# Edaphic Factors and Rhizobia influence the Distribution of Legumes (Fabaceae) in the Core Cape Subregion of South Africa

**Meshack Nkosinathi Dludlu**

Thesis presented for the Degree of

DOCTOR OF PHILOSOPHY

Department of Biological Sciences

UNIVERSITY OF CAPE TOWN

February 2018

**Supervisors:**

A/Prof. A. Muthama Muasya & Dr Samson B.M. Chimphango

# Declaration

I, Meshack Nkosinathi Dludlu know the meaning of plagiarism and declare that all of the work in the thesis, save for that which is properly acknowledged, is my own.

I hereby,

a) grant the University free licence to reproduce the above thesis in whole or in part for the purpose of research;
b) declare that:
    i. the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance from my supervisors, I have received no assistance except as stated below;
    ii. neither the substance nor any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or any other University, except as stated below.

I am now presenting the thesis for examination for the degree of PhD.

The thesis contains three papers, which were published in collaboration with my supervisors and research associates, A. Muthama Muasya, Samson B.M. Chimphango; Gemma Walker and Charles H. Stirton:

1. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2017. Distinct edaphic habitats are occupied by discrete legume assemblages with unique indicator species in the Cape Peninsula of South Africa. *Journal of Plant Ecology* rtx027. doi: 10.1093/jpe/rtx027.
2. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2018. Differential preference of *Burkholderia* and *Mesorhizobium* to pH and soil types in the Core Cape Subregion, South Africa. *Genes* 9, 2. doi: 10.3390/genes9010002.

3. Dludlu, M.N., Chimphango, S.B.M., Walker, G., Stirton, C.H., Muasya, A.M., (accepted). Horizontal gene transfer among rhizobia of the Core Cape Subregion of southern Africa. *South African Journal of Botany*.

My supervisors have testified that I made substantial contributions to the conceptualization and design of the published papers and that I independently ran experiments and wrote the manuscripts, with their support in the form of comments and suggestions (see Appendix D).

**Signature:** Signed by candidate **Date:** February 2018

# Dedication

*I dedicate this thesis to my family, whose unrelenting love and support made this possible.*

# Acknowledgements

I would like to convey my sincere gratitude to my supervisors, A/Prof. Muthama Muasya and Dr Samson Chimphango for their instrumental guidance, motivation and support from the inception of this project throughout the PhD journey. Likewise, I am indebted to Prof. Charles Stirton for his mentorship, which he has been providing cheerfully from the early days of my Master's degree studies to the present day. I cannot imagine having better supervisors and a better mentor for my thesis.

The technicians of the Biological Sciences Department are gratefully acknowledged for their assistance with the sourcing of equipment and reagents and providing access to the relevant laboratory facilities. I am also thankful to Edward Chirwa for his assistance with fieldwork and some of the laboratory work. Dr Benny Lemaire's enthusiastic guidance on some laboratory techniques is gratefully appreciated.

Thanks to Terry Trinder-Smith, the Curator of the Bolus Herbarium, for providing access to datasets on legume occurrence records for the Cape Peninsula. In the same vein, I would like to acknowledge the South African National Biodiversity Institute (SANBI) for providing access to their database (PRECIS), particularly, data on legume occurrence records for the entire Southern Africa region.

I acknowledge the Table Mountain National Parks for providing permits to conduct fieldwork in areas under their management. In the same way, I recognize Mr Giorgio Lombardi, the Warden of the Vogelgat Private Nature Reserve, for granting permission to conduct research on the legumes that occur in their reserve.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Additional financial support towards this study was provided by the University of Cape Town, the University of Swaziland and the Worldwide Universities Network (WUN); to whom I am very grateful.

# Abstract

Fabaceae is the second most speciose plant family in the Core Cape Subregion (CCR) of South Africa, a Mediterranean type ecosystem, with mostly nutrient-poor soils. A majority of the legumes occurring in this region belong to the predominantly nitrogen-fixing subfamily Papilionoideae and they employ a variety of strategies for nutrient acquisition. However, legumes are neither uniformly nor randomly distributed in the CCR landscape. Instead, distinct legume species assemblages tend to occupy particular habitats within the landscape. The drivers of this distribution pattern are yet to be determined. In this thesis, it was hypothesized that edaphic factors (soil chemical and physical characteristics) and the distribution of rhizobia have influenced legume distributions in the CCR landscape.

The influence of edaphic factors on the distribution of legume species assemblages in the Cape Peninsula (a microcosm of the CCR) is the subject of the second chapter of the thesis. It was hypothesized that the composition of legume species assemblages is correlated with soil physical and chemical properties and that the interaction of Phosphorus (P) and the three cations that often bind P, i.e. Aluminium, Calcium and Iron, making it unavailable to plants, drive legume species assemblages in the landscape. Soils from 27 legume sites, spanning all major soil types of the Cape Peninsula, were analysed for 31 chemical and physical properties. Surveys of legume species present at each site were conducted to generate a presence/absence matrix. Canonical correspondence analysis was used to test for a correlation between legume species composition and edaphic factors. The strength of the association between legume species composition and site groupings based on edaphic properties was assessed using indicator species analysis. A significant correlation between edaphic factors and species composition was found and the key edaphic parameters driving the relationship were clay content, iron (Fe), potassium (K), sulphur (S) and zinc (Zn). Indicator species, characteristic of the various edaphic habitats were also identified. These findings indicate that distinct edaphic habitats are occupied by discrete legume species assemblages, implying a significant influence of edaphic factors on the legume distributions.

Chapter three of the thesis sought to determine if the ecological parameters; altitude, pH and soil type influence the distribution of the two main rhizobial genera (*Burkholderia* and *Mesorhizobium*) that nodulate various legumes of the CCR, and to determine the diversity and phylogenetic position of rhizobia that associate with the narrowly distributed and rare *Indigofera superba* in the CCR. The first objective was pursued through molecular characterisation of rhizobial strains isolated from nodules of legume species collected in the field across the Cape Peninsula. DNA sequences for 16S rRNA, *recA* and *nodA* were combined with data from a previous study that sampled broadly within the CCR and phylogenetic analyses were conducted. Tests for phylogenetic signals for the three ecological parameters were conducted, using the D statistic for soil type and Pagel's λ for altitude and pH. These analyses were used to test the hypothesis that closely related species occupy similar habitats with respect to each of the three ecological parameters. For the study of rhizobial symbionts of *Indigofera superba*, field nodules were sampled from multiple populations across its distribution range and a phylogeny of its symbionts was reconstructed in a matrix that included symbionts of diverse legumes from different habitats within the CCR. The results showed that *Burkholderia* is restricted to acidic habitats, while *Mesorhizobium* occurs in both acidic and alkaline habitats. Additionally, both rhizobial genera showed significant phylogenetic clustering for pH and most soil types. However, none of the genera showed a phylogenetic structure with respect to altitude. These findings indicate that pH and soil type influence the distribution of rhizobia in the CCR. Implications of these findings for the distribution of legumes in the landscape are discussed. For the narrowly distributed *I. superba*, the results showed that it associates with diverse strains within the genus *Burkholderia* and such strains are not phylogenetically distinct from strains isolated from localities outside its distribution range. These findings lead to the hypothesis that *I. superba* does not exhibit rhizobia specificity at the intrageneric level. Testing of this hypothesis through analysis of its nodulation capability on soils from outside its distribution range is recommended.

The fourth chapter of the thesis determined the extent of horizontal gene transfer among rhizobial genera in the Core Cape Subregion (CCR) of South Africa and reconstructed the ancestral symbionts of the legumes. Phylogenies of two chromosomal genes (16S rRNA and *recA*) and one nodulation gene (*nodA*) of rhizobia, isolated from diverse legumes in the CCR, were reconstructed using Bayesian and Maximum Likelihood techniques. A cophylogenetic analysis was used to test for congruence between the chromosomal and the *nodA* phylogenies. Five genera of rhizobia (*Bradyrhizobium*, *Burkholderia*, *Ensifer*, *Mesorhizobium* and *Rhizobium*) were studied. A phylogeny of the legumes was reconstructed from *matK* and *rbcL* DNA sequences and it was used to reconstruct their ancestral rhizobia, using Bayesian methods. The chromosomal phylogeny of the rhizobia was mostly incongruent with that of *nodA*, indicating potential horizontal inheritance of the latter. The *nodA* genes of *Burkholderia*, *Mesorhizobium* and *Rhizobium* had different evolutionary histories from their counterparts in other parts of the globe. *Burkholderia* was reconstructed as the ancestral symbionts of the CCR legumes. Evidence of co-diversification between the legumes and their symbionts was observed and this highlights a potential role of the legume-rhizobia interaction to the high diversity of legumes in the CCR. Finally, the availability of compatible rhizobia and their competitive ability are discussed as possible drivers for the lack of shared legumes between the CCR's Fynbos biome and the Kwongan of Australia.

Overall, the study shows that edaphic factors and biotic interactions (rhizobia) have significant influence on the distribution of legumes in the Cape Peninsula and the larger Core Cape Subregion of southern Africa. These findings are consistent with the theory that edaphic factors and biotic interactions have a strong influence on species distributions at local and site spatial scales.

# Table of Contents

# 1. General Introduction

## 1.1 Legume biogeography at global and regional scales

Fabaceae or Leguminosae is the second most economically important plant family after the Poaceae (Duke, 2012; Gepts et al., 2005; Graham and Vance, 2003). In addition to being a source of food, medicine, animal fodder, timber etc., legumes are an environmentally friendly source of nitrogen in agricultural and natural systems due to their ability to convert atmospheric nitrogen into a plant usable form, i.e. nitrogen fixation (Graham and Vance, 2003; Jensen and Hauggaard-Nielsen, 2003; Sprent, 2009). Comprising 770 genera and over 19 000 species, Fabaceae is the third most speciose plant family after the Asteraceae and Orchidaceae (Lewis et al., 2005; LPWG, 2013; LPWG, 2017). Thus, legumes are a highly diverse botanical family with a potential to contribute towards alleviating global poverty and improving soil fertility without the need to use expensive inorganic sources of nitrogen.

Although legumes occupy diverse habitats throughout the globe (Crews, 1999; Schrire et al., 2005), their greatest diversity is found in semi-arid (or seasonally dry) tropical forests and temperate shrublands (Wojciechowski et al., 2004). Examples of such legume-rich areas include the Brazilian Caatinga and Cerrado biomes (Sprent et al., 1996), the Greater Antilles in the Caribbean (Francisco-Ortega et al., 2007; Lavin and Matos, 2008), the Thar Desert in India (Ardley, 2017), the Southwest Australian Floristic Region (Crisp et al., 2004; Hopper and Gioia, 2004) and the Core Cape Subregion (CCR) of South Africa (Goldblatt and Manning, 2000; Manning and Goldblatt, 2012). Notably, these areas are also known for their nutrient-poor soils (Lambers et al., 2011; Lopes and Cox, 1977; Panwar and Tarafdar, 2006; Witkowski and Mitchell, 1987), which suggests that nitrogen-fixation confers an ecological advantage to the legumes in such ecosystems.

The legumes' predilection for arid to semi-arid areas is likely a legacy of their evolutionary history (Sprent and Gehlot, 2010). Until recently, there were two main hypotheses on the origin and spread of legumes to the various parts of the globe and Doyle and Luckow (2003) discuss these comprehensively. One hypothesis was that legumes originated in the humid tropics of Africa, towards the end of the Cretaceous, ~ 80 million years ago (mya), when South America and Africa were still joined, a scenario which would have allowed for migration into South America and subsequently, into North America (Raven and Polhill, 1981). However, with increased understanding of continental drift processes and advances in the reconstruction of molecular phylogenetic relationships and estimation of divergence dates of legumes, it became clear that legumes evolved long after the separation of the continents, prompting an alternative hypothesis (Doyle and Luckow, 2003). The alternative hypothesis invoked the existence of a landbridge that connected Africa, Europe and North

America during the Eocene, i.e. 35-55 mya (Tiffney, 1985), and suggested a northern hemisphere origin of legumes, followed by migration into Africa through this landbridge (Doyle and Luckow, 2003). However, this hypothesis is not supported by the present-day literature.

Contemporary studies support an origin of legumes occurring ~ 55-60 mya (Bell et al., 2010; Lavin et al., 2005) in the semi-arid region on either side of the Tethys seaway (Schrire et al., 2005), which used to partly separate Laurasia and Gondwana (Schettino and Scotese, 2005). Since Laurasia remained intact as the landmasses were drifting apart, while Gondwana comprised several isolated landmasses, it has been hypothesized that the legume flora of the northern and southern hemispheres remained isolated until 3-4 mya when North America merged with South America (Sprent et al., 2013). The merging of North and South America is proposed to have facilitated the dispersal of chiefly Laurasian flora, such as the Fabeae (Schaefer et al., 2012) into the southern hemisphere (Sprent et al., 2017). Meanwhile, Australia's unique legume composition has been hypothesized to be a legacy of its lengthy attachment to Antarctica, which kept it isolated from the rest of the major landmasses until 33 mya (Crisp and Cook, 2013; Sprent et al., 2017). A detailed account of the global distribution patterns of legumes is provided by Sprent *et al*. (2017).

### *Legumes in Southern Africa*

Despite the existing understanding of the global trends in legume biogeography, there is a need for an in depth understanding of the regional, landscape and local scale biogeographic processes in order to better manage and conserve the legumes. This is particularly important as the factors that influence species' distributions vary at the different spatial scales (Pearson and Dawson, 2003; Rahbek, 2005). Southern Africa provides a good study area for such an assessment because out of its ~ 23 000 species of higher plants (Klopper et al., 2007), about 1600 of the species are legumes (Trytsman et al., 2016). Moreover, the distribution patterns of legumes in the region are well documented. The general trend is that distinct habitats (biomes, vegetation units, etc.) are associated with unique legume assemblages (Trytsman et al., 2016). For example, of the four subfamilies of Leguminosae that occur in southern Africa (LPWG, 2017), the fynbos biome (found in the CCR) is almost wholly comprised of the subfamily Papilionoideae, whereas the other subfamilies (i.e. Caesalpinioideae, Cercidoideae and Detarioideae) are mostly restricted to the Grassland and Savanna biomes of the region (Trytsman et al., 2016). Some Mimosoideae *s.s* (indigenous and introduced) have made invasive inroads to the CCR, but not Caesalpinioideae *s.s*. These distribution patterns raise some questions regarding the biogeography of the southern African legumes, such as (i) what has driven legume distributions in southern Africa and its various biomes and (ii) what has driven the variation in legume diversity among the different biomes? These questions have implications for conservation in that if species have limited adaptability to new habitats, their survival could be threatened by shrinkage of suitable habitats due to global change.

## 1.2 Legumes in the Core Cape Subregion of South Africa

Comprising 764 species, legumes are the second most species-rich plant lineage in the Core Cape Subregion (formerly the Cape Floristic Region), a temperate shrubland located on the southwestern tip of South Africa (Manning and Goldblatt, 2012). Although 17 tribes of legumes are represented in the flora of the CCR, most (~ 90 %) of the species belong to four tribes, i.e. the Crotalarieae (382 species), Podalyrieae (115), Psoraleeae (108) and Indigofereae (78) (Schutte, 2012). The CCR is one of the world's floristically rich regions (Cox, 2001; Takhtajan, 1986), with 9 000 plant species in an area of 90 000 km$^2$ and ~ 70 % endemism (Goldblatt and Manning, 2000; Manning and Goldblatt, 2012). It has a Mediterranean type of climate: with cold, wet winters and hot, dry summers. It is dominated by acidic and nutrient-poor soils (about 80 %) derived from quartzitic sandstones, while some limestone, shale and granite derived soils which are slightly higher in nutrients, e.g. N and phosphorus (P), also occur and form about 20 % of the landscape (Campbell, 1986; Witkowski and Mitchell, 1987). The CCR also has a varied topography, with altitudes of up to 2249 m (Verboom et al., 2015) across a range of broadly coastal parallel mountains of increasing aridity. The rainfall gradient is steep, with an annual precipitation ranging from 100 to 2000 mm (Goldblatt and Manning, 2002). Thus, the CCR is an ecologically heterogeneous landscape, providing diverse habitats for various species to adapt to. The high legume diversity suggests that they have adapted to the ecological conditions of the landscape and have diversified their niches to capitalise on its heterogeneity.

