PLACEMENT OF *Syzygium boerlagei* (Merr.) Govaerts (MYRTACEAE) CONFIRMED WITH ATPB-RBCL INTERGENIC SPACER

PUDJI WIDODO[1*], TATIK CHIKMAWATI[2] AND YAYAN WAHYU C KUSUMA[3]

[1]*Faculty of Biology, Universitas Jenderal Soedirman, Purwokerto 53122, Indonesia*
[2]*Department of Biology, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Bogor 16680, Indonesia*
[3]*PKT Kebun Raya Bogor – Indonesian Institute of Sciences (LIPI), Bogor 16003, Indonesia*
[*]*Corresponding author, e-mail: pwidodo@unsoed.ac.id*

Received …….. / Accepted ……..

ABSTRACT

We present a molecular analysis to determine whether *Eugenia boerlagei* Merr. (Myrtaceae) is to be included in *Eugenia* or *Syzygium* based on sequences of cpDNA fragments namely atpB-rbcL intergenic spacer. In this study we used seven specimens of *Syzygium* sect. Jambosa, three of *Syzygium* sect. *Syzygium*, two of *Eugenia s.s.* and one of *Eugenia boerlagei* Merr. and *Baeckea ovalifolia* and *B. tuberculata* as the outgroup. The results showed that *Eugenia boerlagei* is better placed in *Syzygium*.

Keywords: atpB-rbcL intergenic spacer, chloroplast DNA, *Eugenia*, Myrtaceae, *Syzygium*.

INTRODUCTION

Eugenia boerlagei Merr. (type: Robinson 1872, lost; iso: BO, L, elsewhere? Moluccas, Amboina, Liang) (Myrtaceae) is a shrub or small tree. It is characterized by lateral and terminal, slender, 3-flowered inflorescences, long pedicels, and a long, narrowed calyx-tube, which, with sepals and petals is glandular-punctate. The species was dedicated to J.G. Boerlage who contracted a fever while carrying on a botanical exploration of Amboina in the year 1900, which resulted in his untimely death (Cox & Merrill 1916). Govaerts *et al.* (2008) transferred it to *Syzygium boerlagei* Govaerts. In this paper, we would like to verify molecularly whether it should be transferred to *Syzygium* or should be retained in *Eugenia* L.

Although it seems to be pretty rare, at least in L there only is an isotype, but then there are also c. 55 boxes of *Eugenia* and *Syzygium* indet. *Eugenia boerlagei* was selected because it was just transferred in 2008, some months before we started working, and the availability of living materials in Bogor.

Syzygium is the largest genus of the Myrtaceae, comprising ca 1200 species in the Old World (Craven *et al.* 2006) or approximately 1040 species (Govaerts *et al.* 2008). The current concept of Syzygium includes species with and without an intercotyledonary inclusion, inflorescence either solitary, axillary or terminal, calyx either calyptrate or free (Craven *et al.* 2006). Recent revision (Biffin *et al.* 2006) of *Syzygium s.l.* (sensu Hyland 1983) is based on a sub-generic arrangement that distinguishes *Syzygium s.s.* from traditionally associated taxa by the possession of indistinct calyx lobes and coherent petals.

The generic taxonomy of *Syzygium* has long been associated with *Eugenia*, from which it is only weakly distinguished by macro-morphological data. Anatomical and molecular evidence now suggests these two groups are in fact quite distantly related (Biffin, 2005). Most of the species of *Eugenia* in the Old World have been transferred to *Syzygium*.

* Corresponding author: pwidodo@unsoed.ac.id

In this study we use chloroplast DNA, a highly conserved cytoplasmic molecule inherited clonally (without recombination), which has been shown to be a powerful tool to document the parentage of polyploids and the phylogenetic relationships between distinct polyploid taxa in polyploid complexes (*Eugenia* g. Erickson *et al*., 1983; Gauthier *et al*., 1997). Moreover, because cpDNA is maternally inherited in most angiosperms, the use of that marker may be particularly informative for clarifying genetic relationships between the taxa.

In plants, the mitochondrial and chloroplast genomes are evolving too slowly to provide enough variation. For taxonomists, the current strategy is to sequence several DNA regions (Taberlet *et al*., 2007). The chloroplast genome that we used was the atpB-rbcL spacer which is a noncoding region of the genome that has been used in phylogenetic studies of Angiosperms (Manen *et al*. 1994, Manen & Natali 1995). In this case rbcL and atpB are transcribed in opposite directions.

