Biodiv. Res. Conserv. 46: 1-18, 2017

BRC
www.brc.amu.edu.pl

DOI 10.1515/biorc-2017-0007

Submitted 22.03.2017, Accepted 30.06.2017

Phylogenetic analysis of selected representatives of the genus *Erica* based on the genes encoding the DNA-dependent RNA polymerase I

Angelika Maria Gomolińska[1*], Monika Szczecińska[1], Jakub Sawicki[1,2], Katarzyna Krawczyk[1] & Piotr Szkudlarz[3]

[1]Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn, Plac Łódzki 1, 10-727 Olsztyn, Poland
[2]Department of Biology and Ecology, University of Ostrava, Chittusiho 10, Ostrava, Czech Republic
[3]Department of Plant Taxonomy, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland
*corresponding author (e-mail: angelika.gomolinska@uwm.edu.pl)

Abstract. The *rpo* genes are characterized by rapidly-evolving sequences. They encode subunits of plastid-encoded (PEP) polymerase (*rpo*A, *rpo*B, *rpo*C1 and *rpo*C2). This polymerase is one of the most important enzymes in the chloroplasts. The primary aim of the research was to study the rate of molecular evolution in the *rpo* genes and to estimate these genes as phylogenetic markers based on the example of the genus *Erica* (Ericaceae). The tested *rpo* genes demonstrated similarities on multiple levels, for example: phylogenetic informativeness, variation level, intragenic mutation rates and the effect of intragenic mutations on the properties of encoded peptides. This study did not confirm that the analyzed *rpo* genes are reliable markers and may be helpful in understanding phylogenetic relationships between species that belong to the same genus. The *rpo*C2 gene was found to be a most useful phylogenetic marker in the *Erica* genus, while *rpo*C1 was found to be the least promising gene.

Key words: Ericaceae, molecular evolution, phylogenetic informativeness, phylogenetic signal, *rpo*

1. Introduction

Chloroplasts have a unique hybrid transcription system composed of the remaining prokaryotic components, such as prokaryotic RNA polymerase, and nucleus-encoded eukaryotic components. Genes regulating gene expression in chloroplasts are encoded in the nucleus. Chloroplast DNA (CpDNA) in terrestrial plants consists of nearly 100 genes, the majority of which reside in transcription units derived from bacterial operons (Chotewutmontri *et al.* 2016). RNAP – a bacterial multi-subunit – is composed of the core *rpo* complex that recognizes promoter sequences. The bacterial-type RNAP, called plastid-encoded plastid RNAP (PEP), which occurs in chloroplasts, shows a functional similarity to bacterial RNAP (Korczak *et al.* 2004; Yagi *et al.* 2014). The transcription of plastid-encoded genes in plants is arbitrated by two RNA polymerases of different phylogenetic origin: PEP of endosymbiotic origin (a plastid-encoded RNA polymerase – encoded by plastid *rpo* genes) and a nuclear-encoded RNA polymerase – NEP (of nuclear origin). PEP and NEP transcribe many of the same genes using different promoters. Numerous plastid genes and operons have at least one promoter each for PEP and NEP. The subgroup of plastid genes, including photosystem I and II genes, is transcribed from PEP promoters only, while in some plant species, genes are transcribed entirely by NEP (Maddison *et al.* 2006; Blazier *et al.* 2016). By contrast, all genes for chloroplast signal factors were transferred to the nuclear genome, while genes for core subunits are typically retained in the chloroplast genome as *rpo*A, *rpo*B, *rpo*C1 and *rpo*C2. In the present study, the group of *rpo* genes (*rpo*A, *rpo*B, *rpo*C1 and *rpo*C2) was analysed. These genes encode subunits of plastid-encoded polymerase (PEP) and are relatively

©Adam Mickiewicz University in Poznań (Poland), Department of Plant Taxonomy. All rights reserved.

fast-evolving sequences compared to other genes in the genome of plants (Serino *et al.* 1998; Nei & Kumar 2000; Krawczyk & Sawicki 2013). PEP contains four subunits: α – encoded by *rpo*A, and the three largest PEP subunits – β, β', and β'', which are encoded by *rpo*B, *rpo*C1, and *rpo*C2 plastid genes, respectively. The *rpo* genes were used as markers of comparatively fast-evolving sequences of PEP polymerase in several phylogenetic studies (Allison *et al.* 1996; Petersen & Seberg 1997; Korczak *et al.* 2004; Chelo *et al.* 2007; Logacheva *et al.* 2007). These studies indicated the suitability of the studied *rpo* genes for analyses at a high taxonomic rank (such as the division level) and also at the genus level. Single genes of the *rpo* group were used to study plant evolutionary history at the family and genus levels (Little & Hallick 1988; Krause *et al.* 1998; Logacheva *et al.* 2007; Krawczyk & Sawicki 2013) and in bacteria – at the genus level (Korczak *et al.* 2004; Chelo *et al.* 2007).

This study assessed the analyzed genes as phylogenetically-useful markers and analyzed the impact of substitutions that trigger changes in amino acids located in the encoded proteins. The subject of the study were representatives of the family Ericaceae. This family comprise about 4,100 species, many of them endemic and threatened, belonging to 126 genera (Schumann *et al.* 1992; Oliver 1993; Ojeda *et al.* 2015). Ericaceae includes shrubs and climbers that occupy acidic, nutrient-poor habitats, often containing high levels of heavy metals. In response to such conditions, Ericaceae species form symbiosis with mycorrhizal fungi. This type of symbiosis is called ericoid mycorrhiza and ensures successful colonization of acid and poor soils (Bruzone *et al.* 2015). *Ericaceae* range all over the world, except central Greenland, parts of the high Arctic, continental Antarctica, and northern and central Australia. *Erica* is one of the two largest genera in the family, next to *Rhododendron*, and is widespread in Asia (Glen 2002; Linder 2003; Stevens *et al.* 2004; McGuire & Kron 2005). It comprises about 860 species of flowering plants. *Erica* species are generally low shrubs (20-140 cm of high), although sometimes higher – like *E. arborea* that reaches up to 7 m in height. All *Erica* species are evergreen, with needle-like leaves. The region most diverse in the *Erica* species is the Cape Floristic Region (CFR). With more than 700 species and a high level of endemism, *Erica* represents the epitome of plant biodiversity. This genus shows a pattern of high species concentration in fynbos; a fire-prone, heathland-like vegetation and includes re-sprouter and seeder species, like most fynbos woody taxa (Linder 2003; Gillespie & Kron 2010; Yoichi & Tomaru 2014; Ojeda *et al.* 2015). *Erica* is taxonomically well-documented, but still little is known about species-level relationships. The systematic position of *Erica* has been addressed in various studies of phylogenetic relationships within Ericaceae (Gillespie & Kron 2010; Pirie *et al.* 2011; Yoichi & Tomaru 2014). Phylogenetic relationships among species of the genus *Erica* were examined to find whether this genus originates from Africa or Europe. The first phylogenetic study of *Erica* was based on 3 regions of DNA, 2 sequences encoded by chloroplasts *atp*B-*rbc*L spacer and partial *mat*K, and 1 from nuclear genome (internal transcribed spacer region of nuclear ribosomal DNA, ITS) (McGuire & Kron 2005). The second major phylogenetic analysis of *Erica*, using only ITS sequences, included as many as 379 species. It demonstrably shows relatively low variation in nrITS sequences in the broad sample of ca. 45% species assigned to *Erica*. In both cases, the conducted analyses did not support the subgeneric division of the genus *Erica* (Pirie *et al.* 2011). The main goal of this study was to determine the usefulness of tested genes encoding a DNA-dependent RNA polymerase I in phylogenetic analyses of selected representatives of the genus *Erica*. Thus far, no studies have been published on chloroplast genes, particularly the *rpo* genes, in this genus.

2. Material and methods

2.1. Plant material

The study material included 50 *Erica* species that represented the following 5 subgenera: *Syringodea* (18 species), *Euerica* (17), *Chlamydanthe* (7), *Plastoma* (4) and *Stellanthe* (1). Different subgenera were represented by different numbers of species, because the number of representatives was matched to the species richness within the studied subgenera. All examined species are presented in Table 1. The classification of the genus is used according to the division proposed by Bentham (1839), modified by Hansen (1950) and supplemented by Oliver (2000). The nomenclature was adopted from Oliver (Oliver 1993; Oliver *et al.* 2003). Not all tested species provided the final results. For 9 tested species, the results of sequencing were not obtained. Thus the presented data relate to 41 species.

Vaccinum macrocarpon, *Chamaedaphne calyculata* and *Arbutus unedo* were incorporated into the study as an outgroup. Data for those three species were used as an outgroup, because they are closely related to the *Erica* genus and belong to one family Ericaceae. For those species, the entire chloroplast genome sequences were developed and they are available in the NCBI database. The samples were collected by the authors or other scientists during field research. Specimen vouchers are deposited in the herbarium POZ (Department of Plant Taxonomy, Adam Mickiewicz University in Poznań). The origin of analyzed species and collectors of samples are presented in Table 1.