Surprisingly, most vegetation units of the Fynbos biome, which is the largest biome of the CCR, are dominated by Proteaceae, Ericaceae and Restionaceae species (Born et al., 2007), and not the species-rich Fabaceae. The same is true for the other smaller biomes of the CCR, i.e. Albany Thicket, Forests and the Succulent Karoo. This might be because 75 % of CCR legumes are short-lived reseeders (Le Maitre and Midgley, 1992) that emerge shortly after fire and gradually disappear with time after the fire (Kruger, 1983). Nevertheless, the legume reseeders tend to have patchy distributions, whereas the less prevalent resprouters have a more widespread distribution (Schutte et al., 1995). In some legume lineages (e.g. the Psoraleeae) there appears to be some phylogenetic structuring of both fire response traits (Bello et al., 2015). Furthermore, distinct legume species assemblages tend to occupy particular niches at any given locality. The drivers of these distribution patterns are not yet fully understood.

Reseeders tend to grow faster than resprouters (Verdú, 2000) and since this comes at a high nutritional cost (Bell, 2001; Lambers et al., 2008), reseeders tend to require higher nutrient levels than resprouters (Power et al., 2011). Therefore, emerging shortly after fire allows the legumes to capitalise on the flushes of nutrients that are released after the fire event (Brown and Mitchell, 1986; Certini, 2005; Dean et al., 2015). Considering that the distribution of nutrient-enriched sites following a fire event is often patchy (Stock and Lewis, 1986) and that species thrive on soils that meet their nutritional requirements (Richards et al., 1997b), the legumes might be tracking these nutrient enriched patches, hence their patchy distributions. Could the tendency of having distinct legume assemblages occupying particular niches within the CCR landscapes also be due to edaphic factors? Addressing this question is one of the primary objectives of the present study.

### 1.3 Factors that drive species distributions

The distribution of species is mostly influenced by biotic and abiotic factors, as well as dispersal capability (Soberón, 2007). Biotic factors include competition, predation, facilitation, and symbiotic interactions with other organisms (Wisz et al., 2013), while abiotic factors include climate, edaphic conditions, topography, etc. (Randin et al., 2009; Thuiller et al., 2004). Dispersal capability pertains to whether a species can reach its suitable habitats (Boulangeat et al., 2012). Pearson and Dawson (2003) conceptualized a hierarchical framework in which they postulate that different factors that influence the distribution of species act at different spatial scales. Their framework posits that climate has the strongest influence at the global, continental and regional (i.e. > 200 km) scales, while topography is strongest from local to regional scales (i.e. 1-2000 km). Edaphic conditions are more influential at the local (1–10 km) and site (10–1000 m) scales, while the influence of biotic interactions extends from the local to the micro (< 10 m) scales (Pearson and Dawson, 2003). The importance of spatial scale for the differential influence of various factors on species distributions is supported by a number of studies (Guisan and Thuiller, 2005; Rahbek, 2005; Soberón, 2007; Whittaker et al., 2001).

The hierarchical framework of Pearson and Dawson (2003) is supported by numerous studies on the distribution of amphibians, birds, insects, mammals, plants and reptiles (Eiserhardt et al., 2011; Luoto et al., 2006; Pearson et al., 2004; Thuiller et al., 2004). However, deviations from the framework have also been reported. For example, biotic interactions (presence of host plants) were found to influence the distribution of European Apollo butterflies at macro-scales (Araújo and Luoto, 2007). Likewise, the facilitative interaction between owls and woodpeckers (in which the latter provide nesting sites for the former by creating cavities on trees) was found to influence the macro-scale distribution of owls in Northern Europe (Heikkinen et al., 2007). Furthermore, topography has been shown to influence species distributions even at continental scales (Austin and Van Niel, 2011; Luoto and

Heikkinen, 2008) rather than being only important at the local to regional scales as proposed by Pearson and Dawson (2003). These contrasting patterns suggest that the hierarchical framework cannot be generalised to all ecosystems and/or organisms. Therefore, considering that the CCR is unique among Mediterranean ecosystems (a surprisingly high species richness) and given its remarkably high ecological heterogeneity (Cowling, 1990; Linder, 2003), it remains to be determined whether the hierarchical framework holds for such a region.

### 1.4 Legumes and nitrogen fixation

Biological nitrogen fixation is an important adaptive strategy for legumes (Fabaceae), especially those which occur in nutrient-impoverished habitats (Sprent, 2009). This is because it helps sustain their high nitrogen-demanding lifestyle in which legumes retain a higher nitrogen (N) leaf content relative to non-legumes (Reed, 2017; Sprent et al., 1996; Stock and Verboom, 2012; Werner et al., 2015). It has been hypothesized that the high nitrogen-demanding lifestyle evolved in ancestral legumes which occupied nutrient-rich habitats (Herendeen et al., 1992). Since leaf nitrogen content is positively correlated with photosynthetic capacity (Chen et al., 2015; Tuohy et al., 1991), a high leaf nitrogen content facilitates rapid growth and opportunistic production of short-lived leaves, allowing legumes to capitalise on brief periods of resource availability (McKey, 1994; Reed, 2017). McKey (1994) hypothesized that N-fixation evolved later as the legumes were spreading into habitats that have low nitrogen content. This hypothesis has not been tested and there is uncertainty regarding when N-fixation first evolved in angiosperms, let alone Fabaceae (Doyle, 2016; Li et al., 2015; Sprent et al., 2017). However, studies show that all nitrogen-fixing angiosperms belong to a single clade (the Rosid 1 clade), which indicates that the predisposition for N-fixation in angiosperms has a single origin (Soltis et al., 1995; Werner et al., 2014). For the legumes, which originated ~ 55-60 mya and radiated rapidly shortly thereafter (Bell et al., 2010), it has been shown that N-fixation arose multiple independent times during the radiation of the various clades (Doyle, 2011).

Nitrogen fixation occurs through a mutual partnership between legumes and soil bacteria, collectively called rhizobia (Peix et al., 2015; Sprent, 2001). The legumes and rhizobia communicate through chemical signals, leading to the formation of specialised structures in the roots (and sometimes stems) of the legumes, called nodules, in a process called nodulation (Sprent, 2009). The rhizobia reside in the nodules where they convert atmospheric nitrogen into ammonia (a plant usable form) in return for carbohydrates from the host plants (Moulin et al., 2015; Sprent et al., 2009).

Considering that nitrogen and phosphorus (P) are often the major limiting elements in nutrient-poor ecosystems (Lambers et al., 2008; Larimer et al., 2014; Vitousek et al., 2010; Witkowski and Mitchell, 1987), it is not surprising that a majority of the legumes that occur in such environments fix nitrogen (Ardley, 2017; Gehlot et al., 2012; Lafay and Burdon, 1998; Lemaire et al., 2015a; Sankhla et al., 2017; Sprent and Gehlot, 2010). They also form cluster roots, secrete organic acids and phosphatases or partner with mycorrhiza (Cardoso et al., 2017; Lamont, 1982; Larimer et al., 2014; Maistry et al., 2015; Mikola, 1986) to augment their acquisition of P. Therefore, legumes are well-adapted for such harsh environments. Notably, out of the six subfamilies of Fabaceae (LPWG, 2017), nodulation is most common in the Papilionoideae (Sprent, 2007; Sprent et al., 2017) and this is the predominant subfamily in the CCR.

### *Rhizobia diversity and biogeography*

The study of the diversity of rhizobia is ongoing and the current state of knowledge on this subject is comprehensively discussed in several recent reviews (Andrews and Andrews, 2017; Peix et al., 2015; Sprent et al., 2017). Thus far, the known genera of N-fixing bacteria belong to two classes of the phylum Proteobacteria; the Alphaproteobacteria and the Betaproteobacteria, and these are often referred to as ‘α-rhizobia’ and ‘β-rhizobia’, respectively. The main genera that belong to the former class include *Azorhizobium*, *Bradyrhizobium*, *Ensifer* (= *Sinorhizobium*), *Mesorhizobium* and *Rhizobium*, while the latter includes *Burkholderia* and *Cupriavidus* (Masson-Boivin et al., 2009; Moulin et al., 2001; Peix et al., 2015; Sprent et al., 2017).

The taxonomy of the genus *Burkholderia sensu lato* is presently unstable. The genus is highly heterogeneous and includes some pathogenic and some plant-beneficial and environmental species (Depoorter et al., 2016; Estrada-de los Santos et al., 2016). Due to phylogenetic distinctness of these groups, Sawana et al. (2014) proposed splitting the genus into two, such that the pathogenic species retain the name *Burkholderia*, and the plant-associated and environmental species were placed into a new genus, *Paraburkholderia*. Subsequently, several legume-nodulating *Burkholderia* species were transferred to *Paraburkholderia* and a new genus, *Caballeronia*, was proposed for some species of *Burkholderia* and *Paraburkholderia* (Dobritsa and Samadpour, 2016). However, these changes were largely based on inconclusive phylogenetic data such that taxonomic boundaries of the new genera are uncertain. For example, Lopes-Santos et al. (2017) recently described a new genus, *Robbsia*, to accommodate a species that had been classified as *Paraburkholderia andropogonis*. A genome-based phylogenetic study, conducted by Beukes et al. (2017), confirmed the existence of five distinct lineages within *Burkholderia sensu lato* i.e. *Burkholderia sensu stricto*, *Caballeronia*, *Paraburkholderia*, *Robbsia* and a lineage represented by *Paraburkholderia rhizoxinica*. Based on these findings, they proposed a transfer of some species that had been previously classified under *Burkholderia sensu stricto* into *Caballeronia* and the creation of a new genus to accommodate *Paraburkholderia rhizoxinica* (Beukes et al. 2017). Considering that the study of rhizobia is an ongoing exercise, it is yet to be determined if these latter changes (Beukes et al., 2017) help bring about stability to the taxonomy of *Burkholderia sensu lato*.

Most of the main genera of the α-rhizobia are widespread in diverse habitats throughout the globe, where they nodulate a wide range of legumes. For example, despite having only a few (29) described N-fixing species (Thilakarathna and Raizada, 2017), *Bradyrhizobium* is the primary symbiont of diverse legumes in tropical and temperate regions, such as Australia (Lafay and Burdon, 1998; Stępkowski et al., 2012), Brazil (Delamuta et al., 2017), Ethiopia (Degefu et al., 2017), India (Ojha et al., 2017) and South Africa (Beukes et al., 2016). Likewise, *Mesorhizobium* has a wide host range in diverse ecosystems, *inter alia Ademisia* spp. in Chile (Gerding et al., 2017), *Astragalus* and *Oxytropis* in arctic and subarctic Eurasia (Ampomah et al., 2017), various Papilionoid legumes in the CCR of South Africa (Gerding et al., 2012; Lemaire et al., 2015a) and *Sophora* spp. in New Zealand (Tan et al., 2015). On the other hand, β-rhizobia, particularly *Burkholderia*, have just two major centres of diversity, (i) South America, where they are the primary symbionts of Mimosoid legumes (Bontemps et al., 2010; Bournaud et al., 2013) and the CCR of South Africa, where they nodulate diverse legumes of the subfamily Papilionoideae (Beukes et al., 2013; Lemaire et al., 2015a; Lemaire et al., 2016b). Notably, *Burkholderia* has a preference for acidic soils and it gets replaced by α-rhizobia in alkaline habitats in the Americas (Bontemps et al., 2016; Stopnisek et al., 2014). It has also been shown that the diversity of *Burkholderia* strains changes with altitude (Bontemps et al., 2010).

Despite a large number of studies documenting the diversity of rhizobia and their phylogenetic relationships, there are still some unresolved questions regarding the evolution of rhizobia. One such question is: which of the rhizobia genera is the ancestral symbiont of legumes? At present, there are two competing hypotheses pertaining to this question. Because *Bradyrhizobium* is the largest rhizobial genus, has the widest legume host range and is the predominant symbiont of basally branching legumes (e.g. Caesalpinioideae), it has been hypothesized that they are the ancestral symbionts of legumes (Parker, 2015; Sprent et al., 2017). On the other hand, since nodulation genes appear to have originated in β-rhizobia (*Burkholderia*), followed by a horizontal transfer to α-rhizobia (Aoki et al., 2013), and there is evidence suggesting that N-fixing *Burkholderia* may be of ancient origin (Bontemps et al., 2010), *Burkholderia* may well be the ancestral symbionts of legumes.

### *Horizontal gene transfer in rhizobia*

DNA regions that are commonly sequenced for phylogenetic studies of rhizobia include housekeeping genes (responsible for the core functions of the bacteria), nodulation genes (those involved in inducing the formation of nodules in legumes) and nitrogen fixation genes (those involved in the conversion of atmospheric nitrogen into ammonia) (Remigi et al., 2016). The housekeeping genes form part of the chromosomal genome, while the nodulation and nitrogen fixation genes often reside in mobile extra-chromosomal plasmids or symbiotic islands (Flores et al., 2005; Peix et al., 2015; Zahran, 2017). These mobile elements and the genetic material contained therein can be transferred, laterally, from one organism to another (including disparate evolutionary lineages) in a process called horizontal gene transfer (HGT), unlike the usual parent-to-offspring transfer (vertical inheritance) of genetic material across succeeding generations (Juhas et al., 2009; MacLean et al., 2007). The process of HGT can occur through various mechanisms including transformation, transduction and conjugation (Davison, 1999; Paul, 1999).

The organism that is a recipient of transferred genetic material is a chimeric species, i.e. different parts of its genome have different evolutionary histories (Eisen, 2000). Phylogenetic analyses based on these disparate portions of its genome often yield incongruent/conflicting results (Huang and Gogarten, 2006; Paul, 1999). Indeed, several studies on rhizobia have found incongruent phylogenetic trees when analysing chromosomal and nodulation and/or nitrogen fixation genes and such results have been attributed to horizontal gene transfer (Aoki et al., 2013; Huang and Gogarten, 2006; Parker, 2012). Horizontal gene transfer is important in the study of microbial evolution because not only does it lead to the formation of new species (Gogarten et al., 2002), but also to the acquisition of mechanisms for adaptations in extreme environments (Ochman et al., 2000; Wiedenbeck and Cohan, 2011). These consequences of HGT may be the key to the legumes' success in colonising harsh environments, such as acidic and nutrient-poor habitats. Studies have shown that HGT is a common phenomenon among the two main rhizobial symbionts of the CCR legumes (i.e. *Burkholderia* and *Mesorhizobium*) and phylogenetic analyses of their nodulation genes indicate that they have a distinct evolutionary history from that of their counterparts elsewhere in the globe (Beukes et al., 2013; Lemaire et al., 2015b). Considering that the CCR is also a centre of diversity for *Burkholderia*, this provides opportunities for testing the hypothesis that *Burkholderia* are the ancestral symbionts of legumes and to investigate the role of HGT in the diversification of rhizobia in the CCR and its potential contribution to the diversification of the legumes.