The plastid locus most commonly sequenced by plant systematists for phylogenetic purposes is rbcL, followed by the trnL-F intergenic spacer, matK, ndhF, and atpB (Gielly & Taberlet (1994), Shaw *et al*., 2015). rbcL has been suggested as a candidate for plant barcoding (Blaxter, 2004), even though it has generally been used to determine evolutionary relationships at the generic level and above. Besides rbcL and atpB, all of the latter plastid loci have been used at the species level with various degrees of success (Kress *et al*., 2005).

In this study we present a molecular analysis of some species of *Syzygium* and *Eugenia* based on the sequence of the atpB-rbcL intergenic spacer from, what we hope, is a representative sampling of *Syzygium* and *Eugenia*. The objectives were: (1) to determine the placement of Eugenia boerlagei, whether it should be placed in *Syzygium* or in *Eugenia*; (2) to improve an understanding of the relationships between *Eugenia* and *Syzygium* which are morphologically slightly different.

MATERIALS AND METHODS

Writing the Materials and Methods

Samples were obtained from living plants growing in the Bogor Botanic Garden and its vicinity. The ingroup represents a sampling of morphological diversity within *Syzygium*. Ten types of *Syzygium* comprising six specimens of sect. Jambosa, four of sect. *Syzygium*, two of *Eugenia s.s.*, two of *Baeckea ovalifolia* and *B. tuberculata* from GenBank and one of *Eugenia boerlagei* (*Syzygium boerlagei*) were examined (Table 1). Voucher specimens have been stored in the Herbarium Bogoriense (BO and Herbarium Fakultas Biologi Unsoed, PUNS). The sequences of *Eugenia* and *Syzygium* were submitted to GenBank on 22 November 2017 and waiting for accession number.

Table 1 List of specimens examined in this study and the voucher specimen of *Baeckea*, *Eugenia*, and *Syzygium*.

Taxa	Voucher detail	Localities	Accession
Baeckea ovalifolia	*GenBank*	Australia NSW	EF581242
Baeckea tuberculata	*GenBank*	Australia NSW	EF581244
Eugenia boerlagei Merr. – *Syzygium boerlagei* (Merr.) Govaerts	*Widodo 143*	Mollucas, Ambon (KRB)	MG669291
Eugenia pyriformis Cambess	*Widodo 142*	Brazil, Indonesia (KRB)	MH191262
Eugenia uniflora L.	*Widodo 141*	Java Bogor (IPB)	SAMN08056079
Syzygium aqueum (Burm. f.) Alston	*Widodo 132*	Java Bogor (IPB)	MH191263
Syzygium aromaticum (L.) Merr. & L.M. Perry	*Widodo 137*	Java Bogor (IPB)	MH191264
Syzygium littorale (Blume) Amshoff	*Widodo 135*	Borneo (KRB)	MH191265
Syzygium polyanthum (Wight) Walp.	*Widodo 139*	Java Bogor (KRB)	MH191266
Syzygium polycephalum (Miq.) Merr. & L.M. Perry	*Widodo 136*	Java Bogor (KRB)	MH191267
Syzygium samarangense (Blume) Merr. & L.M. Perry	*Widodo 131*	Java Bogor (IPB)	MH191268

Notes: IPB = cultuivated in Institut Pertanian Bogor (Bogor Agricultural University)
KRB = cultuivated in Kebun Raya Bogor (Bogor Botanic Gardens)

DNA extraction, amplification, sequencing, and alignment

Total DNA was extracted from fresh material following the standard CTAB (hexadecyltrimethylammonium bromide) extraction methods (Doyle and Doyle, 1987). Double stranded DNA was directly amplified by PCR for all loci. Reaction volumes were 25 µl and contained 2.5 µl PCR buffer, 1 µl dNTPs, 0.1 µl each of the 10 mM primers, 1.5 µl 25 mM MgCl2, 0.1 µl TaqPol and 15.2 µl ddH2O. Approximately 4.5 µl genomic DNA was added to the PCR mixtur*Eugenia* . The primers used in this study for atpB-rbcL intergenic spacers are (forward: atpB-1: 5'-ACATCKARTACKGGACC AATAA-3' and reverse rbcL-1: 5'-AACACCAGCTTTRAATCCAA-3') (Chiang et al. 1998). A non coding cpDNA fragment namely atpB-rbcL spacer was amplified.

