CHAPTER 7: MOLECULAR STUDY

7.1 Introduction

One of the main goals of biological sciences is to determine life's history of descent (Simpson, 2006), and according to Stuessy (1997), an understanding of the diversity of nature is based on evolutionary relationships, that is, phylogeny (Simpson, 2006; Judd *et al*., 2008). Phylogenetic hypotheses are preferably generated by DNA sequences, as many characters are available and their identity is easy to define (Judd *et al*., 2008).

7.1.1 Phylogeny of the Nyctaginaceae

The family was first studied phylogenetically by Levin (2000). She studied the phylogeny of tribe Nyctagineae Horan, and concluded that the division of this tribe into subtribes Boerhaviinae Benth. & Hook and Nyctaginineae was not justified as *Boerhavia* L. (subtribe Boerhaviinae) and *Allionia* L. (subtribe Nyctaginineae) are closely related; also that *Mirabilis* L. was monophyletic, and that *Acleisanthes* A.Gray and *Selinocarpus* A.Gray formed a monophyletic group. *Selinocarpus* was therefore transferred to *Acleisanthes* (Levin, 2002). This was followed by a study which included nearly all of the accepted genera (Douglas & Manos, 2007) and showed that the classification of the family as compiled by Bittrich and Kühn (1993) wasnot in accordance with the evolutionary relationships demonstrated by molecular evidence, leading to a reclassification of the family into seven monophyletic tribes (Douglas & Spellenberg, 2010) (Table 1.1).

7.1.2 Phylogeny of *Boerhavia* and *Commicarpus*

Boerhavia and *Commicarpus* Standl. were previously classified in subtribe Boerhaviinae (Bittrich & Kühn, 1993). The study of Douglas & Manos (2007) showed the two genera to be monophyletic and they were reclassified into tribe Nyctagineae (Douglas & Spellenberg, 2010).

The phylogenetic relationships of the southern African *Boerhavia* and *Commicarpus* species have not been studied previously. Morphological analyses (Chapters 4–6) have indicated that *Boerhavia* and *Commicarpus* are separate genera. This distinction has been indicated by the phylogenetic study of Douglas and Manos (2007), but this division still needs to be tested at a molecular level for southern African *Boerhavia* and *Commicarpus*.

In addition, some morphological data was inconclusive as to the identity of some species. For instance, *C. squarrosus* and *C. fruticosus* are morphologically indistinguishable except for flower colour and *Boerhavia* specimens that are similar in growth form and foliage to *B. coccinea* var. *coccinea*, have uncharacteristically large, clavate anthocarps. The identity of these two taxa could not be resolved in the morphological chapters, and molecular work could contribute additional variation for the species circumscriptions.

7.1.3 Aim

The aim of this chapter is to construct a phylogeny for the southern African *Boerhavia* and *Commicarpus* species using the ITS nuclear region and chloroplast *ndh*F gene, to gather further evidence that *Boerhavia* and *Commicarpus* are separate genera in southern Africa with well defined species, and to investigate the identity of questionable species.

7.1.4 Molecular regions

The regions used in phylogenetic studies need to be easily amplifiable and evolve relatively rapidly, but they should also be easy to align and provide sufficient and suitable variation within a short sequence segment (Baldwin *et al*., 1995).

Levin (2000) used the internal transcribed spacer (ITS) region of the nuclear rDNA, the 3' end of the *rbcL*, the intergenic region between the 3' *rbcL* gene and the *accD* gene, and the *accD* 5' coding region of the chloroplast genome in her phylogenetic study of the tribe Nyctagineae. She found that the ITS region evolved faster and was phylogenetically more informative than the region between the *rbc*L and *accD* or the

5'*accD* region. Douglas & Manos (2007) used the ITS region as well as the *ndh*F, *rps*16 and *rpl*16 region of the chloroplast genome in their analysis of the phylogeny of the family. They found that the chloroplast regions (*ndh*F, *rps*16 and *rpl*16) were more informative as the ITS1 and ITS2 regions were too variable to align and needed to be excluded (Douglas & Manos, 2007).

The ITS region of the 18S-5.8S-26S nuclear ribosomal DNA region is widely used in phylogenetic studies (Magee *et al*., 2008;Cruz-Mazo *et al*., 2009; Milyutina *et al*., 2010; Pirie *et al*., 2011). It is part of the nuclear rDNA transcript, but is not incorporated into the ribosomes and plays a role in the maturation of nuclear rRNA (Baldwin *et al*., 1995; Soltis & Soltis, 1998). It can be used to infer the phylogenetic relationship between species and closely related genera (Baldwin *et al*., 1995; Soltis & Soltis, 1998) because it is a highly repeated region in the genome. The region undergoes rapid concerted evolution and the small size and the highly conserved sequences flanking each of the spacers make this region easy to amplify (Balwin *et al*., 1995).

The *ndh*F gene is a rapidly evolving chloroplast gene which encodes subunit F of NADH dehydrogenase (Olmstead & Sweere, 1994; Soltis & Soltis, 1998). The 5' region (1380 bp) is different from the 3' region (855 bp) in that the 3' region is A + T rich, has higher levels of non-synonymous base substitutions and higher transversion bias at the codon positions. The different patterns of base substitution at the 5' and 3' region makes this gene ideal for phylogenetic reconstruction as the conserved and variable segments can be used for older and recent groups respectively (Kim & Jansen, 1995).

7.1.5 Phylogenetic methods

There are various methods, models, statistical calculations and algorithms used for the constructing of trees, and a brief overview of only those methods used in this chapter will be given.

Phylogenetic methods are grouped as distance [Neighbour-joining (NJ) or Unweighted Pair-Group Method with Arithmetic mean (UPGMA)] or character-based

methods [Maximum Parsimony (MP) or Maximum Likelihood (ML)] (Page & Holmes, 1998; Wright, 2009). In the distance methods (NJ and UPGMA), the aligned sequences are first converted into a pairwise distance matrix (where an evolutionary distance is calculated for every pair of sequences) and the resulting matrix is then used as the basis for constructing the phylogenetic tree using a clustering algorithm (Page & Holmes, 1998; Wright, 2009). In the character-based methods (MP and ML), each site in the alignment is considered individually and the methods search for the tree that best synthesizes all this information (Page & Holmes, 1998; Wright, 2009).

7.1.5.1 Distance method – Neighbor-joining (NJ)

In this method, the sequences are clustered successively, placing each sequence at that position which will yield the shortest tree. It is a rapid method which produces a single tree (Saitou & Nei, 1987; Nei & Kumar, 2000).

7.1.5.2 Character-based method – Maximum likelihood (ML)

This method calculates a likelihood value (log likelihood score) for each possible tree for a given set of sequences and then selects the tree with the highest likelihood (that is, the tree with the lowest negative log likelihood score) based on a specified model of evolution (Felsenstein, 1981; Wright, 2009).

7.1.5.3 Models of evolution

Evolutionary distances between two sequences become obscure over time due to multiple substitutions at the same site and a count of the distances between two sequences will underestimate how much evolution has actually occurred. Models of evolution try to estimate the true difference between two sequences based on their current states (Nei & Kumar, 2000; Wright, 2009). In this study four models of evolution has been applied (Table 7.1).

7.1.5.4 Heuristic search model –Close-Neighbour-Interchange

This is a method by which a temporary tree is first produced and then all the topologies that differ from the temporary tree are examined with a topological distance of $d_T = 2$ and 4 (Nei & Kumar, 2000).

7.1.5.5 Tree reliability – Bootstrap support

This is a statistical method of measuring confidence intervals in the various branching patterns in the trees. Pseudoreplicates of the data are built and phylogenetic trees are built from these data, and the frequencies with which individual groups are found among all the pseudoreplicates are recorded. A grouping that is found in all the pseudoreplicate trees has a 100% bootstrap support. Low bootstrap support means that the grouping is found in different places in different trees (Felsenstein, 1985).