### 1.5 Rationale and Aims

The overall goal of the study was to determine if edaphic factors and rhizobia have influenced legume distribution patterns in the CCR. The basis of the study is the observation that despite their high nutritional requirements (Chimphango et al., 2015; McKey, 1994; Reed, 2017), legumes are the second most speciose family in the CCR (Manning and Goldblatt, 2012), where nitrogen and phosphorus are particularly limiting (Cramer, 2010; Maistry et al., 2013; Power et al., 2010; Witkowski and Mitchell, 1987). Because the majority of the legumes that occur in the CCR are those which possess key adaptations for N and P acquisition (Allsopp and Stock, 1993; Beukes et al., 2013; Gerding et al., 2012; Lamont, 1982; Lemaire et al., 2015a; Maistry et al., 2015; Power et al., 2010), edaphic factors must be acting as a strong ecological filter in the region and are thus expected to play a significant role in the distribution of species. With N-fixation being the main source of N for the legumes (Sprent, 2009), the availability of suitable rhizobia is expected to also play a significant role in driving the distribution of the legumes. This expectation is based on the observation that, in other ecosystems, rhizobia (particularly *Burkholderia,* which is one of the predominant symbionts in the CCR) has been shown to be constrained by ecological parameters such as altitude, pH and soil type (Bontemps et al., 2010; Stopnisek et al., 2014). Given the heterogeneous nature of the CCR landscape with respect to these parameters (Campbell, 1986; Cowling, 1990; Linder, 2003; Verboom et al., 2015), the distribution of the rhizobia is unlikely to be uniform.

Therefore, the aims of the study were to:

i. determine the key edaphic parameters involved in driving legume distributions in the CCR and identify legume indicator species for the various edaphic habitats;
ii. determine if altitude, pH and soil type influence the distribution of rhizobia in the landscape and hence infer the implications for legume distributions; and
iii. infer the evolutionary history of rhizobia in the CCR and determine their potential role in driving the diversification of legumes in the CCR.

The first aim is pursued in chapter two of the thesis where the chemical and physical properties of soils associated with legume assemblages in the Cape Peninsula, a microcosm of the CCR, were analysed. Two hypotheses were tested: (i) the composition of legume species assemblages is correlated with the chemical and physical properties of their soils (i.e. edaphic factors); and (ii) Phosphorus (P), together with Aluminium (Al), Calcium (Ca) and Iron (Fe), i.e. cations that make P unavailable to plants, are the main edaphic parameters that drive the distribution patterns of the legumes.

The second aim is the subject of chapter three of the thesis which tests the hypothesis that the distribution of rhizobia in the CCR landscapes is phylogenetically structured with respect to the three ecological parameters. The implications of the findings for legume distributions in the CCR are discussed.

The final aim of the study is addressed in chapter four, where the hypotheses that (i) *Burkholderia* is the ancestral symbiont of CCR legumes and (ii) horizontal gene transfer has occurred within and between the various rhizobial genera found in the CCR, are tested. Based on the outcomes of testing these hypotheses, the potential role of HGT in driving the diversification of legumes is discussed.

A general discussion, conclusions and some recommendations for future research form the fifth and final chapter of the thesis.

# 2. Distinct edaphic habitats are occupied by discrete legume assemblages with unique indicator species in the Cape Peninsula of South Africa

## 2.1 Introduction

Legumes generally require higher nutrient levels relative to other plants (McKey, 1994; Reed, 2017) and yet they are the second most speciose plant family in the nutrient-poor CCR of South Africa (Goldblatt and Manning, 2000; Manning and Goldblatt, 2012). Despite being the most speciose lineage, legumes are not the dominant flora in any given vegetation unit of the CCR and they tend to have patchy distributions in which distinct legume assemblages occupy particular niches within a given habitat. The factors that drive these distribution patterns are not yet fully understood. A recent study conducted by Chimphango et al. (2015) showed that CCR legumes generally occupy nutrient-enriched islands within the landscape as compared to their surrounding non-legume vegetation. While this finding provides some clues on the patchy distribution of legumes in the landscape, the pattern of having distinct legume assemblages occupying particular habitats in the CCR remains unresolved and this is the main subject of this chapter.

Considering that nitrogen (N) and phosphorus (P) are the main limiting nutrients in the CCR (Cramer, 2010; Lambers et al., 2011; Maistry et al., 2013; Stock and Verboom, 2012), it is not surprising that the majority of the legumes that occur in this ecosystem are those that have adaptations for acquiring these two nutrients (Allsopp and Stock, 1993; Beukes et al., 2013; Gerding et al., 2012; Kanu and Dakora, 2012; Lamont, 1982; Lemaire et al., 2015a; Maistry et al., 2015; Power et al., 2010). This suggests that edaphic factors could be acting as a selecting force or as an environmental filter for the legumes and thus could play a significant role in influencing their distribution in the landscape. Studies have shown that when exposed to abundant P some legumes of the CCR tend to accumulate excessive amounts of P in their shoots but without a corresponding increase in biomass (Maistry et al., 2013; Power et al., 2010). This feature, which is common among plants from P-limited environments, indicates a poor ability to down-regulate the uptake of P and it can lead to phosphorus toxicity (Shane et al., 2008). Thus, such legumes are likely to be restricted to habitats that have low P levels. Furthermore, most of the soil P is often bound to iron (Fe), calcium (Ca) and aluminium (Al) cations (Mitchell et al., 1984; Payn and Clough, 1989; Witkowski and Mitchell, 1987) and thus not readily available for plants. The concentrations of P and these cations vary widely across the different habitats and soil types of the CCR (Chimphango et al., 2015; Mitchell et al., 1984; Richards et al., 1997b; Witkowski and Mitchell, 1987). Therefore, P and these cations might be important in driving species distributions in the CCR, including the legumes.

The potential influence of edaphic factors on the distribution of legumes within the CCR is especially probable when considering its highly heterogeneous edaphic conditions (Cowling, 1990; Linder, 2003). The region's edaphic heterogeneity includes differences in the parent rock materials from which the soils are formed, soil nutrient content, pH, salinity and texture. For example, the CCR's quartzitic sandstones are acidic, coarse-grained and nutrient-poor, whereas the granite, limestone and shale-derived soils are less acidic, finer textured and relatively more fertile (Stock and Lewis, 1986; Witkowski and Mitchell, 1987). Studies have shown that edaphic heterogeneity can cause selection for certain plant traits and specialisation for particular soil types or extreme edaphic conditions (e.g. acidic conditions, heavy metals and salinity), with the result that particular plant species become adapted to limited conditions (Ellis and Weis, 2006; Pregitzer et al., 2010; Smith et al., 2012). Thus, each legume assemblage could be composed of species that have been selected for the local edaphic conditions. Hence, it is hypothesized that distinct legume assemblages occupy distinct edaphic habitats in the CCR.

Soil is the foremost substrate that anchors most land plants and from which they obtain water and nutrients. Numerous studies from various ecosystems of the world demonstrate the significance of edaphic factors in the distribution of plant species (Bertrand et al., 2012; Clark et al., 1999; Dubuis et al., 2013; Hall et al., 2004; John et al., 2007; Soares et al., 2015). However, the key edaphic parameters that drive species distributions tend to vary with the ecosystem and taxa involved, making it difficult to generalise. Although the edaphic heterogeneity of the CCR is well recognised (Cowling, 1990; Cowling et al., 2009; Goldblatt and Manning, 2002; Linder, 2003; van der Niet et al., 2006; Verboom et al., 2004), the role of edaphic factors in driving plant species distributions has received very little attention from researchers. The only two existing studies examined the role of competition between closely related species, and edaphic factors in structuring plant communities (Esler and Cowling, 1993; Richards et al., 1997a). The study by Esler and Cowling (1993) concentrated mostly on species of the genus *Pteronia* L. in the semi-arid Karoo biome, while that of Richards et al. (1997a) focused on three pairs of Proteaceae species in the Agulhas Plains. Other studies have rather focused on the role of edaphic factors in driving adaptive radiations in CCR plants, e.g. the genus *Argyroderma* L.Bolus (Ellis and Weis, 2006; Ellis et al., 2006); shifts in pollination systems in Geraniaceae, Iridaceae and Orchidaceae (van der Niet et al., 2006) and the evolution of alternative persistence strategies (i.e. reseeding vs. resprouting) in the grass genus *Ehrharta* Thunb. (Verboom et al., 2004). Therefore, a study of the role of edaphic heterogeneity on the distribution of the diverse and speciose family (Fabaceae) will contribute towards a better understanding of the biogeography of the region's flora.

If edaphic factors drive the distribution of legumes in the CCR, it should be possible to identify species or groups of species whose presence or absence in a given habitat signals the prevailing edaphic conditions of the habitat, i.e. indicator species (Siddig et al., 2016). Indicator species are species that can be used as ecological indicators of community types or environmental conditions due to their niche preferences (Niemi and McDonald, 2004). Therefore, the aim of the study was to determine the role of edaphic factors in driving the distribution of legume species in the Cape Peninsula (a microcosm of the CCR) and to identify indicator species within the legume assemblages. Such knowledge will inform conservation planning aimed at kerbing biodiversity loss and allows for prediction of how changes in nutrient deposition, nutrient cycling processes, fire regimes and climate, might impact the composition of the flora. Physical and chemical properties of soils associated with legume species assemblages were analysed to address the following key questions: (i) do edaphic factors (physical and chemical) influence the distribution of legumes in the Cape Peninsula, if so, (ii) what are the key edaphic parameters driving the patterns, and (iii) are there any indicator species for the various soil types in the Cape Peninsula? It was hypothesized that the distribution of legume species is linearly related to edaphic factors and that the interaction of P, Al, Ca and Fe drives the legume species assemblages in the Cape Peninsula.

## 2.2 Materials and Methods

### *Study site*

The study was conducted in the Cape Peninsula, the area located within the southern core of the CCR, west of the 18°30′ E line of longitude (Figure 2.1), as defined by Helme and Trinder-Smith (2006). Topographically, it ranges from the low-lying Cape Flats to gentle slopes and hills, culminating in ridges and plateaus with a maximum altitude of 1113 m on Table Mountain (Cowling et al., 1996). The summits and upper slopes of the Cape Peninsula Mountains are mostly associated with acidic and nutrient-poor sandstone-derived soils, whereas the lower slopes, which are associated with colluvium underlain by granite or shale, are less acidic and nutrient richer (Cowling et al., 1996). The coastal areas are mostly associated with alkaline and poorly consolidated sands (Cowling et al., 1996). The Cape Peninsula receives most of its rainfall during winter (June-August), while its summers are mostly dry. Annual precipitation ranges from 402 mm in Cape point, to over 1000 mm in Maclear's Beacon, the highest point in the Cape Peninsula (Cowling et al., 1996; Higgins et al., 1999). Annual precipitation varies considerably within very short distances owing to the highly heterogeneous altitude, aspect and topography (Cowling et al., 1996).

Floristically, the Cape Peninsula has a total of 2285 species of vascular plants (Trinder-Smith et al., 1996), of which 158 species are endemic (Helme and Trinder-Smith, 2006). The top five species-rich families in the Cape Peninsula are Asteraceae (286 species), Iridaceae (168 species), Fabaceae (162 species), Poaceae (141 species) and Ericaceae (112 species) (Trinder-Smith et al., 1996). The major vegetation types that occur in the Cape Peninsula include fynbos, renosterveld, dune strandveld, and afro-montane forest (Mucina and Rutherford, 2006; Poulsen and Hoffman, 2015; Trinder-Smith et al., 2006). Fynbos, the predominant vegetation type is characterised by sclerophyllous shrubs and is dominated by Restionaceae, Proteaceae and Ericaceae species. Renosterveld occupies only 5 % of the Cape Peninsula (Cowling et al., 1996) and is dominated by Asteraceae (especially Renosterbos: *Elytropappus rhinocerotis* (L.f) Less.), Poaceae species and various geophytic species. Dune strandveld consists of plants that grow on the inland edge of sandy beaches, stabilising the soil. Forests are a minor component of the Cape Peninsula, restricted to moist valleys on eastern slopes and along river banks and consist of typical temperate tree genera such as *Podocarpus* L. (Campbell and Moll, 1977).

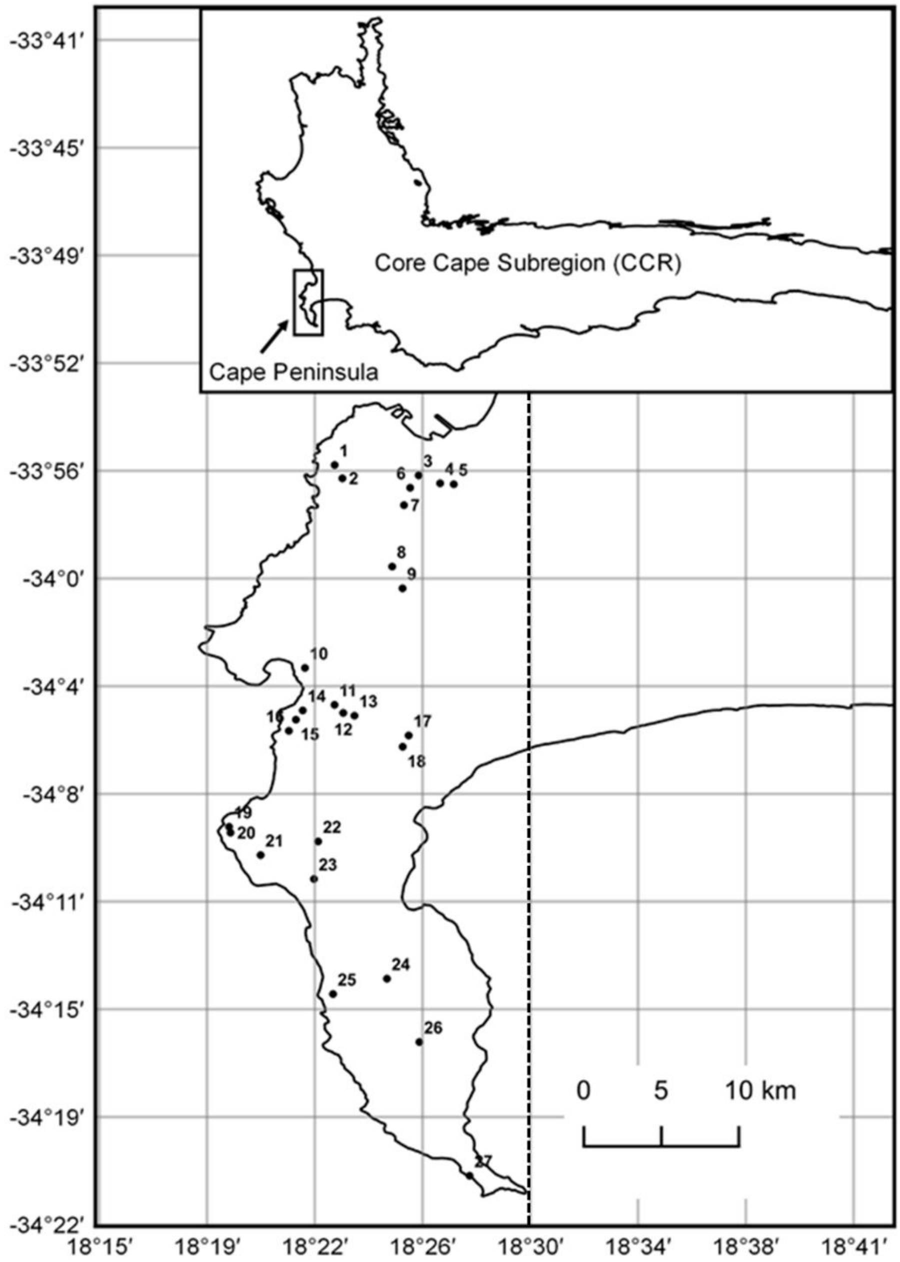


**Figure 2.1.** Location of the Cape Peninsula and the sampling sites within the CCR of South Africa. Grid squares of the Cape Peninsula are those located to the left of the 18°30' line (dashed line). Sampling sites are represented by the closed circles and their names are represented by the numbers as follows: 1 = Lions Head, 2 = Lions Head 1, 3 = Lower Devils Peak G, 4 = Lower Devils Peak UE, 5 = Lower Devils Peak RM, 6 = Tafelberg, 7= Upper Devils Peak, 8 = Cecilia Ravine, 9 = Blue Gums, 10 = Blackburn Ravine, 11= Disa Ridge 1, 12 = Disa Ridge, 13 = Silvermine Dam, 14 = Chapmans Peak foot, 15 = Lower Chapmans Peak, 16 = Chapmans Peak, 17 = Steenberg, 18 = Silvermine East, 19 = Kommetjie, 20 = Slangkop VS, 21= Slangkop Egate, 22 = Redhill, 23 = Kleinplaas Dam, 24 = Smitswinkel Flats, 25 = Cape Point, 26 = Cape Point BF and 27 = Cape of Good Hope.