2.2. Laboratory procedures and sequence alignment

Leaves of representatives of the genus *Erica* were grated in a Mini-Beadbeater-1 tissue disruptor. Isolation of the DNA material was performed using a Zymo Research kit (isolation kit ZR Plant/Seed DNA Kit). Isolated DNA was then used in the subsequent analyses. The DNA fragments of *rpo*A, *rpo*C1 and *rpo*C2 genes

Table 1. The list of tested species within the subgenera and sections of the genus *Erica* with localities

Genus	Subgenus	Section	Species	Locality / collectors
Erica	*Syringodea*	*Gigandra*	*Erica coccinea*	RSA, Cape Aguhlas / SWO
			Erica plukenetii (4 samples)	Bottelary/SWO
				RSA, Cape Aguhlas /SWO
				RSA, Groot Gabel Kraal /SWO
				RSA, near Caledon /SWO
		Pleurocallis	*Erica bauerii*	RSA, Albertinia, W of town /SWO
			Erica hibbertii	Theevaterkloof /SWO
			Erica mammosa	RSA, Kleinmond /SWO
			Erica regia	RSA, Cocsravier near Viljoenshof /SWO
			Erica sessilliflora	RSA, Pilarkop, subalpine heathland /SWO
			Erica vestita	RSA, Jonaskop, lower slopes /SWO
		Evanthe	*Erica cruenta*	RSA, near Caledon /SWO
			Erica curviflora	RSA, Pilarkop, subalpine heathland /SWO
			Erica discolor	RSA, Road to Kreisvalle /SWO
			Erica patersonii	RSA, Betty's Bay /SWO
			Erica versicolor (2 samples)	Bergfontein /SWO
				RSA, Langeberg, Tradous Pass /SWO
		Dasyanthes	*Erica strigilifolia*	RSA, Swartberg Pass /SWO
	Stellanthe	*Euryloma*	*Erica ampullaceae*	RSA, Boskloof /SWO
	Euerica	*Ephebus*	*Erica caffra*	RSA, Swartberg Pass /SWO
			Erica hirtiflora	RSA, Kirstenbosch, slope of the Tablle Montain /SWO
		Gypsocallis	*Erica nudiflora*	Bainskloof/Witelsriver/SWO
			Erica vagans	Spain, Navarra /S
			Eirca multiflora	Spain, Estramadura /S
			Erica globiceps (2 samples)	RSA, Boskloof /SWO
				RSA, Groot Hagelkraal /SWO
		Pachysa	*Erica umbelliflora*	Baviaansmountain /SWO
		Chlorocodon	*Erica coarctata*	Arieskraal, Palmiet River /SWO
			Erica scoparia	Spain, Cap de Creus /S
		Arsace	*Erica hispidula*	Kouga /SWO
			Erica zwartbergensis	RSA, Swartberg Pass /SWO
			Erica arborea (2 samples)	Spain, Montradue /S
				Spain, Cap de Creus /S
		Eremocallis	*Erica tetralix*	Poland, Piaseczno near Węgliniec /S
		Brachycallis	*Erica cinerea* (2 samples)	Spain, Navarra /S
				Scotland, Glen Cloe /S
	Chlamydanthe	*Geissostegia*	*Erica imbricata* (2 samples)	RSA, Elandskloof, Wolseley /SWO
				RSA, Boskloof /SWO
		Eurystegia	*Erica monsoniana*	Cedarberg /SWO
		Adelopetalum	*Erica nabea*	Prince Alfreds Pass /SWO
		Trigemma	*Erica baccans*	Ex. Hort Hermanus, Woodvine s.n. & Hout Bay /SWO
			Erica plumigera	RSA, Kugel Bay, slope of a montain /SWO
			Erica selaginifolia	Touwsberg /SWO
	Platystoma	*Eurystoma*	*Erica floccifera*	Greyton /SWO
			Erica pseudocalycina	Langeberg /SWO
		Gamochlamys	*Erica melanthera*	Tradouw Pass /SWO
			Erica newdigateae	Smutskop /SWO

Explanations: sample collectors, S – Szkudlarz P., W – Wiland-Szymańska J., O – Oliver E. G. H.

were amplified in a volume of 20 µL containing 10 mM Tris-HCl (pH 8.3 at 25°C), 50 mM KCl, 1,5 mM MgCl2, 1 µL BSA, 200 µM each dATP, dGTP, dCTP and dTTP, 1 µM of each primer, one unit (U) of Taq polymerase OpenEx (OpenExome) and 10-20 ng of the DNA template. The PCR reactions for *rpoA*, *rpoC1* and *rpoC2* genes were carried out under the following conditions: (1) initial denaturation 4 min at 94°C; (2) denaturation 45 s at 94°C; (3) annealing 1 min at 51°C; (4) elongation 1.5 min at 72°C; (5) final elongation 7 min at 72°C. Stages 2-4 were repeated 40 times. Above conditions of PCR reactions gave weak amplification for *rpo*B gene, so the conditions and buffer were changed. The DNA fragments of *rpoB* genes were amplified in a volume of 20 µL, containing 10 mM KCl, 10 mM $(NH_4)SO_4$, 20 mM Tris-HCl (pH 8.5 at 25°C), 2 mM $MgSO_4$, 1 µL BSA, 200µM each dATP, dGTP, dCTP and dTTP, 1 µM of each primer, one unit (U) of Taq polymerase (Run DNA Polymerase A&A Biotechnology) and 10-20 ng of the DNA template. The PCR reactions for the *rpoB* genes were performed under the following conditions: (1) initial denaturation 4 min at 94°C; (2) denaturation 35 s at 94°C; (3) annealing 1 min at 48°C; (4) elongation 1 min at 72°C; (5) final elongation 7 min at 72°C. Stages 2-4 were repeated 35 times. Finally, the products of amplification were visualized on 2% agarose gel with GelView (Invitrogen™, Carlsbad, CA, USA) staining. The cleaned, pure PCR products were sequenced in both directions using an ABI BigDye 3.1 Terminator Cycle Kit (Applied Biosystems®, Foster City, CA, USA) and the same primers were used for sequencing. For the procedures of amplification and sequencing of the *rpo*C1 genes, the primers from the website of the Royal Botanical Garden in Kew were used. Primers for the *rpo*A, *rpo*B and *rpo*C2 genes were designed using the chloroplast genome of *Chamaedaphne calyculata* (Szczecińska *et al.* 2014), annotated as KJ463365.1 in the GenBank database (Benson *et al.* 2005). All sequences of the used primers are presented in Table 2.

The program Sequencher 4.1.4 (Gene Codes Corporation, Ann Arbor, MI, USA) was used for editing and assembling electropherograms. The program BioEdit 7 was used for manual adjustment and the program MUSCLE – for sequence alignment of the assembled sequences (Hall 1999; Edgar 2004). The conducted analyses enabled to remove the regions of ambiguous alignment and deficient or incomplete data (i.e., at the end and the beginning of these sequences). The obtained seqences of the *rpo* genes (*rpo*A, *rpo*B, *rpo*C1 and *rpo*C2) are deposited in the public repository – GenBank database – as KY523675-KY523838.

2.3. Distribution of substitutions and phylogenetic informativeness (PI)

The HyPhy program was used to estimate the distribution of substitution rates across the sites by applying a rate class to each site based on a general model characterized by reversibility (Pond & Muse 2005). The DAMBE software was used to evaluate the level of nucleotide substitution saturations by plotting transitions and transversions against pairwise genetic distance (Xia *et al.* 2001; Xia 2013). The phylogenetic informativeness (PI) of *rpo* genes was estimated in the PhyDesign server and the phylogenetic informativeness profiles were plotted with reference to an uncalibrated tree (Lopez-Giraldez *et al.* 2011). The PATHd8 was used for an ultrametrisized tree used to overlay the historic changes in substitution rates, which were obtained with ML (Britton *et al.* 2007). The root of the tree was set at an evolutionary time of 1 and tips at the time of 0 for obtaining the relative ages for the clades. PI calculations of nucleotide data sets were also estimated using the HyPhy program, employing empirical base frequencies and a time-reversible model of substitution (Pond *et al.* 2005). The Rate4Site program was used for the amino acid (AA) data sets. In this program, the JTT model of evolution and ML inference method were used (Mayrose *et al.* 2004). Informativeness was calculated to exclude the influence of gene length on the outcome per site.

2.4. Phylogenetic relationships

Phylogenetic relationships were inferred by applying three methods: Bayesian inference (BI), maximum likelihood (ML) and maximum parsimony (MP). The MrBayes v. 3.2.2 program was used for Bayesian inference phylogenetic analyses with the priors set according to the output of jModelTest 2.1.7 (Hulsenbeck & Ronquist 2001; Posada 2008; Darriba *et al.* 2012). The Bayesian Information Criterion (BIC) results were used as the basis to select optimal models of nucleotide

Table 2. The forward and reverse primer's sequences of DNA applied in the testing procedures

Region	Forward primer's sequence (5'-3')	Reverse primer's sequence (5'-3')
*rpo*A	ATGAGTCGAGAAAACGGAAG	TTGTCGATTCATTTCATTGG
*rpo*B	AAGTGCATTGTTGGAACTGG	CCGTATGTGAAAAGAAGTATA
*rpo*C1	TATGAAACCAGAATGGATGG	GAAAACATAAGTARRCGWGC
*rpo*C2	AAAGAGATACGGCTTTGTACT	GGAATGCCAATCTTTCCTAAT

substitution for *rpo* sequences. The data partitioned for analyses and models for each gene were the same. The likelihood model parameters applied for the region were suitable for a general time-reversible model with a gamma-shaped distribution of rates across sites (TRN2). The estimation of the Bayesian inference was performed by running six incrementally-heated chains (MCMC algorithm) for 1,500,000 generations, sampling one out of every 100 generations of random trees. A test run was done to determine the number of generations that should be eliminated from consensus tree calculations. In the study, the first 30,000 generations were rejected as "burn-in". The Bayesian consensus tree was constructed using the remaining generations. FigTree v1.4.2 software was used to graphically adjust the generated phylogenetic trees.