7.1.5.6 Test for topological similarity between trees – congruency index (I_{cong})

Before matrices obtained from different trees are combined and hypotheses of evolutionary scenarios are made, it is useful to determine the similarity or congruence between the topology of the different trees. The topology of different trees can be compared by testing the null hypothesis (tree topology is more similar than expected by chance) (De Vienne *et al*., 2007). De Vienne *et al*. (2007) developed a topological congruence index which provides a P-value for the null hypothesis. This index tests for the congruence between trees based on the maximum agreement subtree (MAST), which is the largest possible tree compatible with two given trees (Finden & Gordon, 1985; De Vienne *et al*., 2007). The index is quantified by the number of leaves in the MAST for a pair of trees, and the larger the number, the more congruent are the trees (De Vienne *et al*., 2007)

7.2 Materials and Methods

7.2.1 Collecting of plant material

Fresh leaf material collected *in situ* during 2009 and 2010 in Namibia and South Africa were used for the DNA analysis. Young leaves were placed in 50 ml centrifuge tubes with 10 g of silica gel. Voucher specimens of all the collected material were deposited in the A.P. Goossens Herbarium (PUC), Potchefstroom, South Africa (Tabel 7.2). Duplicates of specimens collected in South Africa were deposited in the National Herbarium, Pretoria (PRE), South Africa and duplicates collected in Namibia in the National Herbarium (WIND), Windhoek, Namibia (Tabel 7.2).

DNA extraction, PCR amplification and the sequencing reactions were done with the DNA Sequencer of the Central Analytical Facilities, Stellenbosch University.

7.2.2 DNA extraction

Leaf tissue (1 mm^2) was homogenised with the TissueLyzer (Qiagen) to a fine powder. DNA was extracted using a CTAB (Cetyl trimethyl ammonium bromide) method of Saghai-Maroof *et al.* (1984) with modifications. The powdered leaf tissue was incubated with 500 µl of Buffer PL1 (CTAB-based; NucleoSpin Plant II Kit, Separations) with 2 µl of proteinase K (10 mg/ml; Sigma-Aldrich) at 60 °C overnight. Thereafter the suspension was extracted with chloroform:iso-amylalcohol (24:1). The phases were separated by centrifuging at 20 000 rcf (4 °C) for 10 min and transferred to a new tube. This was repeated twice. DNA was precipitated with ⅔ volume of ice cold isopropanol and incubated overnight at –20 °C. DNA was pelleted at 16 100 rcf (4 °C) for 20 min. The pellets were washed with 200 µl of 70% ethanol twice and dried at 55 °C for 15 min. The DNA was re-suspended in miliQH_2O.

7.2.3 Polymerase Chain Reaction (PCR) amplification and sequencing

PCR amplification and sequencing were done using six primers (Tabel 7.3). Twenty nanogrammes of DNA template were added in a reaction volume of 20 µl containing

1x KAPA ReadyMix (Kapa Biotech/Lasec) and 4 pmol of each primer (ITS4F and ITS5a). PCR was performed using a Verity (Applied Biosystems) with the following cycling conditions: 95 °C for 5 min followed by 30 cycles at 95 °C for 30 sec, 50 °C for 60 sec, and 72 °C for 60 sec and a final extension at 72 °C for 10 min.

Forty nanogrammes of DNA template were added to a reaction volume of 20 µl containing 1x KAPA HiFi ReadyMix (Kapa Biotech/Lasec) and 4 pmol of each primers (Nyct_ndhF1F, Nyct_ndhF13R, Nyct_ndhF972F and Nyct_ndhF22R). PCR was performed using a Verity (Applied Biosystems) with the following cycling conditions: 95 °C for 5 min followed by 35 cycles at 98 °C for 30 sec, 55 °C for 45 sec and 68 °C for 50 sec, and a final extension at 68 °C for 2 min. The cycling conditions for primers Nyct_ndhF1F and Nyct_ndhF22R were as follows: 95 °C for 2 min followed by 35 cycles at 98 °C for 20 sec, 55 °C for 15 sec and 68 °C for 66 sec, and a final extension at 68 °C for 2 min.

Post-PCR purification was done using the NucleoFast Purification System (Separations). Sequencing was performed with BigDye Terminator V1.3 (Applied Biosystems) followed by electrophoresis on the 3730xl DNA Analyser (Applied Biosystems). Sequences were analysed and trimmed using Sequencing Analysis V5.3.1 (Applied Biosystems).

7.2.4 Sequence alignment

Sequence verifications and alignments were done unambiguously with CLC DNA Workbence 6 (CLC bio, Aarhus, Denmark), using the following settings during alignment: gap open cost (10), gap extension cost (1) and end gap cost (as any other). Alignments were also verified manually and ambiguous bases corrected by visual inspection.

7.2.5 Phylogenetic analysis

Both ITS and *ndh*F matrices were analysed separately and in combination using MEGA version 5 (Tamura *et al*., 2011). Missing data and gaps were eliminated. A distance method, Neighbor-Joining, as well as a model based approach, Maximum

Likelihood were used. The congruency index and the P-value for the two matrices were calculated (De Vienne *et al*., 2007) before the matrices were combined. Neighbour-Joining was performed using the Jukes-Cantor model. The Best fit nucleotide substitution models for use with the Maximum Likelihood were calculated with the Bayesian Information Criterion available within Mega 5 and evaluated using the Akaike Information Criterion (Akaike, 1973). The Tamura 3-Parameter +G Model were used for the ITS matrix, the General Time Reversal +G Model for the *ndh*F matrix and the Hasegawa-Kishino-Yano + G model for the combined matrix. The heuristic search model used for the Maximum Likelihood was the Close-Neighbor-Interchange. Bootstrap analysis (1 000 replicates) was performed to determine internal support. A bootstrap percentage of 80 – 100% is considered a high bootstrap support, a bootstrap support of 50 – 80% as moderate and a bootstrap support of less than 50% as weak. The bootstrap consensus trees of the neighbour-joining and maximum likelihoods are reported.

7.2.6 Outgroup, type specimens and combined matrices

Five species were used as outgroups and their ITS and *ndh*F sequences were obtained from National Centre for Biotechnology Information (NCBI), Bethesda, Maryland, United States of America (Tabel 7.4). Outgroup species were chosen as follows: *Pisoniella arboresens* (Lag. & Rodr.) Standl. is a basal taxon of the Nyctaginaceae with *Acleisanthes lanceolata* (Wooton) R.A.Levin, *Acleisanthes longiflora* A.Gray, *Mirabilis jalapa* L. and *Mirabilis multiflora* (Torr.) A.Gray most closely related to *Boerhavia* and *Commicarpus* (Douglas & Manos, 2007).

The type specimen of *Commicarpus* (*Commicarpus scandens* (L.) Standl.) was included to verify that the sequences were indeed of *Commicarpus* origin. ITS and *ndh*F sequences for types were also obtained from NCBI (Tabel 7.3). The type specimen of *Boerhavia* (*Boerhavia erecta* L.) was not included as there are only three, unpublished sequences (and therefore unvalidated) available on NCBI.

Three individuals of *C. squarrosus* (36, 39 and 41) were sent for DNA sequencing, but only the ITS sequences of individual 41 and the *ndh*F sequences of individual 36

could be obtained. As the identity of these two individuals are known and not in question, their ITS and *ndh*F sequences were combined for the combined matrices.

7.3 Results

Forty eight specimens were sent for sequencing, but only 37 specimens could be sequenced with ITS and 26 with *ndh*F. A total of 24 specimens were therefore used in the combined ITS and *ndh*F matrices. The ITS matrix had 537 characters, the *ndh*F matrix 1927 characters and the combined matrices 2 492 characters. The log likelihood score for the Maximum Likelihood tree for the ITS matrix is -2235.5732, for the *ndh*F matrix it is -3761.2873 and for the combined matrices it was -6499.4877.

