HISTORICAL BIOGEOGRAPHY OF BRASSICACEAE

UNRAVELLING THE GEOGRAPHIC ORIGIN OF A FAMILY

Sara van de Kerke

Registration number: 881211 428110

Email: sara.vandekerke@wur.nl

MSc thesis Biodiversity and Evolution - BIS-80436

Supervisors: Setareh Mohammadin and Freek T. Bakker

SUMMARY

Biogeography is the study of the geographic distribution of taxa and their attributes in space and time. There are three main processes that together are considered to pattern the distribution of taxa: dispersal, vicariance and extinction. Over the years, three quite distinct schools have developed within historical biogeography, which all comprehend a number of methodological directions: pattern-based, event-based, and model-based methods.

The Brassicaceae is one of the largest plant families and is considered to comprise around 3660 species distributed over 320 genera. The family was thought to have originated in the New World based on the distribution of either the 'basal' placed tribe Theylopodieae and later the Hesperideae. An Old World origin, based on the genus Aetheionema now considered to be sister to all other Brassicaceae, is now proposed. It is assumed that the origin can be found were at the moment the largest diversity in species is, which would be the Iranoturanian region. In this project, this assumption is investigated.

The most recent phylogenetic tree covering the whole family from Couvreur et al. is based on ITS, *chs, adh, matK, trnL-F, ndhF, rbcL* and *nad4*(Couvreur et al., 2010). This phylogenetic tree is updated with all available data from GenBank and alignent with two setting in the alignment program MAFFT. RAxML Maximum Likelyhood, TNT parsimony and Neighbour Network analyses were run.

GBIF distribution data on genus level was used to define areas of endemism which were assigned to all terminals in the phylogenetic tree.

A Bayesian-DIVA analysis based on 200 TNT phylogenetic trees gave the result that a combination of the regions 'Iberian Peninsula & North Western Africa', 'Central Mediterranean & Southern Balkan', 'Western and Central Anatolia & Levantine coast', and 'Caucasus, Eastern Anatolia & Iranian mountain ranges' forms the ancestral range of all Brassicaceae.

Based on the results of the Bayesian-DIVA analysis, the conclusion can be drawn that the Iranoturanian region is indeed part of the area of origin of the Brassicaceae. However, there are some issues with the underlying alignment wich make both the phylogenetic inference and the historical biogeographical analysis unreliable.

TABLE OF CONTENTS

1. Introduction 1
2. Material and Methods 5
2.1 Phylogenetic Analysis 5
2.2 Geographical Distribution 8
2.3 Biogeographical Analyses 10
3. Results 11
3.1 Phylogenetic Analysis 11
3.2 Geographical Distribution 15
3.3 Biogeographical Analysis 16
4. Discussion and Conclusion 17
4.1 Phylogenetic Analysis 17
4.2 Geographical Distribution 19
4.3 Biogeographical Analysis 20
4.4 General comments 22
Acknowledgements 23
References 24
Appendix A. Overview of GenBank Accessions 28
Appendix B. Exemplary script used in MrBayes 57
Appendix C. Phylogenetic Trees 58

1. INTRODUCTION

HISTORICAL BIOGEOGRAPHY

Biogeography is the study of the geographic distribution of taxa and their attributes in space and time (Morrone, 2010). Biogeographers track the distribution of a (group of species) throughout history (Ree & Smith, 2008). One of the best known examples is the reconstruction of the radiation of humans over the earth (Harcourt, 2012).

In biogeography, three rather different sciences meet; biology, geology and geography (Cox & Moore, 2010; Morrone, 2010). This makes it a quite varied field, which perhaps is why there are so many different approaches and theories developed in the two centuries of research in this field (Morrone, 2009). Morrone lists an exemplary 23 of these approaches for biogeography in general, but concludes that all are part of one of only two disciplines: ecological and historical biogeography (Morrone, 2009).
Ecological biogeography was started by Linnaeus, who attempted to describe all then known animals and plants and recorded the required environmental conditions. This type of biogeography focuses on the biotic and abiotic interactions of a species on a short time span.
After Darwin introduced the theory of evolution, it was accepted that the distribution of species can change over time according to their ecological requirements (Morrone, 2009). These patterns of distribution of species (or higher taxonomic levels) on much larger timescales is the focus of historical biogeography (Morrone, 2009).

There are three main processes that together are considered to pattern the distribution of taxa: dispersal, vicariance and extinction (Figure 1, Morrone 2010).
In dispersal, a species moves from its current area of distribution to another area, usually over a barrier. In this newly colonised area, it develops into another species. This concept is not to be confused with dispersion, which is the movement of a species within its original area of distribution (Cox & Moore, 2010; Crisci, 2001; Morrone, 2009).
With vicariance the original species does not move over a barrier, but one appears splitting the area of distribution and causing speciation (Cox & Moore, 2010; Morrone, 2009).
In extinction, part of the ancestral population falls away, causing it to be split up. Separated sub-populations can remain, which results in speciation (Crisci, 2001; Morrone, 2009).

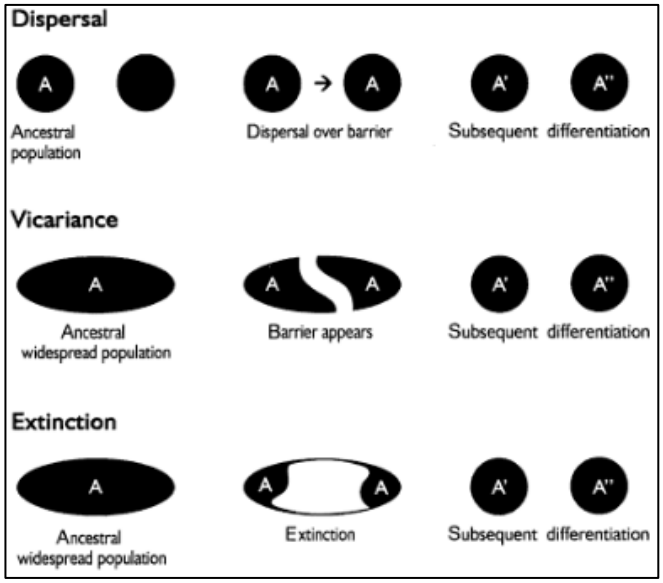


FIGURE 1. THE THREE BIOGEOGRAPHICAL PROCESSES: DISPERSAL, VICARIANCE AND EXTINCTION (AFTER MORRONE, 2010).

The analytical appraoch to biogeography really developed since the acceptance of continental drift in the 1960's. There were quite heated discussions how the above mentioned processes should be implemented in an analytical method and which was the best (Bremer, 1992, 1995; Brooks, 1990; Ronquist, 1994, 1995, 1997; Wiley, 1988). All methods that have been developed are in some way based on the three processes, but differ in how these are treated in the analyses, and in the assumptions concerning patterns and processes of change in geographical distributions.

Over the years, two quite distinct schools of methods have developed within historical biogeography, which both comprehend a number of methodological directions: pattern-based and event-based methods (Ronquist & Sanmartín, 2011). Pattern-based methods 'establish connections between distributional patterns and evolutionary processes after the primary analytical results have been obtained (Ronquist & Sanmartín, 2011). A major field in the pattern-based school is cladistic biogeography. An important part of cladistic biogeography is the constructing of area cladograms based on procedures called Assumption 0, 1, and 2 (**Error! Reference source not found.**). In this Figure, the assigned areas for the examplary cladogram on top are explained. Under Assumption 0, when a taxon occurs in more than one area, they are considered to be a monophyletic group. In contrast, under Assumption 1, they can either be mono- or paraphyletic. Last, under Assumption 2, the areas can mono-, para- or polyphyletic (Crisci, Katinas, & Posadas, 2003).

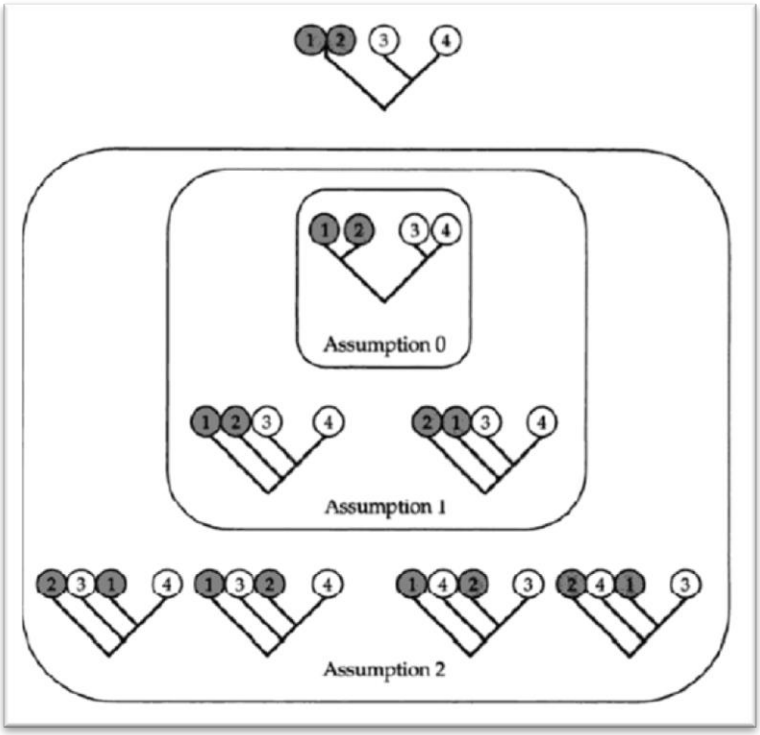
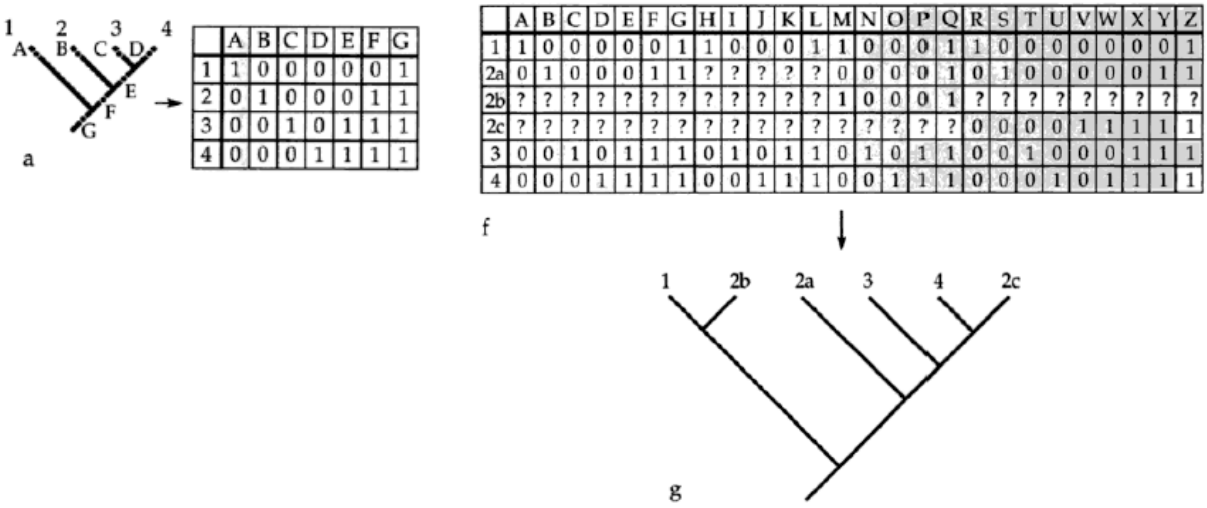


FIGURE 2. AREA CLADOGRAM WITH A WIDESPREAD TAXON IN AREAS 1 AND 2, AND DERIVATION IN RESOLVED AREA CLADOGRAMS UNDER ASSUMPTIONS 0, 1, AND 2 (AFTER CRISCI, KATINAS, & POSADAS, 2003)

One of the best known pattern-based methods is Brooks Parsimony Analysis (BPA, Wiley, 1987). This method is primarily based on Assumption 0 and constists of two analytical steps: primary and secondary BPA. In primary BPA, an area versus cladogram matrix is formed (Figure 3, left). In the cladogram, all terminals and all nodes are given a letter A through G. Then, for each of the areas 1, 2, 3, and 4, the letters (terminals and nodes) are scored they pass when traced back to the base of the cladogram. Area 1, for example, occurs on terminal A and only passes additional node G underway to the root. Area 4 occurs on terminal D and passes nodes E, F, and G, which are thus scored in the matrix. In imaginary outgroup area is defined, for wich all terminals and nodes are scored zero (absent). In secondary BPA, areas that may cause problems in the analysis (for example in the case of extinction, which would mean a loss of the area, or duplication, where multiple terminals are present in the same area), are split up and scored over the separate scenario's individually (Figure 3, right).



	A	B	C	D	E	F	G
1	1	0	0	0	0	0	1
2	0	1	0	0	0	1	1
3	0	0	1	0	1	1	1
4	0	0	0	1	1	1	1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1
2a	0	1	0	0	0	1	1	?	?	?	?	?	0	0	0	0	1	0	1	0	0	0	0	0	1	1
2b	?	?	?	?	?	?	?	?	?	?	?	?	1	0	0	0	1	?	?	?	?	?	?	?	?	?
2c	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	1	1	1	1	1
3	0	0	1	0	1	1	1	0	1	0	1	1	0	1	0	1	1	0	0	1	0	0	0	1	1	1
4	0	0	0	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	0	1	1	1	1

FIGURE 3. (LEFT) AREA VS CLADOGRAM MATRIX WITH COMPLETE DATA AND NO AMBIGUITY ('OUTGROUP' IS NOT SHOWN). (RIGHT) (AFTER CRISCI ET AL., 2003).

Event-based methods on the other hand are based on a cost matrix where each of the vicariance, dispersal and extinction processes is assigned a certain cost. These three processes are then inferred over a phylogenetic tree. This type of methods is based on parsimony, thus the least costly outcome (the most parsimonious) is the final result (Cox & Moore, 2010; Morrone, 2009; Ronquist & Sanmartín, 2011; Sanmartín, 2006).

One of the best known event-based method is Dispersal-Vicariance Analysis (DIVA), developed by Ronquist (Ronquist, 1997). DIVA came as an answer to cladistic biogeography, to which Brooks parsimony belongs (Bremer, 1992, 1995; Ronquist, 1994, 1997). DIVA 'reconstructs ancestral distributions from one given phylogenetic hypothesis, without assuming a particular process a priori' (Morrone, 2009). Vicariance is favoured over dispersal and extinction, which are subsequened considered to be more costly in the parsimony analsis (Cox & Moore, 2010; Morrone, 2009; Yu, Harris, & He, 2010b).

In recent years, with the coming of model based phylogenetic inference, also model-based historical biogeographic methods (both Bayesian inference and Maximum Likelyhood based) have been developed. With these methods, it is now possible to have an statistical based indication for the chance of a particular area occurring on a particular node in the phylogenetic tree.

Examples of model-based methods are S-DIVA and Bayesian DIVA. These are developed in response to some troubles encountered with the original DIVA. DIVA can only handle phylogenetic trees that are completely resolved and it is assumed that the tree topology is known without error. In addition, there is uncertainty with the ancestral area optimisation because the analysis often results in a number of optimal distributions which are similarly parsimonious (Ali, Yu, Pfosser, & Wetschnig, 2012; Nylander, Olsson, Alström, & Sanmartín, 2008).

Therefore, two new programs have been developed that take care of these uncertainties. These are Bayesian-DIVA (Nylander et al. 2008) and Statistical-DIVA (S-DIVA, Yu, Harris, and He 2010a). In Bayes-DIVA, a number of Bayesian inferences are optimised before a DIVA analysis is run. In S-DIVA, the S-DIVA value, which combines the phylogenetic and geographical uncertainties, gives statistical support to the result of the analysis (Yu et al., 2010b). Also, for each node the frequency an ancestral range occurs over all reconstructions are averaged (Ojeda, Novillo, Ojeda, & Roig-Juñent, 2013).

BRASSICACEAE

The Brassicaceae is one of the largest plant families and is considered to comprise around 3660 species distributed over 320 genera (Al-Shehbaz, 2012; Koch, Karl, German, & Al-Shehbaz, 2012). It is also an important plant family containing crop species, weeds and model organisms (Beilstein, Nagalingum, Clements, Manchester, & Mathews, 2010; Couvreur et al., 2010; Koch & Marhold, 2012) which makes it interesting to research the origin of the family.

From as early as the beginning of the 20th century, the origin of the Brassicaceae has been inferred. In the first theory from these early days, the family was thought to have originated in the New World based on the distribution of either the basal placement of the tribe Theylopodieae and later the Hesperideae. With the coming of molecular studies this view shifted and an Old World origin, based on the genus Aetheionema now considered to be sister to all other Brassicaceae, is proposed (Al-Shehbaz, Beilstein, & Kellogg, 2006; Franzke, Lysak, Al-Shehbaz, Koch, & Mummenhoff, 2011; Hauser & Crovello, 1982; Koch, Al-Shehbaz, & Mummenhoff, 2003).
In recent years the phylogeny of the Brassicaceae has thoroughly been revised with the help of DNA-based research (Bailey et al., 2006; Beilstein et al., 2010; Couvreur et al., 2010; Warwick, Mummenhoff, Sauder, Koch, & Al-Shehbaz, 2010). The most recent phylogenetic tree covering the whole family is from Couvreur et al. (2010). This inference is based on ITS, *chs, adh, matK, trnL-F, ndhF, rbcL* and *nad4*. All genomes (nuclear, chloroplast and mitochondrial) are thus covered. Currently, phylogenetic trees are inferred not for the family as a whole but on the tribal level (Jordon-Thaden, Al-Shehbaz, & Koch, 2013; Karl & Koch, 2013). This approach is chosen because the relationships on this level are not yet adequately resolved causing a considerable amount of phylogenetic uncertainty resulting in a lowly supported backbone (Koch, 2013).
At the moment, a number of overarching lineages is recognised within the Brassicaceae (Couvreur et al., 2010; Koch & German, 2013). These major lineages (Lineage I, Lineage II, Extended Lineage II, and Lineage III) each consists of a number of tribes and together are considered to compose the structure for the entire family (Franzke et al., 2011).
Unfortunately, as researchers agree, the Brassicaceae fossil record is rather poor (Franzke, German, Al-Shehbaz, & Mummenhoff, 2009; Franzke et al., 2011). There are fossilised pollen from the early Middle Miocene, Upper Miocene and Latest Cretaceous available, though the last one is doubtful. Also, there is a fossil fruit from the Oligocene, but it remains unclear whether it actually belongs to the family (Franzke et al., 2011).

A lot of historical biogeographical research is being done in recent years. In almost every phylogenetic study, the ancestral areas are inferred alongside the ages of clades.
In this thesis, we chose to perform a historical biogeographical analysis on the entire Brassicaceae because Schranz et al. (2012) proposed the hypothesis that the origin of a family as vastly radiated as the Brassicaceae can be found there were at the moment the largest diversity is (Schranz et al. 2012). In case of the Brassicaceae that would coincide with the area of diversity of the genus *Aethionema* (Schranz, Mohammadin, & Edger, 2012).

Based on the previous, the following question with associated hypothesis is posed:

Where lies the centre of origin of the Brassicaceae?

- The origin can be found were at the moment the largest diversity in species is – Iranoturanian region (Al-Shehbaz et al., 2006)

2. MATERIAL AND METHODS

In this chapter, I give an overview of the methods I used to carry out a historical biogeographical analysis. First, I will explain how I performed the phylogenetic inference, next how I came to the geographical distribution, and last a description of the biogeographical analysis itself. Since this is a MSc thesis project with a lot of trial and error, not all methods I tried were used in the final analysis. These steps are clustered under the heading 'Additional Methods'.

2.1 PHYLOGENETIC ANALYSIS

First, the sequence data of Couvreur et al. (2010) was requested for the phylogenetic inference. This data set was chosen as a starting point because it is the most recent and most complete genus level study to this point. This study includes the nuclear genes ITS, *chs* and *adh,* the chloroplast genes *ndhF*, *rbcL*, *matK*, the *trnL* intron and trnLF spacer and the mitochondrial gene *nad4*. The alignment from Couvreur et al. (2010) was kindly provided by Dr. Couvreur.

To make the taxonomic coverage of this Brassicaceae dataset at the genus level as complete as possible, relevant sequences were sought in GenBank. This included sequences of the above named genes from genera that had not yet been developed, as well as sequences of new genera. In addition, sequences from relevant outgroups from the Cleomaceae, Capparidaceae and genera from other families within the Brassicales were added.

The gathered sequence data are thus species-level data, and it is here assumed that the separate sequences for each species over the eight genes are of one and the same individual. These separate sequences are thus combined to form one concatenated sequence for that one species.

No additional sequences were generated since it is not the intention of this project to update the most recent phylogenetic tree of Couvreur et al. (2010). It is merely used as a starting point for this project.

The sequence data for each gene was gathered in Mesquite (Maddison & Maddison, 2010) and Geneious (Drummond et al., 2011) was used to put these in a rough phylogenetic order with the help of Robin van Velzen. This was done by loading the gene alignments in Geneious, letting it form a 'quick and dirty' phylogenetic tree, and having it order the alignment to match that order. These gene-sets were then aligned using the auto and G-INS-i settings of MAFFT online (http://mafft.cbrc.jp/alignment/software/), giving a total of 16 gene alignments. I refrained from manual adjustments because I wanted to keep the alignment objective and not subjected to subjective interpretation.

