus canescens Phipps.	1	982	AM157410
Crataemespilus grandiflora (Sm) E.G. Camus	1	897	AM157411

* Individuals were sampled by Mr. Paule from the wild in Slovakia.

** Individuals were sampled in the botanical garden of University of Birmingham, UK.

Table 2. **Plant species whose sequences were obtained from GeneBank**

Species	GeneBank accession number
Subfamily Amygdaloideae	
Prunus virginiana	AF348561
Prinsepia sinensis	AF348558
Exochorda racemosa	AF348542
Oemleria cerasiformis	AF348551
Maddenia hypoleuca	AY864827
Subfamily Spiraeoideae	
Spiraea densiflora	AF348571
Aruncus dioicus	AF348536
Physocarpus opulifolius	AY555417
Sorbaria sorbifolia	AF348569
Neillia thyrsiflora	AF348549
Subfamily Rosoideae	
Filipendula vulgaris	AJ416463
Fragaria vesca	AF348545
Potentilla indica	AY634763
Rosa multiflora	AY634764
Rubus ursinus	AF348568
Subfamily Maloideae	
Cotoneaster pannosus	AF348540
Photinia serrulata	AF348552
Pyracantha fortuneana	AF348563
Pyrus caucasica	AF348564
Sorbus californica	AF348570
Vauquelinia californica	AF348573
Kageneckia oblonga	AF348547

versity Botanical Garden as *C. crus-galli* was *C. prunifolia*. The sequence of this particular individual had no indels possessed by the other two *C. crus-galli* individuals. The sequence similarity of this particular individual did not give any doubt as to the species determination.

The obtained consensus sequences, one for each species studied, were submitted to the EMB GeneBank. The length and accession number of each sequence submitted are given in Table 1.

The sequences obtained in this study (Table 1) as well as sequences selected from GenBank (Table 2) were used to generate a phylogenetic tree. The generated maximum parsimony tree is shown in Fig. 1. The bootstrap values supporting each branch are given at the branch nodes. The dendrogram shows the monophyletic origin of the Maloideae subfamily and confirm the data of other authors [2, 3] that the genera *Vauquelinia* and *Kageneckia* should be placed in the Maloideae subfamily. Maximum parsimony analysis divides the Maloideae subfamily into two clusters. One of them contains *Crataegus* and *Mespilus* species, while the other clade contains all the other species that belong to the Maloideae subfamily.

In order to detect any intra-specific differentiation, maximum parsimony and maximum likelihood