7.3.1 Outgroups

Pisoniella aborescence resolved at the base of the neighbour-joining and maximum likelihood trees of the ITS and *ndh*F matrices, except for the maximum likelihood trees of the *ndh*F matrix and the combined matrices. These trees had to be rooted. *Acleisanthes lanceolata* and *A. longiflora* clustered together with high bootstrap support (80–100%) and also *Mirabilis jalapa* and *M. multiflora* with high bootstrap support (90–100%). *Acleisanthes* and *Mirabilis* form a clade between *Boerhavia* and *Commicarpus* in the neighbour-joining (Fig. 7.1) and maximum likelihood trees (Fig. 7.2) of the ITS matrix and the maximum likelihood tree of the combined matrices (Fig. 7.6). In the neighbour-joining tree of the combined matrices (Fig. 7.5) they resolve between *Boerhavia* and *Commicarpus* but not as a clade and in the remaining trees (Fig 7.3; 7.4) at the base of the tree.

7.3.2 Analyses of the ITS matrix

The ITS matrix divides the specimens into three clades: a *Boerhavia* clade, an *Acleisanthes* and *Mirabilis* clade and a *Commicarpus* clade (Fig. 7.1; 7.2). Bootstrap support within the *Boerhavia* clade was weak (4–55%), except for *B. repens* subsp. *repens* (99–100%) and *B. hereroensis* (±86%) which had high bootstrap support. *B. cordobensis* (±68%) had moderate bootstrap support, as well as the subclade comprising *B. coccinea* var. *coccinea* (110 and 120) and *Boerhavia* sp. (±74%

bootstrap). *B*. *coccinea* var. *coccinea* 108 resolves separately from *B*. *coccinea* var. *coccinea* 110 and 120 between *B*. *hereroensis* and *B*. *cordobensis* (Fig. 7.1 and 7.2). *B*. *deserticola* resolves between *B*. *erecta* (Fig. 7.2).

The internal topology of the *Commicarpus* clade differs between the neighbour-joining (Fig. 7.1) and maximum likelihood (Fig. 7.2) trees. In both trees, however, *C*. *plumbagineus* and *C*. *helenae* var. *helenae* form a subclade, *C*. *pentandrus* and *C*. *decipiens* form a subclade and *C*. *squarrosus* forms a subclade with *C*. *fruticosus*.

7.3.3 Analyses of the *ndh*F matrix

The *ndh*F matrix divides the specimens into two well supported clades: a *Boerhavia* clade (99% bootstrap support) and a *Commicarpus* clade (99% bootstrap support) (Fig. 7.3 and 7.4). The *Boerhavia* clade divides into two well supported (99% bootstrap) subclades. The first subclade consists of *B*. *cordobensis*, *B*. *erecta* and *B*. *diffusa* var. *diffusa* which are species alien to southern Africa. The second clade consists of *B*. *deserticola*, *B*. *hereroensis*, and *B*. *repens* subsp. *repens* which are indigenous to southern Africa. *B*. *coccinea* var. *coccinea*, of which the origin is uncertain, groups with the indigenous *Boerhavia* species in the second subclade.

The internal topology of the *Commicarpus* clade differs between the neighbour-joining (Fig. 7.3) and maximum likelihood (Fig. 7.4) trees. *Commicarpus squarrosus* always groups with *C*. *fruticosus* 164. *Commicarpus helenae* var. *helenae*, *C*. *plumbagineus* and *C*. *pilosus* form a subclade. The former two are more closely related to each other than to *C*. *pilosus* in the neighbour-joining tree (Fig. 7.3) while *C*. *plumbagineus* 3970 is more closely related to *C*. *pilosus* in the maximum likelihood tree (Fig. 7.4). *Commicarpus fallacissimus*, *C*. *chinensis* subsp. *natalensis*, *C*. *decipiens* and *C*. *pentandrus* forms a subclade. *C*. *pentandrus* and *C*. *decipiens* and *C*. *fallacissimus* and *C*. *chinensis* subsp. *natalensis* are respectively most closely related in the neighbor joining tree (Fig. 7.3) while in the maximum likelihood tree (Fig. 7.4), *C*. *decipiens* 181 is more closely related to *C*. *chinensis* subsp. *natalensis* and *C*. *fallacissimus*.

7.3.4 ITS and *ndh*F combined matrices

The l_{cong} = 1.375 and the P value = 0.0037 indicate that the tree topology of the ITS and *ndh*F matrices are more similar than can be expected by chance. The ITS and *ndh*F matrices were therefore combined.

The combined matrices resulted in a division into four clades in the neighbour joining tree (Fig. 7.5) and three clades in the maximum likelihood tree (Fig 7.6). In the neighbor-joining tree, a *Boerhavia* clade (100% bootstrap support), separate *Acleisanthes* and *Mirabilis* clades and a *Commicarpus* clade were formed (Fig. 7.5). In the maximum likelihood tree a *Boerhavia* clade, an *Acleisanthes*-*Mirabilis* clade and a *Commicarpus* clade (100% bootstrap support) were formed. In the *Boerhavia* clade, *B*. *repens* subsp. *repens* resolves separate from the other *Boerhavia* species. *B*. *cordobensis*. *B*. *erecta* and *B*. *diffusa* var. *diffusa* group together in a subclade and *B*. *deserticola*, *B*. *hereroensis* and *B*. *coccinea* var. *coccinea* group together in the other subclade.

The internal topology of the *Commicarpus* clade in the neighbor-joining and maximum likelihood trees are the same, except that *C*. *chinensis* subsp. *natalensis* forms part of the *C*. *squarrosus*-*C*. *fruticosus* clade in the maximum likelihood tree (Fig. 7.6), but resolves outside this clade in the neighbor-joining tree (Fig. 7.5). In the neighbor-joining and maximum likelihood trees *C*. *plumbagineus* and *C*. *helenae* var. *helenae*, *C*. *pentandrus* and *C*. *decipiens* and *C*. *squarrosus* and *C*. *fruticosus* form subclades.

7.4 Discussion

7.4.1 Outgroup genera

Douglas & Manos (2007) studied the phylogeny of the Nyctaginaceae and sampled 25 genera, including *Pisoniella*, *Acleisanthes*, *Boerhavia*, *Commicarpus* and *Mirabilis*. In his study only *Mirabilis* resolved sister to *Commicarpus*, while in the present study *Mirabilis* together with *Acleisanthes* resolve as either sister to *Commicarpus* only or sister to the *Boerhavia* and *Commicarpus* clades together. This contradiction is the

result of limited sampling in the present study (five genera), as opposed to the 25 genera in the study of Douglas and Manos (2007).

The morphological and anatomical studies (Chapter 4 and 5) provided characters by which to distinguish between *Boerhavia* and *Commicarpus*. Throughout the molecular analyses, *Boerhavia* and *Commicarpus* formed two separate clades, indicating that *Boerhavia* and *Commicarpus* are two separate genera. The phylogenetic study of Douglas and Manos (2007) also found *Boerhavia* and *Commicarpus* to be two separate genera. The type specimen of *Commicarpus* (*C. scandens*) groups within the *Commicarpus* clade and supports the hypothesis that *Commicarpus* in southern Africa is indeed *Commicarpus* and not *Boerhavia*.

7.4.2 *Boerhavia*

Anthocarp morphology can be used to distinguish seven *Boerhavia* species and the lower coriaceous part of the flower, four species (Chapter 4). The molecular analyses separate these species and none are grouped together consistently, which supports the recognition of two subclades and seven species as defined by the anthocarp morphology.

Boerhavia cordobensis, *B. erecta* and *B. diffusa* var. *diffusa* are alien to southern Africa and native to America (López, 1998; Chen & Wu, 2007; Bromilow, 2010). In the analyses with the *ndh*F matrices and the combined matrices, these three species group together in different combinations in a well supported subclade. The rest of the *Boerhavia* species formed an indigenous subclade as *B. hereroensis*, *B. deserticola* and *B. repens* subsp. *repens* are indigenous to southern Africa (Codd, 1966; Germishuizen & Meyer, 2003). From literature it is not known whether *B. coccinea* var. *coccinea* is indigenous to southern Africa (Codd, 1966) and it has been speculated to have been introduced and naturalized in Africa (Chen & Wu, 2007). However, it continuously groups with *B. deserticola* and *B. hereroensis*, suggesting that it is indigenous to southern Africa. Interestingly, the anthocarps of alien *B. cordobensis*, *B. erecta* and *B. diffusa* var. *diffusa* taper to a pointed tip at the apex, while the anthocarps of indigenous *B. coccinea* var. *coccinea*, *B. hereroensis*, *B.*

deserticola and *B*. *repens* subsp. *repens* are rounded at the apex. This morphological grouping is in accordance with the phylogeny.