Then, the separate alignments were assembled with SequenceMatrix (Vaidya, 2011) into a matrix for these two MAFFT options separately.

Also, because the taxonomic sampling (number of taxa per gene-set) differs quite substantially (from 6% for *adh* to 82% for ITS, Appendix B), additional phylogenetic analyses including only those genes with a completeness of more than 30% was carried out. These are *nad4, ITS, ndhF* and *trnLF*. Table 1 gives an overview of all the matrices formed by the alignment – covering options.

TABLE 1. OVERVIEW OF MATRICES, THE GENES THESE CONSIST OF, AND THE ALIGNMENT METHOD USED.

	Genes	Alignment Method MAFFT
MAFFT_4	ITS, ndhF, trnLF and nad4	Auto
MAFFT_4_slow	ITS, ndhF, trnLF and nad4	Auto
MAFFT_8	ITS, *chs* and *adh,* ndhF, *rbcL*, matK, trnLF and nad4	G-INS-i
MAFFT_8_slow	ITS, *chs* and *adh,* ndhF, *rbcL*, matK, trnLF and nad4	G-INS-i

A number of applications in the phylogenetic inference program TNT (Goloboff et al. 2008) was used on these four matrices. For these analyses, the matrices were exported to TNT format from Mesquite.
A preliminary Traditional Search (500 replications) was run to get insight in the tree topology resulting from the matrix. Then, an analysis with the Sectorial and Ratchet settings combined, using the default settings (200 replications), was run. In a Sectorial tree search, a subset of the data is created based on large clades from a phylogenetic inference. These clades (or subsets) are then optimised and recombined (Goloboff, 1999). With Ratchet, a number of characters from an initial phylogenetic tree is selected and their weight (representing the influence) is altered (Nixon, 1999).
The resulting phylogenetic trees were made suitable for further use by importing them in Mesquite alongside their original matrix and exporting them as a Nexus tree file.

RAxML-HPC2 on XCEDE through the Cypres Science Gateway (Miller, Pfeiffer, & Schwartz, 2010; Stamatakis, 2006) was used for Maximum Likelihood analysis for all separate genes and the four matrixes mentioned above. For this analysis, the matrices were exported to Phyllip format from Mesquite.

SplitsTree4 (Huson & Bryant, 2005) was used to check the ambiguity of the phylogenetic information within the matrices. I did a 'quick and dirty' analysis with the default Uncorrected P-value settings and an analysis with the GTR+Gamma model. In this analysis, the Base Frequencies were set to Empirical frequencies, meaning frequencies of the bases as actually found in the data. The Site Rate Variation was set to Gamma, with an alpha parameter of 0.5 and proportion invariable sites on 0.5.

Additional Methods

I am aware that the MAFFT tool used for alignment of the matrixes does not take the phylogenetic information into account and that this can cause biases in the subsequent analyses (Anisimova, Cannarozzi, & Liberles, 2010; Löytynoja & Goldman, 2008). An example of an aligning program that does use the phylogenetic history is PRANK (Löytynoja & Goldman, 2005). PRANK was used as one of the programs for aligning of the data, but unfortunately the resulting matrix was so pulled apart that no coherence between the characters was left. I therefore decided not to continue with these matrices in the further analysis.

It is desirable to check how 'well' the matrices have been aligned after using an aligning tool as MAFFT or PRANK. Since I wanted to refrain from any subjective interpretation of the data, I used the program GBlocks (Talavera and Castersana 2007). GBlocks finds the poorly aligned positions in the matrix and indicates where they should be omitted (Castresana, 2000). However, I decided not to use this method because too many areas (that looked adequate when checked by eye) where indicated as weak and would have to be removed from the alignment.

A lot of phylogenetic information can be derived from indels, but how to handle them is subject to an extensive debate (Simmons, Müller, & Norton, 2007; Simmons & Ochoterena, 2000). One can either identify and/or interpret indels by eye, or use one of the available programs. Since I wanted to keep the matrix as objective as possible (as with the manual aligning), I chose to use the program SeqState for this analysis (Müller, 2005). Unfortunately it turned out that SeqState over-interpreted the information in the matrixes and assigned substantially more indel characters than could be justified when checked by eye. Therefore I decided not to continue with these analyses and thus did not take into account possible indel information.

In addition to the TNT and RAxML analysis, I used Bayesian methods to infer phylogenetic trees. I used both MrBayes 3.2.1 on XSEDE (Ronquist et al., 2012) and a local version of MrBayes for all above named datasets as well as all sixteen separate gene alignments. Unfortunately, due to time constrains it turned out not to be feasible to successfully carry out these analyses. In Appendix B. Exemplary script used in MrBayesan exemplary script designed for these analyses is provided.

2.2 Geographical Distribution

For the historical biogeographical analysis, it is important to establish the geographic distribution of the terminals in the phylogenetic tree as comprehensive as possible. Instead of using a species level distribution (since the matrices consist of species level sequence data), the geographic dispersal information of the Brassicaceae on genus level was gathered. A further explanation for this decision is provided in the 'Geographical Distribution' section of the Discussion and Conclusion chapter.
To do this, the Global Biodiversity Information Facility (GBIF, The Global Biodiversity Information Facility, 2013) was primarily used. On this website, amongst other things, specimen data from all online published databases from herbaria is collected. The distribution data of all relevant genera (those that are represented in the phylogenetic tree) was downloaded and checked the data in Google Earth (Google, 2014). Suspicious outliers were traced and removed from the dataset when necessary. As can be seen in Appendix A (column 'Percentage'), the percentage of records in GBIF that is geographanced per genus was checked. When this percentage was lower than 20, place marks to unappointed places were manually added in Google Earth based on the data available in GBIF.

With these distribution maps per genus, relevant geographic areas were formed to accurately cover the total distribution of the entire phylogenetic tree. Hereby the example of Karl & Koch was followed, who did the same in their Ancestral Area Reconstruction of the tribe *Arabideae* (Karl & Koch, 2013).

Karl and Koch based their distribution on the global ecoregion map developed by Olson et al. (Figure 4; Olson et al., 2001). Based on geographic distribution data of a large set of plant and animal species, Olson et al. developed a total of 867 ecoregions. They intended for these ecoregions to be used as a 'framework for comparisons among units and the identification of representative habitats and species assemblages' (Olson et al., 2001). The ecoregions thus represent the natural boarders of the species habitats, instead of political boundaries (Kier et al., 2005) and are widely used in conservation studies (Hoekstra, Boucher, Ricketts, & Roberts, 2004; Olson et al., 2001).

FIGURE 4. THE 867 ECOREGIONS OF OLSON ET AL. (2001).

Using these ecoregions, the areas as proposed by Karl and Koch (Figure 5) and based on overlapping distributions in the data of the genera gathered in Google Earth, a total of 14 geographical areas was designed (Figure 6). These include all areas as designed by Karl and Koch, and now include Sub-Saharan Africa and the Australia/New Zealand combination as well.

Then, each genus was assigned to at least one of these geographical areas. In the case of a widespread genus as for example *Alyssum* L. it was necessary to include a number of areas to ensure the entire distribution was taken into account.
For each of the four matrices, the appropriate areas were assigned to all taxa in Excel. These data were then transported to four separate .CSV files; one four each matrix, making it suitable for use in RASP.

FIGURE 5. GEOGRAPHICAL AREAS USED BY KARL & KOCH (2013). A: NORTH & CENTRAL AMERICA, B: SOUTH AMERICA, C: IBERIAN PENINSULA & NW AFRICA, D: CENTRAL MEDITERRANEAN & SOUTHERN BALKANS, E: EUROPE, F: WESTERN AND CENTRAL ANATOLIA & EASTERN MEDITERRANEAN, G: CAUCASUS, EASTERN ANATOLIA & IRANIAN MOUNTAIN RANGES, H: HIGH MOUNTAINS OF THE ARABIAN PENINSULA AND EASTERN AFRICA, I: CENTRAL ASIAN MOUNTAIN RANGES, J: EASTERN HIMALYA AND TIBET-CHINEASE MOUNTAINS, K: EASTERN ASIA, L: SIBERIA & RUSSIAN FAR EAST.

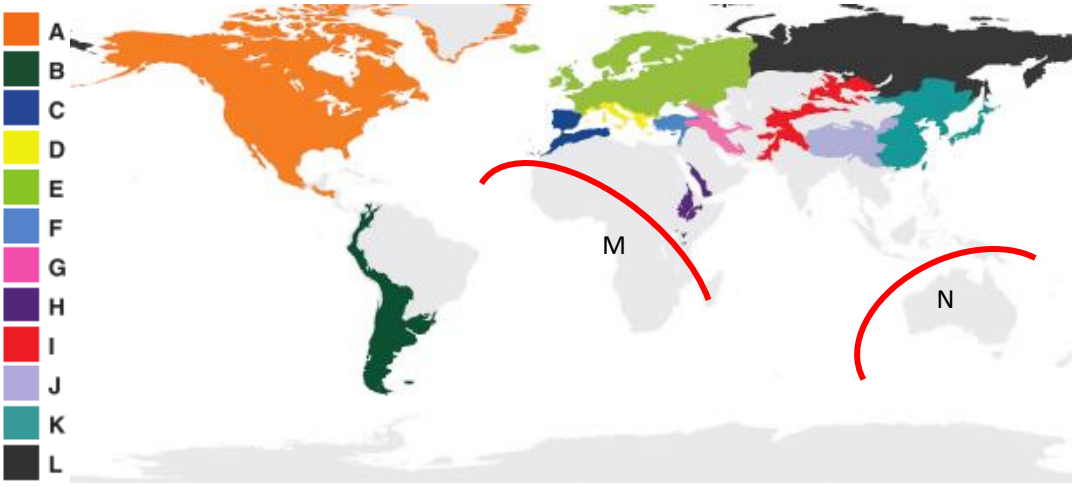


FIGURE 6. PROPOSED GEOGRAPHICAL AREAS (AFTER KARL & KOCH 2013). A: NORTH & CENTRAL AMERICA, B: SOUTH AMERICA, C: IBERIAN PENINSULA & NW AFRICA, D: CENTRAL MEDITERRANEAN & SOUTHERN BALKANS, E: EUROPE, F: WESTERN AND CENTRAL ANATOLIA & EASTERN MEDITERRANEAN, G: CAUCASUS, EASTERN ANATOLIA & IRANIAN MOUNTAIN RANGES, H: HIGH MOUNTAINS OF THE ARABIAN PENINSULA AND EASTERN AFRICA, I: CENTRAL ASIAN MOUNTAIN RANGES, J: EASTERN HIMALYA AND TIBET-CHINEASE MOUNTAINS, K: EASTERN ASIA, L: SIBERIA & RUSSIAN FAR EAST, M: SOUTHERN AFRICA, N: AUSTRALIA AND NEW ZEALAND.

2.3 BIOGEOGRAPHICAL ANALYSES

For each of the four matrixes described above, a historical biogeographical analysis using the Bayesian Binary Method (BBM) implementation in RASP (Nylander et al., 2008; Yu, Harris, & He, 2012) was carried out.

In RASP, a number of background phylogenetic trees has to be provided to form a phylogenetic range. For this, 100 parsimonious TNT phylogenetic trees obtained in the Traditional search were used. Then, a 'Condensed Tree' has to be provided, for which the most parsimonious TNT phylogenetic tree obtained with the Sectorial and Ratchet setting as described above was used. This is the starting tree on which the phylogenetic uncertainty from the background trees is projected. Hereafter, a distribution file is imported wherein the area ranges of all terminals are defined.

In the BMM menu, a number of settings can be chosen by the operator. For example; all settings of a traditional Bayesian MCMC, the root distribution and maximum number of areas per node. Table 2 gives an overview of all combinations of settings used in BBM for each matrix.

From the literature it is know that widespread species can be a possible source of bias in a biogeographic analysis (Karl & Koch, 2013; Morrone & Crisci, 1995). I therefore pruned the results of the TNT analysis and excluded those taxa that were assigned more than six areas. These matrices are named: MAFFT_4_pruned, MAFFT_8_pruned, and MAFFT_8_slow_pruned. This thus gives a total of seven matrices analysed with BBM.

TABLE 2. OVERVIEW OF COMBINATION OF SETTINGS IN RASP.

Number of generations	Chains	Temperature	Root distribution	Number of areas
100.000	4	0.1	Outgroup	4
			Outgroup	6
			Null	4
			Null	6
			Wide	4
			Wide	6

3. Results

3.1 Phylogenetic Analysis

RAxML Bootstrap

In Appendix C.1 through C.4 the results of the RAxML Bootstrap analysis are displayed. While the backbone of all phylogenetic inferences forms a major polytomy and the overall bootstrap support is low, Lineage I (red), Lineage II (purple), Extended Lineage II (green) and Lineage III (blue) can be observed in all four phylogenetic trees. Table 3 gives an overview of the support values of these lineages per matrix. Within these lineages, most clades are fairly well resolved and are fairly well supported. However, the higher we go in the phylogenetic tree, the lower the support becomes. The clades forming Lineage II essentially are part of one major polytomy.

A number of taxa are now retrieved in a different lineage than was the case in Couvreur et al. (2010). These are: *Biscutella didyma, Lunaria rediviva, Magadenia pygmaea, Orychophragmus violaceus, Fourraea alpina, Hornungia petraea, Arabis drummondii, A. fendleri, A. lyalli, A. parishii, A. lignifera, Lunaria annua* (partual), and *Conringia planisiliqua*. Table 4 gives an overview of these taxa per matrix, the lineage they were placed according to Couvreur et al. (2010), the lineage they are placed in the RAxML Bootstrap analysis, and their bootstrap support.

TNT

Appendix C.6 through C.9 show the results of the TNT parsimony analysis for all four matrices. As can be seen from the figures, the major lineages are relatively well retrieved. Only Extended Lineage II (green) is no longer monophyletic except for the MAFFT_8_slow matrix.

Again, a number of taxa have crossed lineage boarders and are retrieved in a different lineage as is the case with the RAxML Bootstrap analysis. These are: *Orychophragmus violaceus, Biscutella didyma, Lunaria rediviva, Magadenia pygmaea, Fourraea alpina, Enarthrocarpus clavatus, Isatis brevipes, Dontostemon dentatus, Dontostemon glandulosus, Hornungia petraea, Arabis glabra, A. lyalii, A. parishii, A. drummondii, A. fendleri, A. lignifera, A. pauciflora, Zuvanda crenulata, Conringia planisiliqua, Iberis procumbens*, and *Sameraria armena*. Table 4 gives an overview of these taxa per matrix, the lineage they were placed according to Couvreur et al. (2010) and the lineage they are placed in the TNT parsimony analysis.

Neighbour Network

Appendix C.10 through C. 13 show the results of the Neighbour Network analysis performed in SplitsTree under the GTR+Γ model. As can be seen from the figures, the alignments resulted in rather complicated networks. In all cases, a large polytomy is found on one side of the network. On the other side, more network-like behaviour can be seen. The results of the MAFFT_4 and MAFFT_8, and the MAFFT_4_slow and MAFFT_8_slow matrices are quite comparable with each other.
Unfortunately, it was not possible to give Lineage colour indications for these results. Therefore, nothing can be said about these Lineages in this analysis.

TABLE 3. OVERVIEW OF LINEAGE NODE SUPPORT PER ANALYSIS. NUMBERS INDICATE BOOTSTRAP SUPPORT VALUE, + SIGN MEANS CLADE IS PRESENT, - MEANS CLADE IS ABSENT.

		Lineage I	Lineage II	Extended Lineage II	Lineage III
RAxML	**MAFFT_4**	53	3	0	54
	MAFFT_4_slow	14	2	1	20
	MAFFT_8	7	0	0	3
	MAFFT_8_slow	10	2	0	1
TNT	**MAFFT_4**	+	+	-	+
	MAFFT_4_slow	+	+	-	+
	MAFFT_8	+	+	-	+
	MAFFT_8_slow	+	+	+	+

TABLE 4. OVERVIEW OF TAXA THAT CHANGED LINEAGES AND THEIR BOOTSTRAP SUPPORT.

	Taxon	Couvreur	RAxML	Bootstrap support value	TNT	Neighbour Network
MAFFT_4	*Orychophragmus violaceus*	II	Extended II	4	Extended II	
	Fourraea alpina	Extended II	II	0	II	
	Biscutella didyma				Extended II	
	Lunaria rediviva				Extended II	
	Magadenia pygmaea				Extended II	
MAFFT_4_slow	*Biscutella didyma*		Extended II	35		
	Lunaria rediviva		Extended II	23		
	Magadenia pygmaea		Extended II	51		
	Orychophragmus violaceus	II	Extended II	0		
	Idahoa scapigera	Extended II				
	Cochlearia acaulis	Extended II				
	Enarthrocarpus vlavatus	II			Extended II	
	Isatis brevipes	II			Extended II	
	Fourraea alpina	Extended II			II	
	Orychophragmus violaceus	II			Extended II	
MAFFT_8	*Idahoa scapigera*	Extended II				
	Dontostemon dentatus, D. glandulosus	III			Extended II	
	Orychophragmus violaceus	II	Extended II	7	Extended II	
	Biscutella didyma				Extended II	
	Lunaria rediviva				Extended II	
	Magadenia pygmaea				Extended II	
	Hornungia petraea	I	III	14	III	
	Arabis glabra, A. lyalii, A. parishii, A. drummondii, A. fendleri, A. lignifera	Extended II	I	32 (A_glabra 51)	I	
	Isatis brevipes	II			Extended II	
	Zuvanda crenulata	Extended II			II	
	Conringia planisiliqua	Extended II			II	
	Arabis pauciflora	Extended II			II	
	Iberis procumbens	Extended II			II	

MAFFT_8_slow	*Lunaria annua (partual)*		Extended II	18	
	Noccaea caerulescens	Extended II			
	Dontostemon dentatus, D. glandulosus	III			Extended II
	Orychophragmus violaceus	II	Extended II	6	Extended II
	Sameraria armena	II			Extended II
	Iberis procumbens	Extended II			II
	Isatis brevipes	II			Extended II
	Hornungia petraea	I			III
	Arabis glabra, A. lyalii, A. parishii, A. drummondii, A. fendleri, A. lignifera	Extended II			I
	Conringia planisiliqua	Extended II	II	36	

3.2 Geographical Distribution

Figure 6 in the 'Geographical Distribution' section of the Materials and Methods chapter shows all the areas I delimited. Table 5 gives an overview of these areas. An overview of the area distribution of per genus is given in Appendix A (column 'Area Proposed').

TABLE 5. OVERVIEW OF AREAS AND THEIR GEOGRAPHICAL RANGE.

Area	Geographic range
A	Northern America - North
B	Southern America - West
C	Iberian Peninsula & North Western Africa
D	Central Mediterranean & Southern Balkan
E	Europe
F	Western and Central Anatolia & Levantine coast
G	Caucasus, Eastern Anatolia & Iranian mountain ranges
H	High mountains of the Arabian Peninsula and Eastern Africa
I	Central Asian mountain ranges
J	Eastern Himalaya and Tibet-Chinese mountains
K	Eastern Asia
L	Siberia & Russian Far East
M	Southern Africa
N	Australia and New Zealand

3.3 Biogeographical Analysis

Because of the large amount of analysis performed in this part of the project, it is not possible to show all the results. The following table (Table 6) gives an overview of the resulting ancestral area for each of the matrices over all types of analyses. Because of time constrains it was not possible to perform a Bayesian DIVA analysis for all matrix-analysis combination.

As can be seen in the table, quite often a * (indicating RASP was not able to determine the ancestral area) is part of the outcome. When possible, the most likely alternative area has additionally been indicated. Next, the combination of areas 'Iberian Peninsula & North Western Africa', 'Central Mediterranean & Southern Balkan', 'Western and Central Anatolia & Levantine coast', and 'Caucasus, Eastern Anatolia & Iranian mountain ranges' (regions C, D, F, and G) is quite often the resulting ancestral area.
Notable is that the ancestral range often consists of a grouping quite a large number of areas.

TABLE 6. OVERVIEW OF ANCESTRAL AREAS PER MATRIX AND LINEAGE. OVERALL IS ANCESTRAL NODE INCLUDING OUTGROUPS; AETHIONEMA IS ANCESTRAL NODE OF ALL BRASSICACEAE; CORE IS ANCESTRAL NODE OF CORE BRASSICACEAE. * NO OUTCOME; -: NOT POSSIBLE TO DETERMINE; LETTERS: PROPOSED AREAS.