PCR was performed with 4 min at 94ºC for the activation of the polymerase, followed by 35 cycles of 45 sec at 94ºC, 45 sec at 55ºC, 2 min at 42ºC, with a final extension period of 10 min at 72ºC The PCR product was checked on 1% agarose gel, and were purified with using a purification kit of Wizard SV Gel and PCR clean up system (PROMEGA), following the manufacturers' protocol prior to sequencing. The DNA concentration was measured with the nanodrop. Cycle sequencing was performed by MACROGEN Korea. The sequences were edited manually and sequently manually adjusted using Sequencher 4.6 and MEGA 3.0 (Kumar *et al*., 2004).

Phylogenetic analysis

Cladistic analyses of the atpB-rbcL IGS sequence data were performed using a maximum parsimony criterion by using MEGA 6.0 (Tamura *et al*., 2013). The methods produced phylogenetic trees that provided insights concerning major general evolutionary trends in the *Eugenia* and *Syzygium*. Interesting findings were: (i) *Eugenia boerlagei* is the sister species to *Syzygium aqueum* (Burm. f.) Alston; (ii) the two *Eugenia* samples are distantly related to all *Syzygium*.

The fit of character data on phylogenetic hypotheses (Swofford 1998) was evaluated by the consistency index, CI (Kluge & Farris 1969), and the retention index, RI (Archie 1989; Farris 1989). The statistical significance of the CI was determined according to the method of Klassen *et al*., (1991). Confidence in the clades was tested by bootstrapping (Effron 1982; Felsenstein 1985) with 100 replicates of heuristic searches on the 50% majority rule trees. The nodes with bootstrap values >0.70, as a rule of thumb, were considered significantly supported with 395% probability (Hillis & Bull 1993).

RESULTS AND DISCUSSION

DNA sequencing and alignment

In our data for some *Syzygium*, the length of atpB-rbcL intergenic spacer varied from 903 to 962 base pairs within Myrtaceae The shortest is *Syzygium lineatum* (903 bp), followed by *Syzygium astronioides* (912 bp), *Syzygium samarangense* (916 bp), and *Syzygium aqueum* (920 bp). While the longest is *Syzygium cumini* (962 bp), followed by *Syzygium malaccense* (955 bp), *Syzygium aromaticum* (942 bp) etc. (Table 1). The position of *Eugenia* uniflora and *Eugenia pyriformis* is in between those of *Syzygium*. Thus, the length of atpB-rbcL spacer does not determine the differences between *Eugenia* and *Syzygium*.

Table 2 Length variation, AT and GC content of *atpB-rbcL* intergenic spacer in *Baeckea, Eugenia,* and *Syzygium*

Taxa	Sequence length (bp)	AT content	GC content
B. ovalifolia	843	110	16
B. tuberculata	844	105	15
E_boerlagei	941	106	16
E_pyriformis	923	106	18
E_uniflora	925	107	18
S_aqueum	920	107	16
S_aromaticum	942	111	16
S_littorale	925	107	16
S_polyanthum	936	105	16
S_polycephalum	921	108	16
S_samarangense	916	107	16

In general *Syzygium* is AT-rich, where the AT content of the spacer ranges from 105 – 108. While the GC content ranges from 16 - 18. The AT content in both *Eugenia* is 106 1nd 107, in

between all *Syzygium*. Thus, the AT content can not be used to determine the differences between *Eugenia* and *Syzygium*. On the other hand, the CG content of both *Eugenia s.s.* is the riches (18) compared to *Syzygium*. Thus GC content of the atpB-rbcL intergenic spacer may indicate the differences between *Eugenia* and *Syzygium* (Figure 1). From this fact, we can conclude that *Eugenia boerlagei* is should be transferred to *Syzygium boerlagei*.

Most of the variation was due to indels and substitutions in atpB-rbcL IGS (Table 2). When aligned, the sequences have 1060 sites for atpB-rbcL IGS In total for two fragments, there are 77 variable characters with parsimony informative sites for atpB-rbcL. The most parsimonious analysis generated six most parsimonius trees with CI = 0.861842, RI = 0.798077, RCI = 0.687816 for all sites), iCI = 0.771739, iRI = 0.798077, iRCI = 0.615907 (for parsimony informative sites).