Only ITS data were available for the three unknown *Boerhavia* species (*Boerhavia* sp. 136, 137, 4420). These species are similar in growth form and foliage to *B*. *coccinea* var. *coccinea*, except that they have very large, clavate anthocarps. These species group together with *B*. *coccinea* var. *coccinea* in both the neighbor-joining and maximum likelihood trees and are either a new species closely related to *B*. *coccinea* var. *coccinea*, or a new *B*. *coccinea* variety introduced to the subtropical eastern parts of southern Africa.

The combinations of *B*. *deserticola* and *B*. *erecta* and the resolving of *B*. *coccinea* var. *coccinea* 108 separate from *B*. *coccinea* var. *coccinea* (110 and 120) in the trees obtained from the ITS matrix is senseless and could be the result of sampling error (Felsenstein, 2004; Martin *et al*., 2005), such as too few genes.

7.4.3 *Commicarpus*

Anthocarp morphology and the lower, coriaceous part of the flowers distinguish eight *Commicarpus* species (Chapter 4). The molecular analyses support these eight species as defined by the anthocarp and flower morphology. Throughout the analyses, *C*. *pentandrus* and *C*. *decipiens* group together and *C*. *plumbagineus* and *C*. *helenae* var. *helenae* group together. *Commicarpus decipiens* and *C*. *pentandrus* are restricted in their distribution to Africa (Meikle, 1978; African Plant Database, 2011), while *C*. *helenae* var. *helenae* and *C*. *plumbagineus* are more widespread, distributed beyond Africa into Europe (Spain) and east through southern Arabia, Iran and Palestine to India (Meikle, 1978; African Plant Database, 2011).

In the trees constructed from the *ndh*F matrix (Fig. 7.3 and 7.4), the indigenous *C*. *fallacissimus* and *C*. *chinensis* subsp. *natalensis* group together with *C*. *pentandrus* and *C*. *decipiens*. *Commicarpus pilosus*, having a distribution extending into Tropical Africa (Klopper *et al*., 2006; African Plant Database, 2011), groups with *C*. *plumbagineus* and *C*. *helenae* var. *helenae*.

Commicarpus fruticosus and *C. squarrosus* are both sub-shrubs and their leaves, flowers and anthocarps are similar in shape, size range and indumentums. The only difference between the two species is the dark purple colour of the flowers of *C. fruticosus* while those of *C. squarrosus* are of a lighter purple (Chapter 4). In all the constructed trees (Fig. 7.1 – 7.6), *C. squarrosus* and *C. fruticosus* group together in close relationship, supporting the morphological findings.

7.4.4 Phylogenetic tree for southern Africa

The topology of the *Boerhavia* and *Commicarpus* clades, constructed from the combined ITS and *ndh*F matrices (Fig. 7.5; 7.6), are very similar to the topology for the clades constructed from the *ndh*F matrix (Fig. 7.3 and 7.4). The topology for the *Boerhavia* clade of the *ndh*F and combined matrices is different from the *Boerhavia* clade in the trees constructed from the ITS matrix (Fig. 7.1 and 7.2), and makes more sense than that of the ITS matrix. This scenario suggests that the *ndh*F are more informative than the ITS for the *Boerhavia* species, although the ITS matrix has more specimens than the *ndh*F and combined matrices. It is not, however, possible to make meaningful comparisons if the matrices do not consist of the same number of specimens, as the number of specimens can influence tree topology (Martin *et al*., 2005). The topology for the *Commicarpus* clade constructed from the ITS, *ndh*F and combined ITS and *ndh*F matrices (Fig. 7.5 and 7.6) is similar.

7.5 Future research

Many authors (Vogl *et al*., 2003; Felsenstein, 2004; Soltis *et al*., 2004; Martin *et al*., 2005) are of opinion that the number and choice of taxa, as well as the number of genes sampled, are central to phylogenetic estimation, as too few taxa or the wrong taxa may lead to incorrect topology and too few genes sampled to sampling error which can result in high bootstrap values for incorrect branches.

Of the 48 specimens submitted for sequencing, only 24 specimens could be sequenced with both the *ndh*F and ITS genes [there are various reasons as to why sequences fail, such as insufficient template DNA, poor annealing between primer and template, inactive DNA polymerase, or contamination (DNAbiotec, 2011)], giving an

incomplete picture of the phylogenetic relationships of the southern Africa n*Boerhavia* and *Commicarpus* species. Before any further analysis and interpretation of the results can be done, the remaining specimens need to be sequenced and more genes added. Once all the molecular data is obtained, the results can then be combined and brought into context with the morphology, anatomy, palynology and biogeography of the southern African *Boerhavia* and *Commicarpus*. A cladistic analysis of a combined molecular and morphology dataset of characters will allow for the construction of a workable, phylogenetic classification.

7.6 Summary

Molecular analyses indicate that *Boerhavia* and *Commicarpus* are separate genera with well defined species as previously indicated by morphological and anatomical characters. The *Boerhavia* species were divided into two clades, namely the *Boerhavia* species alien (*B. cordobensis*, *B. diffusa* var. *diffusa* and *B. erecta*) and the *Boerhavia* species indigenous (*B. deserticola*, *B. hereroensis* and *B. repens* subsp. *repens*) to southern Africa. *B. coccinea* var. *coccinea* groups with the second clade, which provides further support for it to be regarded as indigenous to Africa. In the *Commicarpus* clade, the morphologically similar *C. fruticosus* and *C. squarrosus* group together in close relationship. This provides further support for the view taken in this thesis that the two taxa are the same species. Throughout the analysis, the widespread *C. plumbagineus* and *C. helenae* var. *helenae* group together, as well as *C. pentandrus* and *C. decipiens*. *C. chinensis* subsp. *natalensis*, *C. pilosus* and *C. fallacissimus*, all species of harsh environments (e.g. coastal dunes or arid environments), do not form subclades, except in the *ndh*F matrix where *C. fallacissimus* and *C. chinensis* subsp. *natalensis* group with *C. decipiens* and *C. pentandrus*, and *C. pilosus* groups with the tropical species *C. plumbagineus* and *C. helenae* var. *helenae*. The *Boerhavia* clade forms well defined subclades, but this is not true for the *Commicarpus* clade and is an indication of missing data which is either due to too few species sampled or too few genes.

Tabel 7.1. Evolutionary models of nucleotide substitution used in the phylogenetic analyses.

Model of Evolution	Description	Reference
Jukes-Cantor	Model assumes that all the changes are equally probable	Jukes & Cantor, 1969
Tamura 3-parameter	Model takes the differences in transitional and transversional rates, as well as the G+C content bias, into account	Tamura, 1992
General Time Reversal	Model assumes that the total overall rate into a certain nucleotide is equal to the total rate out of that same nucleotide	Tavaré, 1986
Hasegawa-Kishino-Yano	Model distinguishes between the rate of transitions and transversions and allows unequal base frequencies	Hasegawa *et al*., 1985

Table 7.2. *Boerhavia* and *Commicarpus* specimens used for the phylogenetic study (NP, National Park).