MEGA v.6.06 software was used for conducting the ML and MP analyses (Tamura *et al.* 2013). Advanced software RAxML was used to produce the ML tree. The T92+G model of nucleotide substitution was used as a basis for ML analyses. The tree inference was performed with the Close Neighbor Interchange (CNI in maximum parsimony) in the MP method, the algorithm at the search level of 3 and the number of initial trees equal to 10. The phylogeny was investigated both for the maximum likelihood and maximum parsimony using the bootstrap method with the number of bootstrap replications at the level of 1000 (Felsenstein 1985; Christelová *et al.* 2011).

2.5. Positive selection pressure

TreeSAAP v.3.2 software was used to study the occurrence of positive selection pressure at the protein level (Woolley *et al.* 2003). This program calculated the selective impact on structural and biochemical amino acid properties (31 properties) across a phylogenetic tree. For the classification of each property change, the gradient of eight categories was used, with lower categories denoting more conservative changes and higher categories indicating more radical changes. Next, a Z-test was performed to determine selection on the detected amino-acid properties, when the frequency of radical changes of 6, 7 or 8 category, overstep the frequency prospective by chance, as indicated by positive z-scores. In the TreeSAAP software, each of the analyzed *rpo* genes was inspected using both the whole data set and a sliding window analysis (the width of reading frame equal to 15 codons). The results obtained in the TreeSAAP software allowed to detect positive destabilizing effect for each property contained in the particular amino acid residues.

3. Results

3.1. Gene sequences

Rpo gene fragments ranging from 633 bp (*rpo*C2) to 743 bp (*rpo*A) in length were used in this study (Table 3). The variability of *rpo* genes was caused entirely by substitutions, no indels were identified. The analyzed sequences belonged to 41 different species.

A slightly higher content of purines than pyrimidines and a more numerous amount of AT than GC bonds were present in the *rpo*A, *rpo*B and *rpo*C1. A slightly higher content of pyrimidines over purines and a more numerous amount of AT than GC bonds were observed in the *rpo*C2 gene. In the group of tested genes, the highest level of variation and informative sites was noted in the *rpo*C2 gene (13.9%, 5.68%). The *rpo*A gene has a substitution rate of 0.50 and it was characterized by slightly lower variability; the contribution of variable sites (*V*) reached 9.69% and the parsimony-informative sites (*Pi*) reached 3.49%. The lowest average rate of its evolution was noted in the *rpoC1* gene and it was 0.12 substitutions per nucleotide. The *rpo*C1 had the lowest level of variable sites (6%) and the lowest number of parsimony-informative sites (1.71%) (Table 3).

In this study, the slighter average rate of mutation in the *rpo*B and *rpo*C1 genes in the *Erica* genus corresponded to a higher variability of this rate within

Table 3. Comparison of the analyzed *rpo* gene fragments

	*rpo*A	*rpo*B	*rpo*C1	*rpo*C2
Alignment length	743	654	699	633
Sequence length	742-743	654	699	633
Site rates	0.508	0.348	0.125	0.518
Variable characters (*V*)	72 (9.69%)	56 (8.56%)	42 (6%)	88 (13.9%)
Parsimony – informative sites (*Pi*)	26 (3.49%)	18 (2.75%)	12 (1.71%)	36 (5.68%)
Purines (%)	53.761	54.451	51.876	48.04
Pyrimidines (%)	46.239	45.549	48.124	51.96
GC (%)	32.95	39.254	41.345	32.262
Transition/Transversion bias (*R*)	0.57	1.09	0.68	0.8
Number of the synonymous substitutions	18	28	9	34
Number of the non-synonymous substitutions	85	59	35	69

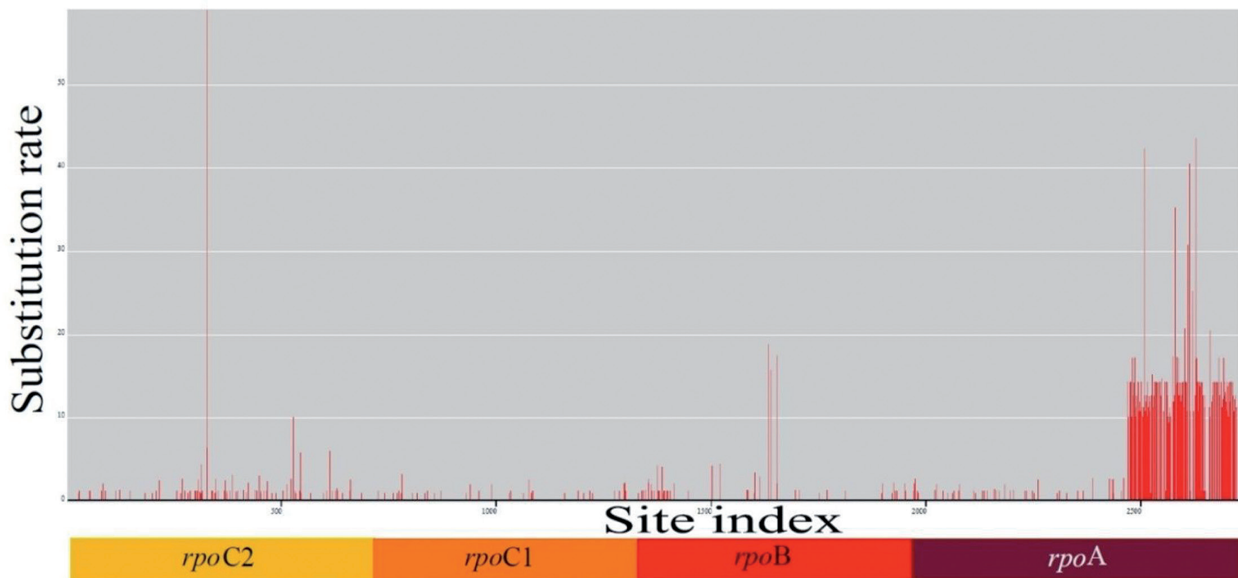

Fig. 1. Result of the distribution of substitution analyzes shown in HyPhy software computed with the TRN model of evolution. A colored slice below the plot represent the proportional position of sites in the analyzed chloroplast gene fragments (yellow – *rpoC2*, orange – *rpoC1*, red – *rpoB*, purple – *rpoA*)

these genes. The *rpo*B and *rpo*C1 genes contain large conserved regions and a few extremely variable sites. A high density of variable sites occurs in the region of the *rpo*A gene (Fig. 1). Site rate variation has an important impact on phylogenetic analysis, especially sequence evolution. A strong variability indicator in *rpo*A shows that some of the substitutions occurred at fast-changing sites.

3.2. Phylogenetic informativeness and the phylogenetic tree

Stating the number of transitions and transversions with the genetic distance for each pair of sequences allowed the state of substitution saturation in examined *rpo* genes to be calculated (Fig. 2). The results received for the *rpo* genes show that the number of substitutions