Matrix	Lineage	Outgroup		Null		Wide	
		6	4	6	4	6	4
MAFFT_4	Overall	ABCEMN	*/ABMN	-	-	-	-
	Aethionema	*/CDFG	*/CDFG	*	*/CDFG	*/CDFGI	*/CDFG
	Core	*/GI/A	*/A/GI	*/GI/A	*/GI/A	*/GI/A	*/GI/A
MAFFT_4_slow	Overall	ABCEMN	*	*	*/ABEM/ABEF	*/ABCEMN	*/AEMN
	Aethionema	*/ABCDFM/ABCDFG/ABCDIM	*/ABEN/ACDG	*/ABCDIM	*/HIJM	*/ABCDIM	*/CDFG
	Core	*/A	*/ABHM	*/ABCIJM	*/ABHM	*/ABCDEN	*/A
MAFFT_8_pruned	Overall	CDN	-	-	*	*/ABCEMN	
	Aethionema	*/CDFG	*/CDFG	-	*/CDFG	*	
	Core	*/N	*/ABEN	-	*/CDFG	*	
MAFFT_8_slow_pruned	Overall	ABCEMN	-	*/ABCEMN	*		
	Aethionema	*	*/ABEN	*	*/CDFG		
	Core	*/N	*/ABEN	*/N	*/N		

4. Discussion and Conclusion

4.1 Phylogenetic Analysis

As indicated in the Results chapter, the bootstrap support values for the four Lineages in all four phylogenetic trees are remarably low. The support values for (almost all) the other nodes are similarly low. This is quite a problem, since the bootstrap support values is a direct reflection of the quality of the underlying alignment. An explanation has thus to be sought there. When looking at the alignment, what strikes is that it is increadibly gappy, in the sence of missing sequences. It appears that one of the reasons for these empty areas have resulted from the large number of species. Because I decided to add all the data available in GenBank, sometimes only one out of a possible eight markers is sampled for a terminal. Also, often a number of species is added per genus which then all comprise one or two markers. This results in a rather pulled apart alignment with not a lot of overlap in species and genera between the separate gene regions. Second, as indicated in the Materials and Methods chapter, there are large differences in taxon sampling completeness. To bypass this problem, I made a separate alignment consisting of only markers with a taxonomic sampling of more than 30%. When we compare the Lineage bootstrap support of the RAxML analysis for MAFFT_8 to MAFFT_4, it can be seen that the MAFFT_4 phylogenetic tree overall does seem better supported. However, the support remains low. In hindsight, it would have been better to eliminate ITS from the matrix along with *adh* and *chl*. These three markers have the highest (ITS, 82%) and lowest (*adh* and chl, 6 and 8%) taxonomic sampling. The others have a sampling raging from 26 through 40%, which are in the same range. It would thus have been better to combine these five genes, because then the alignment would have been far less gappy.
The fact that so many genera seem to be non-monophyletic (i.e. either poly- or paraphyletic) while they 'should' be (Al-Shehbaz, 2012), could be an artefact of the large dissimilarities in gene sampling for each of the species within the genus. It could be that terminals with a similar gene-sampling group together, forcing terminals belonging to the same genus apart. This could be checked by investigating the influence of each of the genes by forming a Phylogenetic Super Network. In such a network (for example formed in SplitsTree) separate gene trees are combined in a network form. It will then be known which taxon corresponds to which line in the network, and the underlying effect of the (lack of) genes can be traced.

What is also of interest for the suitability for phylogenetic analysis of a matrix is how well it is aligned. Usually an alignment tool (i.a MAFFT and PRANK) is used for alignment, which is then checked by eye. At that point, manual adjustments can be made. I decided not to do this last step because I wanted to compare a number of alignment tools, with which manual adjustment would interfere. As explained in the Materials and Methods section, I used two settings in MAFFT and the program PRANK for this comparison. Unfortunately, as explained above, PRANK resulted in such pulled apart alignments that they were no longer usable for further analysis. The default settings of MAFFT have resulted in a slightly better overall bootstrape node support than the G-INS-i settings of MAFFT. This second setting gives a more thorough analysis and is thus expected to give a better resolved matrix. At first sight it is thus surprising that this setting has resulted in phylogenetic trees with lower bootstrap support. However, if we take in mind that the matrix was already quite gappy and the G-INS-i settings have pulled it apart even further, this result is well explainable.
Since both these alignment settings have resulted in badly aligned regions, which in turn affect the phylogenetic inference, I think it would have been better to choose one alignment tool and afterwards optimise the alignment by hand. This is quite common practise in phylogenetics. Indeed, the alignment

will not be completely objective, but as an operator we can take the decision to make a certain change or leave it as specified by MAFFT.

Another cause for problems with this alignment (and resulting in low bootstrap support values) are the indels. As described above, there are a lot of ways to handle them and I made the choice not to include any indel information. Couvreur et al (2010) did include indel information in their analysis, and it seems that this additional step has helped quite a lot in improving the overall support of the phylogenetic tree. Defining indel positions adds a lot of structure to the alignment, which makes it more easy for a phylogenetic analysis to be performed. Since the program now used to define the indel positions did not give sufficient outcome, the alternative would be to define these by hand. In this project, it was however not possible to carry this out.

The support of the phylogenetic trees is thus quite poor and no conclusions can be drawn from these results. These trees are also not suitable for a historical biogeographic analysis. What we can do is compare the four phylogenetic trees with each other and look at their topology.
At first sight there are no large incongruencies between the RAxML Bootstrap phylogenetic trees. All four major lineages are retrieved in each of these phylogenetic inferences. Within these lineages however, there are some differences. The first thing that stands out is the branching order of the lineages. According to Couvreur et al. (2010) and Franzke et al. (2011) the order is: Lineage I (red), Lineage III (blue), Extended Lineage II (green), Lineage II (purple). However, this does not seem to be the case for matrix MAFFT_8 and MAFFT_8_slow. Here the order of lineages is: Lineage III (blue), Lineage 1 (red), Extended Lineage II (green), Lineage II (purple). However, when we zoom in on the node supporting the split between the clades forming Lineage I, III, and (Extended) Lineage II, we see that the support bootstrap is of no significance. These three clades thus form a major polytomy and nothing can be said about their order.
To check the topology of the rest of the phylogenetic tree, I coloured the remaining terminals in colours corresponding to their supposed Lineage based on the tribal composition of Al-Shehbaz (Al-Shehbaz, 2012). As can be seen from the figure in Appendix C.5, almost all newly coloured terminals are found in their corresponding lineage. As a result, almost all clades are now fully coloured.
One of the reasons quite a number of clades is now fully coloured is synonymity. For example, the genus Desideria has become a synonym for the genus Solms-Laubachia. In one case it turned out that three genera splitting up a clade were made synonymous to the fourth genus in that clade (Stubendorffia = Winklera = Cyphocardamum = Lithodraba).
In general, the tree topologies is thus quite acceptable, but there are some causes for concern. One of these is the clade of Outgroup taxa (Moringa ovalifolia, Carica papaya, Reseda lutea) separated from the rest of the Outgroup by the Aethionemeae, forcing them just inside the Ingroup. The bootstrap support for this small Outgroup clade in itself is rather low, but the support for the Aethionema clade is quite acceptable. A possible explanation for this problem is given in the 'General Comments' section of this chapter. Another troubling terminal is the Eutrema_himalaicum right inside the Outgroup while it is supposed to group inside Extended Lineage II.

The results of the Neighbour Network analysis in SplitsTree correspond to the results of the RAxML analysis. As indicated in the 'Neighbour Network' section of the Results chapter, all four matrices display quite a large polytomy on one side of the network. This part corresponds with the large polytomy that is formed by Lineage II and Extended Lineage II in the RAxML phylogenetic trees. Since it was not possible to colour terminals and nodes in the Networks according to their Lineage placement, no further comparisons can be made.

Because the results of the RAxML Maximum Likelyhood analysis are not suitable for a historical biogeographical analysis, I intended to use the results of a Bayesian Inference instead. Unfortunately, as explained in the 'Additional Methods' section of the Materials and Methods chapter, it was not possible to complete these analyses. It seems that MrBayes, used for the Bayesian Inference, had quite a lot of trouble with the amount of missing data. MrBayes uses a lot of parameters, and these have to be 'fed' with information in the form of characters. When there are a lot of gaps, these parameters cannot function optimally and the analysis will be terminated. I therefore run the TNT analyses, as an alternative for the Bayesian Inference.
The results of the TNT parsimony analysis are quite similar to those of the RAxML Bootstrap although a larger number of terminals were not found in the 'right' Lineage. This could be a result of the limited number of replications for the Sectorial - Ratchet combined analysis. As specified in the Materials and Methods chapter, I let this analysis run for 200 replications because in an initial test-run the analysis seemed to reach a platform: it took too much time for the analysis to find shorter trees under these model settings. Since the phylogenetic tree obtained was considerably shorter than those retrieved under the Traditional search, I decided 200 replications were sufficient. However, it can be more parsimonious trees could have been found when the analysis had run for a longer period of time.
The results of the TNT parsimony analysis are thus also not optimal for the historical biogeographical analysis. However, I decided to continue with the analysis using the results of the TNT analysis to get a feel for the data, the type of analysis and the results.

All in all, the combination of normal MAFFT settings and including only the four best sampled genes results in a better tree topology, as well as better lineage boostrap support. Even though this support is still unacceptably low, I feel this matrix shows the most potential.

4.2 Geographical Distribution

For the historical biogeographical analysis, it is important to establish the geographic distribution of all terminals in the phylogenetic tree. In the phylogenetic trees I have inferred in this project, the terminals of the trees represent species. Species level distribution data and corresponding ranges would thus be logical. However, the largest matrices (MAFFT_8 and MAFFT_8_slow) comprise around 570 taxa while the Brassicaceae are considered to comprise around 3660 species (Al-Shehbaz, 2012; Koch et al., 2012). Species level data would thus not give an accurate representation of the true distribution of the Brassicaceae. On the other hand, I do have almost half of all available genera covered in the matrices (391 out of a possible 865 genera). I therefore decided it would be better feasible to use the genus-level distribution data. The effects these distributions might have on the biogeographical analysis are discussed in the 'General Comments' section of this chapter.

To limit subjectivity in assigning areas to the taxa, I took a number of precautions. First, I only used accessions from GBIF with reliable coordinate data as indicated in the 'Geographical Distribution' section of the Material and Methods (exceptions being those genera where no data was available at all). When no specific coordinate data was available, I only used those records with reliable data for, for example, a city or province. In some cases, next to specimens also records specified as 'Human Observation' or 'Living Specimen' were available. I decided not to use these records, because I find a human observation of over 50 years ago without a corresponding specimen unreliable. Also, a living specimen is often part of a botanical garden or collection. In that case, the coordinates of the growing location would be given, and it is questionable whether this location is part of the actual distribution area. For the same reason I

refrained from getting the distribution data from flora. An additional reason for this is that often a quite general distribution area is provided.
In some cases it was not possible to apply this method. Since I intended to base this part of the project solely on data available in GBIF (The Global Biodiversity Information Facility, 2013), I had a problem in the case there were no (acceptable) records available for a genus. In those incidents, I did use information available in flora and other publications.

Even with the above named precautions, I still feel that the assigning of areas could be done more objectively. Forming global areas and letting the operator interpret whether a location falls exactly inside or just on the other side of the boarder is still quite subjective. It would be better to describe the proposed areas by a range of coordinates. For a taxon it can then be determined whether it falls within a particular area based on coordinate data.

4.3 Biogeographical Analysis

As described above, the MAFFT_4 matrix shows the most potential and is thus considered to be the 'winning' alignment. Table 6 shows that the most likely ancestral region of all Brassicaceae (including *Aethionema*) is a combination of the regions 'Iberian Peninsula & North Western Africa', 'Central Mediterranean & Southern Balkan', 'Western and Central Anatolia & Levantine coast', and 'Caucasus, Eastern Anatolia & Iranian mountain ranges' (regions C, D, F, and G). The Iranoturatian region is indeed part of this region, making the hypothesis that the Brassicaceae originated in the Iranoturanian region highly plausible.
If the origin of the family indeed can be found there, it is to be expected that the ancestral region of all Brassicaceae excluding *Aethionema* can also be found around the same area. As can be seen in the same table, this is indeed the case. Here, a combination of the areas 'Caucasus, Eastern Anatolia & Iranian mountain ranges' and 'Central Asian mountain ranges' is in almost all cases the ancestral region. This region corresponds to an area near to the Iranoturanian region. However, in all these cases the area Northern America - North (area A) is almost as likely to be an alternative ancestral area. This outcome is easily explainable, since two of the latest branches joining the rest of the core Brassiceae are *Idahoa scapigera*, and *Iberisprocumbens*, which occur solely in that region. This placement in the RASP tree topology is not in agreement with that in the RAxML Bootstrap phylogenetic tree. There, these taxa clusters within Extended Lineage II, where it belongs according to the tribal division by Al-Shehbaz (Al-Shehbaz, 2012). The outcomes of the ancestral area analysis are clearly heavily influenced by the incorrect placement of these taxa.

In conclusion; based on the outcome of the BBM analysis, the hypothesis that the area of origin of the Brassicaceae is the Iranoturanian region seems accepted. However, I feel there are so many issues with the underlying alignment, that both the phylogenetic inference and subsequently also the historical biogeographical analysis cannot fully be trusted. We are on the right track, but the results obtained in this project are not enough to confidently accept the hypothesis.

Unfortunately, I encountered a number of problems with RASP. Based on the ecoregions developed by Olson et al. (Olson et al., 2001), I had formed the following areas (Figure 7). However, it is not possible in RASP to assign that many areas; the maximum number is fifteen (ranging from A through O; Yu, Harris, & He, 2010a; Yu et al., 2012). Since I find this maximum number rather arbitrary, I contacted the authors of the programme with the question whether this number could be increased. Unfortunately, I did not receive any response. I therefore followed the advice of Pulcherie Bissiengou and grouped the three

North American areas together. This is permitted, because you should also take geographic history into account when forming these areas. For the same reason, it is common practice to group Madagascar together with (Southern) Africa and New Zealand together with Australia. Because of their shared history, we can group them together even though the areas in itself may differ substantially.

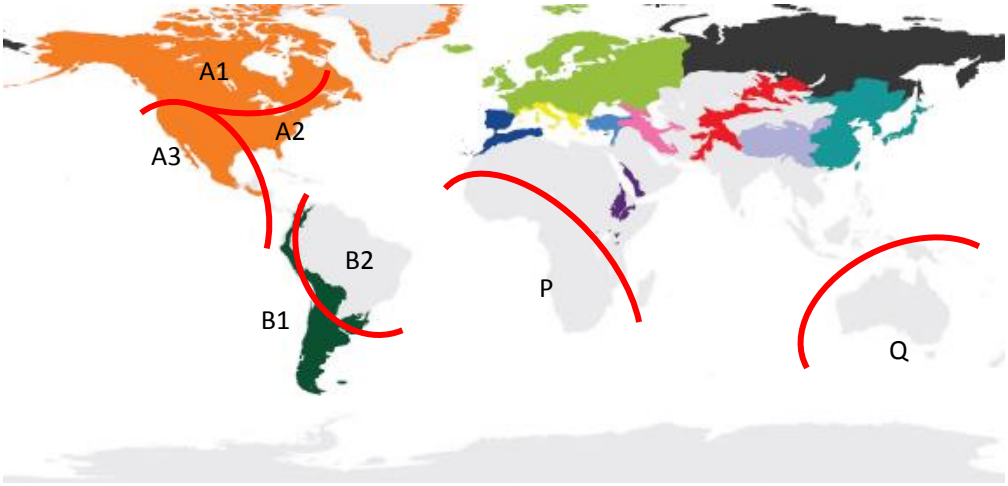


FIGURE 7. ORIGINAL AREA DIVISION (AFTER KARL & KOCH, 2013). A1: NORTHERN AMERICA – NORTH; A2: NORTHERN AMERICA – MIDDLE/EAST; A3: NORTHERN AMERICA - SOUTH/MIDDLE AMERICA; B1: SOUTH AMERICA – WEST; B2: SOUTH AMERICA – EAST; P: SOUTHERN AFRICA; Q: AUSTRALIA AND NEW ZEALAND.

Then, I had planned on using both the S-Diva as well as the Bayesian implementation of RASP (Yu et al., 2010b). Unfortunately, it turned out that one of the errors I got when trying S-Diva was insoluble. Apparently, this error (or any error in RASP for that matter) is quite unique since not a single help forum concerning S-DIVA or RASP is available on the internet. The only available source of help is the manual, where these specific errors are regrettably not addressed.

Then, when I wanted to perform the Bayesian analysis on the full dataset (MAFFT_8 matrix), I got the message that the maximum number of terminals allowed is 512 while my dataset comprises 568 terminals. In the RASP-manual, it is not specified why there is a maximum limit on the number of terminals and why that number is 512. Since two out of four datasets is larger than is allowed, I pruned the phylogenetic trees to be used for the RASP analysis as well as the corresponding alignment by loading all in Mesquite and deleting terminals with a distribution of more than 6 areas. I preferred this method over removing the taxa from the alignment and rerunning the TNT parsimony analysis, because the original larger number of taxa will provide a better phylogenetic result.

Also, when performing any type of Bayesian analysis, it is desirable to check the parameter (.p) files if the analysis has run long enough. RASP itself did not allow the .p files to be opened, and Tracer (normally used to open the files obtained with MrBayes or Beast, (Rambaut & Drummond, 2007)) was not able to open them. This is quite inconvenient, because there is no way of knowing whether the analyses I run with BBM have a sufficient number of generations. This can have influenced the outcome considerably.

Last, when the analysis finishes, a colour coded legend with all possible area combinations is provided. You then have to check which of the eight shades of light green is the one corresponding to the colour in the pie chart in the topology. Of course this is quite an inconvenience, and brings a huge amount of inaccuracy to the interpretation of the results.

4.4 General comments

As explained above, I chose to perform the phylogenetic analysis at the species level while using genus level distribution data for the historical biogeographical analysis. This could be a possible source of bias, for example when the terminals belonging to the same genus do not form a monophyletic clade.

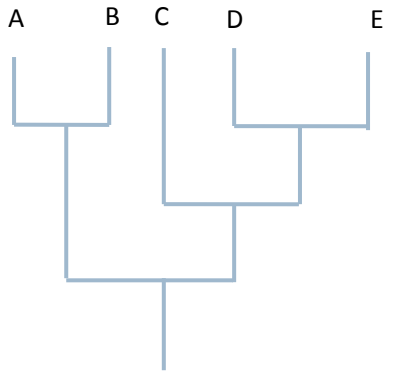


FIGURE 8. IMAGINARY CLADE WITH TAXA A, B, C, D, E.

In Figure 8 for example, all terminals belonging to the same genus with distribution 1, 2 would be the ideal situation. Then, there is no doubt what the ancestral area would be. If, however, terminals A, B, D and E would belong to the one genus (with distribution 1, 2) while misplaced terminal C belongs to another genus (with distribution 3), this terminal causes a bias in the analysis because this single area skews the biogeographical analysis. In this project this is a serious problem because a lot of the genera turned out non-monophyletic as explained above (Phylogetic analysis section of this chapter). Especially with wide-spread species (when terminal C has a distribution of 2, 3, 4, 5) this is a substantial problem, because with one terminal the possible ancestral area range of the clade is dramatically widened. This is one of the reasons why I chose to perform additional biogeographical analysis without the widespread terminals. Another scenario where terminal C would provide a problem in when it has a distribution of 4, 5. In that case, the one terminal also substantially increases areas possible for the ancestral node.

In an ideal world, I would have had a matrix comprising all species of all genera now considered part of the Brassicaceae. With such a complete dataset, it would have been a lot easier to confidently determine the ancestral area of this comprehensive family. In an ideal world, I would also have been able to perform a dating analysis on the phylogenetic tree. We now have an indication of the ancestral area of the Brassicaceae, but it would be helpful if we would be able to place this in a historical context.

ACKNOWLEDGEMENTS

Without the constant support of both my supervisors, Setareh Mohammadin and Freek T. Bakker, it would not have been possible to complete this project in such a happy and stress-free manner; thank you both for that! In addition, I want to thank Lars Chatrou for his help and guidance during the past week and a half of this thesis. Your critical view and dissecting of my thesis have definately helped me a lot in improving it and keepin on thinking further. In addition, I want to thank all staff and students of the Biosystematics group for a wonderful 6 months.

REFERENCES

Ali, S. S., Yu, Y., Pfosser, M., & Wetschnig, W. (2012). Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Annals of Botany*, *109*(1), 95–107. doi:10.1093/aob/mcr274

Al-Shehbaz, I. A. (2012). A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon*, *61*(5), 931–954.

Al-Shehbaz, I. A., Beilstein, M. A., & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution*, *259*(2-4), 89–120. doi:10.1007/s00606-006-0415-z

Anisimova, M., Cannarozzi, G., & Liberles, D. a. (2010). Finding the balance between the mathematical and biological optima in multiple sequence alignment. *Trends in Evolutionary Biology*, *2*(1). doi:10.4081/eb.2010.e7

Bailey, C. D., Koch, M. A., Mayer, M., Mummenhoff, K., O'Kane, S. L., Warwick, S. I., … Al-Shehbaz, I. A. (2006). Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution*, *23*(11), 2142–60. doi:10.1093/molbev/msl087

Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., & Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *PNAS*, *107*(43), 18724–18728. doi:10.1073/pnas.0909766107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.0909766107

Bremer, K. (1992). Ancestral Areas: A Cladistic Reinterpretation of the Center of Origin Concept. *Systematic Biology*, *41*(4), 436–445.

Bremer, K. (1995). Ancestral Areas: Optimization and Probability. *Systematic Biology*, *44*(2), 255–259.

Brooks, D. R. (1990). Parsimony Analysis in Historical Biogeography and Coevolution: Methodological and THeoretical Update. *Systematic Biology*, *39*(1), 14–30.

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol*, *17*(4), 540–552.