Taxon	Specimens examined	Herbarium	Collection locality
Boerhavia coccinea var. *coccinea*	Struwig, M. 108	PUC, PRE	South Africa. Wyllie's Poort
	Struwig, M. 110	PUC, PRE	South Africa. Waterpoort
	Struwig, M. 120	PUC, PRE	South Africa. Mapungubwe NP
Boerhavia cordobensis	Struwig, M. 122	PUC, PRE	South Africa. Tsipise
	Struwig, M. 130	PUC, PRE	South Africa. Pilgrim's Rest
	Struwig, M. 132	PUC, PRE	South Africa. Klerksdorp
Boerhavia deserticola	Struwig, M. 37	PUC, WIND	Namimbia. Brandberg
	Struwig, M. 42	PUC, WIND	Namibia. Twyfelfontein
	Struwig, M. 43	PUC, WIND	Namibia. Sesfontein
Boerhavia diffusa var. *diffusa*	Struwig, M. 125	PUC, PRE	South Africa. Tzaneen
	Struwig, M. 129	PUC, PRE	South Africa. Pilgrim's Rest
	Struwig, M. 140	PUC, KNP	South Africa. Kruger NP
Boerhavia erecta	Struwig, M. 23	PUC, PRE	South Africa. Potchefstroom
	Struwig, M. 135	PUC, KSAN	South Africa. Kruger NP
	Struwig, M. 143	PUC, KSAN	South Africa. Kruger NP
Boerhavia hereroensis	Struwig, M. 34	PUC, WIND	Namibia. Karibib
	Struwig, M. 35	PUC, WIND	Namibia. Klein Spitzkuppe
	Struwig, M. 40	PUC, WIND	Namibia. Twyfelfontein Lodge
Boerhavia repens var. *repens*	Struwig, M. 161	PUC, WIND	Namibia. Büllsport
	Struwig, M. 170	PUC, WIND	Namibia. Maltahöhe
Boerhavia sp.	Struwig, M. 136	PUC, KNP	South Africa. Kruger NP
	Struwig, M. 137	PUC, KNP	South Africa. Kruger NP
	Siebert, S.J. 4420	PUC	South Africa. Kruger NP
Commicarpus chinensis subsp. *natalensis*	Struwig, M. 60	PUC, KZN	South Africa. Tongaat Beach
	Struwig, M. 61	PUC, KZN	South Africa. Umhlanga Rocks
	Struwig, M. 63	PUC, KZN	South Africa. Richards Bay
Commicarpus decipiens	Struwig, M. 47	PUC, WIND	Namibia. Tsumeb
	Struwig, M. 176	PUC, WIND	Namibia. Omaruru
	Struwig, M. 181	PUC, WIND	Namibia. Otjiwarongo
Commicarpus fallacissimus	Struwig, M. 33	PUC, WIND	Namibia. Windhoek
	Struwig, M. 46	PUC, WIND	Namibia. Joubert pass
Commicarpus fruticosus	Struwig, M. 160	PUC, WIND	Namibia. Naukluft Mountains
	Struwig, M. 163	PUC, WIND	Namibia. Naukluft Mountains
	Struwig, M. 164	PUC, WIND	Namibia. Naukluft Mountains
Commicarpus helenae var. *helenae*	Struwig, M. 44	PUC, WIND	Namibia. Khowarib Rest Camp
	Struwig, M. 141	PUC, KNP	South Africa. Kruger NP
	Struwig, M. 183	PUC, WIND	Namibia. Otjimbingwe
Commicarpus pentandrus	Struwig, M. 57	PUC, WIND	Namibia. Aris Farm
	Struwig, M. 131	PUC, PRE	South Africa. Manyaka
	Struwig, M. 138	PUC, KNP	South Africa. Kruger NP
Commicarpus pilosus	Struwig, M. 109	PUC, PRE	South Africa. Waterpoort road
	Struwig, M. 111	PUC, PRE	South Africa. Waterpoort road
	Struwig, M. 114	PUC, PRE	South Africa. Mapungubwe NP
Commicarpus plumbagineus	Struwig, M. 126	PUC, PRE	South Africa. Duiwelskloof
	Siebert, S.J. 3970	PUC	South Africa. Kruger NP
Commicarpus squarrosus	Struwig, M. 36	PUC, WIND	Namibia. Klein Spitskuppe
	Struwig, M. 39	PUC, WIND	Namibia. Brandberg
	Struwig, M. 41	PUC, WIND	Namibia. Twyfelfontein

Table 7.3. Sequences of primers used for PCR amplification and sequencing analyses.

Primer	Sequence	Publication
ITS4	5'TCC TCC GCT TAT TGA TAT GC 3'	White *et al*., 1990
ITS5a	5'CCT TAT CAT TTA GAG GAA GGA G 3'	Stanford *et al*., 2000
Nyct_ndhF1F	5'TGC CTG GAT TAT ACC CTT CA 3'	Douglas & Manos, 2007
Nyct_ndhF13R	5'AAA TGC RGT AAT CCV DCT G 3'	Douglas & Manos, 2007
Nyct_ndhF972F	5'ATG TCT CAA TTG GGT TAT ATG ATG 3'	Olmstead & Sweere, 1994
Nyct_ndhF22R	5'CTT GTA ACG CCG AAA CCA TT 3'	Douglas & Manos, 2007

Table 7.4. NCBI accession numbers of taxa used as outgroups in the phylogenetic study as well as the type specimen for *Commicarpus*.

Taxon	ITS	ndhF
	NCBI Accession Number	NCBI Accession Number
Acleisanthes longiflora	EF079457.1	EF079608.1
Acleisanthes lanceolata	EF079454.1	EF079509.1
Mirabilis jalapa	EF079461.1	EF079515.1
Mirabilis multiflora	EF079452.1	EF079507.1
Pisoniella arborescens	EF079485.1	EF079539.1
Commicarpus scandens	EF079482.1	EF079536.1