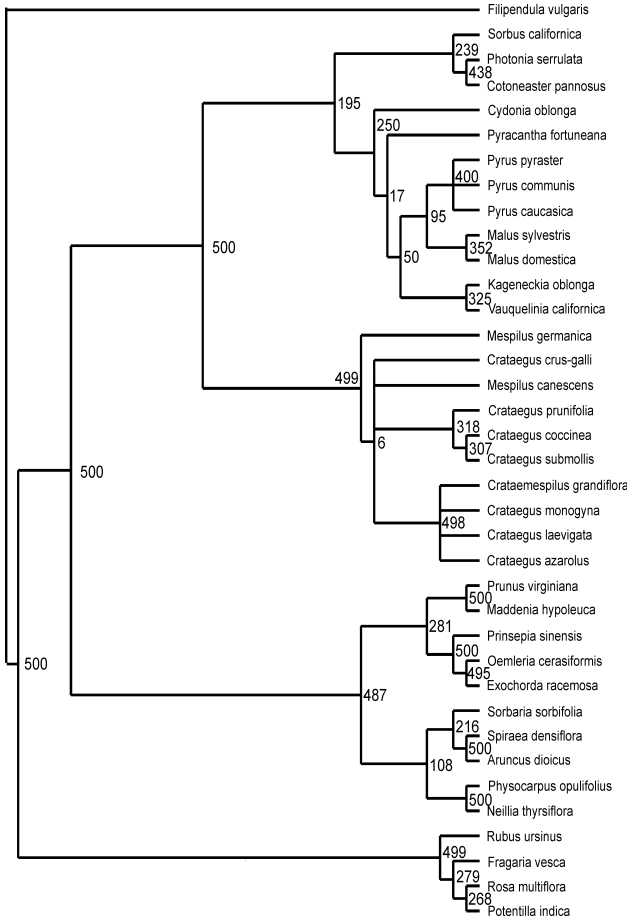

Fig. 1. Consensus of 3971 equally parsimonious trees resulting from maximum parsimony analysis using data from the partial *trnL* gene and inter-generic spacer between *trnL-trnF* (775 aligned nucleotide positions). Bootstrap values were generated by 500 replicate re-samplings. Bootstrap values are given at branch nodes

analyses were performed using only the sequences obtained during this study. However, no significant intra-specific differentiation was obtained. Most of the subdivisions generated during this analysis (dendrogram not shown) are weakly supported statistically and should be treated as not significant [21].

The most extensively studied genus in this study was *Crataegus*. Seven species from the genus were sequenced. To test the hypothesis that endemic to Arkansas *Mespilus canescens* is of hybrid origin, all *Crataegus* sequences of *trnL-trnF* loci were clustered together with both *Mespilus* sequences and one known inter-generic hybrid (*Crataemespilus grandiflora*) using *Prunus virginiana* as an out-group. The closest genus to the *Crataegus* is *Mespilus;* this was shown also by [4]. *Crataemespilus grandiflora* is an inter-generic hybrid between *C. laevigata* and *M. germanica*. The dendrogram generated with maximum parsimony analyses based on the *Crataegus–Mespilus* data subset is presented in Fig. 2. The results obtained in this study do not deny the hypothesis that the endemic Arkansas *M. canescens* could be of hyb-

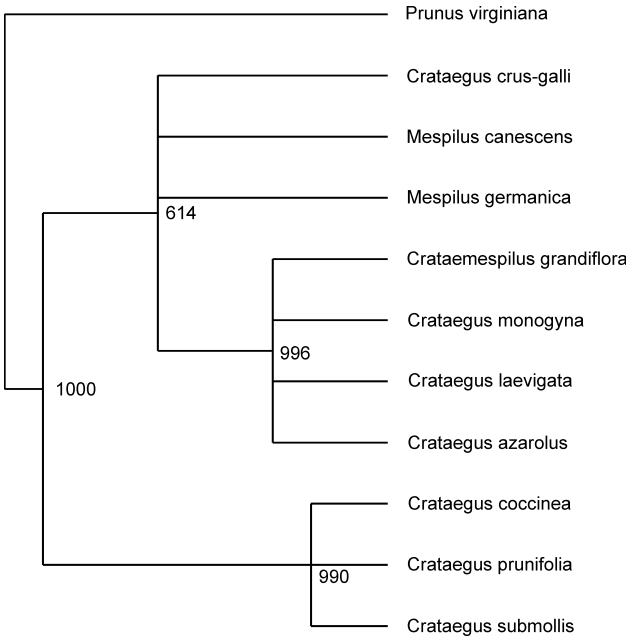

Fig. 2. Consensus of 1002 equally parsimonious trees resulting from maximum parsimony analysis using data from the partial *trnL* gene and inter-generic spacer between *trnL-trnF* (840 aligned nucleotide positions). Bootstrap values were generated by 1000 replicate re-samplings. Bootstrap values are given at branch nodes

rid origin. However, the final conclusion about the *M. canescens* origin could not be drawn from this study. To confirm or deny the hypothesis, more extensive genetic studies based not only on plastid DNA but also on nuclear DNA sequences are needed.

As a concluding remark it should be mentioned that various species sequence information employed in this study was useful in determining phylogenetic relationships within the Maloideae subfamily. Such information will be also of great use when there will be more information on different sequences available in public databases.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Allen Combes from Hilliers arboretum for providing the plant material and Fay Hughes for laboratory assistance.