Generally, the Cape Peninsula is a well-collected area of the CCR due to its location in the vicinity of two major herbaria (Bolus and Compton) and the fact that several botanists and general plant collectors have extensively studied its flora over the last two centuries. This suggests that herbarium specimen-based records regarding the geographic distribution of species are likely to match their actual distributions. Its small size allows for detailed fine-scale sampling and its high legume diversity (162 species) has all the major CCR legume lineages represented. Together with its high ecological heterogeneity, these attributes make the Cape Peninsula a microcosm of the CCR and thus, a suitable area to study the distribution of CCR legumes.

### *Sampling*

Herbarium specimen records, dating as far back as the 1700s, were sourced from the Bolus Herbarium (BOL) and the Pretoria Computerised Information System (PRECIS) database of South African plants to determine the broad-scale distribution of legumes in the Cape Peninsula. These two sources carry the most comprehensive records of the CCR flora and the PRECIS database aggregates collections from all herbaria under the South African National Biodiversity Institute (SANBI). However, the PRECIS records come georeferenced to 0.25° grid square, which is not ideal for this study, where finer scale geographic coordinates were required to accurately characterise species habitats. Therefore, all specimens obtained from the PRECIS database were further georeferenced using GEOLocate version 3.2.2 (Rios and Bart, 2010) as the primary georeferencing tool and some detailed maps of the Cape Peninsula (Slingsby Maps). Most of the BOL specimens had already been accurately georeferenced, hence only those that had not been done were subjected to georeferencing.

To identify patterns of legume distributions within the Cape Peninsula (i.e. which areas have similar legume species composition) and areas of high legume species richness, the map of the Cape Peninsula was subdivided into 0.0625° grid squares. Thus, 19 such grid squares fall within the boundaries of the Cape Peninsula (Figure 2.1). Using the georeferenced data, a matrix of presence (1) or absence (0) of all known (162) Cape Peninsula legume species in each of the 19 grid squares was generated. The matrix was analysed by cluster analysis (clustering algorithm: UPGMA, distance: Raup-Crick) using the program PAST version 3.12 (Hammer et al. 2001).

Results of the cluster analysis informed field sampling, which sought to capture the diversity in legume species assemblages, geological substrates, altitude, aspect and topography of the Cape Peninsula. Accordingly, at least one site was selected from each cluster, and where members of the same cluster fell into different geological substrate types (inferred from overlaying a geology layer on the grids), at least one grid per geological type was sampled. Overall, a total of 27 legume sites, representing 12 of the 19 grid squares were sampled (Figure 2.1). During fieldwork, each site was surveyed for legume species present through transect walks within 20 m × 20 m quadrats. All legume species observed at each site were recorded and voucher specimens collected. Soil samples (three replicates) were randomly taken from the top layer (10–15 cm depth) of the soil profile by means of a soil corer or garden trowel. At the laboratory, the soil samples were air-dried at room temperature, all plant debris and roots removed and sieved through a 2 mm mesh.

### *Analysis of chemical and physical properties of soils*

The dried soil samples were analysed for 31 chemical and physical characteristics including macro and micro-elements, beneficial and toxic elements. The macronutrients studied were: C, Ca, K, Mg, N (total N and $NH_4$-N), P (total P and P Bray II) and S (Maathuis, 2009). Micronutrients were: Cu, Fe, Mn, Mo, Ni and Zn (Hänsch and Mendel, 2009); while the beneficial elements studied were: Al, Co, Ga, Na, Se, Si and V (Pilon-Smits et al., 2009; Simon et al., 1989). Toxic elements were: As, Cd, Hg, Pb and Sn (Nagajyoti et al., 2010) and the physical characteristics analysed were pH, sand, silt and clay contents. Each soil sample was partitioned into three portions. One portion was sent to the Elsenberg Laboratory of the Western Cape Government's Department of Agriculture for the analysis of Ammonium ($NH_4$-N), Calcium (Ca), Carbon (C), clay, Magnesium (Mg), pH, Potassium (K), sand, silt and Sodium (Na) content. Water-soluble concentrations of the elements (i.e. available to plants) were measured for Ca, K, Mg, and Na. Another portion was sent to Bemlab (Somerset West, South Africa) for the analysis of Total P, P Bray II (available for plants) and total N. The final portion was crushed into a fine powder using mortar and pestle and analysed for total concentrations of the following 22 chemical elements: Al, As, Ca, Cd, Co, Cu, Fe, Ga, Hg, K, Mg, Mn, Mo, Ni, P, Pb, S, Se, Si, Sn, V and Zn using an X-Ray Fluorescence Spectrophotometer.

### *Determining the species composition of sites*

Although herbarium specimen records were initially used to identify the broad scale legume species assemblages in the Cape Peninsula to locate sampling areas, such records are subject to spatial biases and georeferencing errors (Stropp et al., 2016), which compound accurate determination of the fine-scale species composition of an area. Therefore, the data collected during the field surveys that were conducted as part of this study were used to compile a matrix of legume species compositions of the sampled sites. Each species was scored for presence (1) or absence (0) in each of the sites to develop a presence/absence matrix.

### *Statistical analyses*

The soil data were natural log-transformed where necessary to achieve normality. A hierarchical cluster analysis (clustering algorithm: Ward's, distance: Euclidean) of all soil parameters was conducted to group the sites based on overall similarity of soil characteristics. A canonical discriminant function analysis (DFA) was used to determine the soil parameters that discriminate between the groups established from the cluster analysis. A stepwise DFA, whereby the variables are entered into the model one after the other, each time choosing the variable that maximises the discriminatory power of the model was used. Additionally, analysis of variance (ANOVA) was performed on the individual variables that had higher discriminatory contributions (based on their standardised coefficients) in the DFA. Tukey HSD tests were used to identify significantly different means ($p < 0.05$). Correlations between total and plant available concentrations of Ca, K, Mg, and P were determined using Pearson correlation analysis. Except for the DFA, which was performed in Statistica (Statistica version 13, StatSoft, Inc., Tulsa, OK, USA), all analyses were performed in R (R version 3.3.2, R-core team 2016).

To test for correlation between species composition and edaphic factors, canonical correspondence analysis (CCA) was employed. The CCA is a multivariate technique that uses gradients in environmental attributes to depict species' habitat preferences on ordination graphs in which the axes are linear combinations of the environmental factors (Ter Braak, 1986). This technique was chosen over other ordination methods because of its robustness to skewed species distributions, unusual sampling approaches and cases where not all variables driving species composition are known (Palmer, 1993). Analyses were performed using the 'vegan' package in R. To select variables that best explain the species-environment relationship while accounting for correlation among variables (multicollinearity), stepwise variable selection was performed using the 'ordistep' function and the variance inflation factors (VIF) of the variables were examined (Ter Braak, 1987). Variables with VIF > 10 were excluded from the analyses. Permutation tests (10 000 permutations) were run to evaluate the statistical significance of the CCA model, its terms (soil variables) and the CCA axes. Preliminary analyses showed that sites from dune sands had exceptional values for most variables, which skewed the rest of the sites in the CCA analyses, thus they were excluded from subsequent analyses. Additionally, three granite-shale sites were ordinated separately from the main group of granite-shale sites, but closer to the sandstone sites. Thus additional cluster analysis, DFA and CCA were performed on the three granite-shale sites and all the sandstone sites to determine if there were any distinct groupings within this subset of the data.

Partial Mantel tests were conducted to test for spatial autocorrelation in the species-environment relationships using the 'ecodist' package (Goslee and Urban, 2007) in R. The partial Mantel test examines the correlation between two distance matrices while controlling for a third one. Consequently, it is widely used to identify spatial autocorrelation in ecological studies where the effects of environmental variables on species composition are determined while taking into account geographic distance (Giraldo et al., 2016; Martiny et al., 2011; Talbot et al., 2014). The statistical significance of the partial Mantel test results was assessed using 10 000 permutations.

The strength of association between legume species composition and site groupings based on edaphic parameters was determined using Indicator values, which measure the predictive value of a species as an indicator of particular site groups (De Cáceres et al., 2010). Analyses were performed in R using the ‘indicspecies’ package. The ‘IndVal.g’ function of ‘multipatt’, which corrects for unequal group sizes was employed. Permutation tests (10 000) were performed to assess the statistical significance of the association between species and groups of sites. The species presence and absence matrix for the 27 sites was used and the sites were grouped based on their distinctness in overall edaphic characteristics as depicted by the results of the DFA performed on the soil data.

## 2.3. Results

### *Legumes species composition of sites*

The total number of legume species encountered during the surveys was 67, representing 22 genera (out of 25), and 9 tribes (out of 10). The list of species observed in each of the sampled sites is found in Appendix A1. Most species had restricted distributions e.g. 38 % of the species occurred on only one site; 26 % on two sites; 14 % on three sites; and 22 % of the species occurred on more than three sites (Figure 2.2a). The few widespread species encountered include; *Bolusafra bituminosa* Kuntze and *Aspalathus ericifolia* Willd. ex. Walp., both occurring on seven sites, *Otholobium virgatum* C.H. Stirt. was recorded on 11 sites and *Psoralea pinnata* L. occurred on 12 sites (Appendix A1). A majority of the species (66 %) occurred in only one soil type, 21 % on two soil types and only 13 % of the species occurred on three different soil types (Figure 2.2b). Although four soil types were studied (i.e. dune sands, granite, sandstone and shale), no single species occurred on all of them.

### *Chemical and physical properties of soils*

The cluster analysis of the overall soil characteristics grouped the 27 sites into three main groups (Appendix A2). One group comprised sixteen sites, which are sites associated with Table Mountain sandstone parent rock material. Another group comprised two sites that are associated with coastal dune sands, while the third group (nine sites) comprised sites from both shale and granite substrates, henceforth referred to as granite-shale sites (Appendix A2). Of the 31 soil characteristics studied, ten had significant contributions towards the discriminant function, thus the ten were used for the DFA of the three groups obtained from the cluster analysis. The standardised coefficients of the ten variables on the first and second roots of the DFA are shown in Appendix A3. Chi-square tests showed that the two roots (discriminant functions) were significant ($p < 0.001$).

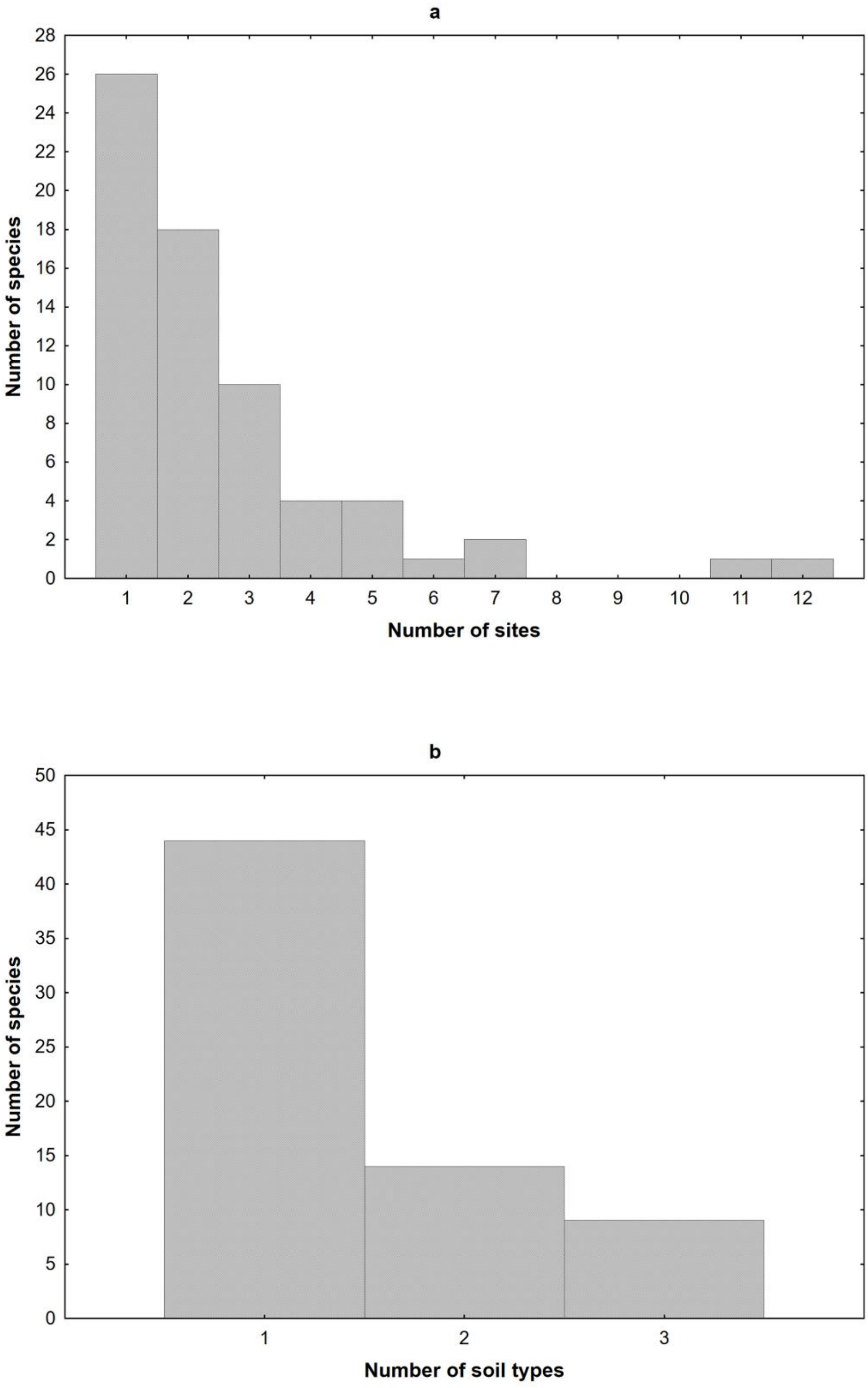


**Figure 2.2.** Bar plot of the total numbers of legume species encountered in different numbers of sites (**a**) and soil types (**b**).

Root 1, which explained 69.09 % of the variance in the model, was strongly influenced by Na, Fe, Al and K, in descending order, while for Root 2, which explained 30.91 % of the variance, C and P were the most important elements (Appendix A3). A scatterplot of the canonical scores confirmed the separation of the three groups along both roots (Figure 2.3). In terms of univariate differences between the groups, sites from granite-shale substrates had significantly higher Al, Fe, K, Mn and N, but lower Sand content than the sites from Sandstone and Dune sands (Table 2.1). In contrast, the sites from Dune sands had higher concentrations of P Bray II and Na than sites from the other two soil types (Table 2.1).