Couvreur, T. L. P., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., & Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution*, *27*(1), 55–71. doi:10.1093/molbev/msp202

Cox, C. B., & Moore, P. D. (2010). *Biogeography: An Ecological and Evolutionary Approach* (Eight.). John Wiley and Sons.

Crisci, J. V. (2001). The voice of historical biogeography. *Journal of Biogeography*, *28*(2), 157–168. doi:10.1046/j.1365-2699.2001.00523.x

Crisci, J. V., Katinas, L., & Posadas, P. (2003). *Historical Biogeography; an Introduction*. Harvard University Press.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., … Wilson. (2011). Geneious v5.4.

Franzke, A., German, D., Al-Shehbaz, I. A., & Mummenhoff, K. (2009). Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon*, *58*(2), 425–437.

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., & Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science*, *16*(2), 108–16. doi:10.1016/j.tplants.2010.11.005

Goloboff, P. A. (1999). Analyzing Large Data Sets in Reasonable Times: Solutions for Composite Optima. *Cladistics*, *15*, 415–428.

Goloboff, P. A., Farris, J. S., & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, *24*(5), 774–786.

Google. (2014). Google Earth.

Harcourt, A. H. (2012). *Human biogeography.* Univ of California Press.

Harris, A. J., & Xiang, Q.-Y. (Jenny). (2009). Estimating ancestral distributions of lineages with uncertain sister groups: a statistical approach to Dispersal-Vicariance Analysis and a case using Aesculus L. (Sapindaceae) including fossils. *Journal of Systematics and Evolution*, *47*(5), 349–368. doi:10.1111/j.1759-6831.2009.00044.x

Hauser, L. A., & Crovello, T. J. (1982). Numerical Analysis of Generic Relationships in Thelypodieae (Brassicaceae). *Systematic Botany*, *7*(3), 249–268.

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., & Roberts, C. (2004). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, *8*(1), 23–29. doi:10.1111/j.1461-0248.2004.00686.x

Huson, D. H., & Bryant, D. (2005). Estimating phylogenetic trees and networks using SplitsTree 4. *Manuscript in Preparation, Software Available from Www. Splitstree. Org.*

Jordon-Thaden, I. E., Al-Shehbaz, I. A., & Koch, M. A. (2013). Species richness of the globally distributed, arctic–alpine genus Draba L. (Brassicaceae). *Alpine Botany*, *123*(2), 97–106. doi:10.1007/s00035-013-0120-9

Karl, R., & Koch, M. A. (2013). A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. *Annals of Botany*, *112*(6), 983–1001. doi:10.1093/aob/mct165

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., & Barthlott, W. (2005). Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, *32*(7), 1107–1116. doi:10.1111/j.1365-2699.2005.01272.x

Koch, M. A. (2013). Personal Communication.

Koch, M. A., Al-Shehbaz, I. A., & Mummenhoff, K. (2003). Molecular Systematics, Evolution, and Population Biology in the Mustard Family (Brassicaceae). *Annals of the Missouri Botanical Garden*, *90*(2), 151–171.

Koch, M. A., & German, D. A. (2013). Taxonomy and systematics are key to biological information: Arabidopsis, Eutrema (Thellungiella), Noccaea and Schrenkiella (Brassicaceae) as examples. *Frontiers in Plant Science*, *4*, 267. doi:10.3389/fpls.2013.00267

Koch, M. A., Karl, R., German, D. A., & Al-Shehbaz, I. A. (2012). Systematics , taxonomy and biogeography of three new Asian genera of Brassicaceae tribe Arabideae: An ancient distribution circle around the Asian high mountains. *Taxon*, *61*(5), 955–969.

Koch, M. A., & Marhold, K. (2012). Phylogeny and systematics of Brassicaceae — Introduction. *Taxon*, *61*(5), 929–930.

Lomolino, M. V., Riddle, B. R., Whittaker, R. J., & Brown, J. H. (2010). *Biogeography*. Sinauer Associates, Inc.

Löytynoja, A., & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10557–62. doi:10.1073/pnas.0409137102

Löytynoja, A., & Goldman, N. (2008). Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science*, *320*, 1632–1635.

Maddison, W. P., & Maddison, D. R. (2010). Mesquite: A modular system for evolutionary analysis. Version 2.74. Retrieved from http://mesquiteproject.org

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. *Proceedings of the Gateway Computing Environments Workshop*, 1–8.

Morrone, J. J. (2009). *Evolutionary Biogeography, An Integrative Approach with Case Studies*. Columbia University Press.

Morrone, J. J. (2010). *Evolutionary Biogeography: Principles and Methods*. (M. Gailis & S. Kalnins, Eds.)*Biogeography* (pp. 1 – 63). Nova Science Publishers.

Morrone, J. J., & Crisci, J. V. (1995). Historical biogeography: introduction to methods. *Annual Review of Ecology and Systematics.*, 373–401.

Müller, K. (2005). SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics*, *4*, 65–69.

Nixon, K. C. (1999). The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics*, *15*, 407–414.

Nylander, J. A. A., Olsson, U., Alström, P., & Sanmartín, I. (2008). Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: Turdus). *Systematic Biology*, *57*(2), 257–68. doi:10.1080/10635150802044003

Ojeda, A. A., Novillo, A., Ojeda, R. A., & Roig-Juñent, S. (2013). Geographical distribution and ecological diversification of South American octodontid rodents. *Journal of Zoology*, *289*(4), 285–293. doi:10.1111/jzo.12008

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933. doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Rambaut, A., & Drummond, A. J. (2007). Tracer v1. 4.

Ree, R. H., & Smith, S. a. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*(1), 4–14. doi:10.1080/10635150701883881

Ronquist, F. (1994). Ancestral Areas and Parsimony. *Society of Systematic Biologists*, *43*(2), 267–274.

Ronquist, F. (1995). Ancestral Areas Revisited. *Society of Systematic Biologists*, *44*(4), 572–575.

Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quatification of Historical Biogeography. *Systematic Biology*, *46*(1), 195–203.

Ronquist, F., & Sanmartín, I. (2011). Phylogenetic Methods in Biogeography. In *Annual Review of Ecology, Evolution, and Systematics* (pp. 441–464).

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–42. doi:10.1093/sysbio/sys029

Sanmartín, I. (2006). Event-based biogeography: Integrating patterns processes, and time. In *Biogeography in a Changing World* (pp. 135–159).

Schranz, M. E., Mohammadin, S., & Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology*, *15*(2), 147–53. doi:10.1016/j.pbi.2012.03.011

Simmons, M. P., Müller, K., & Norton, A. P. (2007). The relative performance of indel-coding methods in simulations. *Molecular Phylogenetics and Evolution*, *44*(2), 724–40. doi:10.1016/j.ympev.2007.04.001

Simmons, M. P., & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, *49*(2), 369–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12118412

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.

Talavera, G., & Castersana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577.

The Global Biodiversity Information Facility. (2013). BGIF Backbone Taxonomy. Retrieved July 08, 2014, from http://www.gbif.org/species/3112

Vaidya, G. . L. D. J. . & M. R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180.

Warwick, S. I., Mummenhoff, K., Sauder, C. A., Koch, M. A., & Al-Shehbaz, I. A. (2010). Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Systematics and Evolution*, *285*(3-4), 209–232. doi:10.1007/s00606-010-0271-8

Wiley, E. O. (1987). Methods in Vicariance Biogeography. In P. Hovenkamp (Ed.), *Systematics and Evolution: A Matter of Diversity* (pp. 283–306). Institute of Systematic Botany, Utrecht University.

Wiley, E. O. (1988). Parsimony analysis and vicariance biogeography. *Systematic Zoology*, *37*(3), 271–290.

Yu, Y., Harris, A. J., & He, X. (2010a). A Rough Guide to S-DIVA.

Yu, Y., Harris, A. J., & He, X. (2012). A Rough Guide to RASP 2 . 1 (Beta).

Yu, Y., Harris, A. J., & He, X. J. (2010b). S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Mol Phylogenet Evol*, *56*(2), 848–850. doi:10.1016/j.ympev.2010.04.011

APPENDIX A. OVERVIEW OF GENBANK ACCESSIONS

Genus	Species	ITS (82%)	ndhF (38%)	nad4 (36%)	rbcL (21%)	matK (26%)	trnLF (40)	adH (6%)	percentage	Area prop
Aethionema	arabicum	AY254539					DQ180218		26	CDFG
Aethionema	saxatile		DQ288726 [saxatile]		AY483262 [saxatile]					
Aethionema	spinosum	GQ424545	DQ288798							
Aethionema	grandiflorum				AY167983 [grandiflorum]	AF144354 [grandiflorum]				
Agianthus										
Alliaria	petiolata	AJ862703/AJ862704	DQ288727		JQ933212 [petiolata]*	AF144363	JN189781 [petiolata]		92	ACDEGI
Alyssoides	utriculata	EF514593							12	DE
Alyssopsis	mollis	GQ424523							14	A
Alyssum	chalcidicum	GQ284876 [chalcidicum]							51	ABCDEFG
Alyssum	canescens		DQ288728 [canescens]							
Alyssum	simplex					JF926641 [simplex]-				
Alyssum	mollis						FJ188076 [mollis]			
Alyssum	lenense						FN677633 [lenense]*			
Ammosperma	cinerea	GQ424606							2	C
Anastatica	hierochuntica	GQ424524				GQ424573			25	CFH
Anchonium	elichrysifolium	DQ357516			FN594834 [elichrysifolium]*-+				7	F
Andrzeiowskia										
Anelsonia	eurycarpa	DQ452059	DQ288729				JX146081 [eurycarpa]*	JX146666 [eurycarpa]*	32	A
Aphragmus	oxycarpus	DQ165337 [oxycarpus]				DQ518350 [oxycarpus]	DQ518350 [oxycarpus]		62	IJ
Aphragmus	involucratus									
Aphragmus	obscurus									
Aplanodes	doidgeana	GQ497847 [doidgeana]							37	M

Arabidella	glaucenscens								89	N
Arabidella	trisecta	JX630158 [trisecta]								
Arabidella	eremigena				FN597049 [eremigena]-					
Arabidopsis	suecica			GQ438794 [suecica]-					86	ABCDEFG
Arabidopsis	lyrata		DQ288730 [lyrata]		FN594842 [lyrata]*-					
Arabidopsis	halleri					AF144341 [halleri]				
Arabidopsis	arenosa						GQ386495 [arenosa]*			
Arabidopsis	thaliana	AJ232900			Full cp genome NC_000932	AF144348, AF144328		AF110456		
Arabis	alpina	DQ060109	DQ288731	EU931347 [alpina]			DQ518351	AF110429	64	ABCDEFG
Arabis	hirsuta				JX848435 [hirsuta]					
Arabis	stelleri				D88903 [stelleri]-					
Arabis	glabra				DQ310542 [glabra]					
Arabis	pauciflora					AF144335				
Arabis	drummondii					AF144350 [drummondii]				
Arabis	parshii					AF144349 [parishii]				
Arabis	lyalli					AF144332 [lyallii]				
Arabis	fendleri					AF144351 [fendleri]				
Arabis	glabra					AF144333 [glabra]				
Arabis	turrita					AF144347 [turrita]				
Arcyosperma	primulifolium	GQ424525		GQ424780 [primulifolium]*			JQ919863 [primulifolium]*		54	I
Armoracia	rusticana	AF078032/AF078031 [rusticana]		GQ424684 [rusticana]*	AF020323 [rusticana]	FN597648 [rusticana]*-	EF426785 [rusticana]-		95	A
Aschersoniodoxa	mandoniana			GQ424784 [mandoniana]*					32	B
Aschersoniodoxa	cachensis	EU620282 [cachensis]								

Asperuginoides	axillaris	EF514626 [axillaris]					GU181984 [axillaris]		27	G
Asta	schaffneri	GQ424526	DQ288733 [sp.]						39	A
Atelanthera	perpusilla	FM164518/FM164519 [perpusilla]*							48	I
Athysanus	pusillus	EF514629					GU246241 [pusillus]		42	A
Athysanus	unilateralis			GQ424804 [unilateralis]*						
Aubrieta	deltoidea	AJ232909	DQ288734			AF144352	DQ180303	AF110425	66	CDEFG
Aurinia	saxatilis	EF514630	KF022950 [saxatilis]		JQ412329 [saxatilis]*-	JQ412213 [saxatilis]*-	DQ518349		37	CDEFN
Baimashania	pulvinata	DQ523426	DQ288736			DQ409251	DQ523325 [pulvinata]		60	J
Ballantinia	antipoda				FN597048 [antipoda]*-				33	N
Barbamine										
Barbarea	verna	X98631	DQ288737		NC_009269	AF144330 [vulgaris]	DQ518352 [vulgaris]	AF110458 [vulgaris]	86	ABCDEFG
Berteroa	incana	EF514631	AY330097		KF613070 [incana]-	GQ424574	KF022814 [incana]		82	ADEGI
Berteroella	maximowiczii	GU182052 [maximowiczii]					GU181985 [maximowiczii]		1	J
Biscutella	didyma	DQ452058	DQ288738	GQ424732 [didyma]*		GQ424575			51	ACDEFG
Bivonaea	lutea	HQ327490 [lutea]					HQ327491 [lutea]		1	CD
Blennodia	pterosperma	DQ357519 [pterosperma]					JX134205 [pterosperma]-		93	N
Blennodia	canescens				JX630173 [canescens]					
Boechera	stricta	AF137575				AF144343		AF110437	41	AM
Boechera	parishii							AF110450 [parishii]		
Boechera	lyallii							AF110448 [lyallii]		
Boechera	lignifera							AF110447 [lignifera]		
Boechera	fendleri							AF110438 [fendleri]		
Boechera	laevigata		DQ288739 [laevigata]							

Boechera	holboellii				FN594845 [holboellii]*-		AY257788 [holboellii]				
Boechera	divaricarpa				JX848436 [divaricarpa]						
Boleum											
Boreava	orientalis	DQ249859					DQ518353 [orientalis]		7	CFGI	
Bornmuellera	baldaccii	EF514635	KF022959 [baldaccii]				KF022818 [baldaccii]		38	DF	
Borodinia	macrophylla	JX146999 [macrophylla]*					JX146151 [macrophylla]*		0	KL	
Botschantzevia	karatavica	EF514690 [karatavica]					GU181986 [karatavica]		#DIV/0!	I	
Brachycarpaea	juncea	AJ862707/AJ862708 [juncea]							0	M	
Brassica	juncea	AF128093			AY167979 [juncea]		DQ180232		57	ABCDEFG	
Brassica	oleracea		DQ288742 [oleracea]					AF110434 [oleracea]			
Brassica	spinescens					JN584953 [spinescens]*					
Brassica	napus							JQ796372 [napus]			
Braya	rosea	AY353129	DQ288743 [rosea]				DQ518354		69	AEIJK	
Braya	glabella				JQ933246 [glabella subsp purpusascens]*						
Brayopsis	monimocalyx			GQ424808 [monimocalyx]*					46	B	
Brayopsis	colombiana	EU620283 [colombiana]*	EU718525 [colombiana]-				UE620339 [colombiana]*				
Bunias	orientalis	DQ249863	DQ288744			GQ424576	FN677645 [orientalis]*			CDEFGI	
Cakile	maritima	DQ249830 [maritima]	DQ288745 [maritima]		AY167981 [maritima]	GQ424577	DQ180247 [maritima]		73	ABCDEFG	
Calepina	irregularis	AY722504			HE616642 [irregularis]		DQ518356		43	CDEF	
Calymmatium	draboides	GQ497854 [draboides]							33	I	
Camelina	microcarpa	AF137574	DQ288746		JN847825 [microcarpa]		DQ821412 [microcarpa]-			ABCDEFIL	
Camelina	sativa					GQ424578 [sativa]			61		

Camelinopsis	campylopoda		DQ288817						0	G
Camelinopsis	glaucophylla	DQ357581 [glaucophylla]								
Capsella	bursa-pastoris	AF531561	DQ288748		D88904	GQ424578		AF110435 [rubella]	88	ABCDEFH
Capsella	rubella						DQ180225 [rubella]	DQ343324, DQ343314, DQ343313		
Capsella	sp.				AB586239 [sp.SH-2010]-					
Cardamine	scutata	DQ268478		HQ616526 [hirsuta]	D88905 [flexuosa]	GQ424604 [flexuosa]			84	ABCDEFH
Cardamine	lacustris	AF100683								
Cardamine	hirsuta				HE616645 [hirsuta]					
Cardamine	angustata		AF198139 [angustata]							
Cardamine	flexuosa						AB247985 [flexuosa]*			
Cardamine	amara					AF144337 [amara]		AF110430 [amara]		
Cardamine	rivularis					AF144365 [rivularis]				
Cardamine	penzesii					AF144364 [penzesii]				
Cardamine	microzyga					JF926662 [microzyga]				
Cardaria	pubescens	AJ628279/AJ628280 [pubescens]					AY015920 [pubescens]		29	ABCDEFG
Carinavalva	glauca	GQ424527							75	N
Carrichtera	annua	DQ249829				GQ424579	AY751761		81	ACDEFN
Catadysia	rosulans	EU620284 [rosulans]*	EU718526 [rosulans]-				EU620340 [rosulans]*		75	B
Catenulina	hedysaroides	GQ424607							0	I
Catolobus	pendulus	AF137572	DQ288732	GQ424666 [pendulus]*		DQ406758 [pendulus]	FJ188022 [pendulus]-		20	JKL
Caulanthus	crassicaulis		DQ288750	GQ424650 [crassicaulis]*			EU620341 [crassicaulis]*		31	A
Caulanthus	amplexicaulis	AF346630 [amplexicaulis var. amplexicaulis]	EU718527 [amplexicaulis]-							

Caulanthus	inflatus		EU718533 [inflatus]-						
Caulanthus	hallii		EU718531 [hallii]-						
Caulanthus	stenocarpus		EU718535 [stenocarpus]-						
Caulanthus	lasiophyllus			EU931391 [lasiophyllus]					
Caulostramina	jaegeri		DQ288751 [jaegeri]					0	A
Ceratocnemum	rapistroides	AY722429						14	C
Ceriosperma									
Chalcanthus	renifolius	GQ424528	DQ288752	GQ424705 [renifolius]				0	G
Chamira	circaeoides	AJ862719/AJ862720			AM234932			35	M
Chartoloma	platycarpum	GQ424529		GQ424793 [platycarpum]*				0	GL
Chaunanthus	acuminatus	GQ497855 [acuminatus]	EU718536 [acuminatus]-	GQ424779 [acuminatus]*			EU620344 [acuminatus]*	58	A
Cheirinia									
Chilocardamum									
Chlorocrambe	hastata	GQ497856 [hastata]	EU718538 [hastata]-				EU620346 [hastata]*	35	A
Chorispora	tenella	DQ357526	DQ288753		FN594833 [tenella]*-		DQ518384	52	ABEFGIJN
Christolea	crassifolia	DQ523423	DQ288754			DQ409256 [crassifolia]	DQ23322 [crassifolia]	80	IJ
Christolea	niyaensis				JN847808 [niyaensis]				
Chrysochamela	velutina	DQ249856 [velutina]						4	FG
Cithareloma	lehmannii	EF514641		GQ438795 [lehmannii]				13	GI
Cithareloma	vernum					JF926677 [vernum]-			
Clastopus	vestitus		KF022966 [vestitus]				KF022825 [vestitus]	5	G
Clausia	aprica	DQ357529						3	IKL
Clausia	trichosepala				JN847815 [trichosepala]	JF926653 [trichosepala]-			

Clausia	podlechii						FN677719 [podlechii]*			
Clypeola	aspera	EF514642 [aspera]	EU907360 [aspera]				KF022826 [aspera]		41	CDFG
Cochlearia	danica			EU931354 [danica]		AF174531			84	ADE
Cochlearia	acaulis	HQ268659 [acaulis]	DQ288785	EU931369 [acaule]	FN594827 [acaulis]*-		HQ268714 [acaulis]*-			
Cochlearia	officinalis		AY514390 [officinalis]							
Cochlearia	prolongoi					AF144369 [prolongoi]				
Coelophragmus	auriculatus	EU620290 [auriculatus]	EU718539 [auriculatus]-				EU620347 [auriculatus]*		32	A
Coincya	monensis				JN892991 [monensis]-				59	ACDE
Coincya	longirostra					JN584960 [longirostra]*				
Coluteocarpus	vesicaria	GQ497857 [vesicaria]							12	FG
Conringia	planisiliqua	GQ424570	DQ288756		JN847840 [planisiliqua]	JF926665 [planisiliqua]-	AY751762 [planisiliqua]		65	ACDEFGI
Cordylocarpus	muricatus	DQ249827				JN584963 [muricatus]	AY751759		4	C
Crambe	tataria								62	ACDEFGH
Crambe	filiformis	AY722434 [filiformis]								
Crambe	maritima				JN893554 [maritima]-	GQ424580 [maritima]				
Crambe	kotschyana						EF426778 [kotschyana]-			
Crambella	teregifolia	AF039986/AF040029 [teregifolia]-				JN584967 [teretifolia]*				
Cremolobus	chilensis	GQ424530		GQ424774 [chilensis]					22	B
Cremolobus	subscandens		DQ288757 [subscandens]				EU620348 [subscandens]*			
Crucihimalaya	mollissima	AF137552			FN594843 [mollissima]-		DQ518358		69	GIJ
Crucihimalaya	himalaica		HM120274 [himalaica]		D88902 [himalaica]	AF144356 [himalaica]		AB015503 [himalaica]		
Crucihimalaya	bursifolia				FN598779 [bursifolia]-	DQ406759 [bursifolia]				