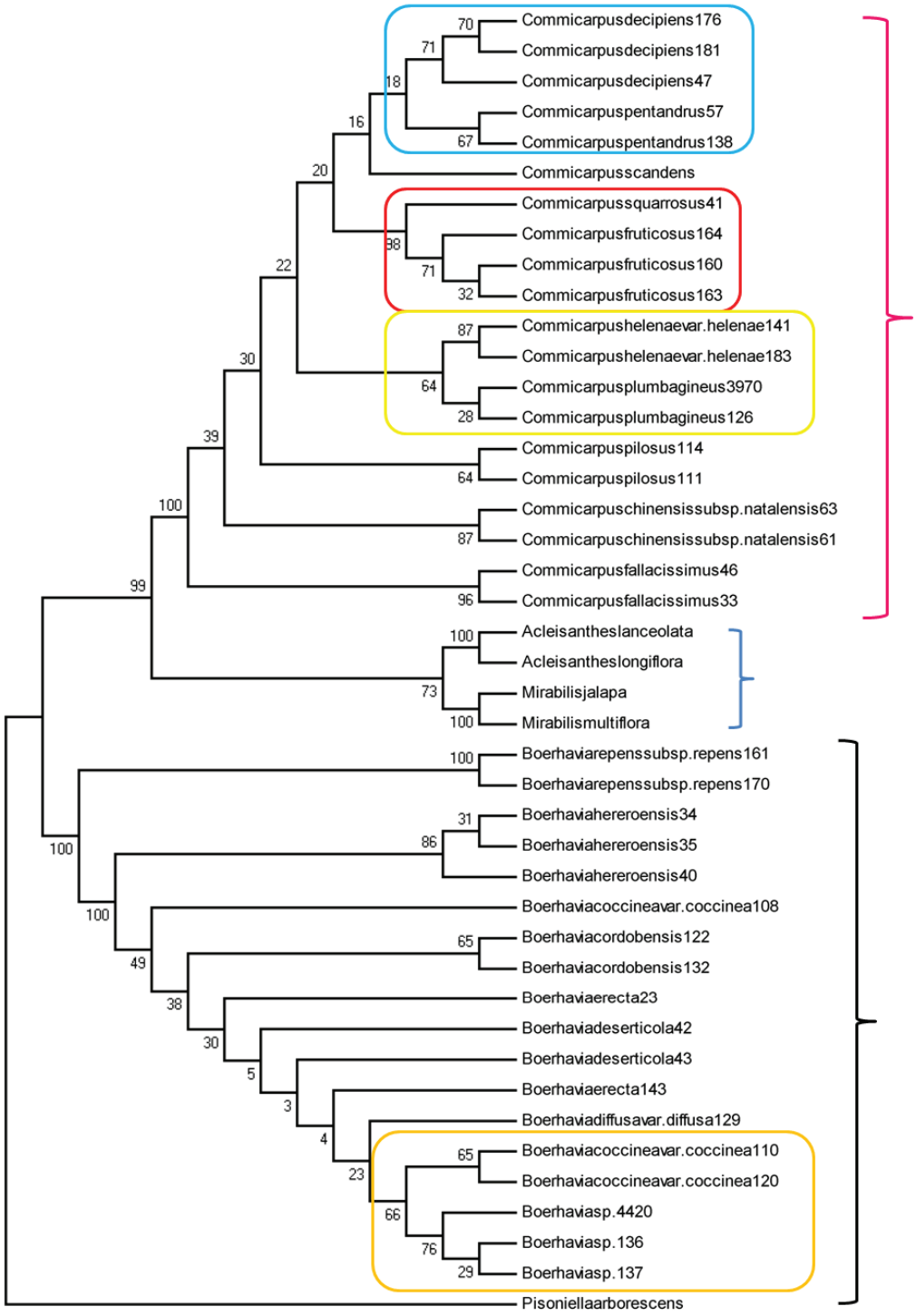

Figure 7.1: Bootstrap consensus tree of the neighbor-joining method used on the ITS matrix. The *Boerhavia* clade is indicated by the black bracket, the *Commicarpus* clade with the pink bracket and the *Mirabilis* + *Acleisanthes* clade by the blue bracket. The clade formed by *C. decipiens* and *C. pentandrus* is indicated by the blue circle, the clade formed by *C. squarrosus* and *C. fruticosus* by the red circle, *C. helenae* var. *helenae* and *C. plumbagineus* by the yellow circle and *B. coccinea* var. *coccinea* and *Boerhavia* species by the orange circle. Numbers above or below the nodes are bootstrap support values.

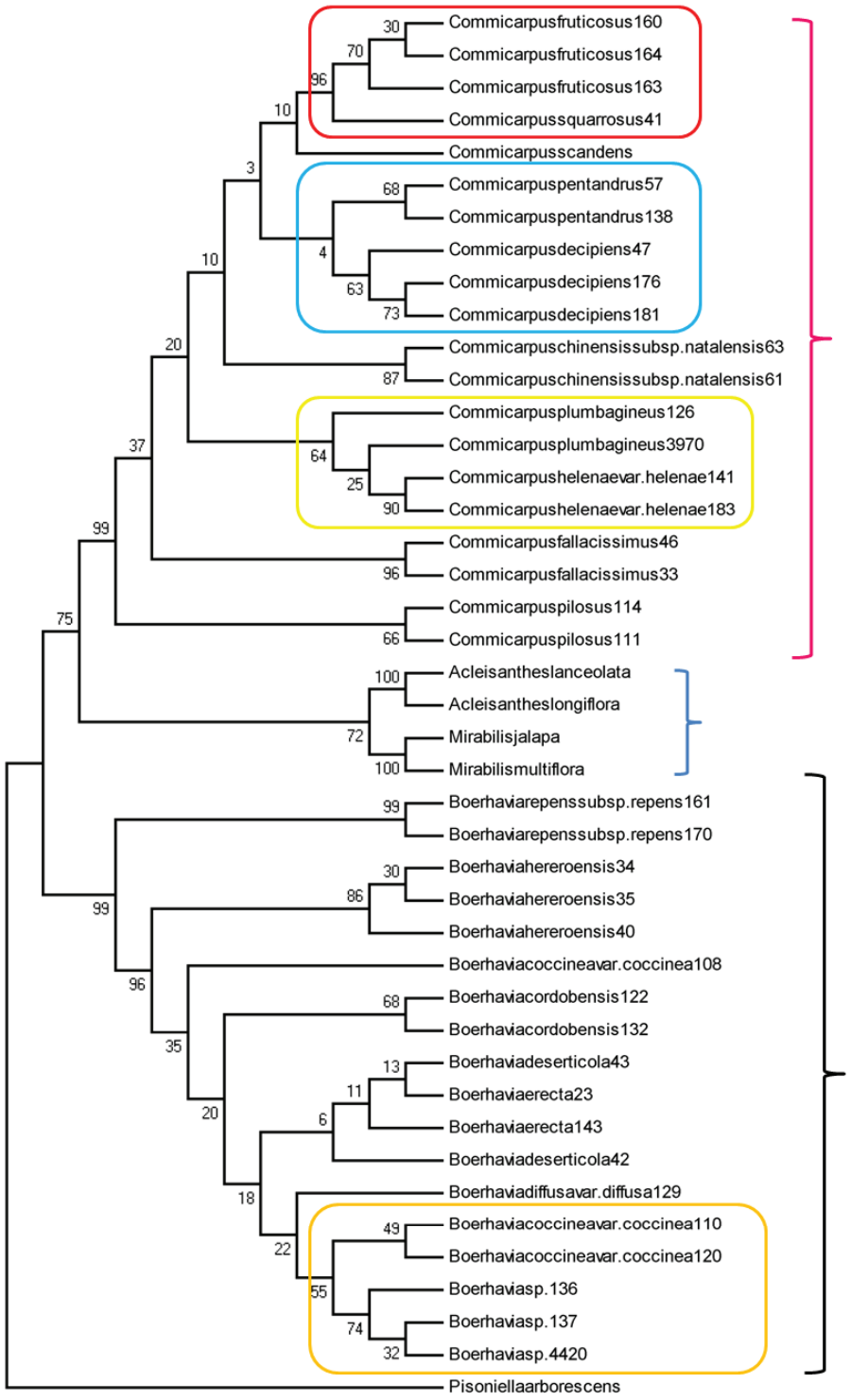

Figure 7.2: Bootstrap consensus tree of the maximum likelihood method used on the ITS matrix. The *Boerhavia* clade is indicated by the black bracket, the *Commicarpus* clade by the pink bracket and the *Mirabilis* + *Acleisanthes* clade by the blue bracket. The clade formed by *C. decipiens* and *C. pentandrus* is indicated by the blue circle, the clade formed by *C. squarrosus* and *C. fruticosus* by the red circle, *C. helenae* var. *helenae* and *C. plumbagineus* by the yellow circle and *B. coccinea* var. *coccinea* and *Boerhavia* species by the orange circle. Numbers above or below the nodes are bootstrap support values.