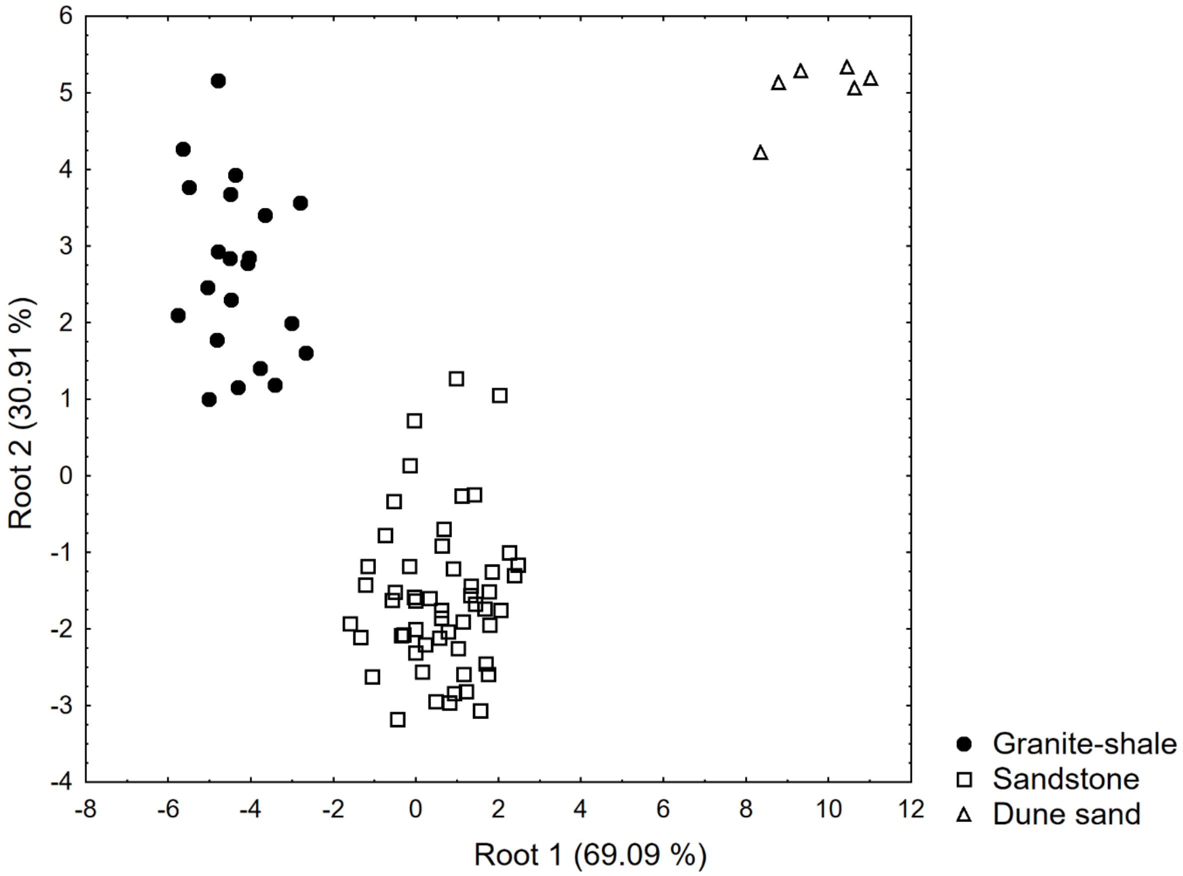


**Figure 2.3.** Scatterplot of canonical scores for the two discriminant functions of the soil data for all sites showing separation of the three site groups.

An assessment of correlation between total concentrations of Ca, K, Mg and P, as measured using the x-ray fluorescence technique and their corresponding plant available portions showed a positive relationship: Pearson r values = 0.82, 0.85, 0.54 and 0.81, respectively ($p < 0.0001$).

**Table 2.1.** Mean±SE of the concentrations of the ten variables separating the three main soil groups associated with the 27 sites. Values with different letters are significantly different at $p < 0.05$.

| **Soil variable** | **Granite-shale** | **Sandstone** | **Dune sands** |
|---|---|---|---|
| Al (%) | 8.632±0.468$^a$ | 2.174±0.292$^b$ | 2.633±0.876$^b$ |
| P (%) | 0.048±0.007$^a$ | 0.027±0.004$^b$ | 0.155±0.013$^c$ |
| Mn (%) | 0.049±0.004$^a$ | 0.003±0.002$^b$ | 0.001±0.007$^b$ |
| Fe (%) | 3.035±0.154$^a$ | 0.342±0.096$^b$ | 0.109±0.287$^b$ |
| P Bray II (mg kg$^{-1}$) | 8.591±3.166$^a$ | 2.083±1.974$^a$ | 77.878±5.923$^b$ |
| Na (mg kg$^{-1}$) | 79.381±14.428$^a$ | 33.556±8.998$^b$ | 222.167±26.993$^c$ |
| K (mg kg$^{-1}$) | 160.714±8.337$^a$ | 35.037±5.199$^b$ | 23.333±15.597$^b$ |
| N (%) | 0.123±0.01$^a$ | 0.078±0.006$^b$ | 0.033±0.018$^b$ |
| C (%) | 2.218±0.293$^a$ | 2.083±0.183$^a$ | 1.467±0.548$^a$ |
| Sand (%) | 86.952±0.715$^a$ | 94.704±0.446$^b$ | 96.0±1.338$^b$ |

### *Correlation between legume species composition and edaphic factors*

Soil characteristics that best explained the species-environment relationship after removing collinear variables were clay content, Fe, K and Zn, hence they were used for the CCA. The CCA confirmed a significant linear relationship between species composition and soil characteristics (permutation $F = 1.910$, $p = 0.001$). The first and second canonical axes, which were both significant ($p < 0.001$), jointly explained 82 % of the constrained inertia (i.e. variance explained by the soil variables). The sites formed two main groups which separated along the first CCA axis, one comprised of six granite-shale sites and the other comprising three granite-shale sites ordinated next to the sandstone sites (Figure 2.4a). The six granite-shale sites were associated with higher values of all four key soil characteristics. Legume species were distributed throughout all four quadrants of the first two CCA axes (Figure 2.4b), indicating distinct edaphic optima for the different legume species assemblages.

Due to the clustering of three granite-shale sites with the sandstone sites (Figure 2.4a), separate cluster analysis, DFA and CCA were performed on this group to determine if there were any distinct sub-groups within it. The cluster analysis recovered two main clusters of the sandstone sites (denoted Sandstone_1 and Sandstone_2) and one cluster containing all three granite-shale sites and one sandstone site (denoted Granite-shale_1) (Appendix A4). Six soil variables (C, Fe, K, N, P Bray II and pH) contributed significantly to the discrimination of the three clusters (Appendix A5) and the DFA confirmed significant differences between the clusters ($p<0.0001$). Their ordination along the two roots is presented in Figure 2.5. Root 1 (accounting for 84.55 % of the variance) was strongly associated with Fe and Root 2 (15.45 %) was strongly associated with C (Appendix A5). For the univariate analysis of variance among the groups, Fe and P Bray II were significantly higher in the granite-shale sites than the two groups of Sandstone sites (Table 2.2). The two groups of sandstone sites differed significantly in their C, K and N concentrations (Table 2.2).

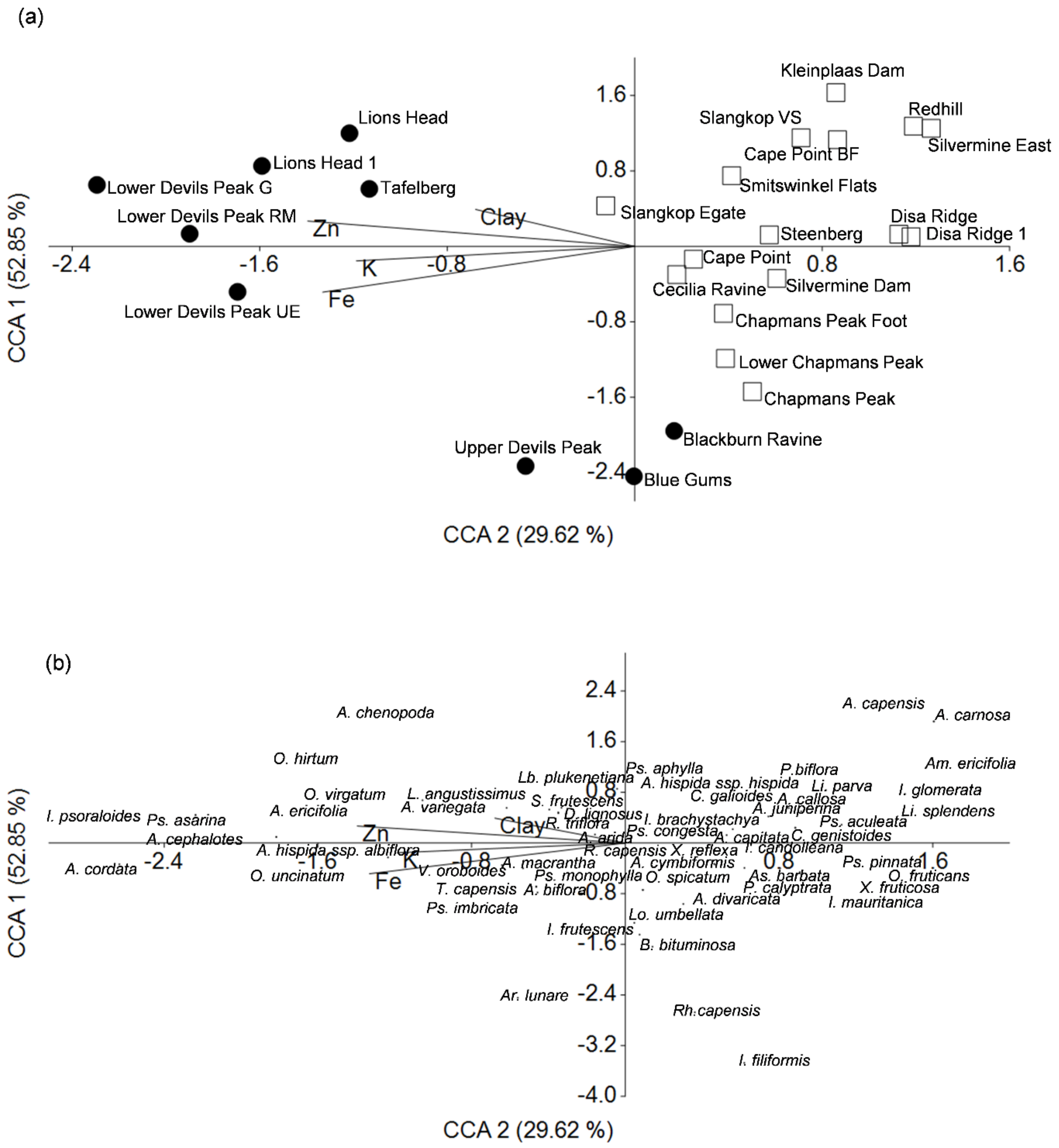


**Figure 2.4.** (a) species-conditional biplot based on a CCA of soil characteristics and legume species composition (for all sites except the dune sand sites), showing the ordination of legume sites along the first two CCA axes. Sandstone sites are represented by open squares, while the closed circles represent granite–shale sites. Soil characteristics are indicated by lines whose lengths indicate the strength of their correlations with the ordination axes. (b) species-conditional biplot based on a CCA of soil characteristics and legume species composition (for all sites except the dune sand sites), showing the ordination of legume species along the first two CCA axes. Soil characteristics are indicated by lines, whose lengths indicate the strength of their correlations with the ordination axes. Genera are abbreviated as follows: A = *Aspalathus*, Am = *Amphithalea*, Ar = *Argyrolobium*, B = *Bolusafra*, C = *Cyclopia*, D = *Dipogon*, I = *Indigofera*, L = *Lotus*, Le = *Lessertia*, Li = *Liparia*, Lo = *Lotononis*, O = *Otholobium*, P = *Podalyria*, Ps = *Psoralea*, R = *Rafnia*, Rh = *Rhynchosia*, T = *Tephrosia*, V = *Virgilia*, W = *Wiborgia*, X = *Xiphotheca*.

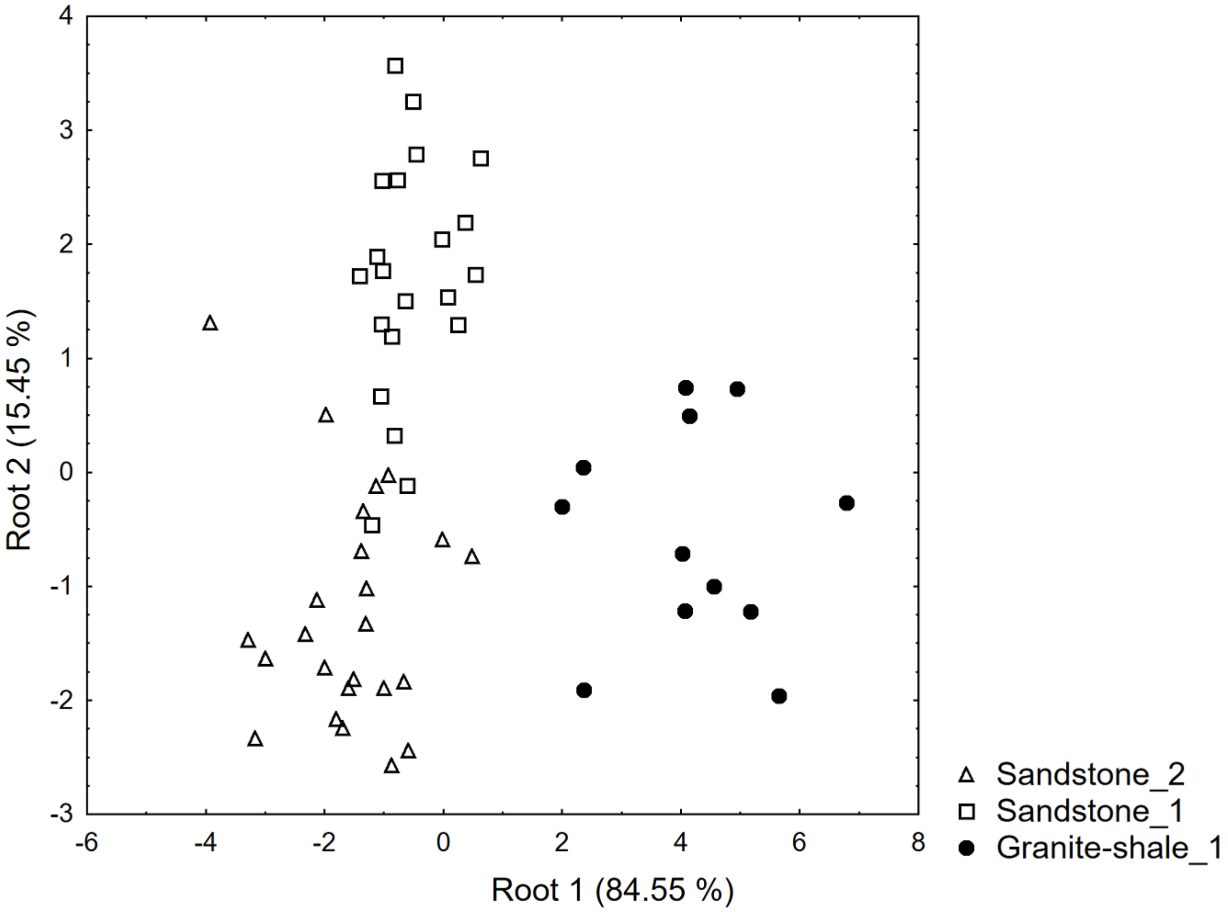


**Figure 2.5.** Scatterplot of the canonical scores for the two discriminant functions of the subset of the data comprising three granite–shale sites and all the sandstone sites.