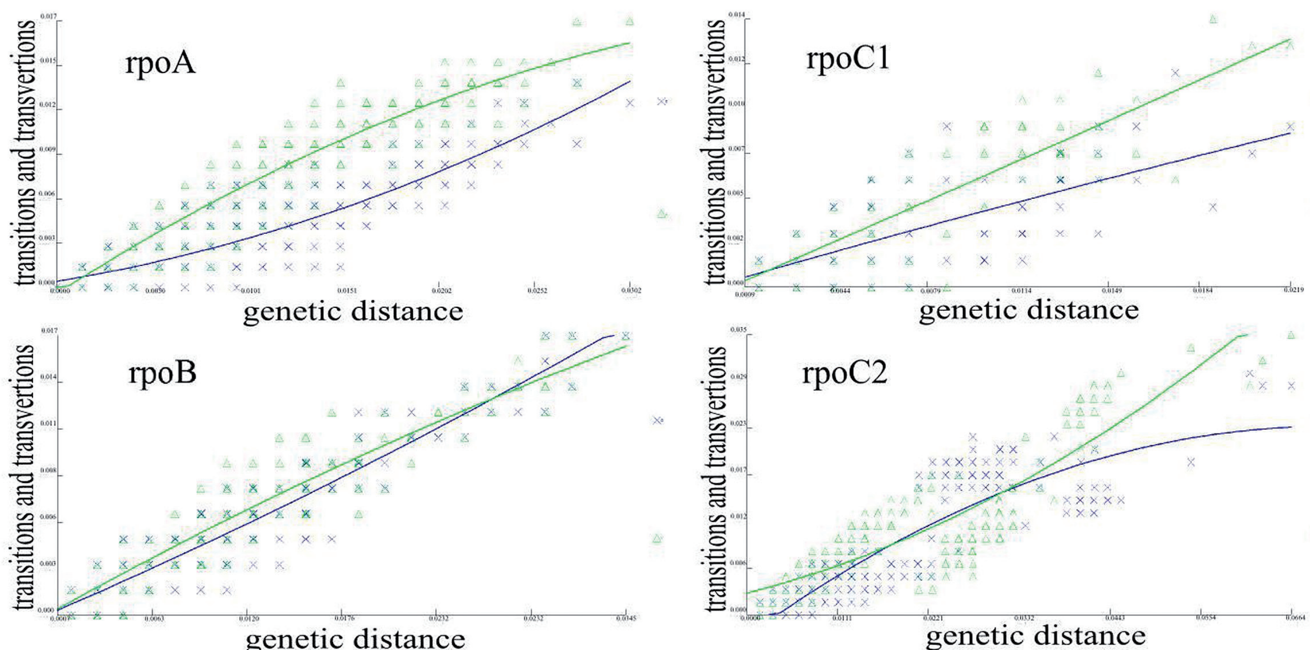

Fig. 2. Transitions and transversions vs divergence plots (Xs – transitions, Δv – transversions; lines show the main trend). The estimated number of transitions and transversions for each pairwise comparison was plotted against the genetic distance

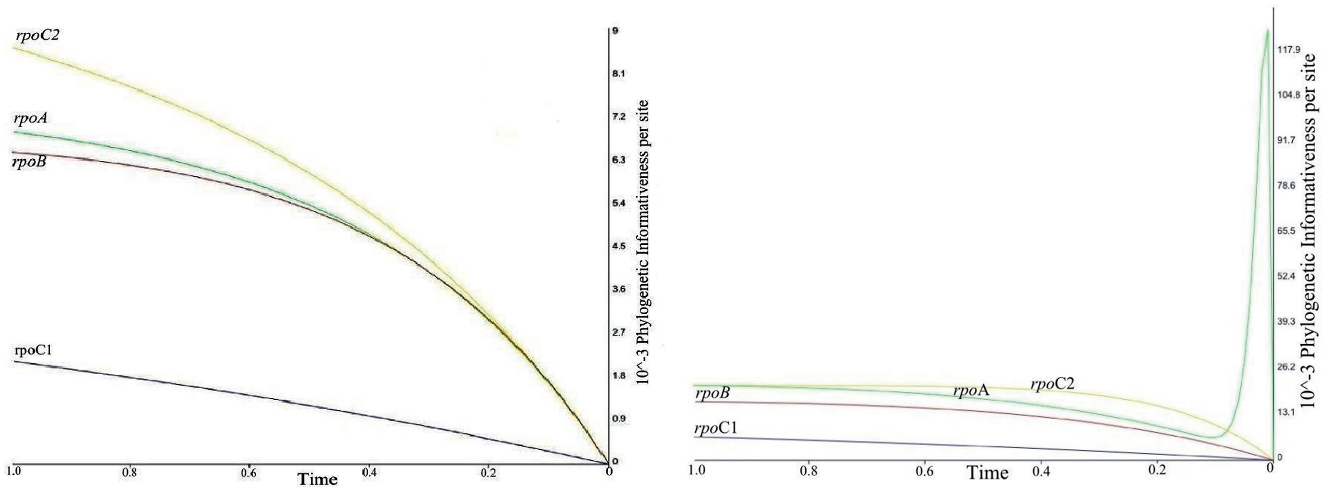

Fig. 3. The charts obtained by analysis using PhyDesign software correspond to phylogenetic informativeness per site for the studied *rpo* genes measured from the nucleotide (A) and amino acid sequences (B) – the ultrametrisized ML tree derived from the dataset of all tested *rpo* genes was used for the calibration of the time axis

progressed with the increase in genetic distance. A lower increasing tendency was only noted in the number of transitions in the *rpo*C2 gene. For the *rpo*A and *rpo*C1 genes, a distinct advantage of transversions over transitions was noted.

The study also shows that the difference in the load of information carried by two types of data depends on the comparison of the PI profile based on nucleotide sequences with the profile obtained for the amino acid sequences. The PI analysis revealed that among the tested *rpo* genes, the *rpo*C2 gene shows a higher usefulness as a phylogenetic marker for the genus *Erica*. In the *rpo*C2 gene, the high informativeness of the nucleotide sequence (Fig. 3A) was correlated with a strong phy-

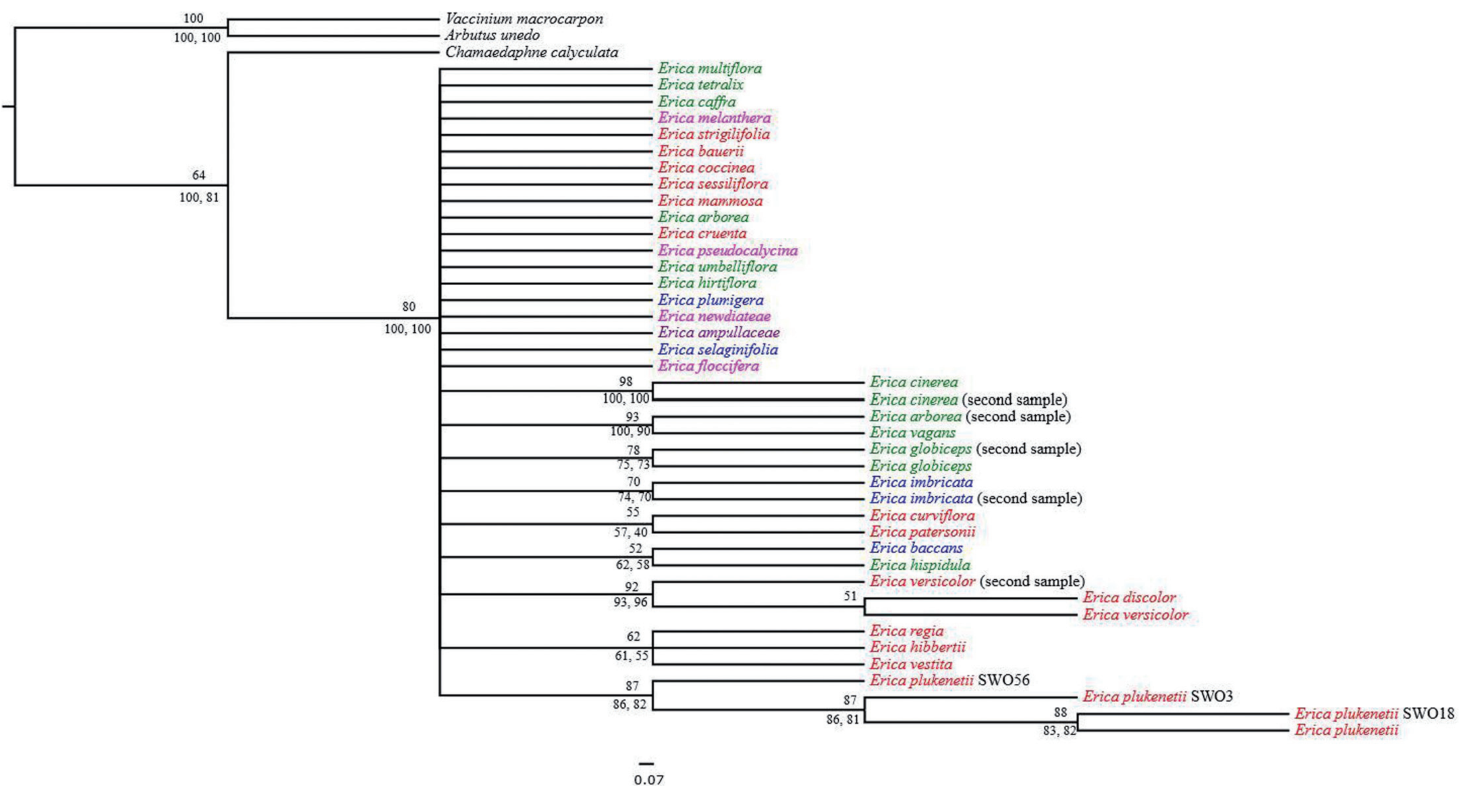

Fig. 4. A consensus phylogenetic tree obtained by Bayesian analysis of the *rpo* genes. Colors of the species names denote subgeneric affiliation: red - *Syringodea*, purple – *Stellanthe*, green – *Euerica*, blue – *Chlamydanthe*, fuchsia – *Platystoma*. Names in black denote the outgroup. Bootstrap values lower than 50% are hidden. Numbers under the branches represent bootstrap support obtained by maximum likelihood and maximum parsimony methods. The MP value is placed before the comma and ML value – after the comma

logenetic signal at the amino acid level (Fig. 3B). This shows a high convergence of protein sequence evolution with the phylogenetic tree topology. The level of informativeness at the nucleotide and amino acid levels was similar.