Crucihimalaya	wallichii	JX146960 [wallichii]*						JX146689 [wallichii]*		
Cryptospora	falcata	DQ357531							4	GI
Cuphonotus	humistratus	JX630160 [humistratus]			JX630170 [humistratus]				93	N
Cusickiella	quadricostata	DQ452066	DQ288758	GQ424617 [quadricostata]*					43	A
Cusickiella	douglasii					DQ406761 [douglasii]	JX146113 [douglasii]*	JX146718 [douglasii]*		
Cymatocarpus	pilosissimus	GQ497858 [pilosissimus]							0	G
Cyphocardamum	aretioides	GQ497859 [aretioides]							0	I
Dactylocardamum										
Decaptera										
Degenia	velebitica	EF514646 [velebitica]					DQ518359 [velebitica]		0	E
Delpinophytum	patagonicum	GQ497887 [patagonicum]							70	B
Dendroarabis	fruticulosa	GU182058 [fruticulosa]-					FN677634 [fruticulosa]*		#DIV/0!	IJL
Descurainia	sophia	AY230619	DQ288759	EU931355 [stricta]	JX848439 [sophia]	GQ424581	DQ518361		68	ABCDEGIJ
Descurainia	stricta				FN594838 [stricta]*-					
Descurainia	pinnata				JX848438 [pinnata]					
Descurainia	californica	GU246181 [californica]								
Desideria	linearis	DQ523417 [linearis]*	DQ288760 [linearis]						#DIV/0!	J
Desideria	baiogoinensis					DQ409252 [baiogoinensis]	DQ523315 [baiogoinensis]			
Desideria	stewartii					DQ409265 [stewartii]				
Desideria	linearis					DQ409254 [linearis]				
Diceratella	elliptica					GQ424582			20	GHM
Diceratella	inernis	DQ357533 [inermis]								
Dichasianthus	subtilissimus	AF137594 [subtilissimus]							#DIV/0!	F
Dictyophragmus	englerianus	EU620293 [englerianus]*							88	B

Dictyophragmus	punensis					EU620349 [punensis]*		
Didesmus	aegypticus	GQ424531					4	CD
Didesmus	bipinnatus				GQ424605 [bipinnatus]			
Didymophysa	fedtschenkoana	EF514648					35	GI
Didymophysa	aucheri					FJ188023 [aucheri]		
Dielsiocharis	kotschyi	GQ424532					6	G
Dilophia	salsa	FM164649 [salsa]*	DQ288761	JQ933304 [salsa]*		DQ518364	81	J
Dimorphocarpa	wislizenii	AF137593, AF137592	DQ288763	FN594839 [wislizeni]*-			28	A
Diplotaxis	erucoides	DQ249826 [erucoides]				DQ180245 [erucoides]	100	ABCDEFG
Diplotaxis	tenuifolia			JN892620 [tenuifolia]-				
Diplotaxis	assurgens				JN584970 [assurgens]*			
Dipoma	iberideum	GQ497861 [iberideum]					12	J
Diptychocarpus	strictus	DQ357534	DQ288762		JF926637 [strictus]-	FN677716 [strictus]*	5	GI
Dithyrea	californica	AF137592 [californica]					29	A
Dithyrea	maritima		JQ323048 [maritima]					
Dontostemon	integrifolius	DQ357536 [integrifolius]					24	IJK
Dontostemon	senilis		DQ288764 [senilis]	JN847816 [senilis]				
Dontostemon	glandulosus				JF926648 [glandulosus]-			
Dontostemon	intermedius					FN677644 [intermedius]*		
Douepea	tortuosa	GQ497862 [tortuosa]					0	I
Draba	altaica	AY134115	DQ288765				73	ABCDEFG
Draba	nemorosa			NC_009272 [nemorosa]				
Draba	ellipsoidea					DQ180236 [ellipsoidea]		
Draba	sp.				GQ424583 [sp]			

Drabastrum	alpestra	JX630161 [alpestre]			JX630176 [alpestre]			88	N
Dryopetalon	runcinatum	AF531634						61	A
Dryopetalon	palmeri		EU718541 [palmeri]-						
Dryopetalon	paysonii		EU718542 [paysonii]-	GQ424813 [paysonii]*			EU620350 [paysonii]*		
Eigia	longistyla	GQ497863 [longistyla]						0	F
Elburzia	fenestrata	GQ424533						0	G
Enarthrocarpus	clavatus					GQ424584		10	CDF
Enarthrocarpus	acruatus	AY722456 [acruatus]							
Enarthrocarpus	lyratus						AB670026 [lyratus]-		
Englerocharis	pauciflora	EU620295 [pauciflora]*	EU718543 [pauciflora]-				EU620351 [pauciflora]*	44	B
Eremobium	aegyptiacum	DQ357537		GQ424663 [aegyptiacum]*		GQ424585		3	CFGI
Eremoblastus	caspicus	FN821522 [caspicus]					FN677643 [caspicus]*	#DIV/0!	F
Eremodraba	intricatissima	GQ424534		GQ424654 [intricatissima]*				10	B
Eremophyton	chevallieri	GQ424535						24	C
Erophila	verna	DQ467575 [verna]-			HQ619740 [verna]-		DQ467135 [verna]	12	CDF
Erophila	majuscula					JN894943 [majuscula]-			
Eruca	vesicaria	DQ249821 [sativa]				GQ424586	AY751765 [sativa]	49	ABCDEFG
Erucaria	boveana	AY722495						32	CDFG
Erucaria	erucarioides					JN584976 [erucarioides]*			
Erucastrum	varium	AF531614						60	ACDEHMI
Erucastrum	gallicum				JX520951 [gallicum]		AY751766 [gallicum]		
Erucastrum	canariense					JN584982 [canariense]*			
Erysimum	capitatum				AY167980	GQ424587 [repandum]	KF022857 [cuspidatum]	68	ABCDEFG

Erysimum	mongolicum			GQ424670 [mongolicum]*					
Erysimum	canum	DQ357539 [canum]							
Erysimum	canescens		DQ288766						
Erysimum	inconspicuum				JX848441 [inconspicuum]				
Erysimum	sisymbioides				JN847824 [sisymbioides]	JF926666 [sisymbrioides]-			
Erysimum	siliculosum				JN847822 [siliculosum]				
Erysimum	cheiranthoides					JF926663-			
Erysimum	perofskianum					DQ406762 [perofskianum]			
Euclidium	syriacum	DQ357543	DQ288767				DQ180251	269	AEGI
Eudema	hauthalii			GQ424620 [hauthalii]*				36	B
Eudema	nubigena	EU620299 [nubigena]*	EU718545 [nubigena]-				EU620352 [nubigena]*		
Eurycarpus									
Eutrema	heterophyllum	DQ165352	DQ288768	EU931358 [heterophyllum]				43	AIJK
Eutrema	altaicum	DQ165364	DQ288836	EU931389 [altaicum]					
Eutrema	salsugineum	AF531626		GQ424704 [salsugineum]*		DQ406771 [salsugineum]	JN387821 [salsugineum]*-		
Eutrema	himalaicum				JQ933332 [himalaicum]*				
Exhalimolobos	palmeri	AF137569 [palmeri]	JQ323086 [palmeri]					35	AB
Farsetia	aegyptiaca	EF514649	DQ288769	GQ424687 [aegyptia]*			KF022851 [aegyptia]	24	CFGHM
Fezia	pterocarpa	GQ424536						20	C
Fibigia	suffruticosa 1	FM164657	EU907361 [suffruticosa]					20	CDEFG
Fibigia	clypeata						DQ518368 [clypeata]		
Foleyola	billotii	GQ497866 billotii]						16	C
Fortuynia	garcini	AF263398 [garcini]						9	G
Fourraea	alpina	DQ518395		GQ424721			DQ180226	58	CDEM

				[alpina]*						
Galitzkya	macrocarpa	EF514655							0	IKL
Galitzkya	potaninii		KF022983 [potaninii]				FN677635 [potaninii]			
Geococcus	pusillus	GQ424571							95	N
Glastaria										
Glaucocarpum	suffrutescens		DQ288770 [suffrutescens]						0	A
Goerkemia										
Goldbachia	laevigata	DQ357545 [laevigata]	DQ288771 [laevigata]			JF926643 [laevigata]-			39	GIJ
Gorodkovia	jacutica	AY230646 [jacutica]							0	AIJL
Graellsia	saxifragifolia	GQ424572	DQ288772						4	CFG
Guillenia	lasiophylla	EU620287 [lasiophylla]*	EU718534 [lasiophylla]-				EU620343 [lasiophylla]		89	A
Guillenia	flavescens		EU718548 [flavescens]-							
Guiraoa	arvensis	AY722468				JN584987 [arvensis]*			58	C
Gynophorea										
Halimolobos	diffusus	AF307645	EU718621	EU931359 [diffusus]	FN594846 [diffusus]*-				19	A
Halimolobos	perplexa					AF144346 [perplexa]		AF110441 [perplexa]		
Halimolobos	jaegeri					DQ406763 [jaegeri]	**JX146114 [jaegeri]**	JX146691 [jaegeri]*	0	
Harmsiodoxa	blennodioides	JX630162 [blennodioides]			JX630172 [blennodioides]-				93	N
Hedinia										K
Heldreichia	bupleurifolia	FN397988 [bupleurifolia]*							1	F
Heliophila	hurkana	AJ864823/AJ863573		GQ424962 [hurkana]*					43	MN
Heliophila	dregeana			GQ424709 [dregeana]*						
Heliophila	juncea			GQ424613 [juncea]*						

Heliophila	linearis	AJ863573/AJ864823		EU931361 [linearis]					
Heliophila	sp.		DQ288775 [sp.]	EU931362 [sp.]*					
Heliophila	pubescens				AM234933 [pubescens]				
Heliophila	variabilis					GQ424588 [variabilis]			
Heliophila	coronopifolia						DQ518369 [coronopifolia]		
Hemicrambe	fruticulosa	AY722469				JN584988 [fruticulosa]*		31	H
Hemilophia									
Henophyton	deserti	GQ424537				JN584989 [deserti]*		4	C
Hesperidanthus	suffrutescens	GQ424567						34	A
Hesperidanthus	sp.			GQ424727 [sp.]					
Hesperidanthus	barnebyi						EU620356 [barnebyi]*		
Hesperidanthus	linearifolia	AF531612	DQ288821						
Hesperidanthus	jaegeri	GQ424569	DQ288751						
Hesperis	sibirica	DQ357549			FN594835 [sibirica]*-		FN677642 [sibirica]	83	ACDEFG
Hesperis	matronalis		DQ288776 [matronalis]			HQ593319 [matronalis]*-			
Heterothrix	debilis				EF455920 [debilis]				
Hilliella	fumarioides	AF100851/AF100852 [fumarioides]-						#DIV/0!	K
Hirschfeldia	incana	AY722470 [incana]	DQ288778 [incana]			JN584990 [incana]*	EU620407 [incana]*	72	ABCDEFG
Hollermayera									
Hormathophylla	longicaulis	EF514660						39	CE
Hormathophylla	spinosa		KF022987 [spinosa]						
Hormathophylla	purpurea						FN677738 [purpurea]		
Hornungia	petraea	AJ440303						58	ABCDEFI
Hornungia	procumbens		DQ288779						

			[procumbens]						
Hornungia	alpina				DQ310538 [alpina]		DQ310515 [alpina]		
Hornungia	petraea					JN893991 [petraea]-			
Horwoodia	dicksoniae	GQ424538						3	H
Hugueninia									
Ianhedgea	minutiflora	AF137568	DQ288780	EU931366 [minutiflora]	FN594825 [minutiflora]*-			61	I
Iberidella									
Iberis	amara	AJ440311		EU931367 [amara]	FN594828 [amara]*-	GQ424589		45	A
Iberis	procumbens			EU931368 [procumbens]					
Iberis	sempervirens		DQ288781 [sempervirens]					36	
Iberis	umbellata				HE616648 [umbellata]				
Icianthus									
Idahoa	scapigera	GQ497867 [scapigera]	DQ288783	GQ424761 [scapigera]*				50	A
Iodanthus	pinnatifidus	GQ424539	DQ288784					33	B
Ionopsidium									CE
Irenepharsus	magicus				JX630175 [magicus]			92	N
Isatis	brevipes		DQ288786 [tinctoria]	EU931370 [brevipes]				55	ACDEFGIK
Isatis	tinctoria	DQ249851 [tinctoria]				AB354278 [tinctoria]-	DQ518370 [tinctoria]		
Isatis	pachycarpa				FN594830 [pachycarpa]*-				
Iskandera	hissarica	DQ357553 [hissarica]*						0	I
Kernera	saxatilis	AJ440313		GQ424715 [saxatilis]*				35	CE
Kotschyella	stenocarpa	GQ497888 [stenocarpa]						#DIV/0!	G
Kremeriella	cordylocarpus	AY722471 [cordylocarpus]				JN584991 [cordylocarpus]*		4	C
Lachnoloma	lehmannii	GQ497889 [lehmannii]			JN847812 [lehmannii]			0	GIL

Lachnocapsa	spathulata	GQ424540							29	H
Leavenworthia	crassa	GQ424541	DQ288787 [crassa]				AM072871 [crassa]*	AF037563	9	A
Leiospora	eriocalyx	DQ357554	DQ288788 [eriocalyx]						67	I
Leiospora	exscapa									
Leiospora	pamirica									
Lepidium	latifolium			EU931371 [latifolium]					73	ABCDEFG
Lepidium	leptopetalum			GQ438806 [leptopetalum]						
Lepidium	pedicellosum			GQ438805 [pedicellosum]						
Lepidium	sagittatum			EU931388 [sagittatum]						
Lepidium	africanum				AM234934 [africanum]					
Lepidium	draba		DQ288790 [draba]							
Lepidium	pamirica					DQ409255 [pamirica]	DQ180250 [pamirica]			
Lepidium	peroliatum					DQ406766 [perfoliatum]				
Lepidium	capitatum						DQ518371 [capitatum]			
Lepidium	campestris	AF055197 [campestris]				AF144359 [campestre]				
Lepidostemon	glaricola	GQ424542 [glaricola]							4	J
Leptaleum	filifolium	DQ357556			JN847811 [filifolium]	GQ424590			34	I
Lesquerella	fendleri	AF055198 [fendleri]							52	A
Lexarzanthe										
Lithodraba	mendocinensis	GQ497890 [mendocinensis]		GQ4524812 [mendocinensis]*					76	B
Litwinowia	tenuissima	FN821591 [tenuissima]*					FN677713 [tenuissima]*		26	IJ
Lobularia	maritima	EF514681	DQ288791	GQ424689 [maritima]*	NC_009274 [maritima]	GQ424591			68	ABCDEFG
Lobularia	libyca						DQ518372 [libyca]			

Lunaria	rediviva	GQ424543		GQ424733 [rediviva]*				83	ABCDEN
Lunaria	annua		DQ288792 [annua]		HE963547 [annua]*	GQ424592 [annua]			
Lycocarpus									
Lyrocarpa	coulteri	AF137591		EU931372 [coulteri]			GU246240 [coulteri]	54	A
Lyrocarpa	xantii		JQ323051 [xantii]						
Macropodium	nivale			GQ424656 [nivale]*			FN677638 [nivale]*	20	KL
Macropodium	pterospermum	GU182055 [pterospermum]-							
Malcolmia	littorea	DQ357559		EU931373 [littorea]				32	ACDEFGI
Malcolmia	africana		DQ288793 [africana]		JN847814 [africana]	JF926644 [africana]-	EU170625 [africana]-		
Mancoa	hispida		DQ288794	GQ424628 [hispida]*				49	B
Mancoa	foliosa	AF307632 [foliosa]					AF307552 [foliosa]		
Maresia	nana	DQ357562	KF023006 [nana]	EU931374 [nana]		GQ424593		41	CDFG
Mathewsia	auriculata		EU874868 [auriculata]-	GQ424646 [auriculata]*				14	B
Mathewsia	foliosa	DQ357563 [foliosa]							
Mathewsia	nivea		EU718556 [nivea]-				EU620361 [nivea]*		
Matthiola	fragrans		DQ288796 [farinosa]					39	ABCDEFG
Matthiola	maderensis						DQ180302 [maderensis]-		
Matthiola	incana	DQ249848 [incana]			HM850161 [incana]	AF144361 [incana]			
Matthiola	sp.	GQ424566							
Megacarpaea	megalocarpa	FM164564/FM164565 [megalocarpa]*		GQ424722 [megalocarpa]*				24	IJ
Megacarpaea	delavayi		GU174514 [delavayi]*						
Megacarpaea	polyandra				JQ933404 [polyandra]*				

Megadenia	pygmaea	GQ424544		GQ424726 [pygmaea]*				75	J
Menkea	sphaerocarpa	JX630164 [sphaerocarpa]			JX630171 [sphaerocarpa]			94	N
Menonvillea	hookeri	JX470565 [hookeri]*	DQ288797	GQ424611 [hookeri]*				37	B
Microlepidium	pilosum	GQ497869 [pilosulum]						99	N
Microstigma	deflexum						FN677641 [deflexum]*	0	KL
Microstigma	brachycarpum	DQ2357569 [brachycarpum]*							
Microthlaspi	perfoliatum					AF144362		50	ACDEFGI
Morettia	philaeana	DQ357572 [philaeana]	KF023007 [philaeana]					28	CFGH
Moricandia	arvensis	EF601899				GQ424594	AY751767	46	ACDFH
Moricandia	sinaica				JN375996 [sinaica]*				
Moriera	spinosa	GQ424545	DQ288798 [spinosa]					2	G
Morisia	monanthos	AY722476				JN584993 [monanthos]*		3	D
Mostacillastrum	orbignyanum	AF531583						53	B
Mostacillastrum	gracile		EU874869 [gracile]-						
Mostacillastrum	andinum		EU718557 [andinum]-				EU620363 [andinum]*		
Mostacillastrum	elongatum		DQ288799 [elongatum]						
Murbeckiella	huetii	GQ424546						58	CEFG
Muricaria	prostrata	AF039992/AF040035 [prostrata]				JN584994 [prostrata]*		5	CG
Myagrum	perfoliatum	GQ424547	DQ288800	EU931376 [perfoliatum]				45	CDEF
Nasturtiopsis	coronopifolia	GQ424548				GQ424595		12	CF
Nasturtium	officinale	X98643	DQ288801		AF020325	AY483225 [officinale]	AY030271 [officinale]	85	ABCDEFH
Neotorularia	korolkowii	AY353155	DQ288803		JN847813 [korolkowii]			22	FGIJ
Neotorularia	torulosa	AY353166 [torulosa]				GQ424596 [torulosa]			

Neotorularia	humilis						DQ180252 [humilis]			
Nerisyrenia	linearifolia	AF137587 [linearifolia]							41	A
Nerisyrenia	johnstonii		EU907362 [johnstonii]							
Neslia	paniculata	AF137576	KF023019 [paniculata]		DQ310541	GQ424597, DQ406767 [paniculata]	DQ310518 [paniculata]	DQ343325		ABCDEFG
Neslia									70	
Neuontobotrys	elloanensis		DQ288802 [elloanensis]	GQ424652 [elloanensis]*					30	B
Neuontobotrys	linearifolia	EU620306 [linearifolia]*					EU620367 [linearifolia]*			
Neuontobotrys	lanata		EU718559 [lanata]							
Neurotropis										
Nevada	holmgrenii	DQ452061	DQ288829					JX146652 [holmgrenii]*	24	A
Noccaea	cochleariforme		DQ288804		FN594826 [caerulescens]*-	GQ424598	DQ180219 [caerulescens]		66	ABCDEFH
Noccaea	fendleri	AY154806 [fendleri]								
Noccidium	hastulatum	AF336164/AF336165 [hastulatum]							#DIV/0!	G
Notoceras	bicorne	DQ357573 [bicorne]*				GQ424599			36	CFH
Notothlaspi	rosulatum	AF100690							50	N
Notothlaspi	australe				JQ933421 [australe]*					
Ochthodium	aegyptiacum	GQ497870 [aegyptiacum]		GQ424781 [aegyptiacum]*					60	F
Octoceras	lehmannianum	GQ424609							0	GI
Olimarabidopsis	pumila	AF137549	DQ288807		DQ310543	AF144345	DQ180224	AF110440	24	GI
Onuris	papillosa			GQ424619 [papillosa]*					74	B
Onuris	graminifolia	EU620307 [graminifolia]*								
Oreoblastus										
Oreoloma	violaceum	DQ357576 [violaceum]*	DQ288808 [violaceum]		JN847818 [violaceum]				50	J