Received 1 December 2005
Accepted 27 February 2006

References

1. Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA. Flora Europaea Cambridge University Press, Cambridge. 1964.
2. Morgan DR, Soltis DE, Robertson KR. Amer J Bot 1994; 81: 890–903.
3. Judd SW, Campbell CS, Kellogg EA, Stevens PF. Plant Systematics: a Phylogenetic Approach Sunderland, Massachusetts, USA. 1999.
4. Dickinson TA, Evans RC, Campbell CS. Amer J Bot 2000; 87(6, Suppl): 122–3.
5. Eriksson T, Hibbs MS, Yoder AD, Delwiche CF, Donoghue MJ. International J Plant Sci 2003; 164: 197–211.
6. Bortiri E, Oh SH, Jiang J, Baggett S, Granger A, Weeks C, Buckingham M, Potter D, Parfitt DE. Systematic Botany 2001; 26(4): 797–807.
7. Shaw J, Small RL. Amer J Bot 2004; 91(6): 985–96.
8. Oh SH, Potter D. Molecular Phylogenetics and Evolution 2003; 29(2): 203–15.
9. Doyle JJ, Doyle JL. Phytochem Bull 1987; 19: 11–5.
10. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. Amer J Bot 2005; 92(1): 142–66.
11. Chiang TY, Schaal BA, Peng CI. Bot Bull Acad Sin 1998; 39: 245–50.
12. Bell CD, Edwards EJ, Kim ST, Donoghue MJ. Harvard Papers in Botany 2001; 6(2): 481–99.
13. Navarro E, Bousquet J, Moiroud A, Munive A, Piou D, Normand P. Plant and Soil 2003; 254: 207–17.
14. Petit RJ, Demesure B, Dumolin-Lapegue S. In: Molecular Tools for Screening Biodiversity: Plants and Animals. Ed. Karp A, Isaac PG, Ingram D. Chapman & Hall, London. 1998.
15. Taberlet P, Gielly L, Pautou G, Bouvet J. Plant Mol Biol 1991; 17: 1105–9.
16. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. Nucleic Acids Res 1997; 24: 4876–82.
17. Felsenstein J, Churchill G. A. Mol Biol Evol 1996; 13: 93–104.
18. Eck RV, Dayhoff MO. National Biomedical Research Foundation 1966; Silver Spring, Maryland.
19. Kluge AG, Farris JS. Systematic Zoology 1969; 18: 1–32.
20. Simmons MP, Ochoterena H. Systematic Biology 2000; 49: 369–81.
21. Zharkikh A, Li W-H. Mol Biol Evolut 1992; 9: 1119–47.

R. Verbylaitė, B. Ford-Lloyd, J. Newbury

SUMEDĖJUSIŲ MALOIDEAE (ROSACEAE) POŠEIMIO RŪŠIŲ FILOGENETINIŲ RYŠIŲ TYRIMAI PAGAL CHLOROPLASTŲ *TRNL-TRNF* SEKŲ DUOMENIS

Santrauka

□io tyrimo metu buvo nustatytas tinkamiausias DNR išskyrimo metodas Maloideae pošeimio rūšims, taip pat išsiaiškinta, kad mažiausio galimo pokyčių skaičiavimo (angl. *maximum parsimony*) metodas yra tinkamausias filogenetinių ryšių analizei.

Tyrimo metu buvo patvirtinta monofiletinė Maloideae pošeimio kilmė, įskaitant *Vauquelinia* ir *Kageneckia* gentis, taip pat nustatyta, kad *Crataegus* bei *Mespilus* gentys yra artimai giminingos. Analizuojant *trnL-trnF* chloroplasto regiono sekas nenustatyta Maloideae pošeimio rūšių vidurūšinių skirtumų. Šio tyrimo duomenys neleidžia galutinai patvirtinti ar atmesti hibridinę *Mespilus canescens* kilmę.