Iron, K and S were the variables that contributed significantly to the species-environment relationship for this subset of the data. The CCA showed a significant correlation between species composition and soil characteristics (permutation F = 1.764, p = 0.001). The first two CCA axes explained 69.29 % (p = 0.001) and 30.69 % (p = 0.002) of the constrained inertia, respectively. The granite-shale_1 sites clustered together and separately from the sandstone sites, and were associated with higher Fe and K (Appendix A6). With the exception of two sites, the sites of the Sandstone_2 group formed a distinct cluster, while those from the Sandstone_1 group had no distinct grouping (Appendix A6). The species were distributed on all four quadrants of the CCA plot, with various assemblages having distinct edaphic optima (Appendix A6).

**Table 2.2.** Mean±SE of the concentrations of the six key variables separating the three main soil groups associated with the sandstone and three granite-shale sites only. Values with different letters are significantly different at $p < 0.05$.

| **Soil variable** | **Sandstone_1** | **Sandstone_2** | **Granite-shale_1** |
|---|---|---|---|
| Fe (%) | $0.275±0.118^a$ | $0.129±0.110^a$ | $1.508±0.156^b$ |
| K (mg $kg^{-1}$) | $46.286±4.099^a$ | $22.583±3.834^b$ | $52.167±5.423^a$ |
| N (%) | $0.085±0.008^a$ | $0.055±0.008^b$ | $0.117±0.011^a$ |
| C (%) | $2.632±0.263^a$ | $1.120±0.246^b$ | $2.956±0.347^a$ |
| P Bray II (mg $kg^{-1}$) | $1.410±0.372^a$ | $1.597±0.348^a$ | $4.522±0.492^b$ |
| pH (KCl) | $3.790±0.166^a$ | $4.196±0.155^a$ | $4.142±0.220^a$ |

The partial Mantel test on species composition vs. geographic distance while accounting for edaphic factors was not significant (Mantel $r = 0.090$, $p = 0.097$), whereas the correlation between edaphic factors and species composition while taking into account geographic distance, was significant (Mantel $r = 0.280$, $p = 0.0001$), indicating that there was no significant spatial autocorrelation in the data. Similar results were obtained for the analysis of the subset of the data (comprising three granite-shale and all sandstone sites). Thus the observed correlations in the species-environment relationships were not compounded by spatial autocorrelation.

Based on the results of the DFA, the following groups were used for indicator species analysis: Granite-shale, Dune sand (Figure 2.3), Sandstone_1, Sandstone_2, and Granite-shale_1 (Figure 2.5). A total of 13 species were identified s significant indicator species, three for Sandstone_2, two species for Dune sands, two for Granite-shale_1, and six species for the main Granite-shale sites (Table 2.3). No species were associated with more than one group, and only one site group (Sandstone_1) did not have any significant indicator species assigned to it.

**Table 2.3.** Results of Indicator species analysis for the legume sites of the Cape Peninsula. Presented for each species are the probability that the surveyed site belongs to the target site group (A), the probability of finding the species in sites belonging to the site group (B), the Indicator value (IndVal), which is a product of A and B (De Cáceres *et al*. 2010) and the statistical significance of the association (p-value).

| **Species** | **Group** | **A** | **B** | **IndVal.** | **p-value** |
|---|---|---|---|---|---|
| *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Sandstone_2 | 0.889 | 1.00 | 0.943 | 0.00001 |
| *Aspalathus capensis* (Walp.) R. Dahlgren | Sandstone_2 | 1.00 | 0.875 | 0.935 | 0.00001 |
| *Aspalathus carnosa* Eckl. & Zeyh. | Sandstone_2 | 1.00 | 1.00 | 1.00 | 0.0002 |
| *Otholobium bracteolatum* (Eckl. & Zeyh.) C.H. Stirt. | Dune sand | 1.00 | 1.00 | 1.00 | 0.0392 |
| *Psoralea repens* P.J. Bergius | Dune sand | 1.00 | 1.00 | 1.00 | 0.0392 |
| *Argyrolobium lunare* Druce | Granite-shale_1 | 0.857 | 1.00 | 0.926 | 0.0012 |
| *Indigofera filiformis* Thunb. | Granite-shale_1 | 0.800 | 1.00 | 0.894 | 0.0021 |
| *Aspalathus cordata* (L.) R. Dahlgren | Granite-shale | 1.00 | 1.00 | 1.00 | 0.0001 |
| *Indigofera psoraloides* L. | Granite-shale | 1.00 | 1.00 | 1.00 | 0.0001 |
| *Otholobium hirtum* (L.) C.H. Stirt. | Granite-shale | 1.00 | 1.00 | 1.00 | 0.0001 |
| *Psoralea asarina* (P.J. Bergius) T.M. Salter | Granite-shale | 1.00 | 1.00 | 1.00 | 0.0001 |
| *Aspalathus hispida* ssp. *albiflora* (Eckl. & Zeyh.) R. Dahlgren | Granite-shale | 1.00 | 0.833 | 0.913 | 0.0004 |
| *Aspalathus chenopoda* L. | Granite-shale | 0.800 | 1.00 | 0.894 | 0.0015 |

## 2.4 Discussion

The results of the DFA on the soil data showed that soils of the Cape Peninsula differ in chemical and physical properties that are known to influence plant growth and metabolism. Although not all soil characteristics were analysed for water-soluble portions (i.e. plant available concentrations), the findings of significant positive correlations between total and available concentrations of four elements (where concentrations were measured from both total and plant available forms) validate inferences of the role of the studied edaphic factors in driving legume species distributions. The results indicate that the soils differ in their physical and chemical properties according to their parent geology, except for the granite and shale, which were grouping together. Additionally, variation exists within the different soil types as indicated by the observation of a further significant split within the sandstone and granite-shale sites (Figure 2.5). Importantly, this variation within soil types can be associated with varying species composition on the landscape, as was shown in the present study.

The CCA confirmed a linear relationship between legume species composition of sites and edaphic factors, and this relationship was not due to spatial autocorrelation, as indicated by the results of the partial Mantel tests. Thus, habitats that differ in their soil characteristics are associated with different legume species assemblages. This is consistent with broad vegetation patterns in the CCR, in which landscapes of a particular geology are occupied by specific taxa. For example, Ericaceae, Proteaceae and Restionaceae tend to be restricted to sandstone-derived soils, while *Mesembryanthemum*, *Oxalis* and Scrophulariaceae mostly occur on shale (Linder, 2003). This suggests that edaphic factors play a significant role in determining what species can occupy particular habitats in the CCR. Therefore, the findings confirm the hypothesis that edaphic factors influence the biogeography of legume species in the Cape Peninsula and it corroborates the hierarchical framework of Pearson and Dawson (2003), which highlights edaphic factors as important for species distributions at such local spatial scales. Similar findings have been reported from several ecosystems e.g. temperate forests (Bertrand et al., 2012), tropical rainforests (Clark et al., 1999; John et al., 2007), Alps (Dubuis et al., 2013), Mediterranean shrublands (Richards et al., 1997a) and Savannas (Soares et al., 2015). Therefore, studies on the distribution of species at local spatial scales need to consider the effects of soils.

The observation of several species being restricted to fewer sites or soil types (Figure 2.2) and the observation of distinct optima for the various legume species (Figure 2.4b and Appendix A6) indicate that the species differ in their edaphic requirements and are differentially adapted to the various edaphic conditions. These differences in edaphic preferences could have allowed the legumes to take advantage of the heterogeneous nature of the CCR’s edaphic environment through diversifying their niches, leading to speciation. This could account for the high legume species richness of the CCR and it would be consistent with the hypothesis that the high ecological heterogeneity of the Cape Peninsula promotes higher *beta* diversity, thereby leading to its high species richness (Simmons and Cowling, 1996).

The main soil characteristics involved in driving the distribution of the Cape Peninsula legumes were clay, Fe, K, S and Zn (Figure 2. and Appendix A6). The involvement of multiple elements underscores the idea that focusing on a single soil parameter for species distributions may be misleading (Chimphango et al., 2015). These soil parameters are known for their roles in plant growth and their involvement in driving plant species distributions is well documented. Chimphango *et al.* (2015) reported K among the nutrients which were higher in legume patches than nearby non-legume vegetation. In this study, K was over four times higher in the granite-shale sites than the sandstone or dune sand sites (Table 2.1). Moreover, the CCA separated granite-shale sites from sandstone sites (Figure 2.4 and Appendix A6), indicating that species on these habitats have different nutritional optima. Considering that four of the six indicator species for the Granite-shale group (Table 2.3) occur exclusively on granite-shale habitats (Appendix A1), it is possible that such species cannot survive on the low K sandstone and dune sand habitats, hence their restriction to the high K shale and granite habitats. Therefore, K must be important for legume distribution in the CCR.

Sulphur is a macronutrient involved in the synthesis of proteins, vitamins, chlorophyll and defence compounds against biotic and abiotic stress (Rausch and Wachter, 2005). Likewise, Zn is an essential component of many proteins in plants and while its deficiency may inhibit plant growth and metabolism, excess amounts of Zn can be toxic to plants (Broadley et al., 2007). Therefore, the involvement of S and Zn in the legume distributions suggests that variations in their availability in the landscape could impact species distributions. Clay content is known to influence a number of soil properties e.g. organic matter content, water retention and infiltration capacity (Diamantis et al., 2017; Zavala et al., 2014), thus, its involvement here highlights its potential significance for species distributions in the CCR.

It was hypothesized that the interaction of P with Al, Ca and Fe are the main drivers of legume distributions in the Cape Peninsula. The Fe content of granite-shale sites was up to ten times higher than that of sandstone and dune sand sites (Table 2.1 and Table 2.2) and it was the strongest driver of the legume species composition-soil relationship in the CCA (Figure 2.4 and Appendix A6). Its involvement might be linked to its tendency to bind P (the main limiting nutrient), making it unavailable to plants (Witkowski and Mitchell, 1987). Apart from its role in respiration, photosynthesis, hormone structure and DNA synthesis (Graziano and Lamattina, 2005; Jeong and Guerinot, 2009; Rout and Sahoo, 2015), Fe is needed for nodulation and N-fixation (Rotaru and Sinclair, 2009; Tang et al., 1990). Considering that CCR soils are poor in N (Kruger, 1983; Stock and Lewis, 1986; Witkowski and Mitchell, 1987), whereas legumes have a higher N requirement (McKey, 1994), N-fixation must be the primary source of N for the legumes. Indeed, a majority of the observed native legumes in the CCR are capable of N-fixation (Beukes et al., 2013; Gerding et al., 2012; Lemaire et al., 2015a). Therefore, because of its involvement in N-fixation, Fe must be important for legumes in the CCR, thus influencing their distribution. However, in excess amounts, Fe can be toxic to plants through generating oxidative stress (Kampfenkel et al., 1995), thus it could also be acting as an environmental filter, allowing only those species that can tolerate or counteract its toxicity to survive in a given habitat.

The indicator species analysis identified 13 significant associations of legume species with particular groups of sites based on their soils' chemical and physical properties. This accentuates the significance of the correlation between edaphic factors and legume species composition of sites. Thus, by observing these indicator species on a given site, one can potentially predict the edaphic conditions of the site and similarly, given the list of species occurring at a site, one can predict the prevailing edaphic conditions of the site. This observation holds true outside the Cape Peninsula, where a number of the indicator species occupy sites with (potentially) similar edaphic conditions. For example, *O. bracteolatum* is widespread within the CCR where it is found in the strandveld, a vegetation type characterising the Dune Sand ecosystem. Thus the results on indicator species have broader utility. Indicator species are important for conservation planning in that their presence or absence in a particular habitat could guide delineation of ecoregions or provide a signal of changes in the state of the environment such as nutrient deposition from air pollution. Therefore, the findings provide a basis for further studies incorporating more environmental attributes to improve understanding of the interplay between the environment and community assembly processes in the CCR.

## 2.5 Conclusions

The study has shown that the Cape Peninsula is edaphically heterogeneous and that differences in soil characteristics of sites are associated with differences in legume species assemblages. In addition, multiple soil parameters, rather than a single soil parameter are involved in driving the legume species distributions. The study also showed that some legume species can serve as indicator species for the edaphic conditions of the sites they occupy. Therefore, soil chemical and physical factors contribute significantly towards the biogeography of legumes in the Cape Peninsula. Considering that the Cape Peninsula is a microcosm of the CCR in terms of legume species diversity and edaphic habitat types, it is predicted that similar results can be obtained with a further sampling of legumes across the CCR. These findings imply that conservation planning and studies seeking to predict future distributions of the legumes, such as those relating to the impacts of global change, need to consider patterns and amounts of nutrient deposition that could affect the survival of some species in an area or their migration out of these areas. It is recommended that further studies investigate the effects of climate, dispersal and biotic interactions in order to assess their relative contributions towards the distribution of legumes in the CCR.

# 3. Differential preference of *Burkholderia* and *Mesorhizobium* to pH and soil types in the Core Cape Subregion, South Africa

## 3.1 Introduction

There is an open debate in the microbial biogeography literature regarding whether or not microorganisms are biogeographically structured (Martiny et al., 2006; Queloz et al., 2011; Yang et al., 2010), thanks to the Baas Becking hypothesis that "everything is everywhere, but the environment selects" (O'Malley, 2007). The premise of the hypothesis is that since microorganisms are small, they reproduce rapidly, they have dormancy stages and they have high dispersal potential; it follows that they should not be limited by geographical barriers and distances (Fontaneto et al., 2008; Fontaneto and Hortal, 2012). However, there is a growing body of evidence from studies on archaea, bacteria, fungi and protists, which points to the existence of microbial biogeographic structure (Geml, 2017; Rout and Callaway, 2012; Sánchez-Ramírez et al., 2017; Shen et al., 2013; Telford et al., 2006; Whitaker et al., 2003).

Like the other microorganisms alluded to above, the various rhizobial genera exhibit some notable biogeographic structuring at local, regional, continental and global scales (Sprent et al., 2017). For example, while *Burkholderia* is the predominant symbiont of Mimosoid legumes in the Brazilian Caatinga and Cerrado biomes (Bontemps et al., 2010), the Mimosoid legumes occurring in Mexico are predominantly nodulated by Alphaproteobacteria, particularly the genera *Rhizobium* and *Ensifer* (Bontemps et al., 2016). Genome level studies have also shown that the *Burkholderia* species that nodulate Mimosoid legumes in South America are genetically distinct from those that nodulate Papilionoid legumes in the CCR of South Africa, such that they are incapable of nodulating each other's hosts (De Meyer et al., 2016; Lemaire et al., 2016a; Zheng et al., 2017). Furthermore, a recent study of the symbionts of legumes found in the sub-Himalayan region of India showed that they are nodulated by distinct *Bradyrhizobium* strains that represent new species to science (Ojha et al., 2017). Likewise, legumes of the Core Cape Subregion (CCR) of southern Africa are predominantly nodulated by unique *Burkholderia* and *Mesorhizobium* strains (Beukes et al., 2013; Gerding et al., 2012; Lemaire et al., 2015a), whereas those from the Grassland and Savannah biomes of the region are largely nodulated by unique strains of *Bradyrhizobium* (Beukes et al., 2016). Therefore, the distribution of rhizobia is as prone to biogeographic limitations as other living organisms.