The phylogenetic tree received with the MP method determined a consensus of the 2,727 most parsimonious trees with a length of 382 steps each (CI = 0.743455 and RI = 0.617188). The topology of the phylogenetic tree is completely compatible with the Bayesian inference results. The ML tree obtained from analyses of concatenated nucleotide sequences of *rpo* genes has similar topology to the MP and Bayesian trees. The compiled results of the phylogenetic analyses are presented in Fig.4. Phylogenetic analysis based on the *rpo* gene sequence variation does not show the whole picture of the genus *Erica* systematics. Although it shows valuable information about the phylogenetic relationships between the examined *Erica* species, there are several exemptions from the taxonomic division introduced by Oliver (Oliver 1993; Oliver *et al.* 2003).

3.4. Positive Selective Pressure

The examined genes showed differences in the content of silent mutations and non-synonymous mutations. For the *rpo* genes, the predominance of variable sites containing non-synonymous substitutions (Table 3) were noted. All genes had fewer synonymous sites. The highest number and highest percentage of non-synonymous sites in relation to all variable sites was found in the *rpo*C2 and *rpo*A genes, where the number of variable sites was equal. The *rpo*C2 had 69 non-synonymous substitutions, while *rpo*A had 85 non-synonymous substitutions.

Some of the observed non-synonymous mutations resulted in important changes in physicochemical properties of amino acids.

4. Discussion

There are insufficient data in the available literature to compare the variation of *rpo* genes in *Erica* species with the variation in other genera of vascular plants. Thus, our results based on the analyses of *rpo* gene fragments, can be related to studies of the *rpo* gene variability in *Lamium* species. The data published by Krawczyk and Sawicki (2013), indicate that the level of variable sites observed in the analyzed genes in the genus *Erica* was higher than in the genus *Lamium*. The variability of *rpo*A, *rpo*B, *rpo*C1 and *rpo*C2 genes observed in the genus *Erica* (Table 3) was visibly higher than in the genus *Lamium*, where the degree of variability of these genes amounted to V=5.91%, V=2.52%, V=2.85% and V=5.03%, respectively. Similarly, the comparison of results of the *rpo* gene variability observed in the tribe Triticeae and family Myristicaceae and in the genus *Erica* showed that variability in the two former taxa is lower than in the genus analyzed in this work, although, it was parallel to other phylogenetic markers, which indicates that it is still comparatively high (Petersen & Seberg 1997; Newmaster *et al.* 2008).

The inherence of highly variable sites increases the likelihood of reverse mutation or homoplasy, which hinders a faultless phylogenetic relationship reconstruction (Little & Hallick 1988). Our study supports the results of research conducted at higher taxonomic levels, as in the division Magnoliophyta. In this division, the *rpo*A and *rpo*C2 genes proved to be better phylogenetic markers than the *rpo*B and *rpo*C1 genes (Logacheva *et al.* 2007). Similar conclusions can be drawn from the studies of the genus *Lamium* (Krawczyk & Sawicki 2013). The diversity profile of *rpo*A and *rpo*C2 genes makes their phylogenetic signal more precise. With respect to molecular taxonomy, a steady rate of substitution is tendered in a given genome section (Yang 1996; Lió & Goldman 1998).

Among all four graphs, illustrating the state of substitution saturation in the examined *rpo* genes, none took the form of a plateau, which is typical of the state of saturation with substitutions. If the separate fragments of genome have not yet reached the level of substitution saturation, they can provide a reliable phylogenetic signal. However, determining whether the similarity in the pair of sequences is a homoplasious trait or a homologous trait derived from a common ancestor is more difficult (Jeffroy *et al.* 2006; Townsend 2007). The signal is not sufficient to resolve the phylogeny at the genus level in the genus *Erica*. The obtained results suggest that the *rpo* genes not seem to be useful in phylogenetic inference in the genus *Erica*. Finding a suitable molecular marker to resolve relationships within a genus will be very useful for ecological and evolutionary research in the genus *Erica*.

The phylogenetic informativeness analysis of individual *rpo* genes allows a congruous inference to be drawn. The obtained profiles of PI indicated that the tested genes reached the apex of informativeness for the events evolutionarily older than the speciation within the genus *Erica*. This indicates that the information noise that arose from the accumulation of mutations carrying non-phylogenetic signal was diminished (Townsend *et al.* 2008; Townsend & Leuenberger 2011).

The PI analysis of the PI profile based on nucleotide sequences with the profile based on the amino acid sequences, revealed that among the tested *rpo* genes, the *rpo*C2 gene shows a higher usefulness as a phylogenetic marker for the genus *Erica*. The level of informativeness at the nucleotide and amino acid levels was suchlike. The interpretation of both graphs shown in Fig. 3 has to be cautious, because the PI values are presented per

site. If we want to confront the PI values for nucleotide and amino acid types of data, the scores demonstrated on the Y-axis of the graph in Fig. 3B should be divided by three, because the amino acid sequences are three times shorter than the nucleotide sequences. The results obtained based on this principle prove that the phylogenetic signal carried by examined *rpo* genes was quite similar at the amino acid and nucleotide levels. The spike in phylogenetic informativeness in the *rpo*A gene, inferred from amino acid sequences for recent divergences, might coincide with the high density of variable sites occurring in the region of the *rpoA* gene received in HyPhy. This spike probably results from the maximum likelihood estimate for the rate of several sites and has its peak at infinity. The result of this type of analysis was highly reliable, even though the amount of information was lower than in the case of nucleotides – as in the case of the genus *Lamium* analysis. In these studies, the most useful phylogenetic marker proved to be *rpo*A, in which high informativeness at the nucleotide level corresponds to a strong signal at the amino acid level. In other genes, there was a discrepancy recorded between informativeness at the nucleotide and amino acid levels (Krawczyk & Sawicki 2013). This may be explained by the fact that the analysis based on the amino acid sequence was less obscured by the information noise compared to the analysis involving nucleotide sequences (Townsend *et al.* 2008).

The phylogenetic tree of the genus *Erica*, calculated based on the *rpo* genes, was generally in agreement with the results based on an analysis of six, mainly non-coding, chloroplast markers and is helpful in determining the phylogenetic relationships within the studied taxon. These results are similar to those obtained in the study of the *rpo* genes in *Lamium* (Krawczyk & Sawicki 2013). This work illustrates the considerable genetic distance between the representatives of the outgroup: *Vaccinium macrocarpon, Arbutus unedo* and *Chamaedapne calyculata*. This is indicative of the early separation of this phylogenetic line. The topology of this phylogenetic tree indicates a closer relationship between *Chamaedaphne calyculata* and the *Erica* species examined, than between *Erica* and two others species from the outgroup. The representatives of all subgenera of the genus *Erica* are not grouped in one ordination and they do not form the clades that validate the division of the genus *Erica* into subgenera and sections. The subgenera *Euerica* and *Chlamydanthe* were interesting because of the relationships between individual species. *E. baccans* (subgenus *Chlamydanthe)* is phylogenetically closer to *E. hispidula* (subgenus *Euerica*) than to other species from the subgenus *Chlamydanthe*. The subgenus *Syrignodea* is also interesting because of relationships between individual sections. The noted changes in the *rpoA*, *rpoB* and *rpoC2* genes in the *Syringodea* subgenus were also noticeable at the protein level (Appendix 1). The analysis of the phylogenetic tree may indicate that the ancestor of *Erica* genus was bicontinental *Erica multiflora* from the subgenus *Euerica*. Two individuals of the same species, belonging to the subgenus *Euerica* are of interest because of the problem associated with *E. arborea*. The second specimen of *E. arborea* is phylogenetically closer to *E. vagans* than to *E. arborea*. The result obtained with the ML method also showed similar separation between the first and second specimen of *E. arborea*. A similar phenomenon was observed in the subgenus *Syringodea*, i.e., *E. versicolor* was phylogenetically a little closer to *E. discolor* than to the second specimen of *E. versicolor*. The difference occurs in the case of *E. versicolor*, which has a greater similarity to the second sample of *E. versicolor* than to *E. discolor*. This probably results from the wrong identification of *E. arborea* or, alternatively, we are witnessing the separation of a new species. This relationship was also noted at the nucleotide level and was confirmed by the presence of an adaptive mutation in the protein encoded by the *rpo*B gene, where many changes in the properties of the amino acid were detected in *E. arborea* and *E. vagans*. The phylogenetic tree reveals a long distance between the representatives of one species (as for *E. arborea*).

An analysis of the changes in amino acid properties indicated that, with a few exceptions, these changes were not reflected in the construction of the phylogenetic tree. Minor exceptions were common changes in the *rpo*C2 gene in *E. baccans* and *E. hispidula*, *E. cinerea* and the second sample of *E. cinerea*, and *E. multiflora* and *E. tetralix*. The changes in the *rpo*C1 gene occurred in *E. versicolor*, the second sample of *E. versicolor* and *E. discolor*, and in *E. imbricata* and the second sample of *E. imbricata*. The common changes in the *rpo*A gene were observed in *E. cinerea* and the second sample of *E. cinerea*, and also *E. multiflora* and *E. tetralix*. Examples of these common changes in the *rpo*B gene were not found (Appendix 1).

The strong conclusion one can make based on the presented tree is that the genus *Erica* is monophyletic. There are no clades grouping together taxa belonging to different subgenera and almost no well-supported clades at all. Some species fall into one clade, but some not (for example *E. arborea*).