Oreophyton	falcatum	GQ424549						7	H
Ornithocarpa	torulosa	GQ424550		GQ424789 [torulosa]*				27	A
Orychophragmus	violaceus	EU306541		GQ424778 [violaceus]*		JF926671 [violaceus]-	GQ261977 [violaceus]-	34	K
Otocarpus	virgatus	AY722477 [virgatus]						3	C
Pachycladon	fastigiata	AF100680			EF015666			9	N
Pachycladon	exilis	EF015673			EF015667				
Pachycladon	novae-zelandiae	EF015677			EF015666		EF015661		
Pachycladon	enysii				EF015668 [enysii]-				
Pachycladon	latisiliqua				EF015665 [latisiliqua]-				
Pachycladon	stellata				EF015664 [stellata]-				
Pachycladon	wallii				EF015663 [wallii]-				
Pachycladon	stellatum		HM120287 [stellatum]						
Pachycladon	cheesemanii					JQ806762 [cheesemanii]			
Pachymitus	cardaminoides	JX630165 [cardaminoides]			JX630174 [cardaminoides]			83	N
Pachyneurum	grandiflorum	DQ467584					DQ518374	9	I
Pachypterygium	multicaule	GQ424551			JN847842 [multicaule]	JF926652 [multicaule]-		0	G
Parlatoria	rostrata	GQ424552	DQ288809					0	FG
Parodiodoxa	chionophila	JX971121 [chionophila]*		GQ424806 [chionophila]*			JX971122 [chionophila]*	81	B
Parolinia	ornata			GQ424752 [ornata]*				57	C
Parolinia	intermedia	DQ357577 [intermedia]							
Parrya	nudicaulis					DQ409253 [nudicaulis]	DQ180253 [nudicaulis]	72	AL
Parrya	asperrima	DQ357578 [asperrima]							
Paysonia	stonensis		JQ323062 [stonensis]					6	A
Paysonia	densipila	AF137586 [densipila]							

Pegaeophyton	scapiflorum	DQ518398		GQ424731 [scapiflorum]*			DQ180254		41	IJ
Pegaeophyton	nepalense				JQ933435 [nepalense]*					
Peltaria	alliacea	DQ249855	KF023033 [alliacea]				DQ518375		16	CDEF
Peltariopsis	planisiliqua	GQ424553							4	G
Pennellia	micrantha	AF307629			FN594847 [micrantha]*-				50	AB
Pennellia	brachycarpa		DQ288811 [brachycarpa]							
Pennellia	longifolia						AF307549 [longifolia]			
Petiniotia										
Petrocallis	pyrenaica	GQ497871 [pyrenaica]							29	CE
Petroravenia										
Phaeonychium	villosum	FJ026827 [villosum]-							56	IJ
Phaeonychium	jafrii					DQ409261 [jafrii]				
Phaeonychium	kashgaricum						FN677739 [kashgaricum]*			
Phlebolobium	maclovianum	GQ497873 [maclovianum]							12	B
Phlegmatospermum	eremaeum	JX630166 [eremaeum]			JX630169 [eremaeum]				88	N
Phlegmatospermum	cochlearinum						JX134220 [cochlearinum]-			
Phoenicaulis	cheiranthoides	DQ399121	DQ288812			DQ406768 [cheiranthoides]	JX146092 [cheiranthoides]	JX146663 [cheiranthoides]*	49	A
Physaria	acutifolia	AF137582 [acutifolia]							37	ABKLM
Physaria	fendleri				FN594840 [fendleri]*-					
Physaria	arctica					JN966346 [arctica]*				
Physaria	floribunda		DQ288813 [floribunda]							
Physaria	purpurea			GQ424677 [purpurea]*						
Physaria	brassicoides			EU931382 [brassicoides]						
Physoptychis	caspica	EF514682	KF022994				KF022850		7	G

			[caspica]				[caspica]		
Physorhynchus	chamaerapistrum	JQ911318 [chamaerapistrum]*-							
Planodes	virginica	GQ424554	DQ288814	GQ424681 [virginica]*				27	A
Pleiocardia									
Polyctenium	fremontii	AY230647	DQ288816			DQ406769 [fremontii]		45	A
Polypsecadium	harmsianum	EU620310 [harmsianum]*	EU718561 [harmsianum]-				EU620370 [grandiflorum]*	45	A
Polypsecadium	grandiflorum		EU718560 [graniflorum]-						
Polypsecadium	rusbyi		EU718562 [rusbyi]-						
Polypsecadium	fremontii						JX146116 [fremontii]*		
Pseudanastatica									
Pseuderucaria	teretifolia	GQ497891 [teretifolia]				JN584995 [teregifolia]*		2	CF
Pseudoarabidopsis	toxophylla	AF137558						0	EL
Pseudocamelina	glaucophylla	DQ357581 [glaucophylla]*						3	G
Pseudocamelina	campylopoda		DQ288817 [campylopoda]						
Pseudoclausia	gracillima	FN821530 [gracillima]*					FN677652 [gracillima]*	0	GIL
Pseudofortuynia	esfandiarii	GQ497876 [esfandiarii]						8	G
Pseudosempervivum									
Pseudovesicaria	digitata	GQ497877 [digitata]						15	G
Psychine	stylosa	DQ249835				JN584996 [stylosa]*	AY751768	3	C
Pterygiosperma	tehuelches	EU620311 [tehuelches]*					EU620374 [tehuelches]*	0	B
Pterygostemon									
Ptilotrichum	canescens	EF514687 [canescens]					GU181990 [canescens]	79	I
Pugionium	dolabratum	JF978171 [dolabratum]*			JF943786 [dolabratum]*	JF955862 [dolabratum]*		0	KL

Pugionium	cornutum					JF926645 [cornutum]-				
Pycnoplinthopsis	bhutanica	GQ497878 [bhutanica]							0	J
Pycnoplinthus	uniflorus	GQ497879 [uniflorus]							63	IJ
Quezeliantha										
Quidproquo										
Raffenaldia	primuloides	AY722478				JN584997 [primuloides]*			9	C
Raphanus	raphanistrum					JN584998 [raphanistrum]*	GQ268043 [raphanistrum]-		78	ABCDEFG
Raphanus	sativus	JQ911325 [sativus]*			GQ184382 [sativus]*-					
Rapistrum	rugosum	DQ249825			HM850300 [rugosum]	JN584999 [perenne]*	AY751769		69	ABCDEFG
Rhammatophyllum	pseudoparrya	DQ357583 [afghanicum]	DQ288818 [erysimoides]				FN677742 [kamelinii]*		10	I
Rhizobotrya										
Ricotia	cretica	GQ497880 [cretica]		GQ424766 [cretica]*					71	F
Ricotia	lunaria					GQ424600 [lunaria]				
Robeschia	schimperii	GQ497881 [schimperii]	EU907364 [schimperii]						8	FG
Rollinsia	paysonii									
Romanschulzia	sp.		DQ288819 [sp.]	GQ424653 [sp.]*					54	A
Romanschulzia	costaricensis	AF531636 [costaricensis]								
Romanschulzia	arabiformis						AY958538 [arabiformis]			
Rorippa	amphibia					AF174530 [amphibia]		JQ582800 [amphibia]*	86	ABCDEFG
Rorippa	palustris						EF426789 [palustris]-			
Rorippa	islandica					DQ406770 [islandica]				
Rorippa	palustris					AF144355 [palustris]				
Rorippa	curvipes		AF198138 [curvipes]							

Rorippa	sylvestris							JQ582801 [sylvestris]*		
Rorippa	indica	AF128108			D88907					
Rytidocarpus	moricandioides	AY722483 [moricandioides]				JN585001 [moricandioides]*			7	C
Sameraria	armena			GQ424630 [armena]*					3	G
Sameraria	nummularia	GQ424555 [nummularia]								
Sandbergia	whitedii	JX146965 [whitedii]*				DQ406765 [whitedii]	**JX146117 [whitedii]***	JX146694 [whitedii]*	33	A
Sandbergia	perplexa					DQ406764 [perplexa]				
Sarcodraba	dusenii	GQ424568		GQ424811 [dusenii]*					64	B
Savignya	parviflora	AF263399	KF023017 [parviflora]						11	CFG
Scambopus	curvipes	JX630167 [curvipes]			JX630177 [curvipes]				55	N
Schimpera	arabica	GQ424556		GQ424794 [arabica]*					7	FGH
Schivereckia	podolica	AY134136							2	DF
Schizopetalon	rupestre		DQ288820	GQ424618 [rupestre]*			EU620376 [rupestre]*		13	B
Schizopetalon	angustissimum	KC174369 [angustissimum]*-								
Schoenocrambe	linifolia	AF183089 [linifolia]	DQ288821 [linifolia]						100	A
Schoenocrambe	linearifolia			EU931383 [linearifolia]			AY958541 [linifolia]			
Schouwia	purpurea	AY722500				JN585002 [purpurea]*			4	H
Scoliaxon	viereckii	HQ541180 [viereckii]								
Selenia	dissecta	GQ424557	DQ288822				AM072852 [dissecta]		16	A
Shangrilaia	nana	GQ424558	DQ288823						100	J
Sibara	rosulata	AF531648							34	AB
Sibara	pectinata		EU718570 [pectinata]-							

Sibara	deserti		EU718568 [deserti]-						
Sibara	laxa		EU718569 [laxa]-						
Sibara	angelorum						EU620379 [angelorum]*		
Sibaropsis	hammittii	GQ424559	EU718571 [hammitii]-				EU620380 [hammitii]*	24	A
Sinapidendron	angustifolium					JN585003 [angustifolium]*		4	C
Sinapidendron	frutescens	DQ249823					AY751771		
Sinapis	arvensis	DQ249828 [arvensis]-						84	ABCDEFG
Sinapis	alba				HM849823 [alba]	AB354277 [alba]-	JQ041854 [alba]*		
Sinosophiopsis	bartholomewii	AY230609 [bartholomewii]						#DIV/0!	J
Sisymbrella	aspera	GQ424560			HM850357 [aspera]	HM850739 [aspera]		46	CE
Sisymbriopsis	mollipila	AY353157	DQ288824					82	IJ
Sisymbriopsis	yechengnica				JN847809 [yechengnica]	JF926640 [yechengnica]			
Sisymbrium	heteromallum			GQ424703 [heteromallum]*				85	ABCDEFG
Sisymbrium	irio				AY167982 [irio]	AF144366 [irio]	AY167982 [irio]		
Sisymbrium	altissimum		DQ288826 [altissimum]			JN585004 [altissimum]			
Sisymbrium	altissimum					JN585004 [altissimum]*			
Sisymbrium	frutescnes		DQ288827 [frutescens]						
Sisymbrium	officinale				HM850358 [officinale]				
Smelowskia	jacutica	AY230646	DQ288774					61	AIJL
Smelowskia	altaica			EU931360 [altaica]	FN594836 [altaica]*-				
Smelowskia	calycina	AY230640	DQ288828	EU931386 [calycina]	FN594837 [calycina]		DQ180249		
Smelowskia	flavissima	AY230611							
Smelowskia	bartholomewii	AY230609 [bartholomewii]		GQ424695 [bartholomewii]*	JQ933355 [tibetica]*				

Smelowskia	tibetica				JN847827 [Hedinia taxkargannica=Smelowskia tibetica]					
Smelowskia	tibetica					JF926647 [taxkargannica]-				
Smelowskia	taxkargannica									
Sobolewskia	caucasica	GQ424561								
Sobolewskia	annua		DQ288831 [annua]							
Sobolewskia	jacutica			GQ424786 [jacutica]*						
Sobolewskia	tibetica	AY230607 [tibetica]				JF953887 [tibetica]*				
Solms-laubachia	jafrii	DQ523422			JQ933442 [jafrii]*	DQ409261		64	IJ	
Solms-laubachia	eurycarpa						DQ523304 [eurycarpa]*			
Solms-laubachia	linearis				JQ933299 [linearis]*					
Solms-laubachia	zhongdianensis	DQ523415	DQ288830			DQ409250				
Solms-laubachia	flabellata	GQ424562								
Sophia										
Sphaerocardamum	stellatum		JQ323091 [stellatum]				AF307531 [stellatum]	41	A	
Sphaerocardamum	macropetalum	AF137589 [macropetalum]								
Spirorhynchus	sabulosus	FM164600/FM164601 [sabulosus]*						2	G	
Spryginia	winkleri	GQ424563						0	GI	
Spryginia	falcata						FN677740 [falcata]*			
Stanleya	pinnata	AF531620	DQ288832	GQ424647 [pinnata]*	AJ235809 [pinnata]	AY483226	EU620381 [pinnata]*	31	A	
Stanleya	tomentosa		EU718576 [tomentosa]-							
Stenopetalum	decipiens	GQ424564		GQ424743 [decipiens]*				95	N	
Stenopetalum	nutans		DQ288833 [nutans]		FN594848 [nutans]*-					

Stenopetalum	velutinum				FN594850 [velutinum]*-				
Sterigmostemum	violaceum	DQ357576	DQ288808		JN847818 [violaceum]		FN677640 [violaceum]*	3	G
Sterigmostemum	matthioloides					JF926655 [matthioloides]			
Sterigmostemum	acanthocarpum	DQ357591	DQ288834 [acanthocarpum]						
Stevenia	cheiranthoides	GU182059 [cheiranthoides]-						0	I
Stevenia	canescens		KF023032 [canescens]						
Stevenia	axillaris						FN677639 [axillaris]*		
Straussiella									
Streptanthella	longirostris	AF531621	EU718579 [longirostris]-				EU620383 [longirostris]*	33	A
Streptanthus	cordatus	EU620323 [cordatus]*						32	A
Streptanthus	barbatus		EU718581 [barbatus]-						
Streptanthus	barbiger		EU718583 [barbiger]-						
Streptanthus	carinatus				HE616651 [carinatus]*				
Streptanthus	squamiformis		DQ288835 [squamiformis]						
Streptanthus	sparsiflorus			EU931387 [sparsiflorus]					
Streptanthus	glandulosus				FN594831 [glandulosus]*-				
Streptanthus	tortuosus						EU620387 [tortuosus]*		
Streptoloma	desertorum	FM164618, FM164619						0	G
Strigosella	africana	DQ357557					DQ518373	45	ABCEFGIJ
Stubendorffia	lipskyi			GQ424682 [lipskyi]*				17	I
Stubendorffia	gracilis	DQ780944/DQ780945 [gracilis]					AM991729 [gracilis]		
Subularia	monticola	GQ424565 [monticola]						85	ACEHL
Succowia	balearica	AF263395						36	CDN

Synstemon	petrovii	DQ357599 [petrovii]*						#DIV/0!	H
Synthlipsis	greggii	AF137590	JQ323083 [greggii]-	GQ424753 [greggii]*	FN594841 [greggii]*-			42	A
Syrenia	cuspidata	DQ249864 [cuspidata]-					DQ518376 [cuspidata]	0	I
Tauscheria	lasiocarpa	DQ249843				JF926657 [lasiocarpa]	DQ518377	30	I
Tchihatchewia	isatidea	GQ497882 [isatidea]						0	F
Teesdalia	nudicaulis	AF336214/AF336215 [nudicaulis]				GQ424601 [nudicaulis]		82	ABCDEFN
Teesdaliopsis	coronopifolia				HE616649 [coronopifolia]				
Tetracme	quadricornis	DQ357602	DQ288837 [pamirica]		FN594832 [quadricornis]		DQ518378	44	GI
Thelypodiopsis	purpusii			GQ424758 [purpusii]*				39	A
Thelypodiopsis	ambigua	AF531625 [ambigua]							
Thelypodiopsis	vermicularis		EU718604 [vermicularis]-						
Thelypodiopsis	elegans						EU620391 [elegans]*		
Thelypodium	laciniatum		DQ288838					42	A
Thelypodium	wrightii	EU620329 [wrightii]*							
Thelypodium	sagittatum						EU620393 [sagittatum]*		
Thlaspi	arvense		DQ288839		FN594829 [arvense]*-	GQ424602	DQ821410 [partial sequence used]	86	ABCDEFG
Thlaspi	perfoliatum	AY154810 [perfoliatum]							
Thysanocarpus	curvipes	AY254542	EU718613 [curvipes]-	GQ424655 [curvipes]*			EU620394 [curvipes]*	35	A
Torularia									
Trachystoma	ballii	AY722492					AB670034 [ballii]-	8	C
Trachystoma	labasii					JN585006 [labasii]*			
Transberingia									A
Trichotolinum								52	

Tropidocarpum	gracile	GQ497883 [gracile]	AY514394 [gracile]						35	A
Turrita										
Turritis	glabra	AJ232922	DQ288840		HQ589958 [glabra]*-	AF144333	DQ518389	AF110439	80	ACDEFGH
Turritis	brassica							AF110451 [brassica]		
Vania	campylophylla	AF336168/AF336169							0	FG
Vella	aspera					JN585007 [aspera]*			64	C
Vella	pseudocytisus	AF263393			GQ248705 [pseudocytisus]*-					
Vella	spinosa						AY751773 [spinosa]			
Veselskya										
Warea	sessilifolia	AF531644								
Warea	cuneifolia		EU718615 [cuneifolia]-				EU620398 [cuneifolia]*		9	A
Weberbauera	colchaguensis	KC174373 [colchaguensis]*-							32	B
Weberbauera	herzogii		EU718616 [herzogii]-							
Weberbauera	dillonii			GQ424816 [dillonii]*						
Weberbauera	peruviana						EU620402 [peruviana]*			
Werdermannia	mendocina	EU620338 [mendocina]*	EU718620 [mendocina]-				EU620404 [mendocina]*		42	B
Winklera	silaifolia	AJ628321/AJ628322 [silaifolia]-					AM991732 [silaifolia]		62	I
Xerodraba	pecinata	GQ497884 [pecinata]							56	B
Yinshania										
Zilla	spinosa	AY722501				JN585009 [spinosa]*	DQ984122 [spinosa]		38	CFGH
Zuvanda	crenulata	DQ357606								CF
Batis	maritima		EU002199		M88341	AY483219				
Capparis	flexuosa	JQ229789 [tomentosa]-	EU002208		M95754 [hastata]	EU371760				
Carica	papaya	AY461547 [papaya]	AY483248		M95671	AY483221	DQ061124 [papaya]			ABHJKMN

Cleome	viscosa	GQ470549 [spinosa]*-	DQ288755 [rutidosperma]	AY483268 [viridiflora]	EU371806		
Floerkea	proserpinacoides			L12679	EU002178		
Moringa	oleifera	JX092069 [oleifera]-	AY122405	L11359	AY483223	JX091844 [ovalifolia]	ABKMN
Reseda	lutea	DQ987089 [crystallina]*	EU002256	AY483273	AY483241	DQ987017 [lutea]*	
Tovaria	pendula	DQ987073 [pendula]*	AY122407	M95758	AY483242		
Tropaeolum	minus	JN115053 [majus]*-	EU002270	AB043534	AY483224		

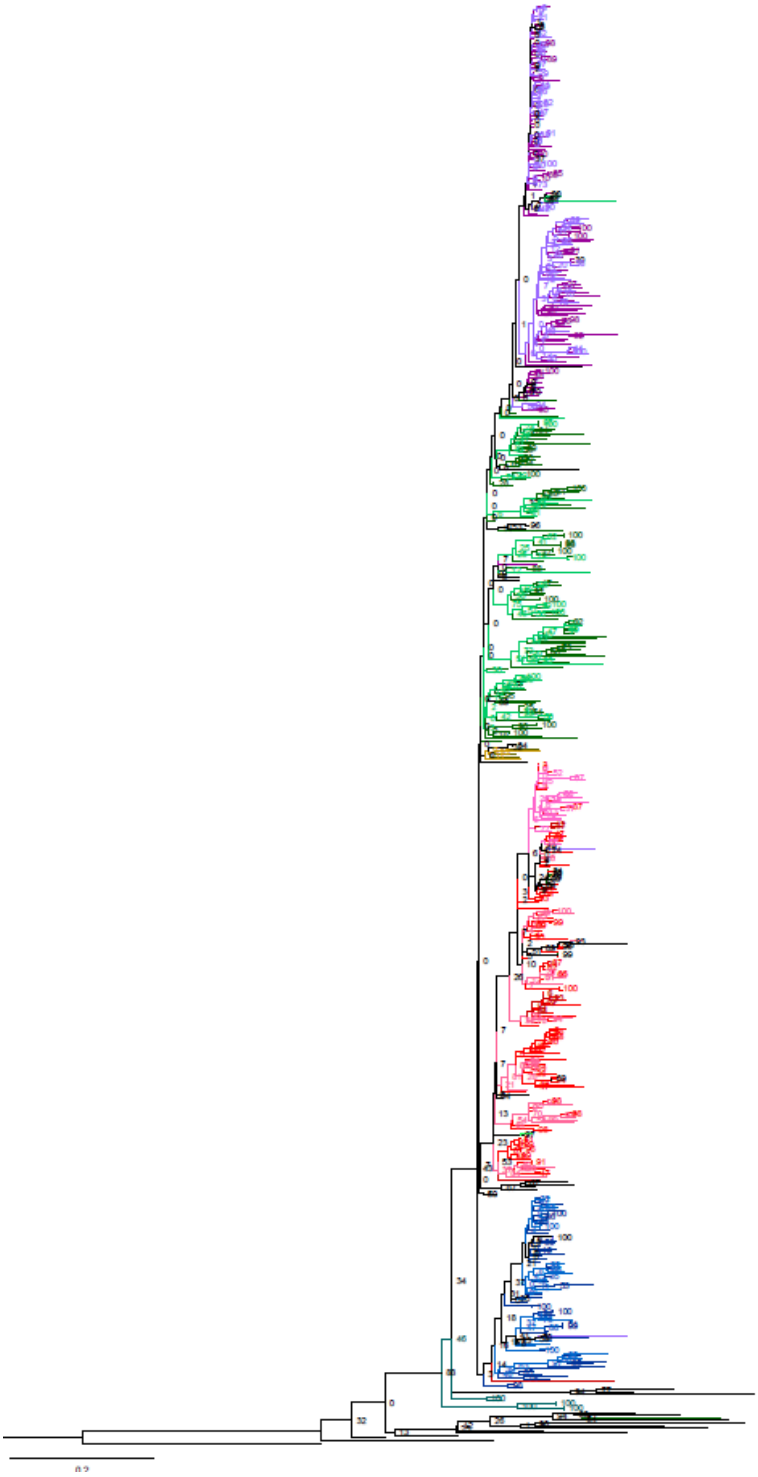
APPENDIX B. EXEMPLARY SCRIPT USED IN MRBAYES

```
begin mrbayes;

        [set autoclose=no nowarn=yes; ]

        [outgroup ; ]

 [codon positions if you wish to use these]

        charset 1st_pos =  3 - 1959\3;

        charset 2nd_pos =  1 - 1960\3;

        charset 3rd_pos =  2 - 1961\3;

partition by_codon = 2:1st_pos 2nd_pos, 3rd_pos;

        set partition = by_codon;

        lset applyto=(all) nst=mixed rates=gamma;

        [prset clockvarpr = ccp; ]

        unlink shape=(all) pinvar=(all) statefreq=(all) revmat=(all);

        prset applyto=(all) ratepr=variable;

        mcmc samplefreq=10000 printfreq=10000 temp=0.2 ngen= 50000000 relburnin=yes
burninfrac=0.25 savebrlens=yes;

        mcmc;

        sump;

        sumt;

end;
```