Molecular evolution of the chloroplast noncoding region between atpB-rbcL genes in both *Eugenia* and some Syzygium showed that most variations amongst *Syzygium* were contributed by insertion and only a few nucleotide substitutions were found. Interesting findings were as follows: (i) The main characters distinguishing *Eugenia s.s.* from *Syzygium* are the substitutions. *Eugenia s.s.* is characterized by the high number of substitutions namely ca. 33 of 1060 or around 3%. On the other hand, *Syzygium* is characterized by the low number of substitutions where the average is 0.4%.

The three morphologically distinct taxa, *Baeckea*, *Eugenia, Syzygium* are distantly related and clearly separated. *Syzygium boerlagei* is closely related to *S. aqueum*. *S. malaccense* is closely related to *S. samarangense*. *S. littorale* is a sister group of *S. aromaticum* and *S. polycephalum* (Figure 2). *Eugenia boerlagei* is better placed in *Syzygium*, so the correct name becomes *Syzygium boerlagei*.

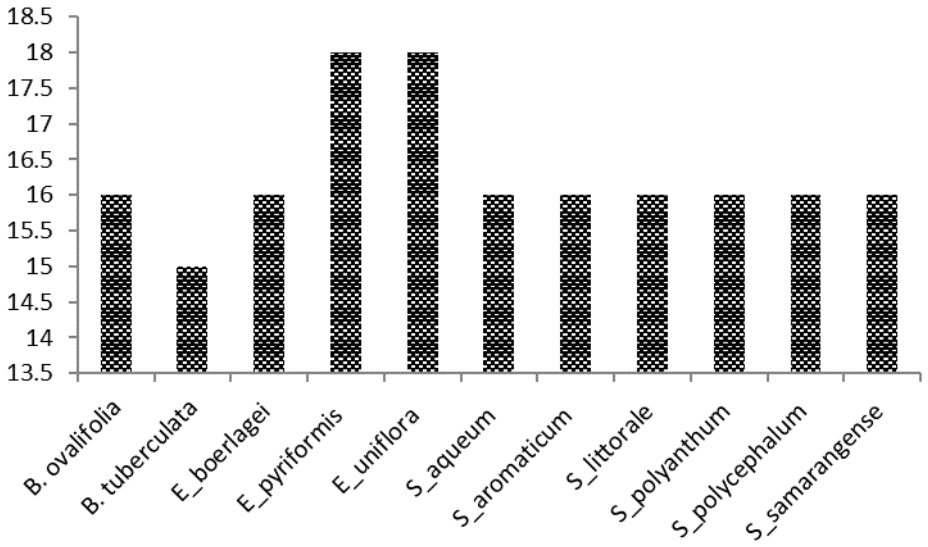

Figure 1 GC content of *atpB-rbcL* spacer sequences in *Baeckea, Eugenia* and *Syzygium*

Table 3 Insertion, deletion and substitution on DNA sequence of each taxa

No	Taxa	Insertion	Deletion	Substitution
1	*Baeckea ovalifolia*	4	123	5
2	*Baeckea tuberculata*	3	124	7
3	*Eugenia uniflora*	24	137	33
4	*Eugenia pyriformis*	20	138	33
5	*Eugenia* or *Syzygium boerlagei*	25	19	8
6	*Syzygium aqueum*	5	131	3
7	*Syzygium aromaticum*	30	18	2
8	*Syzygium littorale*	10	130	5
9	*Syzygium polyanthum*	31	124	1
10	*Syzygium polycephalum*	8	139	6
11	*Syzygium samarangense*	1	144	3