Two specimens of *E. arborea* are not grouped together and one of them is sister to another species. Based on the presented tree, it is difficult to conclude about genetic relationships. Both studied samples originated from Spain, but not from the same location – the first sample was collected in Montradue and the second in Cap de Creus. This difference in origin might explain the separate grouping of *E. arborea* specimens in the phylogenetic tree. A similar situation was observed

for some species of the genus *Lamium*. The phylogenetic analysis showed that *L. album* 218 was closer to *L. tomentosum* than *L. album* 220 and *L. album* 222 (Krawczyk & Sawicki 2013).

The phylogenetic tree and juxtaposition of amino acid property changes did not allow the various representatives of *Erica* to be organized or sorted into clades. Radical amino acid property changes under the positive destabilizing selection were observed in all *rpo* genes. Specific changes in the properties of amino acids encoded by these genes often occurred in a number of codons simultaneously. Most amino acid property changes were observed in the *rpo*B gene (7 from "8" category of change, 5 from "7" and 7 from "6"), while the fewest property changes were noted in the *rpo*C1 gene (2 from "8" category of changes, 5 from "7" and 2 from "6"). Numerous property changes from the category "8" occurred in the *rpo*A gene (9 properties). In relation to the *rpo*C2 gene, most amino acid properties changes occurred in the subgenera *Chlamydanthe*, *Euerica* and *Syringodea*. There were no changes in the subgenus *Stellanthe* and only one in the subgenus *Platystoma*. For the *rpo*C1 gene, the most numerous changes occurred in the subgenus *Syringodea*. There were no changes in the subgenus *Stellanthe* and only a few changes recorded in other subgenera. For the *rpo*B gene, numerous amino acid property changes were noted in the subgenera *Syringodea* and *Euerica*, while no changes were observed in the subgenus *Stellanthe* and only a few changes in other subgenera. For the *rpo*A gene, numerous changes were noted in the subgenera *Syringodea*, *Euerica* and *Chlamydanthe*. For the subgenera *Stellanthe* and *Platystoma*, only a few individual amino acid properties changes were observed. Some changes in amino acids were specific to subgenera, but those are not sufficient to classify species to subgenera. As far as we know, the discussed *Erica* genus has never been addressed in molecular studies.

The results obtained for *Erica* show that the phylogeny based on more than 2,000 nucleotide long alignment is poorly resolved (the tree contains large polytomies and multiple clades with low support). The phylogenetic signal has not been obscured by multiple substitutions (no saturation), but there is clearly no evidence that variability of PEP sequences enables estimation of a robust phylogeny of *Erica*. The results of this study show that the *rpo* genes are not useful phylogenetic markers that can be helpful in the reconstruction of evolution of the genus *Erica*. Coding regions contained a high, but not reliable phylogenetic signal. The value of this signal may be increased by an analysis of the changes in the amino acid chain. Of the all examined *rpo* genes, the most useful phylogenetic marker was the gene *rpo*C2, while the least suitable gene turned out to be *rpo*C1. In four tested genes, the destabilizing effects of positive pressure were noted. This illustrated the relations between mutations fixed by positive selection and the separation of phylogenetic lines within the genus tested. A notable advantage is the ease of alignment of the analyzed sequences.

Acknowledgements. This work was made possible thanks to financial support from the Polish State Committee for Scientific Research (KBN grant N303 807840).

References

Allison L. A., Simon L. D. & Maliga P. 1996. Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. EMBO Journal 15: 2802-2809.

Benson D. A., Karsch-Mizrachi I., Lipman D. J., Ostell J. & Wheeler D. L. 2005. GenBank. Nucleic Acids Research. 40: 34-38.

Bentham G. 1839. Ericaceae. In A. P. De Candolle, Prodromus systematis naturalis regni vegetabilis, 7: 580-733. Treutel & Weiss, Paris.

Blazier J. C., Ruhlman T. A., Weng M. L., Rehman S. K., Sabir J. S. M. & Jansen R. K. 2016. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Scientific Reports. DOI: 10.1038/srep24595

Botanical Garden in Kew website. Available online: http://www.kew.org/barcoding/protocols/ (accessed on 21 January 2017).

Britton T., Anderson C. L., Jacquet D., Lundqvist S. & Bremer K. 2007. Estimating divergence times in large phylogenetic trees. Systematic Biology. 56: 741–752.

Bruzone M. C., Fontenla S. B. & Vohník M. 2015. Is the prominent ericoid mycorrhizal fungus *Rhizoscyphus ericae* absent in the Southern Hemisphere's Ericaceae? A case study on the diversity of root mycobionts in Gaultheria spp. From northwest Patagonia, Argentina. Mycorrhiza 25(1): 25-40. DOI 10.1007/s00572-014-0586-3

Chelo I. M., Zé-Zé L. & Tenreiro R. 2007. Congruence of evolutionary relationships inside the Leuconostoc–Oenococcus–Weissella clade assessed by phylogenetic analysis of the 16S rRNA gene, dnaA, gyrB, rpoC and dnaK. International Journal of Systematic and Evolutionary Microbiology 57: 276-286.

Chotewutmontri P. & Barkan A. 2016. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. PLoS Genet 12 (7): e1006106. doi:10.1371/journal.pgen.1006106

Christelová P., Valárik M., Hřibová E., de Langhe E. & Doležel J. 2011. A multi-gene sequence-based phylogeny of the Musaceae (banana) family. BMC Evolutionary Biology. 11:103. doi: 10.1186/1471-2148-11-103

Darriba D., Taboada G. L., Doallo R. & Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9(8): 772.

Edgar R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792-97.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791.

Fig Tree, version 1.4.2. Software for molecular evolution, phylogenetics and epidemiology. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 29 January 2017).

Gillespie E. & Kron K. 2010. Molecular phylogenic relationships and a revised classification of the subfamily *Ericoidae* (Ericaceae). Molecular Phylogenetics and Evolution 56: 343-354.

Glen H. F. 2002. Cultivated Plants of Southern Africa. 428 pp. Jacana, Johannesburg.

Hall T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuclear Acids Symposium Series 41: 95-98.

Hansen I. 1950. Die europäischen Arten der Gattung *Erica* L. Bot. Jarhb. 75: 1-81.

Hulsenbeck J. P. & Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

Jeffroy O., Brinkmann H., Delsuc F. & Philippe H. 2006. Phylogenomics: The beginning of incongruence? Trends in Genetics 22: 225-231.

Korczak B., Christensen H., Emler S., Frey J. & Kuhnert P. 2004. Phylogeny of the family Pasteurellaceae based on rpoB sequences. International Journal of Systematic and Evolutionary Microbiology 54: 1393-1399.

Krause K., Falk J., Humbeck K. & Krupinska K. 1998. Responses of the transcriptional apparatus of barley chloroplasts to a prolonged dark period and to subsequent re-illumination. Journal of Plant Physiology 104: 143-152.

Krawczyk K. & Sawicki J. 2013. The Uneven Rate of the Molecular Evolution of Gene Sequences of DNA-Dependent RNA Polymerase I of the Genus *Lamium* L. International Journal of Molecular Science 14: 11376-11391.

Liere K., Weihe A. & Borner T. 2011. The transcription machineries of plant mitochondria and chloroplasts: composition, function and regulation. Journal of Plant Physiology 168: 1345-1360.

Linder H. P. 2003. The radiation of the Cape flora, southern Africa. Biological Review. 78: 597-638.

Lió P. & Goldman N. 1998. Models of molecular evolution and phylogeny. Genome Research 8: 1233-1244.

Little M. C. & Hallick R. B. 1988. Chloroplast rpoA, rpoB, and rpoC genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. Journal of Biological Chemistry 263: 14302-14307.

Logacheva M. D., Penin A. A., Samigullin T. H., Vallejo-Roman C. M. & Antonov A. S. 2007. Phylogeny of flowering plants by the chloroplast genome sequences: In search of a "Lucky Gene". Biochemistry 72: 1324-1330.

Lopez-Giraldez F. & Townsend J. P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evolutionary Biology. doi:10.1186/1471-2148-11-152.

Maddison W. P. & Knowles L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematical Biology 55: 21-30.

Mayrose I., Graur D., Ben-Tal N. & Pupko T. 2004. Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. Molecular Biology and Evolution 21: 1781-1791

McGuire A. F. & Kron K. A. 2005. Phylogenetic relationships of European and African Ericas. International Journal of Plant Science 166: 311-318.

Nei M. & Kumar S. 2000. Molecular Evolution and Phylogenetics. xiv+333 pp. Oxford University Press, New York, NY, USA.

Newmaster S. G, Fazekas A. J, Steeves R. A. D. & Janovec J. 2008. Testing candidate plant barcode regions in the *Myristicaceae*. Molecular Ecology Research 8: 480-490.

Ojeda F., Budde K. B. & Heuertz M. 2015. Biogeography and evolution of seeder and resprouter forms of *Erica coccinea* (Ericaceae) in the fire-prone Cape fynbos. Plant Ecology. DOI: 10.1007/s11258-015-0539-8.

Oliver E. G. H. 1993. Studies in the *Ericoideae* (Ericaceae). XI. The generic relationship between *Erica* and *Blaeria*. Kew Bulletin 48: 771-780.