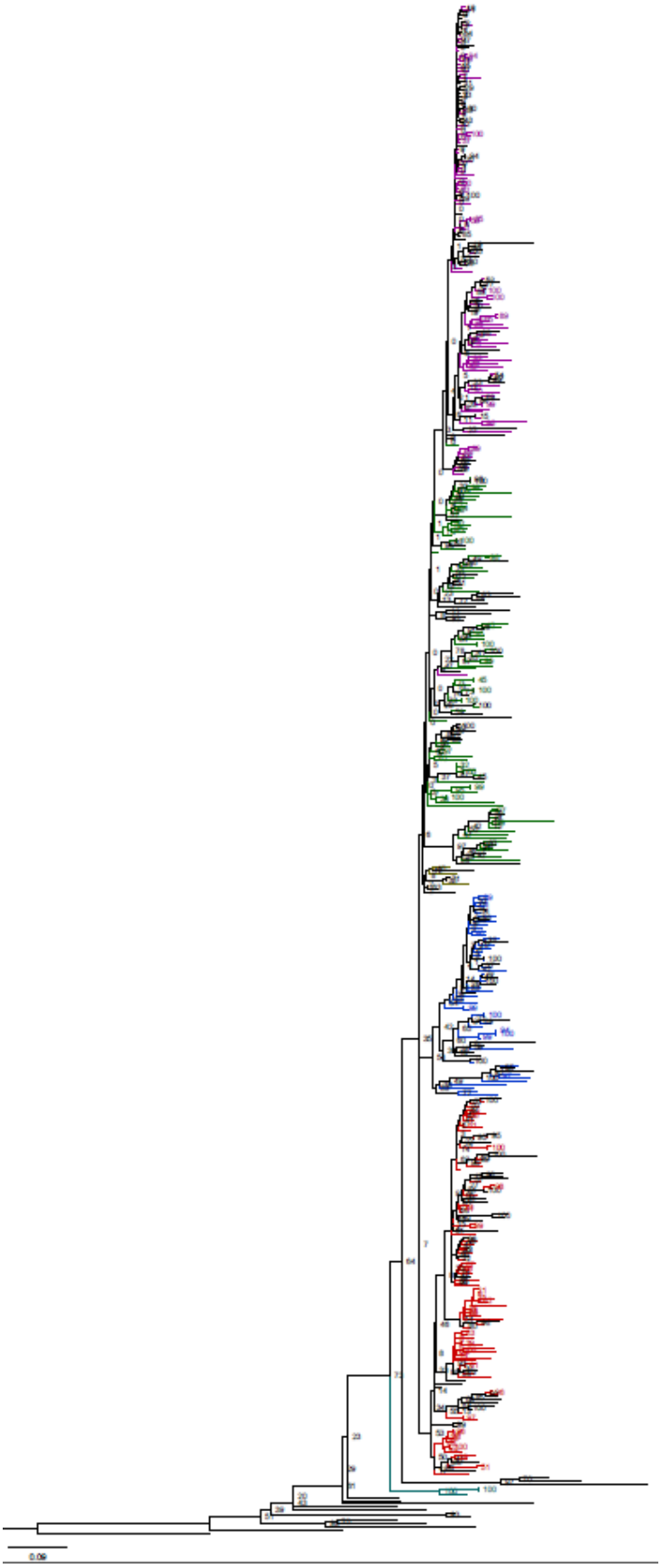
APPENDIX C. PHYLOGENETIC TREES

Colour indicates lineage: Lineage I: red; Lineage II: purple; Extended Lineage II: green; and Lineage III: blue.

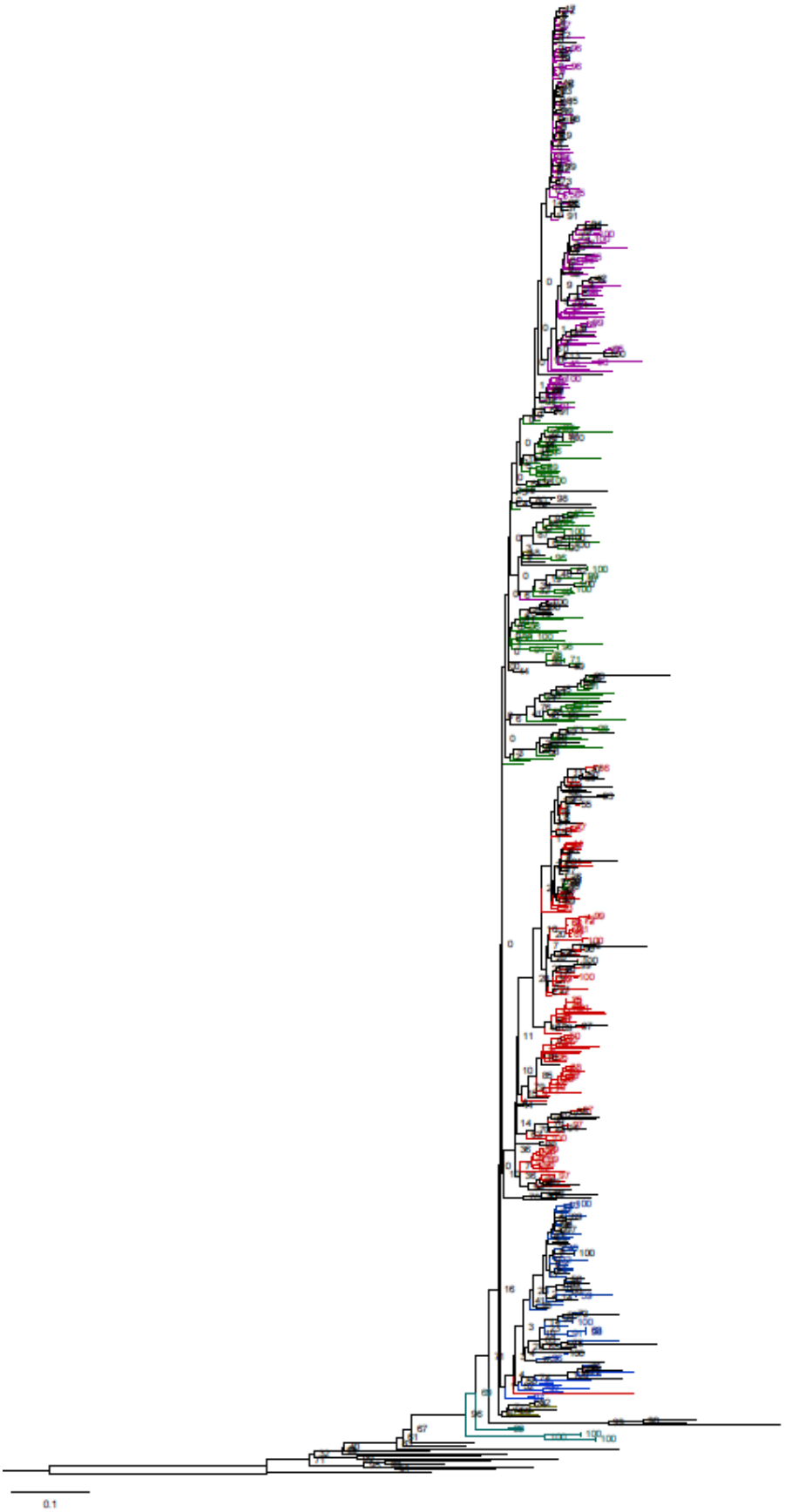
1. RAXML BOOTSTRAP MAFFT – EIGHT GENES



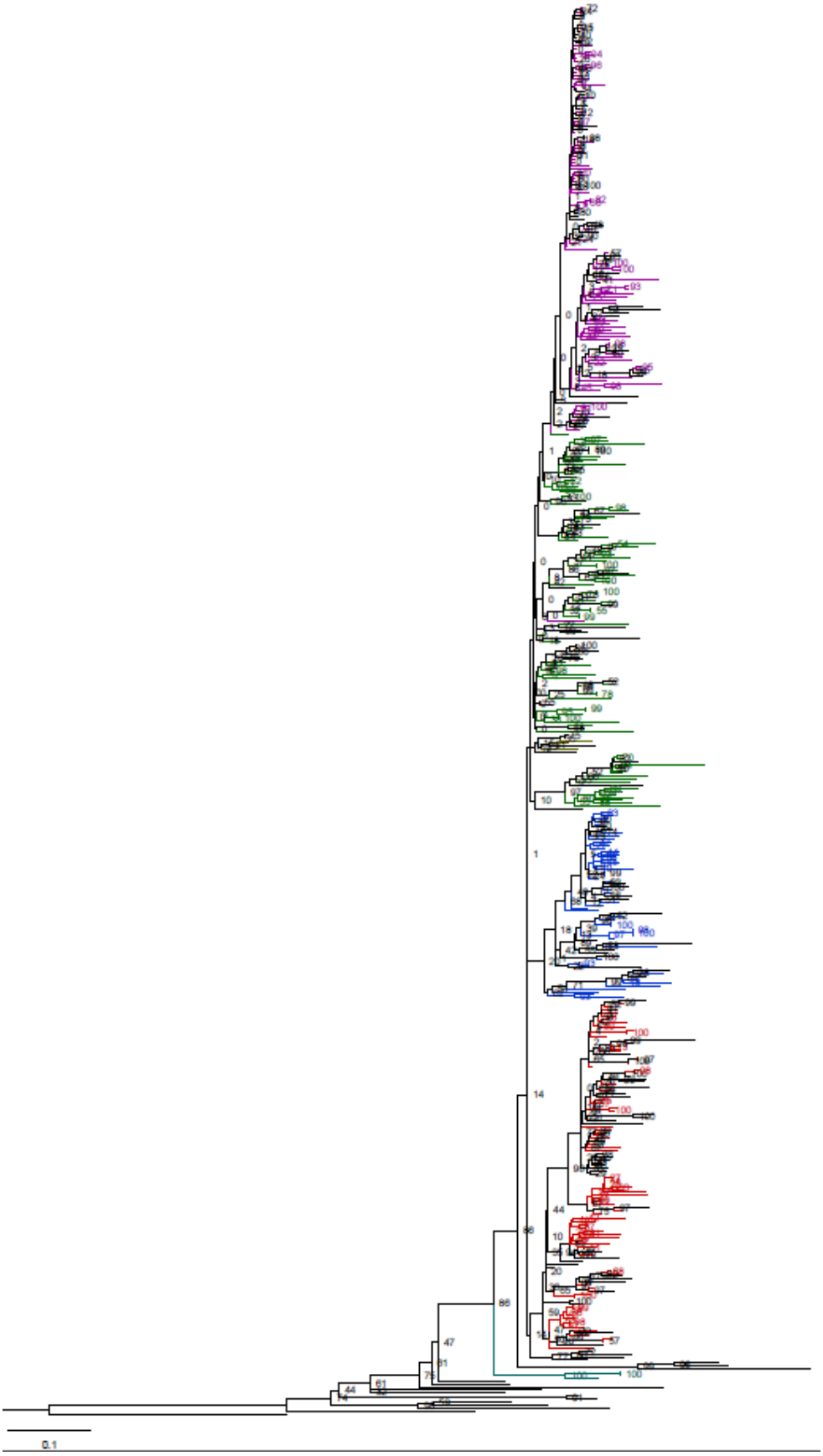
2. RAxML Bootstrap MAFFT – Four genes



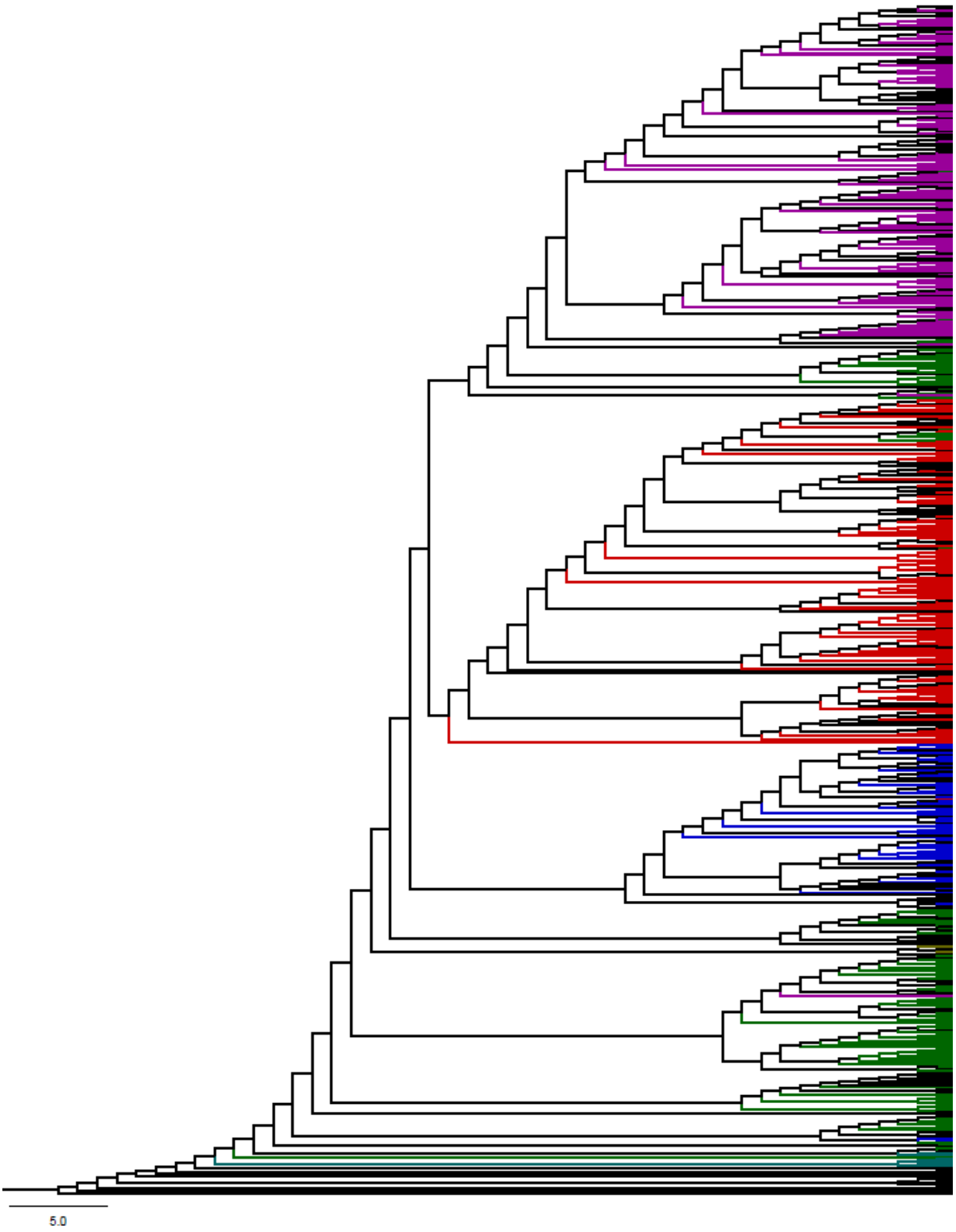
3. RAxML Bootstrap MAFFT_Slow – Eight genes



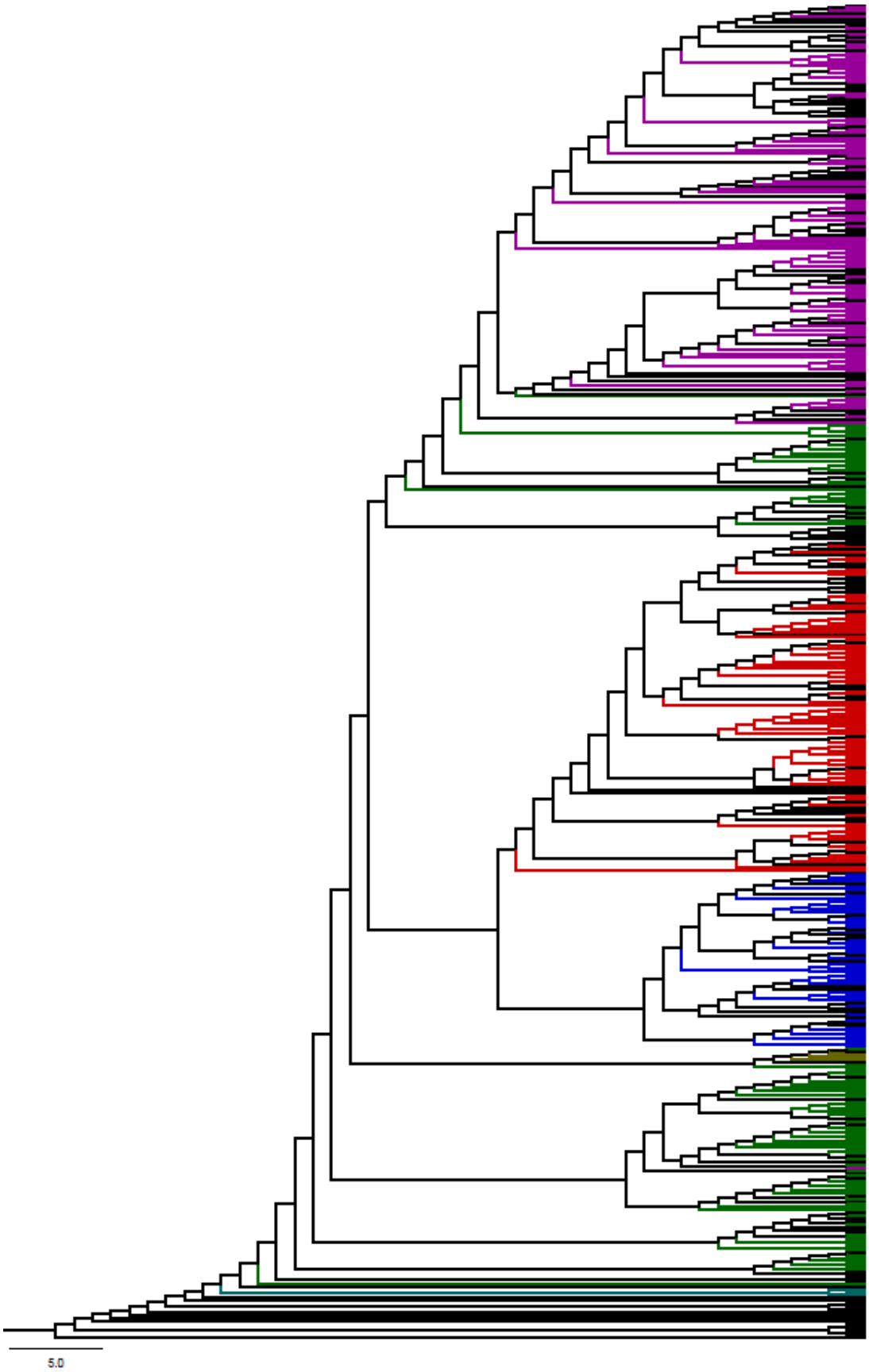
4. RAxML Bootstrap MAFFT_Slow – Four genes