Maximum Parsimony analysis of taxa

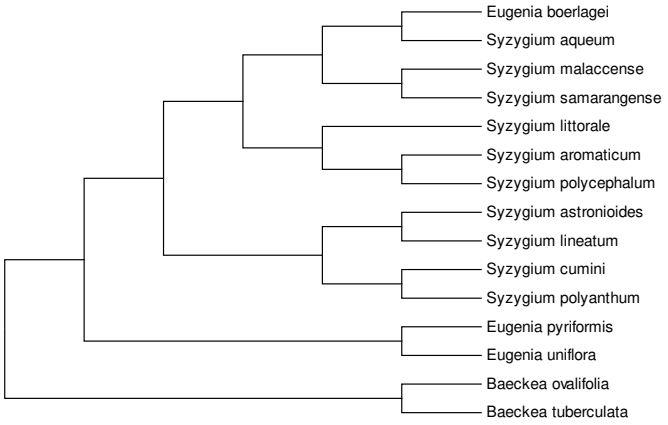

Figure 2 Maximum parsimony tree of *Syzygium, Eugenia* and *Baeckea* based on *atpB-rbcL* intergenic spacer sequence. *Eugenia boerlagei* is nested in *Syzygium*

The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length = 142 is shown. The consistency index is 0.859155 (0.780220), the retention index is 0.850746 (0.850746), and the composite index is 0.730923 (0.663769) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in ref. Nei & Kumar, 2000) with search level 0 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 735 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013).

The maximum probability of correct phylogenetic inference increases with the number of variable (or informative) characters and their consistency index and decreases with the number of taxa, when the consistency index has been standardized to eliminate its dependence on the number of taxa. Equations for the probability of correct phylogenetic inference and for the standardized consistency indices (including or excluding autapomorphies) are derived. Given that actual studies based on DNA restriction sites and sequences generate more characters with a higher level of consistency than comparable studies based on morphology, calculations suggest that such molecular studies may often provide a more precise guide to phylogenetic relationships (Givnish and Sytsma 1997).

Morphologically, *Eugenia boerlagei* is closer to *Syzygium* than to *Eugenia* s.s. because it is characterized by: 1) shoot sylleptic (not proleptic); 2) leaf bud smooth (not papillous); 3) inflorescence panicle (not solitary and clustered at nodes); and 4) fruits with 1-2 seeds (not many). Either morphologically or molecularly, *Eugenia boerlagei* is very much closer to *Syzygium* than to Eugenia. Thus, the transfer to *Syzygium* by Govaerts et al. (2008) is acceptable. These were supported by the facts that on one hand, the leaf buds are smooth and not papillose, on the other hand it has a low number of substitutions (<15) compared to the "real" *Eugenia* which have >30 substitutions in terms of DNA sequences.

Based on atpB-rbcL data, *Syzygium cumini* and *S. polyanthum* are closely related. Morphologically, both plants have similar bark patterns that are whitish, and close to each other. *S. lineatum* is closely related to *cumini* because they have the same number of GC content namely 17. With regards to the position of *S. malaccense* in relation to *S. samarangense*, our results showed that they are much closer compared to Biffin's results. The *Eugenia* group is separated from *Syzygium* because *Eugenia* has much more substitutions or mutations in some sites than *Syzygium*.

Both samples of *Eugenia* are clearly characterised by (1) substitution of C to T at

position of 143 and followed by insertion TAC from position of 144-146; (2) substitution of T to C at 359 and followed by insertion of ATTGCC from 360-365. On the other hand, *E. pyriformis* and *E. uniflora* are distinctly marked by deletion from position of 762 to 776. Molecularly, there are still some more differences between *Eugenia* and *Syzygium*, however, they are not significant enough to discuss here.

CONCLUSION

The results of maximum parsimony analysis with two species of *Baeckea* as the outgroup showed evidence that *Eugenia boerlagei* is nested in *Syzygium*, so it should be transferred to *Syzygium* as was done by Govaerts et al. (2008). The two samples of *Eugenia pyriformis* and *Eugenia uniflora* are distantly related to all *Syzygium*.

ACKNOWLEDGEMENTS

Dr. Irawati former head the Bogor Botanical Garden, and Ir. Mustaid Siregar, MSi, head of the Bogor Botanical Garden are thanked for the permission to observe and sample the collections. Mr Yayan Wahyu Chandra Kusuma is appreciated for providing materials and assistance. Dr. Barbara Gravendeel and Marcel Eurlings, Van der Klaauw Laboratory Leiden University, The Netherlands are thanked for assistance, chemicals, software, and facilities. Dr. Tim Fulcher, Jodrell Laboratory of the Royal Botanic Gardens Kew, UK is thanked for assistance and facilities. Dr. J.F. Veldkamp (L) critically read an advance draft.

REFERENCES

Archie, J.W. (1989). Homoplasy excess ratio: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. 38: 253-269.