Oliver E. G. H. 2000. Systematics of *Ericeae* (Ericaceae: *Ericoideae*): species with indehiscent and partially dehiscent fruits. Contributions from the Bolus Herbarium 19: 1-483.

Oliver E. G. H. 1993. Studies in the *Ericoideae* (Ericaceae). XII. The placing of the genus *Blaeria* into synonymy under *Erica*; nomenclatural and taxonomic changes for the southern African region. Bothalia 23: 1-7.

Oliver E. G. H. & Oliver I. M. 2003. Ericaceae. In: G. Germishuizen & N. L. Meyer (eds.). Plants of southern Africa: an annotated checklist, pp 424-451. Strelitzia. National Botanical Institute, Pretoria.

Petersen G. & Seberg O. 1997. Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA sequence data. Molecular Phylogenetics and Evolution 7: 217-230.

Pirie M. D., Oliver E. G. H. & Bellstedt D. U. 2011. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggest numerous shifts in floral macro-morphology. Molecular Phylogenetics and Evolution 61: 593-601.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution 25: 1253-1256.

Pond S. L. K. & Muse S. V. 2005. HyPhy: Hypothesis testing using phylogenies. Bioinformatics 21: 676-679.

Schumann D., Kirsten G. & Oliver E. G. H. 1992. Ericas of South Africa. 272 pp. Fernwood Press, Cape Town.

Serino G. & Maliga P. 1998. RNA polymerase subunits encoded by the plastid rpo genes are not shared with

the nucleus-encoded plastid enzyme. Plant Physiology 117: 1165-1170.

Stevens P. F., Luteyn J., Oliver E. G. H., Bell T. L., Brown E. A., Crowden R. K., George A. S., Jordan G. J., Ladd P., Lemson K., McClean C. B., Menadue Y., Pate J. S., Stace H. M. & Weiller C. M. 2004. Ericaceae. In: K. Kubitzki (ed.). The Families and Genera of Vascular Plants, pp. 145-194. Springer-Verlad, Berlin-Heidelberg New York.

Szczecińska M., Gomolińska A., Szkudlarz P. & Sawicki J. 2014. Plastid and nuclear genomic resources of a relict and endangered plant species: *Chamaedaphne calyculata* (L.) Moench (Ericaceae). Turkish Journal of Botany 38: 1229-1238.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. 2013. MEGA 6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729.

Townsend J. P. 2007. Profiling phylogenetic informativeness. Systematic Biology 56: 222-231.

Townsend J. P., López-Giráldez F. & Friedman R. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. Journal Molecular of Systematic and Evolution 67: 437-447.

Townsend J. P. & Leuenberger C. 2011. Taxon sampling and the optimal rates of evolution for phylogenetic inference. Systematic Biology 60: 358-365.

Woolley S., Johnson J., Smith M. J., Crandall K. A. & McClellan D. A. 2003. TreeSAAP: Selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671-672.

Xia X. & Xie Z. 2001. DAMBE: Software package for data analysis in molecular biology and evolution. Journal of Heredity 92: 371-373.

Xia X. 2013. DAMBE 5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. Molecular Biology and Evolution 30(7): 1720-1728.

Yagi Y. & Shiina T. 2014. Recent advances in the study of chloroplast gene expression and its evolution. Front Plant Science 5: 61-68.

Yang Z. 1996. Among site rate variation and its impact on phylogenetic analyses. Trends in Ecology and Evolution 11: 367-372.

Yoichi W. & Tomaru N. 2014. Patterns of geographic distribution have a considerable influence on population genetic structure in one common and two rare species of *Rhododendron* (Ericaceae).Tree Genetics & Genomes 10: 827-837.

Appendix 1. Characteristics of radical amino acid properties changes under positive destabilizing selection in the *rpo* genes (an outgroup not included)

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
rpoC2	112	*Erica hirtiflora*	L into E	Chromatographic index	8	0.001
	125	*E. baccans*	F into C			
	180	*E. bauerii*, *E. imbricata*, *E. imbricata2*, *E. vestita*, *E. coccinea*, *E. mammosa*, *E. sessiliflora*, *E. hibbertii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*	F into C			
	125	*E. baccans*	F into C	Composition (C)	8	0.050
	180	*E. bauerii*, *E. imbricata*, *E. imbricata2*, *E. vestita*, *E. coccinea*, *E. mammosa*, *E. sessiliflora*, *E. hibberii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*	F into C			
	114	*E. baccans*	W into G	Molecular volume	8	0.050
	114	*E. baccans*	W into G	Molecular weight	8	0.050
	114	*E. baccans*	W into G	Partial specific volume (v_i)	8	0.050
	114	*E. baccans*	W into G	Refractive index	8	0.050
	107	*E. baccans*, *E. hirtiflora*	N into I	Buriedness	7	0.050
	111	*E. selaginifolia*	Q into L			
	112	*E. hirtiflora*	L into E			
	123	*E. baccans*, *E. hispidula*	N into I			
	148	*E. baccans*	N into I			
	107	*E. baccans*, *E. hirtiflora*	N into I	Coil tendencies	7	0.050
	114	*E. baccans*	W into G			
	123	*E. baccans*, *E. hispidula*	N into I			
	148	*E. baccans*	N into I			
	93	*E. cinerea*, *E. cinerea2*	H into Y	Power to be at the middle of alpha-helix (a_m)	7	0.050
	90	*E. hirtiflora*	L into S	Coil tendencies	6	0.050
	16	*E. baccans*	T into K	Isoelectric point (pH_i)	6	0.050
	59	*E. baccans*	P into R			
	102	*E. multiflora*, *E. tetralix*	G into R			
	107	*E. caffra*, *E. mammosa*, *E. patersonii*, *E. selaginifolia*, *E. sessiliflora*, *E. strigilifolia*	K into N			
	107	*E. plumigera*	K into T			
	114	*E. selaginifolia*	W into R			
	121	*E. plumigera*	K into N			
	127	*E. baccans*	T into K			
	176	*E. pseudocalycina*	N into K			
	179	*E. cinerea*, *E. cinerea2*	K into N			
	68	*E. multiflora*	D into G	Polar requirement	6	0.050
	98	*E. cruenta*	I into K			
	107	*E. hirtiflora*, *E. baccans*	N into I			
	121	*E. baccans*, *E. hispidula*	K into I			
	123	*E. baccans*, *E. hispidula*	N into I			
	148	*E. baccans*	N into I			
	16	*E. baccans*	T into K	Total non-bonded energy	6	0.050
	68	*E. multiflora*	D into G			
	102	*E. multiflora*, *E. tetralix*	G into R			
	107	*E. baccans*, *E. hirtiflora*	N into I			
	107	*E. plumigera*	K into T			
	112	*E. hirtiflora*	L into E			
	114	*E. selaginifolia*	W into R			
	123	*E. baccans*, *E. hispidula*	N into I			

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
	127	*E. baccans*	T into K			
	148	*E. baccans*	N into I			
rpoC1	42	*E. pseudocalycina*	A into P	Alpha-helical tendencies (P_α)	8	0.050
	48	*E. discolor*, *E.versicolor*, *E. versicolor2*	G into E			
	60	*E. imbricata*, *E. imbricata2*	A into G			
	147	*E. globiceps*, *E.globiceps2*, *E. versicolor*, *E. versicolor2*	E into G			
	73	*E. multiflora*	W into C	Chromatographic index	8	0.050
	98	*E. melanthera*	W into C			
	195	*E. coccinea*	F into C			
	133	*E. sessiliflora*	D into V	Beta-structure tendencies	7	0.050
	48	*E. discolor*, *E.versicolor*, *E. versicolor2*	G into E	Compressibility	7	0.050
	147	*E. globiceps*, *E.globiceps2*, *E. versicolor*, *E. versicolor2*	E into G			
	73	*E. multiflora*	W into C	Partial specific volume	7	0.050
	98	*E. melanthera*	W into C			
	102	*E. coccinea*, *E. sessiliflora*	R into G			
	197	*E. bauerii*	G into R			
	45	*E. melanthera*	P into S	Power to be at the C-terminal	7	0.050
	48	*E. discolor*, *E.versicolor*, *E. versicolor2*	G into E			
	132	*E. vagans*	P into S			
	133	*E. sessiliflora*	D into V			
	147	*E. globiceps*, *E.globiceps2*, *E. versicolor*, *E. versicolor2*	E into G			
	233	*E. baccans*	E into K			
	118	*E. patersonii*, *E. vagans*	T into I	Solvent accessible reduction ratio	7	0.050
	133	*E. sessiliflora*	D into V			
	73	*E. multiflora*	W into C	Molecular volume	6	0.050
	98	*E. melanthera*	W into C			
	102	*E. coccinea*, *E. sessiliflora*	R into G			
	197	*E. bauerii*	G into R			
	73	*E. multiflora*	W into C	Molecular weight	6	0.050
	98	*E. melanthera*	W into C			
rpoB	108	*E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	I into R	Buriedness	8	0.001
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Coil tendencies	8	0.050
	27	*E. regia*	G into V			
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	189	*E. sessiliflora*	A into G			
	39	*E. cinerea2*	R into G	Helical contact area	8	0.050
	103	*E. caffra*	G into R			
	215	*E. sessiliflora*, *E. plukenetii2*	R into G			
	108	*E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	I into R	Hydropathy	8	0.010