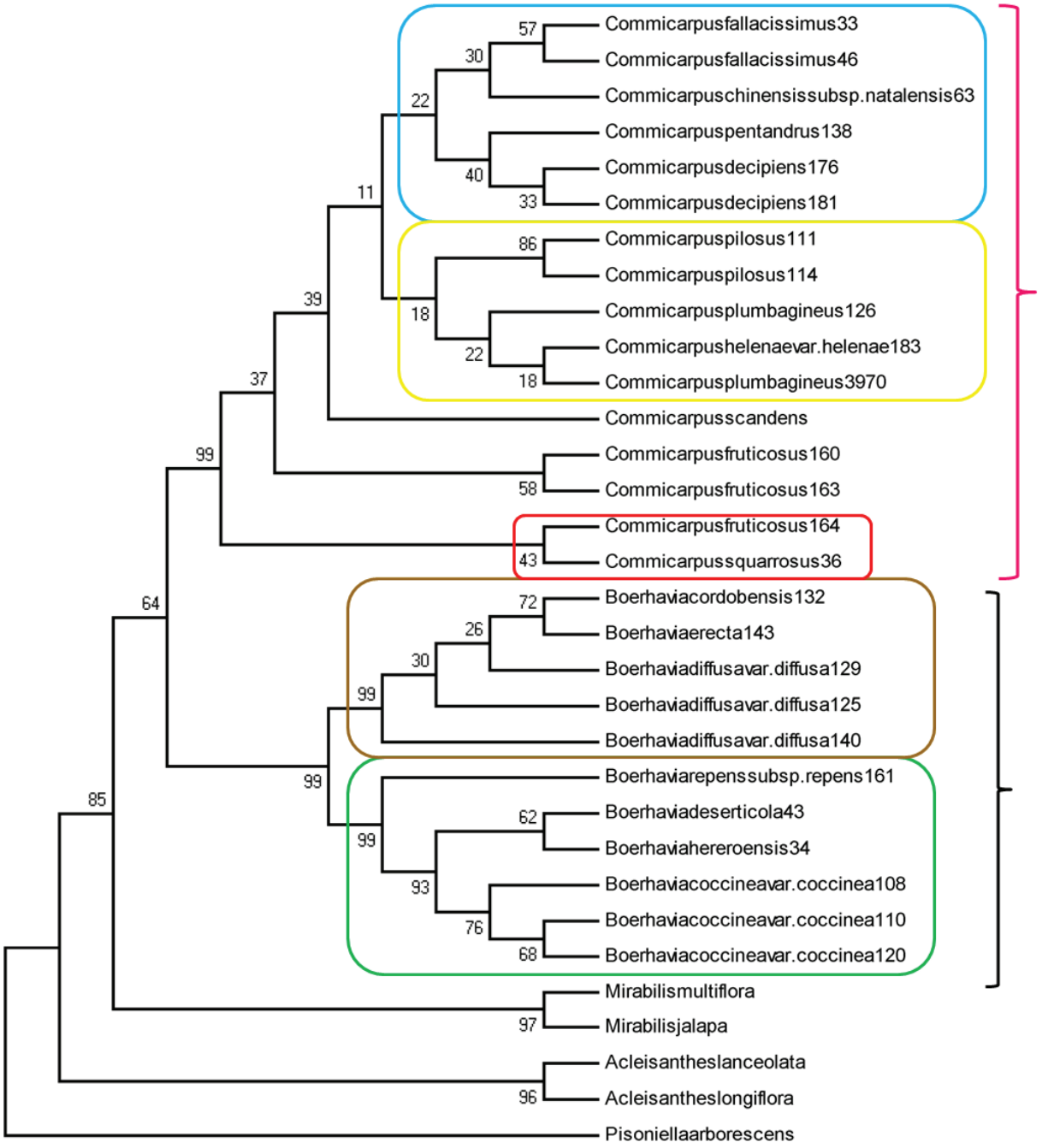

Figure 7.3: Bootstrap consensus tree of the neighbor-joining method used on the *ndh*F matrix. The *Boerhavia* clade is indicated by the black bracket and the *Commicarpus* clade by the pink bracket. The indigenous *Boerhavia* clade is indicated by the green circle and the alien *Boerhavia* clade by the brown circle. The *C. fallacissimus*, *C. chinensis* subsp. *natalensis*, *C. pentandrus* and *C. decipiens* clade is indicated by the blue circle and the *C. pilosus*, *C. plumbagineus* and *C. helenae* var. *helenae* clade by the yellow circle. The close relationship between *C. squarrosus* and *C. fruticosus* is indicated by the red circle. Numbers above or below the nodes are bootstrap support values.

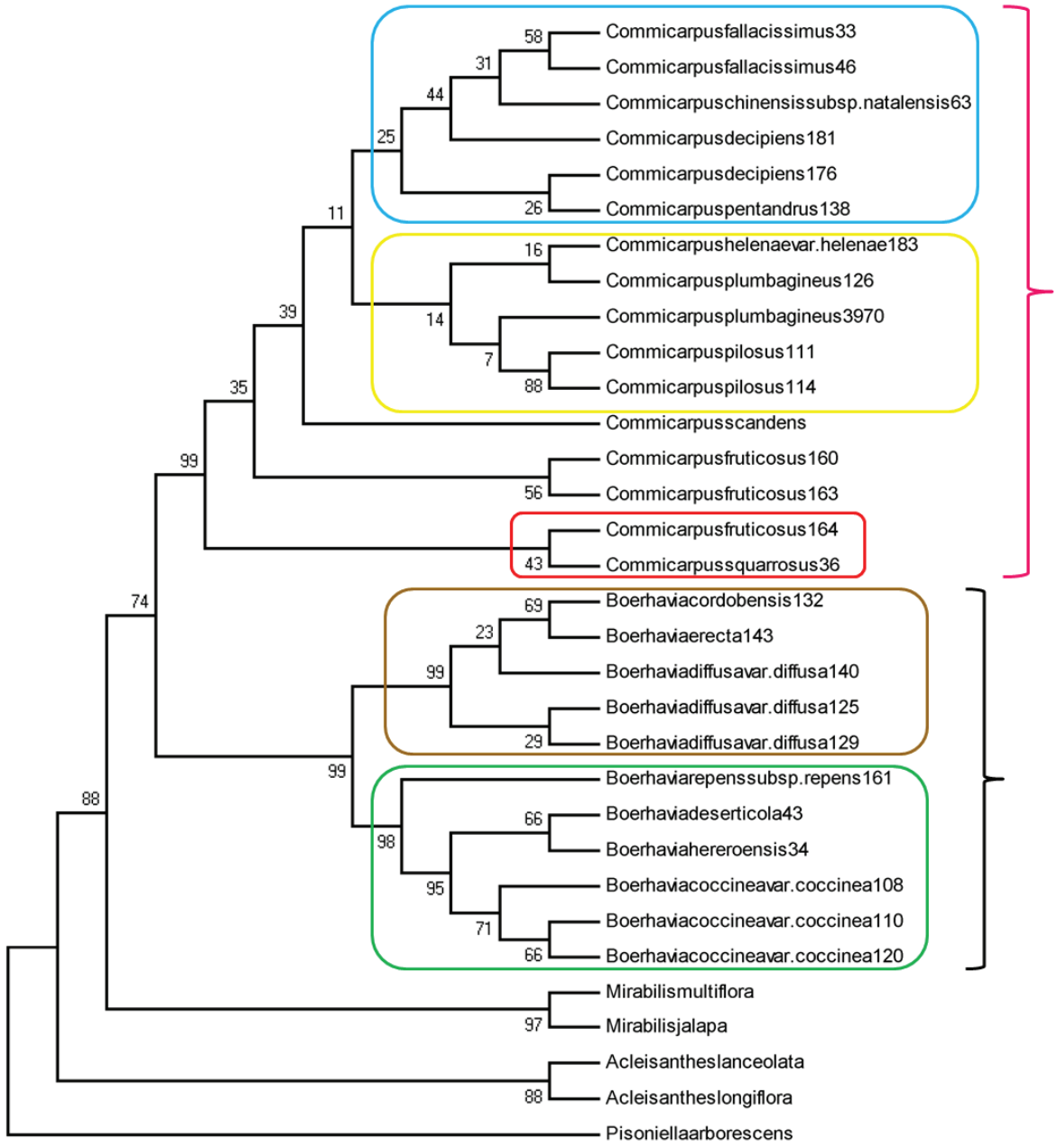

Figure 7.4: Bootstrap consensus tree of the maximum likelihood method used on the *ndh*F matrix. The *Boerhavia* clade is indicated by the black bracket and the *Commicarpus* clade by the pink bracket. The indigenous *Boerhavia* clade is indicated by the green circle and the alien *Boerhavia* clade by the brown circle. The *C. fallacissimus*, *C. chinensis* subsp. *natalensis*, *C. pentandrus* and *C. decipiens* clade is indicated by the blue circle and the *C. pilosus*, *C. plumbagineus* and *C. helenae* var. *helenae* clade by the yellow circle. The close relationship between *C. squarrosus* and *C. fruticosus* is indicated by the red circle. Numbers above or below the nodes are bootstrap support values.