Some of the factors that influence the growth and distribution of rhizobia species include pH, temperature, salinity and the distribution of suitable hosts (Abdelmoumen et al., 1999; Bordeleau and Prévost, 1994; Kulkarni et al., 2000; Laranjo and Oliveira, 2011; Parker et al., 2006; Platero et al., 2016; Pires et al., 2018; Silva et al., 2018; Zahran, 1999). These factors also affect general plant growth and nodule development (Ferguson et al., 2013; Graham et al., 1982); hence, they can influence levels of nitrogen fixation. Notably, rhizobia species differ in their sensitivity to these factors. For example, species of the genus *Burkholderia* can tolerate acidic soil conditions, whereas they are replaced by alpha-rhizobia in alkaline habitats (Estrada-de Los Santos et al., 2011; Garau et al., 2009; Stopnisek et al., 2014). This could explain the predominance of *Burkholderia* in the acidic soils of the Cerrado, Caatinga biomes and other parts of South America (Bontemps et al., 2010; Bournaud et al., 2013), and in South Africa's CCR, where it associates with diverse legume tribes including the Crotalarieae, Hypocalypteae, Indigofereae, Phaseoleae and Podalyrieae (Beukes et al., 2013; Lemaire et al., 2015a). However, unlike in South America's Caatinga and Cerrado biomes, *Burkholderia* is not the only dominant rhizobial symbiont in the CCR. *Mesorhizobium* is also an abundant symbiont that associates with a wide range of legumes in the tribes Crotalarieae, Galegeae, Genisteae and Psoraleeae (Gerding et al., 2012; Lemaire et al., 2015a), and the reasons for its dominance are yet to be determined.

Contrary to *Burkholderia*'s genus-wide predilection for acidic soils (Stopnisek et al., 2014), *Mesorhizobium* species exhibit differential tolerance to environmental stress, including heavy metals, pH, salinity and temperature (Brígido and Oliveira, 2013; Brígido et al., 2017; Laranjo and Oliveira, 2011). In terms of pH, *Mesorhizobium* species can tolerate a wide range of pH conditions (3–10), despite an optimal range of pH 6–8 (Brígido et al., 2007; Chen et al., 2005). For example, *Mesorhizobium* was found to be the dominant symbiont of *Cicer arietinum* L. (chickpea) plants growing on alkaline soils in China (Zhang et al., 2012). On the other hand, a study of *Mesorhizobium* strains nodulating chickpea plants in Portuguese soils showed that some strains were able to tolerate acidic conditions down to a minimum of pH 3 (Brígido et al., 2007). This suggests that the predominance of *Mesorhizobium* in the CCR (in addition to *Burkholderia*) might be linked to its wide-ranging tolerance to different pH conditions. Notably, while acidic soil conditions are more prevalent in the CCR, particularly in the sandstone-derived soils, patches of near neutral and alkaline soils (e.g., granite, limestone and shale) also exist (Manning and Goldblatt, 2012; Mucina and Rutherford, 2006). Based on the discussion above, it appears that *Burkholderia* is more sensitive to pH, and hence, soil type, than *Mesorhizobium*. Therefore, in the case of the CCR, it is hypothesized that the distribution of *Burkholderia* species is structured by soil type and pH; while *Mesorhizobium* should be more dispersed. Moreover, *Burkholderia* should only dominate in the acidic soils, being replaced by *Mesorhizobium* species in neutral and alkaline soils.

Apart from the effects of edaphic factors on the growth and distribution of rhizobia, some studies have found correlations between turnover in the diversity of rhizobia and altitude. For example, Bontemps et al. (2010) observed that discrete *Burkholderia* species complexes were restricted to specific altitudes in the Brazilian Caatinga and Cerrado biomes. Likewise, turnover in *Ensifer* community assemblages along elevation gradients were observed in Northern China (Zhao et al., 2014). Since differences in altitude are directly related to changes in humidity and temperature (Körner, 2007), the correlations between altitude and rhizobial diversity suggest that rhizobial lineages vary in their sensitivity and tolerance to these attributes. The evident influence of altitudinal gradients on microbial diversity is not unique to rhizobia as similar patterns have been reported for other microorganisms, e.g., non-rhizobial bacteria and fungi (Bryant et al., 2008; Shen et al., 2013; Wang et al., 2011; Wang et al., 2017). Considering that altitude is highly variable in the CCR and that it is one of the major drivers of the diversification of the CCR flora (Verboom et al., 2015), it is hypothesized that altitude influences rhizobial diversity and turnover in CCR landscapes.

Considering that the soils of the CCR are generally oligotrophic (Witkowski and Mitchell, 1987) and the observation that legumes have a high nitrogen-demanding lifestyle (McKey, 1994; Werner et al., 2015), nitrogen fixation must be a key strategy for their success in the region. Since the distribution of rhizobia is constrained by environmental factors (as previously discussed), legumes might fail to establish in habitats where their rhizobial symbionts are lacking (Parker, 2001; Simonsen et al., 2017). Therefore, legumes that are highly specific in the kinds of rhizobia that they associate with might be restricted to habitats where their specific symbionts are present. A study by Lemaire et al. (2015a) showed that CCR legumes of the tribe Podalyrieae are exclusively nodulated by *Burkholderia* species. A subsequent study, which sampled multiple disjunct populations of the widespread *Podalyria calyptrata* Willd., found high levels of genetic diversity between the *Burkholderia* strains that nodulate the species (Lemaire et al., 2016b). This indicates that while *P. calyptrata* exhibits symbiotic specificity towards the genus *Burkholderia*, it associates with diverse lineages within *Burkholderia*, and this could explain its widespread distribution within the CCR. Studies on the diversity of symbionts that nodulate geographically-restricted taxa are lacking for the CCR, yet such studies could shed light on the potential influence of rhizobia specificity on legume distributions. For the CCR, one such taxon is *Indigofera superba* C.H. Stirt., a rare legume species that is restricted to the Kleinrivier Mountains within the Fynbos biome of the CCR (Raimondo et al., 2009). It occurs on sandstone-derived soils, at altitudes of 100–300 m (Stirton, 1982a). It occurs in sympatry with some widespread legume species, such as *Aspalathus carnosa* Eckl. & Zeyh., *Indigofera filifolia* Thunb. and *Psoralea pullata* C.H. Stirt. Its rhizobial symbionts are presently unknown, and it is hypothesized that rhizobia specificity contributes to its limited distribution.

The main objectives of the present study were to determine if the ecological parameters; altitude, pH and soil type influence the distribution of rhizobial symbionts that nodulate various legumes of the

Cape Peninsula as a microcosm of the CCR and to determine the diversity and phylogenetic position of rhizobia that associate with the narrowly-distributed *I. superba* in the CCR. The first objective was pursued through molecular characterization of rhizobial strains isolated from nodules of legume species collected in the field across the Cape Peninsula. These were analyzed together with the data from a previous study (Lemaire et al., 2015a) that sampled broadly within the CCR. It was postulated that if an ecological parameter limits the distribution of symbionts within the landscape, then each habitat type should predominantly harbor symbionts that are suitably adapted to the local conditions. Such symbionts would likely be genetically similar. Therefore, a significant phylogenetic signal would be expected for that parameter, i.e., closely-related species would occupy similar habitats (Blomberg and Garland, 2002). Thus, tests for phylogenetic signals for the three ecological parameters were conducted based on phylogenies of housekeeping and nodulation genes of the rhizobial strains. For the study of rhizobial symbionts of the rare *I. superba*, field nodules were sampled from multiple populations across its distribution range, and a phylogeny of its symbionts was reconstructed in a matrix that included symbionts of diverse legumes from diverse habitats within the CCR.

## 3.2 Materials and Methods

*Study site, nodule sampling and rhizobia isolation*

The primary study area was the Cape Peninsula, which is located on the south westernmost tip of the Core Cape Subregion of South Africa. Details of its climatic, edaphic, physiographic and vegetation characteristics, and the selection of sampling sites are as described in Chapter 2 of this thesis. Root nodules of legume species occurring at each site were collected and transported to the laboratory, where they were kept at 4 °C before the isolation of rhizobia, which took place within 2 to five days of sampling. Rhizobia were isolated and cultured using standard protocols (Vincent, 1970) on Yeast Extract Mannitol Agar (YEMA), with the exception that for the surface sterilization of the nodules, a 4 % solution of sodium hypochlorite (NaOCl) was used instead of acidified mercuric chloride. Rhizobial isolates were incubated at 28 °C for three to ten days depending on their growth rates and pure cultures were obtained by sub-culturing on fresh YEMA plates. Purified cultures were suspended in 20 % (v/v) glycerol solution and stored in a -80 °C freezer for long term storage. This method of obtaining rhizobial cultures was chosen over the direct sequencing of DNA from the nodules because it allows one to produce a pure culture that can be authenticated for nitrogen fixing properties. Furthermore, previous studies on CCR rhizobia have shown that each nodule is occupied by a single dominant rhizobial strain (Lemaire et al., 2016b), but nodules may be colonized by non-rhizobial bacteria.

### *DNA extraction, amplification and sequencing*

DNA was extracted from the pure rhizobial cultures using a modified version (Dludlu, 2010) of the cetyl trimethylammonium bromide (CTAB) DNA Extraction protocol (Doyle and Doyle, 1987). Polymerase chain reactions (PCR) were conducted to amplify 16S ribosomal RNA (rRNA), recombinase A (*recA*) and N-acyltransferase (*nodA*) using an Applied Biosystems GeneAmp 2700 thermal cycler (Applied Biosystems, Foster City, CA, USA). Primer pairs used were 16S-f27 and 16S-r1485 (Lane, 1991; Weisburg et al., 1991) for 16S rRNA; recA-63F and recA-504R (Gaunt et al., 2001) for *recA*; and nodA-1F and nodA-2R (Haukka et al., 1998) for *nodA*. Each PCR reaction had a total volume of 25 µL: comprising 19.92 µL of water, 2 µL of 10× buffer (Buffer A) that contained 1.5 mM $Mg^{2+}$, 0.4 µL of 10 mM dNTP, 0.8 µL each of forward and reverse primers (10 µM), 0.08 µL of *Taq* polymerase (Kapa Biosystems, Cape Town, South Africa) and 1 µL of template DNA. All DNA regions were amplified according to the reaction conditions described by the authors of the primers, i.e.: Weisburg et al. (1991) for 16S rRNA, Gaunt et al. (2001) for *recA* and Haukka et al. (1998) for *nodA*. PCR products were loaded onto ethidium bromide - stained agarose gels (1%) and subjected to electrophoresis using 0.5× Tris Borat EDTA (TBE). The gels were observed under UV light (Wavelength = 365 nm) to identify successfully amplified samples. Amplified products were enzymatically purified using the Exo/SAP protocol (Werle et al., 1994) and sent to Macrogen (Macrogen, Amsterdam, The Netherlands) for sequencing with the same primers used for PCR amplification. Newly generated sequences were deposited in the GenBank database, and the accession numbers range from MG593870 to MG593941 for 16S rRNA, MG704159 to MG704225 for *recA* and MG704226 to MG704280 for *nodA*.

### *Contig assembly and phylogenetic analyses*

The forward and reverse DNA sequence contigs were assembled using the Staden package Version 2.0.0 (Staden et al., 1998) and aligned using the online version of the Multiple Alignment using Fast Fourier Transform (MAFFT) program (Katoh et al., 2002). Identification of the isolated strains was achieved by comparing individual sequences with publically available sequences on GenBank, using the Basic Local Alignment Search Tool (BLAST) of Altschul et al. (1990). The highest matching (% similarity) GenBank sequences for the various strains are provided as part of the supplementary materials (Appendix B1). The newly-generated sequences were combined with those from the study by Lemaire et al. (2015a), which sampled various legume species throughout the CCR to allow for a broader representation. The alignments were viewed in Bioedit Version 7.1.9 (Hall, 1999), and equivocally aligned fragments were adjusted manually. Phylogenetic analyses of the aligned matrices were performed on the Cyberinfrastructure for Phylogenetic Research (CIPRES) web portal (https://www.phylo.org), through a maximum likelihood (ML) approach, using RaxML Version

8.2.10 (Stamatakis, 2014) and Bayesian inference (BI), as implemented in MrBayes Version 3.2.6 (Ronquist and Huelsenbeck, 2003). The ML analyses employed the GTR + GAMMA substitution model, and statistical support on nodes was evaluated using the non-parametric rapid bootstrapping technique (Stamatakis et al., 2008), with 1000 replicates. For the BI analysis, the best model of nucleotide substitution was determined using jModelTest2 Version 2.1.6 (Posada, 2008), employing the Bayesian Information Criterion (BIC). The BI analyses were run for as many generations as necessary to achieve chain convergence (5–10 million generations). A conservative burn-in of 25% was applied to all BI analyses, and convergence of the chains was assessed using Tracer Version 1.6 (Rambaut et al., 2014).

To determine the combinability of the different DNA data partitions, the approach used by Pirie et al. (2008; 2009) was employed. The DNA sequence data for the different genes were first analyzed separately by ML techniques as described above, and the resulting tree topologies were examined for conflicting nodes with ≥ 70 % bootstrap support. Nodes that had < 70% bootstrap support were considered unsupported, and thus, when no supported conflict was observed, the partitions were considered combinable. This approach was chosen over the widely used incongruence length difference (ILD) test (Mickevich and Farris, 1981) because the ILD only tests for overall incongruence between partitions without detecting local conflict that is due to specific taxa or clades (Pirie et al., 2009). There was no conflict observed between 16S rRNA and *recA*, and therefore, these partitions were combined in subsequent analyses. However, the *nodA* partition had significantly supported conflict with both chromosomal markers, and therefore, it was analyzed separately.

For the study of the diversity of rhizobia associated with *Indigofera superba,* root nodules were sampled from six populations of the species across its distribution range in Vogelgat Private Nature Reserve (Hermanus, Western Cape, South Africa), sampling multiple (at least five) individuals per population to capture any potential genetic variation within and between populations. Root nodules from other legumes (i.e. *Aspalathus carnosa* Eckl. & Zeyh., *Indigofera candolleana* Meisn., *Psoralea pullata* C.H. Stirt. and *Psoralea restioides* Eckl. & Zeyh.) that occur in the same locality as *I. superba* were also sampled to determine phylogenetic relationships between their symbionts. One chromosomal gene (*recA*) and one nodulation gene (*nodC*) were sequenced for this study. The use of *nodC,* instead of the *nodA* that was used in the larger study, was due to difficulty in amplifying the *nodA* gene for the *I. superba* study. Additional sequences from previous studies (Lemaire et al., 2015b; Lemaire et al., 2016b) on CCR legumes were incorporated into the dataset to determine the phylogenetic position of *I. superba* strains relative to strains nodulating other legumes in the CCR. Some sequences for reference strains, downloaded from Genbank were also included (Appendix B2).

*Determination of phylogenetic signals*

Analyses of phylogenetic signals for the various ecological parameters were conducted in R (R Core Team, 2017) using the phylogenetic trees constructed above as input, and the corresponding parameters' data as described below. Data for soil types of the sampling sites were extracted from a geological map of the CCR (shapefiles were kindly provided by the Geology Department, University of Cape Town) using the site GPS information collected during fieldwork. Soil type was coded as a binary character for each of the four soil types from which the legumes had been sampled (granite, limestone, sandstone and shale), as follows: 1 when the site belonged to a particular soil type, and 0 if it did not (Appendix B3). Phylogenetic structuring of rhizobial strains by soil type was tested using the D statistic, which measures phylogenetic signal for a binary trait (Fritz and Purvis, 2010). This was implemented using the 'phylo.d' function of the 'Caper' package, which calculates the value of D and tests for its significant departure from a random association and the clumping expected under a Brownian motion model (Orme et al., 2012). The statistic $D = 0$ denotes a phylogenetically conserved trait under a Brownian model, while $D = 1$ indicates a random distribution of traits on the tips of the phylogeny; and $D < 0$ indicates a strong phylogenetic signal, while $D > 1$ points toward phylogenetic overdispersion (Fritz and Purvis, 2010). Significance testing was conducted using 10, 000 permutations.