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
	215	*E. regia*	R into L			
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Power to be at the C-terminal	8	0.050
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Power to be at the N-terminal (a_n)	8	0.001
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Compressibility	7	0.001
	15	*E. floccifera*	E into G			
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	197	*E. cinerea2*	G into E			
	39	*E. cinerea2*	R into G	Molecular weight	7	0.050
	103	*E. caffra*	G into R			
	215	*E. sessiliflora*, *E. plukenetii2*	R into G			
	5	*E. globiceps*	I into T	Solvent accessible reduction ratio	7	0.001
	19	*E. selaginifolia*	T into I			
	108	*E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	I into R			
	14	*E. hirtiflora*	M into R	Total non-bonded energy	7	0.001
	84	*E. globiceps*	Q into L			
	108	*E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	I into R			
	215	*E. regia*	R into L			
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Turn tendencies	7	0.010
	9	*E. bauerii*, *E. vestita*	S into L			
	15	*E. floccifera*	E into G			
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	189	*E. sessiliflora*	A into G			
	197	*E. cinerea2*	G into E			
	201	*E. regia*	L into S			
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L	Alpha-helical tendencies	6	0.050
	14	*E. plukenetii SWO56*	M into T			
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	90	*E. newdigate*	A into T			
	5	*E. globiceps*	I into T	Average number of surrounding residues	6	0.001
	7	*E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans*	P into L			
	9	*E. bauerii*, *E. vestita*	S into L			
	19	*E. selaginifolia*	T into I			
	57	*E. bauerii*, *E. hibbertii*, *E. vestita*	L into P			
	84	*E. globiceps*	Q into L			
	108	*E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii*	I into R			

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
		SWO18, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*				
	201 215 39 101 103 108 122 214 215	*E. regia* *E. regia* *E. cinerea2* *E. arborea2*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. cruenta*, *E. discolor*, *E. globiceps*, *E. imbricata2*, *E. mammosa*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. regia*, *E. sessiliflora*, *E. strigilifolia*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. vestita* *E. caffra* *E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora* *E. cinerea2* *E. patersonii* *E. sessiliflora*, *E. plukenetii2*	L into S R into L R into G T into K G into R I into R K into Q K into N R into L	Isoelectric point (pHi)	6	0.001
	84 108	*E. globiceps* *E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	Q into L I into R	Mean r.m.s. fluctuation displacement	6	0.050
	14 108	*E. hirtiflora* *E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	M into R I into R	Polarity	6	0.001
	5 7 19 57 108	*E. globiceps* *E. multiflora*, *E. arborea*, *E. arborea2*, *E. vagans* *E. selaginifolia* *E. bauerii*, *E. hibbertii*, *E. vestita* *E. regia*, *E. cinerea2*, *E. multiflora*, *E. imbricata2*, *E. globiceps2*, *E. arborea2*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. strigilifolia*, *E. sessiliflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. discolor*, *E. cruenta*, *E. curviflora*	I into T P into L T into I L into P I into R	Surrounding hydrophobicity	6	0.001

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
	27	*E. cinerea*, *E. cinerea2*	G into D	Total non-bonded energy	6	0.010
	39	*E. cinerea2*	R into G			
	101	*E. arborea2*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. cruenta*, *E. discolor*, *E. globiceps*, *E. imbricata2*, *E. mammosa*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E.plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. regia*, *E. sessiliflora*, *E. strigilifolia*, *E. vagans*, *E. versicolor*, *E. versicolor2*, *E. vestita*	T into K			
	103	*E. caffra*	G into R			
	215	*E. sessiliflora*, *E. plukenetii2*	R into G			
rpoA	50	*E. pseudocalycina*	E into G	Alpha-helical tendencies	8	0.050
	67	*E. floccifera*, *E. plumigera*	G into A			
	181	*E. ampullaceae*	P into A			
	237	*E. umbelliflora*	G into E			
	13	*E. tetralix*	W into G	Bulkiness	8	0.010
	19	*E. umbelliflora*	V into G			
	189	*E. imbricata*	V into G			
	19	*E. umbelliflora*	V into G	Coil tendencies	8	0.010
	67	*E. floccifera*, *E. plumigera*	G into A			
	181	*E. ampullaceae*	P into A			
	189	*E. imbricata*	V into G			
	239	*E. hispidula*, *E. versicolor2*	P into L			
	44	*E. hispidula*	R into L	Hydropathy	8	0.050
	13	*E. tetralix*	W into G	Molecular volume	8	0.050
	13	*E. tetralix*	W into G	Molecular weight	8	0.050
	13	*E. tetralix*	W into G	Partial specific volume	8	0.050
	82	*E. umbelliflora*	H into P	Power to be at the C-terminal	8	0.001
	176	*E. cruenta*, *E. discolor*, *E. hispidula*, *E. patersonii*, *E. plumigera*, *E. sessiliflora*, *E. tetralix*, *E. versicolor*	H into P			
	211	*E. ampullaceae*, *E. arborea2*, *E. cinerea*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. floccifera*, *E. globiceps*, *E. globiceps2*, *E. imbricata*, *E. imbricata2*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. pseudocalycina*, *E. selaginifolia*, *E. vagans*, *E. versicolor2*	P into H			
	239	*E. hispidula*, *E. versicolor2*	P into L			
	82	*E. umbelliflora*	H into P	Power to be at the middle of alpha-helix (a_m)	8	0.001
	176	*E. cruenta*, *E. discolor*, *E. hispidula*, *E. patersonii*, *E. plumigera*, *E. sessiliflora*, *E. tetralix*, *E. versicolor*	H into P			
	211	*E. ampullaceae*, *E. arborea2*, *E. cinerea*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. floccifera*, *E. globiceps*, *E. globiceps2*, *E. imbricata*, *E. imbricata2*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii*	P into H			

Gene	Codon	Branch	Amino acid change	Property	Category of change	Statistical error
		SWO56, *E. plukenetii2*, *E. pseudocalycina*, *E. selaginifolia*, *E. vagans*, *E. versicolor2*				
	13	*E. tetralix*	W into G	Refractive index	8	0.050
	44	*E. hispidula*	R into L	Chromatographic index	7	0.050
	136	*E. pseudocalycina*	C into R	Composition (C)	7	0.050
	151	*E. coccinea*, *E. imbricata*, *E. imbricata2*	C into R			
	50	*E. pseudocalycina*	E into G	Compressibility	7	0.001
	82	*E. umbelliflora*	H into P			
	176	*E. cruenta*, *E. discolor*, *E. hispidula*, *E. patersonii*, *E. plumigera*, *E. sessiliflora*, *E. tetralix*, *E. versicolor*	H into P			
	211	*E. ampullaceae*, *E. arborea2*, *E. cinerea*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. floccifera*, *E. globiceps*, *E. globiceps2*, *E. imbricata*, *E. imbricata2*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. pseudocalycina*, *E. selaginifolia*, *E. vagans*, *E. versicolor2*	P into H			
	237	*E. umbelliflora*	G into E			
	239	*E. hispidula*, *E. versicolor2*	P into L			
	28	*E. tetralix*	S into R	Isoelectric point	7	0.050
	19	*E. umbelliflora*	V into G	Mean r.m.s. fluctuation displacement	7	0.001
	82	*E. umbelliflora*	H into P			
	176	*E. cruenta*, *E. discolor*, *E. hispidula*, *E. patersonii*, *E. plumigera*, *E. sessiliflora*, *E. tetralix*, *E. versicolor*	H into P			
	211	*E. ampullaceae*, *E. arborea2*, *E. cinerea*, *E. cinerea2*, *E. coccinea*, *E. curviflora*, *E. floccifera*, *E. globiceps*, *E. globiceps2*, *E. imbricata*, *E. imbricata2*, *E. melanthera*, *E. multiflora*, *E. patersonii*, *E. plukenetii SWO3*, *E. plukenetii SWO18*, *E. plukenetii SWO56*, *E. plukenetii2*, *E. pseudocalycina*, *E. selaginifolia*, *E. vagans*, *E. versicolor2*	P into H			
	216	*E. imbricata*	M into K			
	239	*E. hispidula*, *E. versicolor2*	P into L			
	160	*E. globiceps*, *E. globiceps2*, *E. tetralix*,	T into A	Alpha-helical tendencies	6	0.050
	168	*E. strigilifolia*	M into T			
	182	*E. cinerea*, *E. cinerea2*	A into T			
	203	*E. regia*	T into A			
	239	*E. hispidula*, *E. versicolor2*	P into L			
	43	*E. globiceps2*	I into T	Buriedness	6	0.050
	216	*E. imbricata*	M into K			
	13	*E. tetralix*	W into G	Chromatographic index	6	0.050
	60	*E. multiflora*, *E. tetralix*	F into S			
	216	*E. imbricata*	M into K			
	31	*E. umbelliflora*, *E. melanthera*	N into K	Isoelectric point	6	0.050
	44	*E. hispidula*	R into L			
	71	*E. cinerea*, *E. cinerea2*	K into T			
	198	*E. versicolor2*	T into K			
	222	*E. caffra*	K into Q			