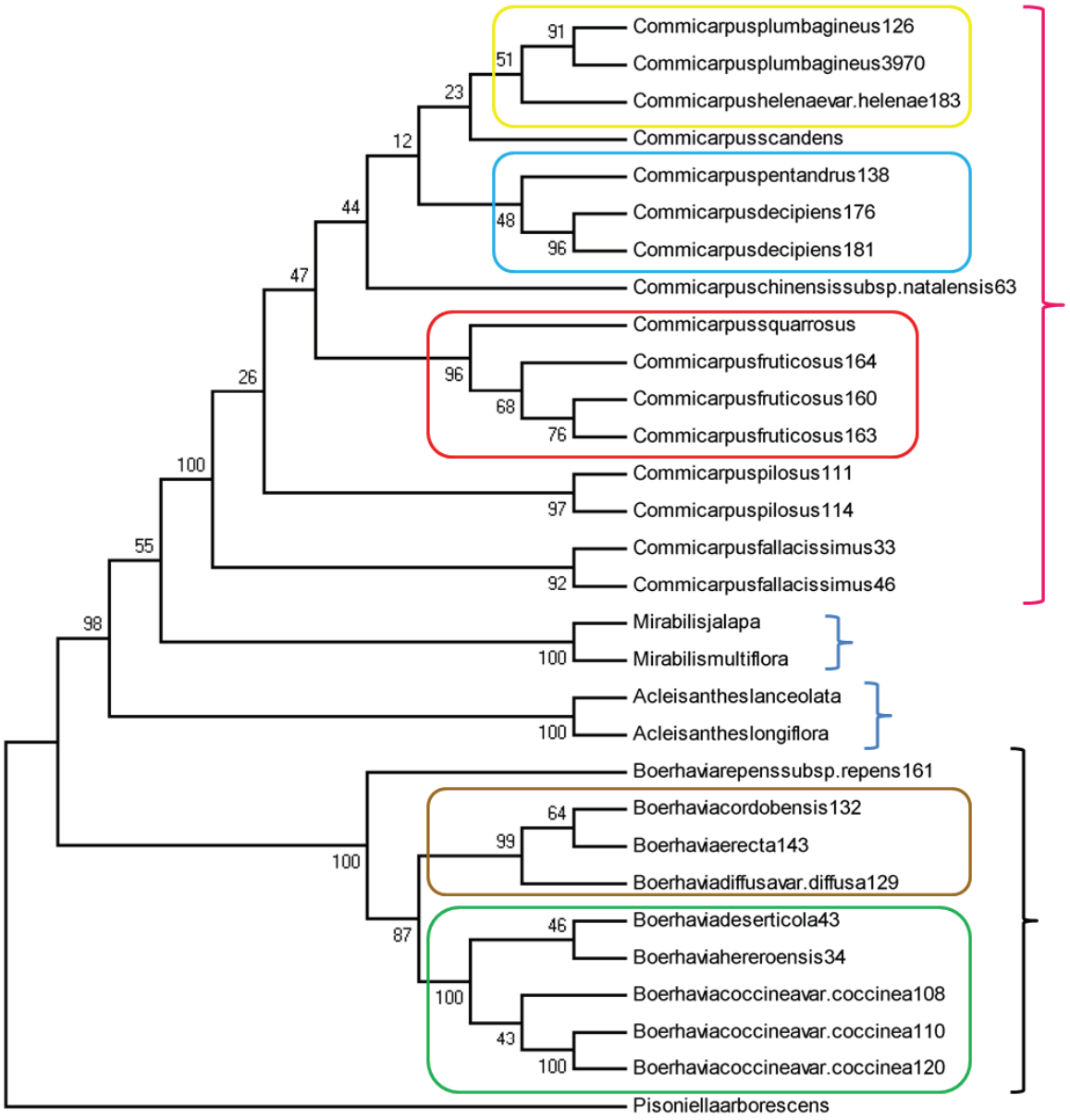

Figure 7.5: Bootstrap consensus tree of the neighbor-joining method used on the ITS and *ndh*F matrices combined. The *Boerhavia* clade is indicated by the black bracket and the *Commicarpus* clade by the pink bracket. The two smaller *Mirabilis* and *Acleisanthes* clade is indicated by the blue brackets. The indigenous *Boerhavia* clade is indicated by the green circle and the alien *Boerhavia* clade by the brown circle. The clade formed by *C. decipiens* and *C. pentandrus* is indicated by the blue circle, the clade formed by *C. squarrosus* and *C. fruticosus* by the red circle and the clade formed by *C. helenae* var. *helenae* and *C. plumbagineus* by the yellow circle. Numbers above or below the nodes are bootstrap support values.

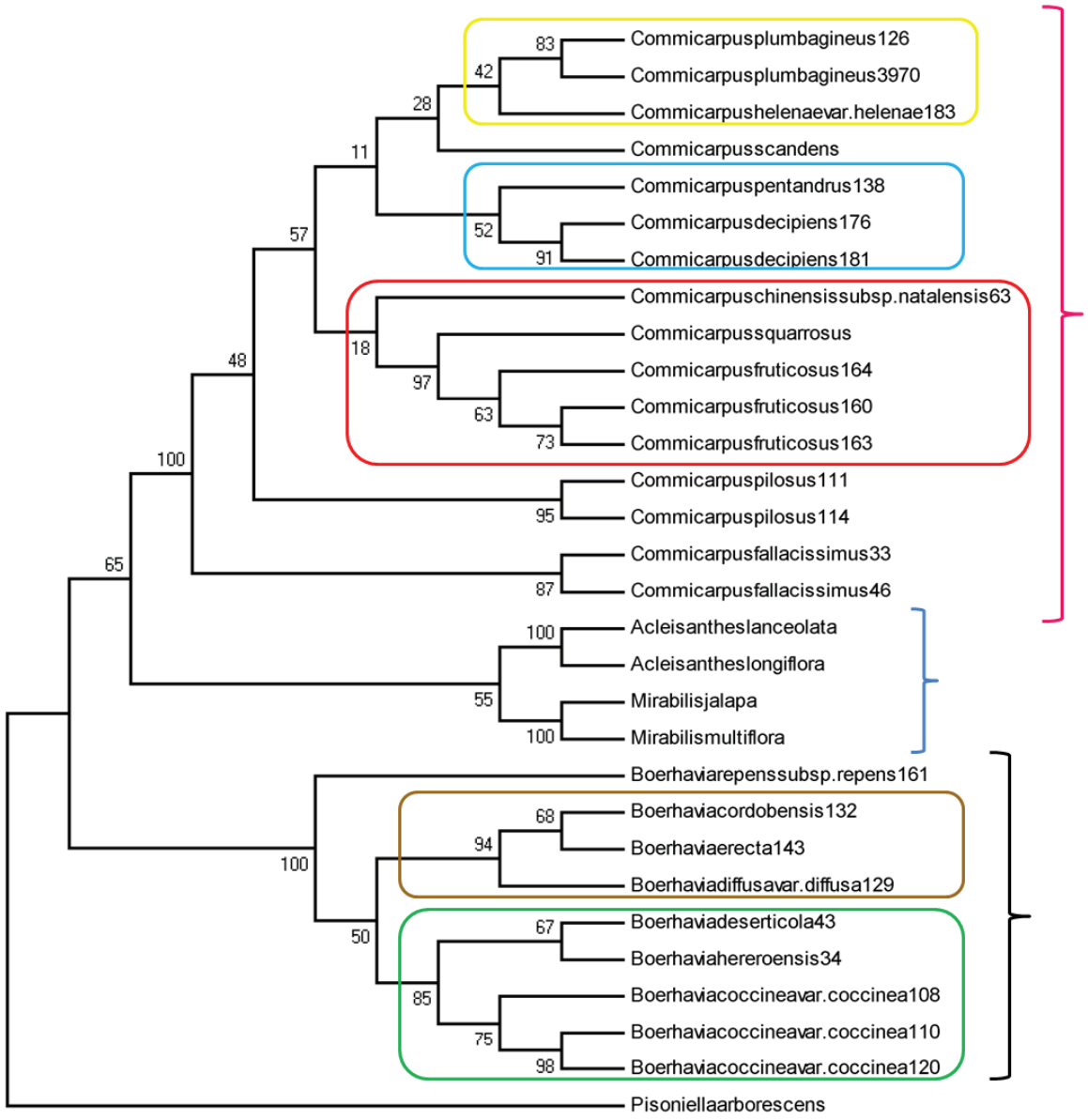

Figure 7.6: Bootstrap consensus tree of the maximum likelihood method used on the ITS and *ndh*F matrices combined. The *Boerhavia* clade is indicated by the black bracket and the *Commicarpus* clade by the pink bracket and the *Mirabilis* and *Acleisanthes* clade by the blue brackets. The indigenous *Boerhavia* clade is indicated by the green circle and the alien *Boerhavia* clade by the brown circle. The clade formed by *C. decipiens* and *C. pentandrus* is indicated by the blue circle, the clade formed by *C. chinensis* subsp. *natalensis*, *C. squarrosus* and *C. fruticosus* by the red circle and the clade formed by *C. helenae* var. *helenae* and *C. plumbagineus* by the yellow circle. Numbers above or below the nodes are bootstrap support values.