Biffin, E., L. A. Craven, M. D. Crisp, and P. A. Gadek, (2006). Molecular systematics of *Syzygium* and allied genera (Myrtaceae): evidence from the chloroplast genome. Taxon 55: 79-94.

Blaxter M.L. (2004). The promise of a DNA taxonomy. Philosophical Transaction B. Apr 29; 359(1444): 669–679.

Chiang, T.Y, Schaal B.A and Peng C.I. (1998). Universal Primers for amplification and sequencing a noncoding spacer between atpB and rbcL genes of chloroplast DNA. Bot. Bull. Acad. Sin. 39: 245-250.

Chiang, T.Y. and Schaal, B.A. (2000). Molecular evolution of the atpB-rbcL noncoding spacer of chloroplast DNA in the moss family Hylocomiaceae. Bot. Bull. Acad. Sin. 41: 85-92.

Cox, A.J. and Merrill E. D. (1916). Myrtaceae. The Philippine Journal of Science 11: 296-297.

Craven LA, Biffin E, Ashton P.S. 2006. Acmena, Acmenosperma, Cleistocalyx, Piliocalyx and Waterhousea formally transferred to *Syzygium* (Myrtaceae). Blumea 51: 131–142

Doyle, J. J. & Doyle, J. L. (1990). The isolation of plant DNA from fresh tissue Focus 12: 13–15.

Effron, B. (1982). The jackknife, the bootstrap, and other resampling plans. Conf. Board Math. Sci. Soc. Ind. Appl. Math. 38: 1-92.

Erickson, L. R., Strauss, W. A. and Bevensdorf, W. D. (1983). Restriction patterns reveal origins of chloroplast genomes in Brassica amphidiploids. Theor Appl Genet, 65: 201–206.

Farris, J.S. (1989). The retention index and homoplasy excess. Syst. Zool. 38: 406-407.

Felsenstein, J. (1985). Confidence limits on phylogenetics: an approach using the bootstrap. Evolution 39: 783-791.

Gauthier, P., Lumaret, R. and Bédécarrats, A. (1997). Chloroplast-DNA variation in the genus Lotus (Fabaceae) and further evidence regarding the maternal parentage of Lotus corniculatus L. Theor Appl Genet, 95: 629–636.

Gielly, L. & Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. Mol. Biol. Evol. 11 , 769-777.

Givnish TJ and Sytsma KJ. (1997). Consistency, Characters, and the Likelihood of Correct Phylogenetic Inference. Molecular Phylogenetics and Evolution. 7(3): 320-330

Govaerts, R., Sobral, M., Ashton, P., Barrie, F. Holst, B.K.,Landrum, L., Nic Lughadha, E., Matsumoto, K., Mazine, F., Proença, C., Soares-Silva, L.H., Wilson, P.G. & Lucas, E. (2008). World Checklist of Myrtaceae. The Board of Trustees of the Royal Botanic Gardens, Kew.

Hillis, D.M. and Bull, J.J. (1993). An empirical test of bootstrapping as a method assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182-192

Klassen, G.J., Moy R.D., and Locke, A. (1991). Consistency indices and random data. Syst. Zool. 40: 446-457.

Kluge, A.G. & Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. Syst. Zool. 18: 1-32.

Kress W.J., Kenneth J. Wurdack K.J., Zimmer E. A., Lee A. Weigt, and Daniel H. Janzen D.H. (2005). Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A. Jun 7; 102(23): 8369–8374.

Kumar S., K. Tamura, and M. Nei. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 2

Manen J, Savolainen V, and Simon P. (1994). The atpB and rbcL promoters in plastid DNAs of a wide dicot range J. Mol. Evol. 38: 577-244

Manen J. & Natali A. (1995). Comparison of the evolution of ribulose-1, 5-biphosphate carboxylase (rbcL) and atpB-rbcL spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). J. Mol. Biol. 41: 920-927.

Nei M. and Kumar S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Shaw, J., Lickey, E.B., Beck, J. T., Farmer, S.B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E.E. & Small, R. L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis 1. Am. J. Bot. 92 , 142-166.

Swofford, D. L. (1998). PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.08. Sinauer, Massachusetts.

Taberlet, P, Coissac E, Pompanon, F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, and Willerslev E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acid Research. 35(3): e14

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution30: 2725-2729.