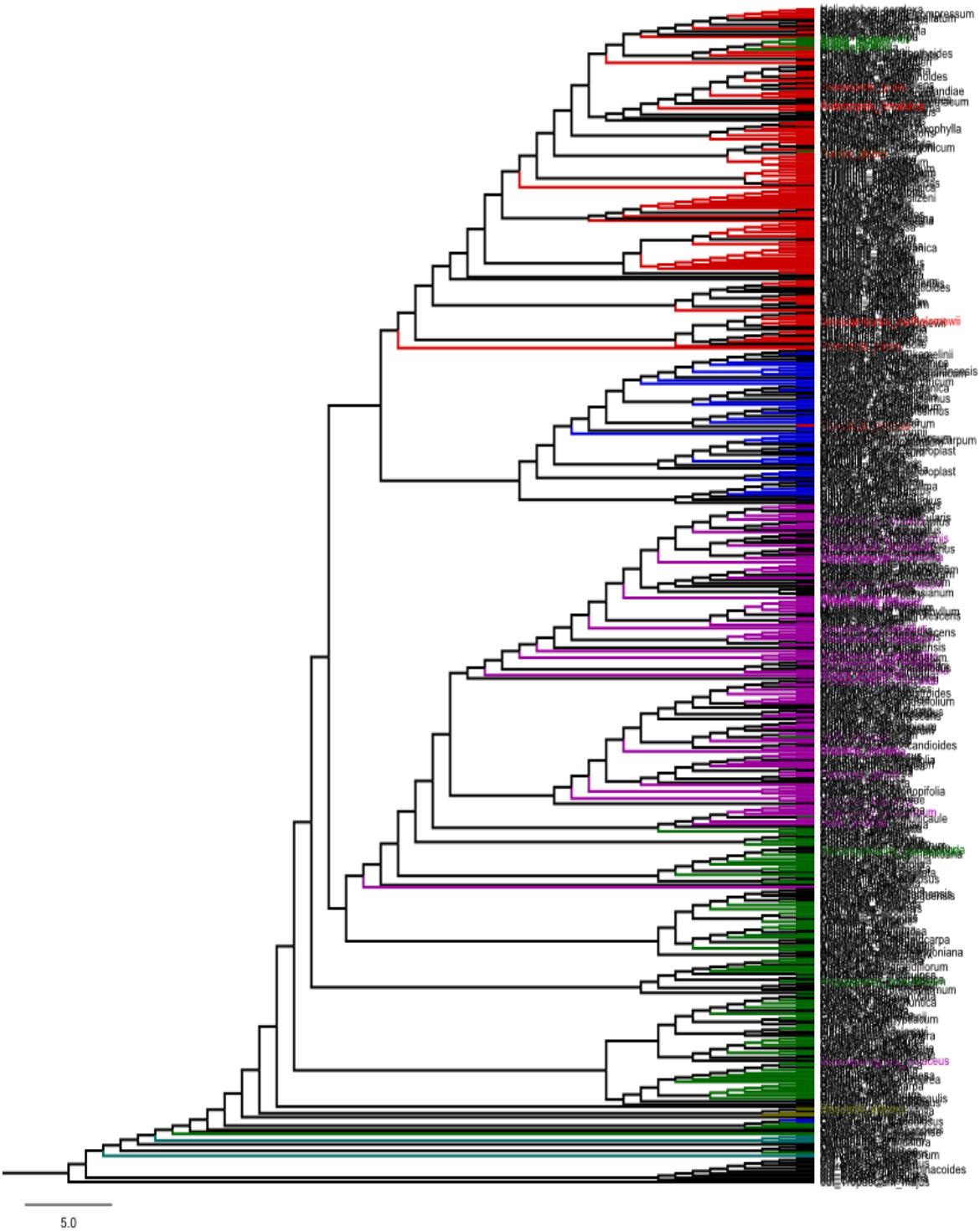
5. TNT Sectorial & Ratchet MAFFT – eight genes



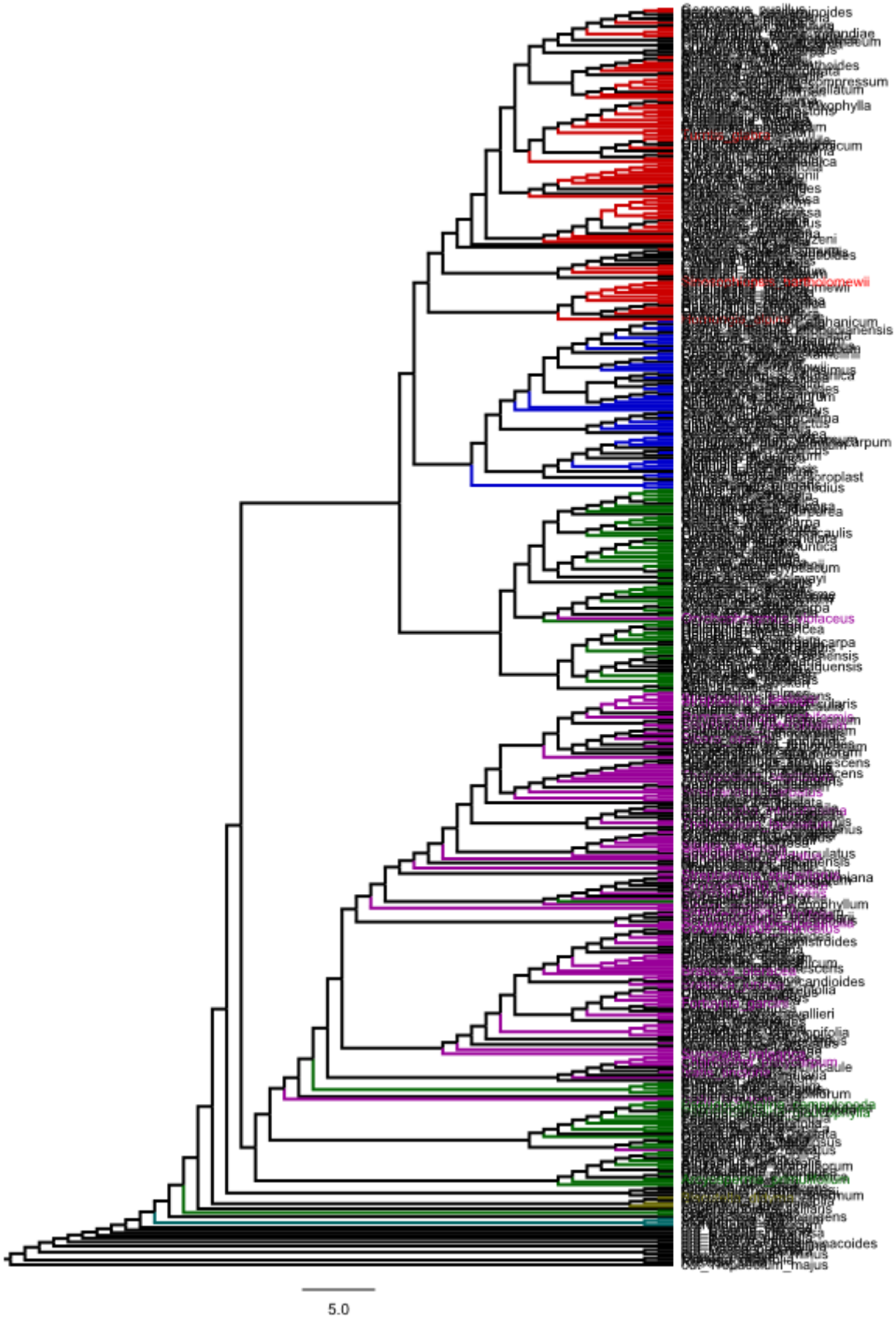
6. TNT Sectorial & Ratchet MAFFT – Four genes



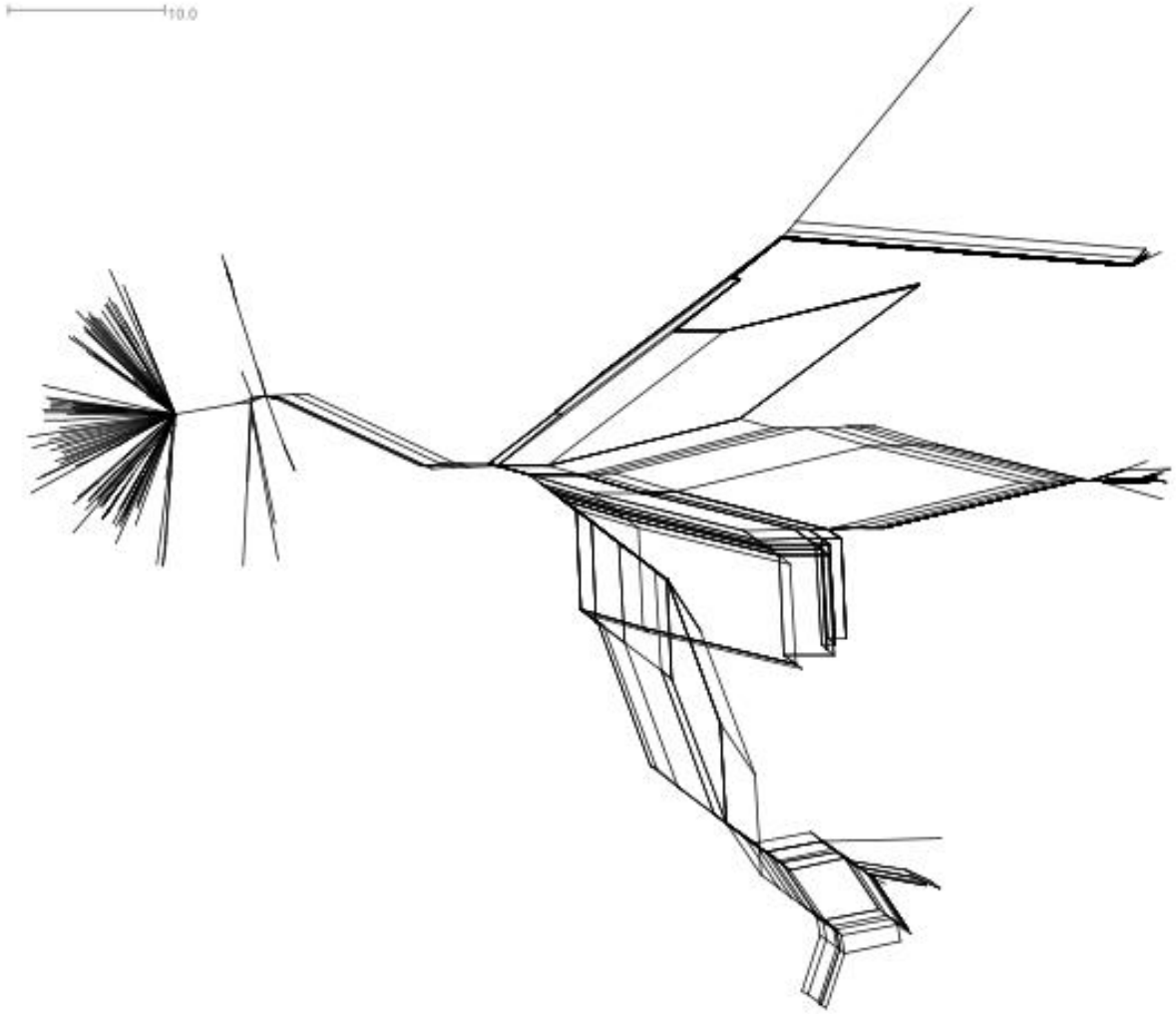
7. TNT Sectorial & Ratchet MAFFT_slow – Eight genes



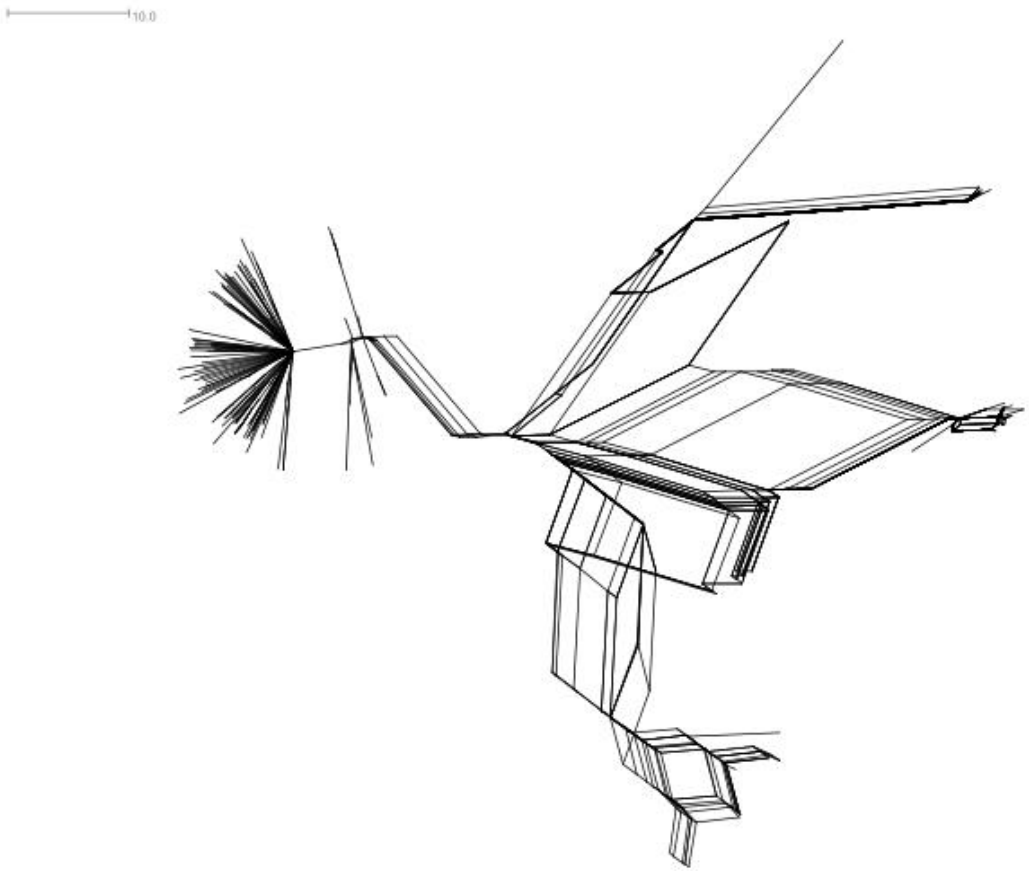
8. TNT Sectorial & Ratchet MAFFT_slow – Four genes



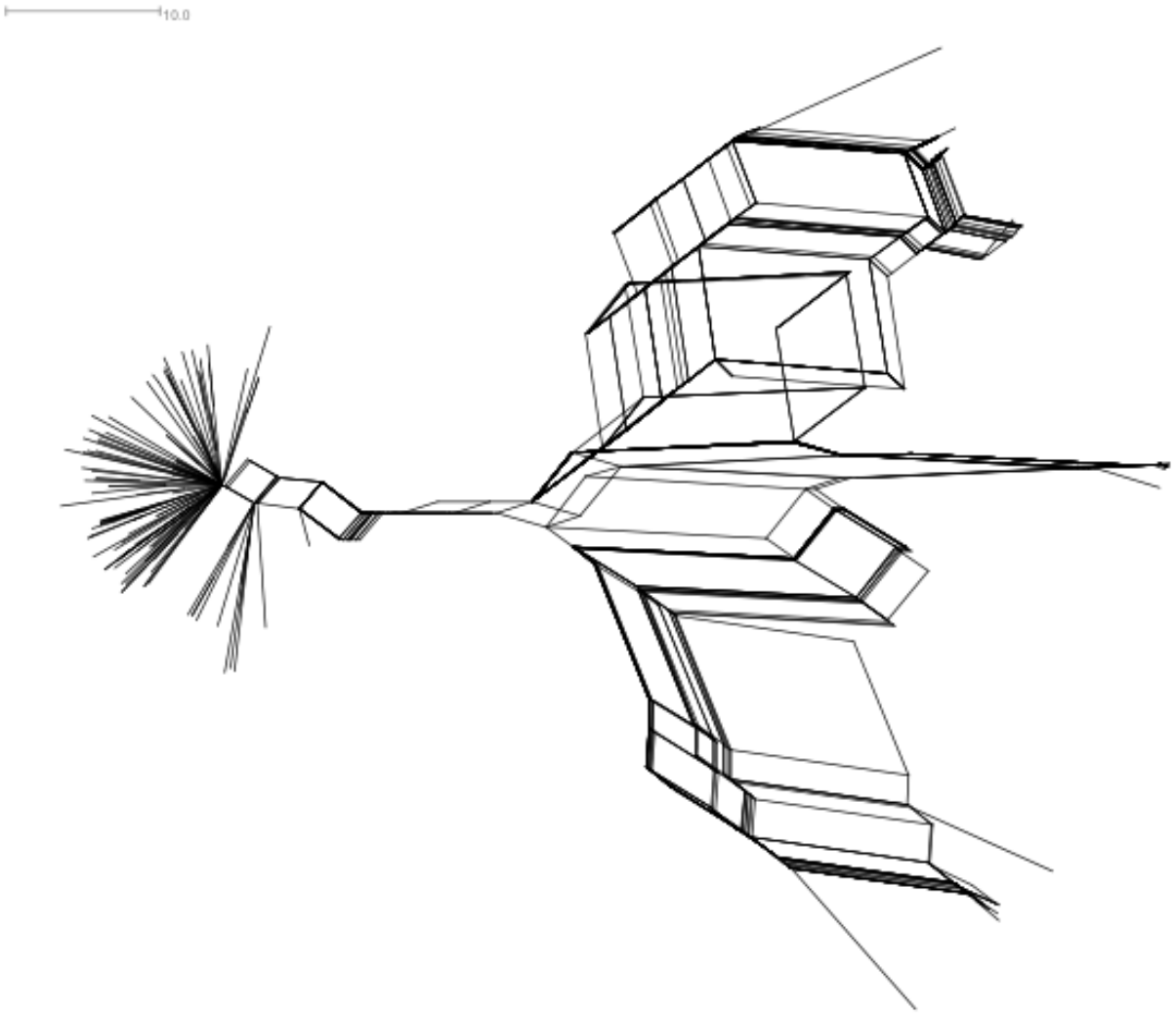
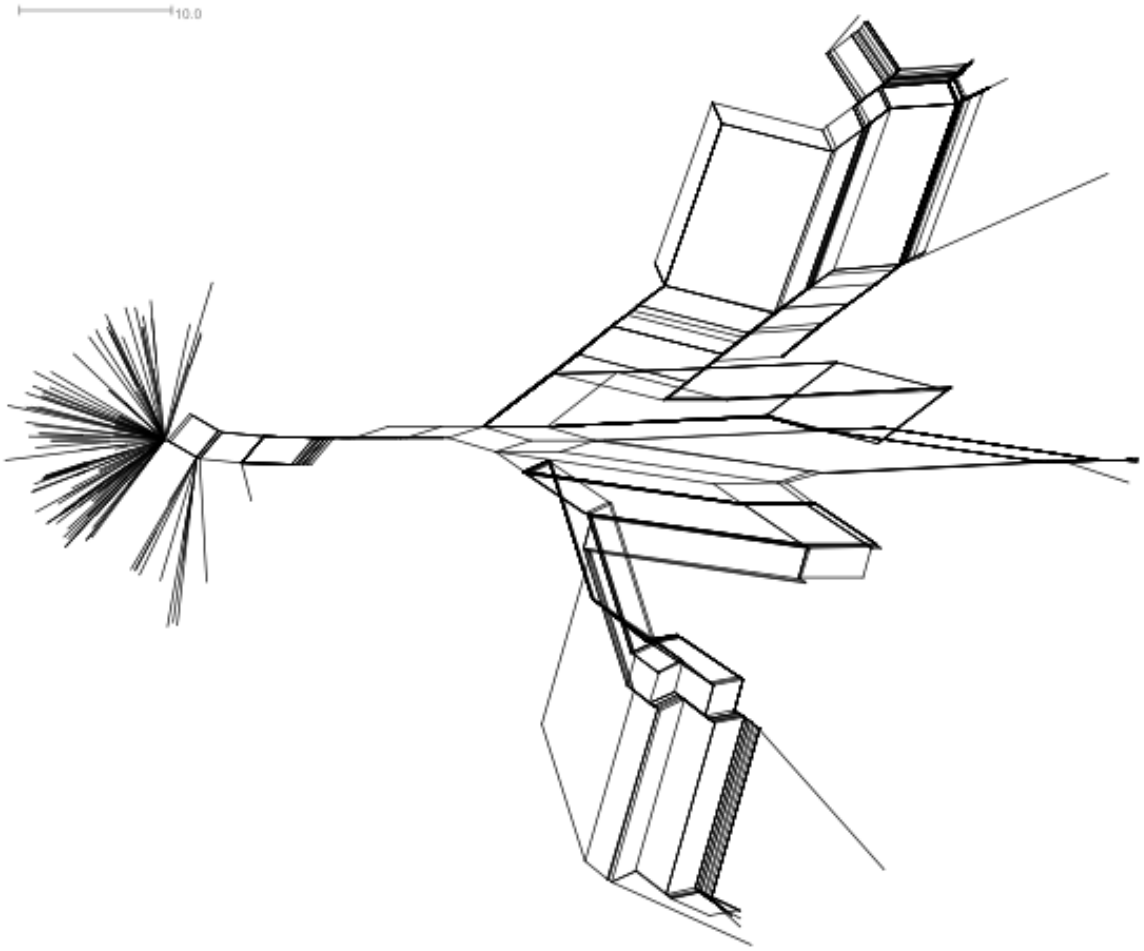
9. Neighbour Network GTR+Gamma MAFFT – Eight genes



10. Neighbour Network GTR+Gamma MAFFT – Four genes



11. Neighbour Network GTR+Gamma MAFFT_slow – Eight genes



12. Neighbour Network GTR+Gamma MAFFT_slow – Four genes