Altitude data for the sampled sites were recorded during field surveys using a GPS and the soil pH was determined using the methods described in Chapter 2 of this thesis. The raw data for altitude and pH are provided as part of the supplementary materials (Appendix B4). Pagel's $\lambda$ (Pagel, 1999) was used to test for the presence of phylogenetic signals for these two parameters. This metric ranges from 0 to 1, where 0 indicates that the trait evolves independently of the phylogeny and 1 indicates that the trait evolves according to the shared evolutionary history of the phylogeny's tips, i.e. presence of phylogenetic signal (Freckleton et al., 2002). The metric has proven to be robust to incomplete phylogenetic information and the presence of polytomies in the phylogenetic tree (Diniz-Filho et al., 2012; Molina-Venegas and Rodríguez, 2017), making it appropriate for the present study. The analyses were conducted using the 'phylosig' function of the 'phytools' package (Revell, 2012), employing 10, 000 simulations for significance testing.

## 3.3 Results

*Strain identification and phylogenetic analyses*

All strains isolated as part of this study were identified to the genus level, based on BLASTn search results, as belonging to either *Burkholderia* or *Mesorhizobium*. All *Burkholderia* strains had at least 97 % similarity to known South African strains, while the *Mesorhizobium* strains were similar to rhizobial type strains from various parts of the world. Strains isolated from the legume genera *Aspalathus* L. (except for *A. callosa, A. capensis and A. carnosa*), *Argyrolobium* Eckl. & Zeyh., *Otholobium* C.H. Stirt. and *Psoralea* L. were identified as *Mesorhizobium*. *Burkholderia* strains were from *Amphithalea* Eckl. & Zeyh., *Aspalathus* L., *Bolusafra* Kuntze., *Dipogon* Liebm., *Indigofera* L., *Lebeckia* Thunb., *Podalyria* Willd., *Rafnia* Thunb. and *Virgilia* Poir. Phylogenetic analyses were conducted separately for each of the two genera to allow for independent analyses of phylogenetic signals within each genus. The aligned 16S rRNA matrix of *Burkholderia* consisted of 67 strains and 1540 characters (total aligned length), while that of *Mesorhizobium* had 73 strains and 1520 characters. The *recA* matrix for *Burkholderia* had 67 strains and 951 characters, while that of *Mesorhizobium* had 67 strains and 886 characters. The Bayesian and ML analyses of the individual chromosomal gene regions produced trees of similar topologies and in all cases the *recA* tree was better resolved than that of the 16S rRNA. The trees from the concatenated matrices were better resolved and more strongly supported (Figure 3.1 and Figure 3.2) than the individual gene trees. The aligned *nodA* matrix for *Burkholderia* had 74 rhizobial strains and 734 characters, while that of *Mesorhizobium* had 41 strains and 674 characters. The Bayesian and ML trees had similar topologies and were well supported (Figure 3.3 and Figure 3.4). However, for both *Burkholderia* and *Mesorhizobium*, the *nodA* topologies were incongruent to those of the chromosomal gene trees, suggesting disparate evolutionary histories between the chromosomal and nodulation genes.

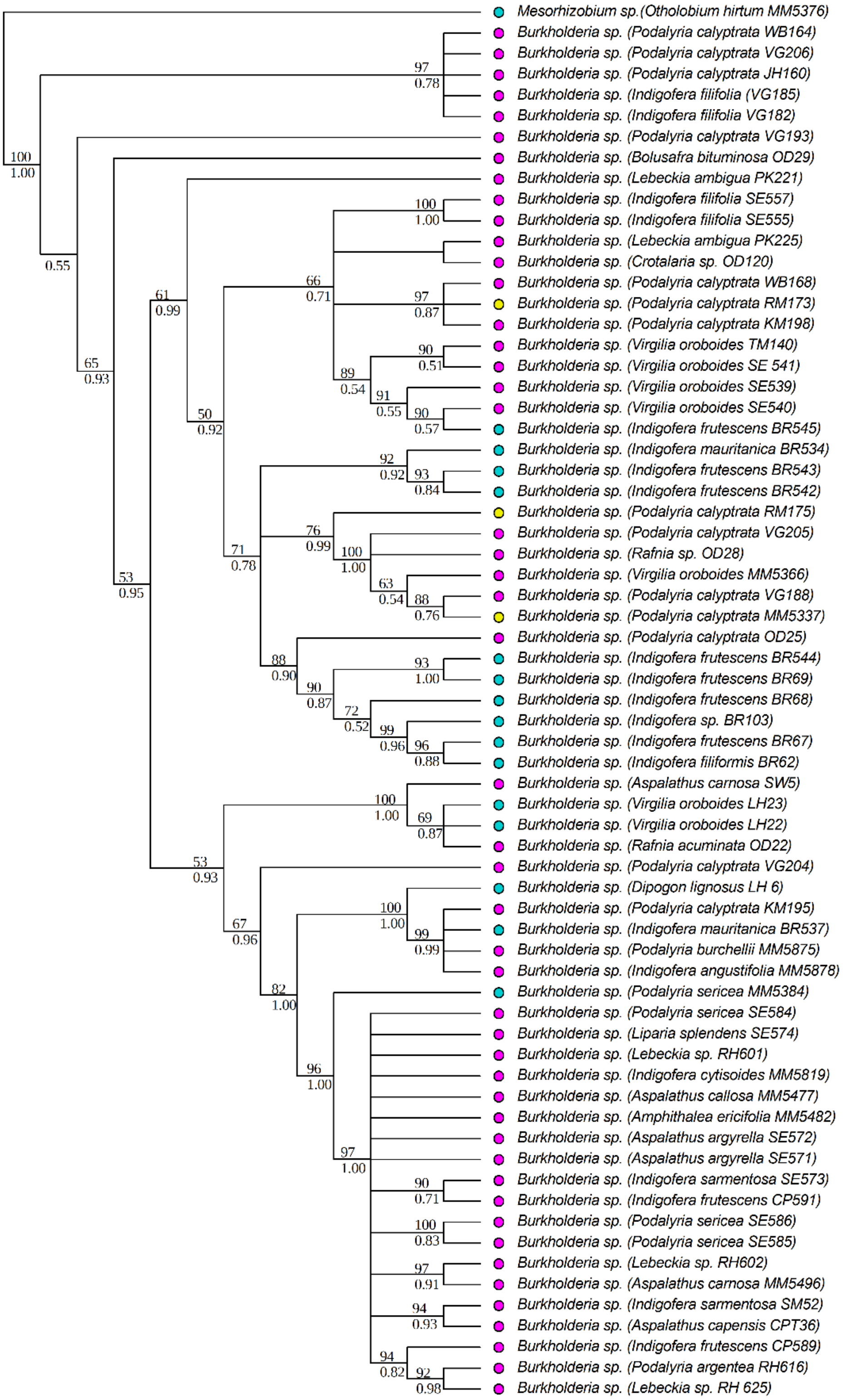


**Figure 3.1.** Phylogenetic relationships of *Burkholderia* strains based on 16S ribosomal RNA (rRNA) and recombinase A (*recA*) data. Names of the legume hosts and the rhizobial strain numbers are in parentheses. Strain numbers with the prefixes OD- (i.e collector name: Oscar Dlodlo) and MM- (i.e. collector name: Muthama Muasya) are from the study by Lemaire et al. (2015a). All other strains were newly generated in this study. Maximum likelihood (ML) bootstrap (%) and Bayesian Inference (BI) posterior probabilities are shown above and below nodes, respectively. Colored circles indicate the soil types of the sites where the legumes and their symbionts were collected, blue: granite, pink: sandstone, yellow: shale.

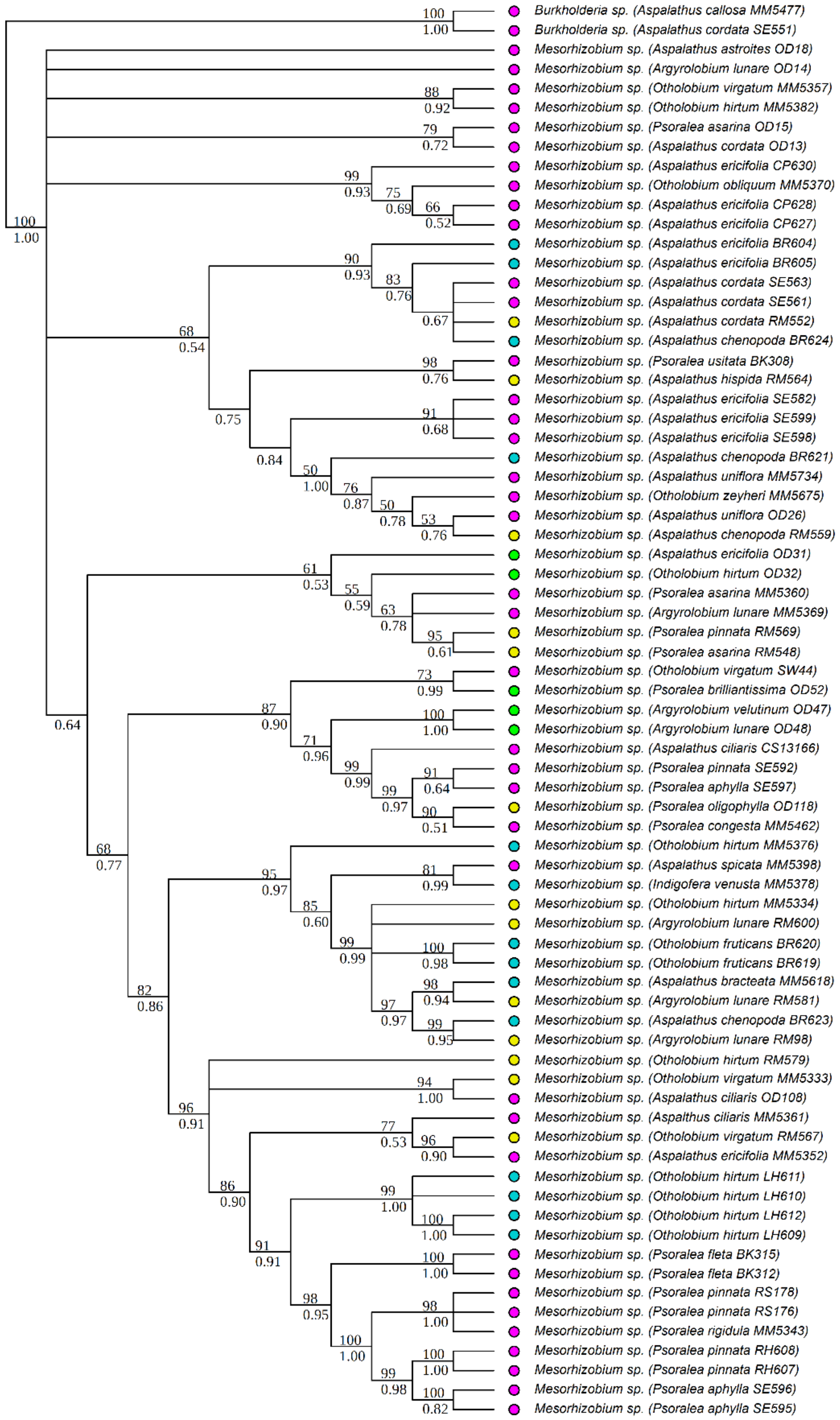


**Figure 3.2.** Phylogenetic relationships of *Mesorhizobium* strains based on 16S rRNA and *recA* data. Names of the legume hosts and rhizobial strain numbers are in parentheses. Strain numbers with the prefixes OD- and MM- are from the study by Lemaire et al. (2015a). The rest were newly generated in this study. ML bootstrap (%) and BI posterior probabilities are shown above and below the nodes, respectively. Colored circles indicate the soil types of the sites where the legumes and their symbionts were collected, blue: granite; green: limestone; pink: sandstone, yellow: shale.

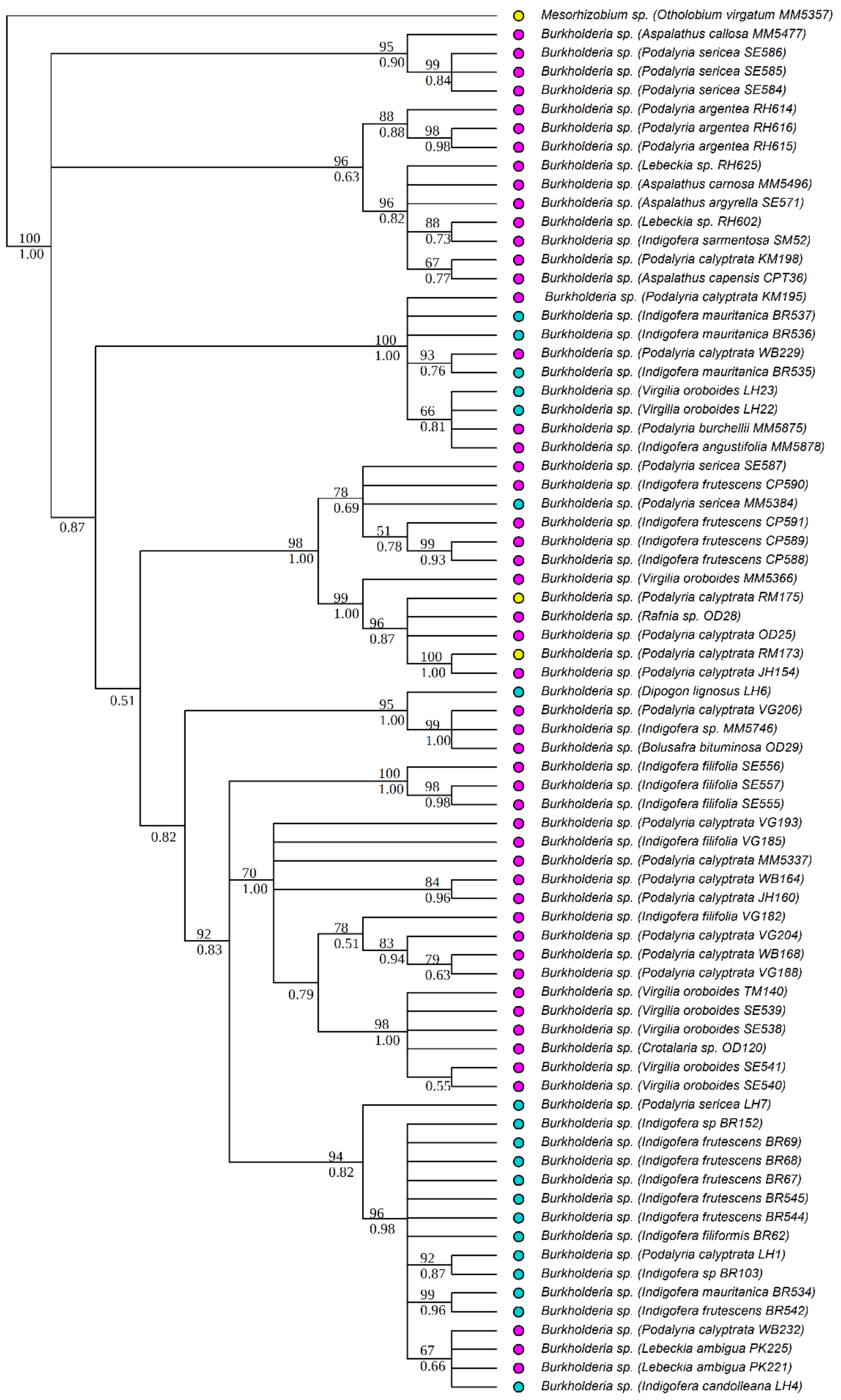


**Figure 3.3.** Phylogenetic relationships of *Burkholderia* strains based on N-acyltransferase (*nodA*) data. Names of the legume hosts and rhizobial strain numbers are in parentheses. Strain numbers with the prefixes OD- and MM- are from the study by Lemaire et al. (2015a). The rest were newly generated in this study. ML bootstrap (%) and BI posterior probabilities are shown above and below the nodes, respectively. Colored circles indicate the soil types of the sites where the legumes and their symbionts were collected, blue: granite; pink: sandstone, yellow: shale.

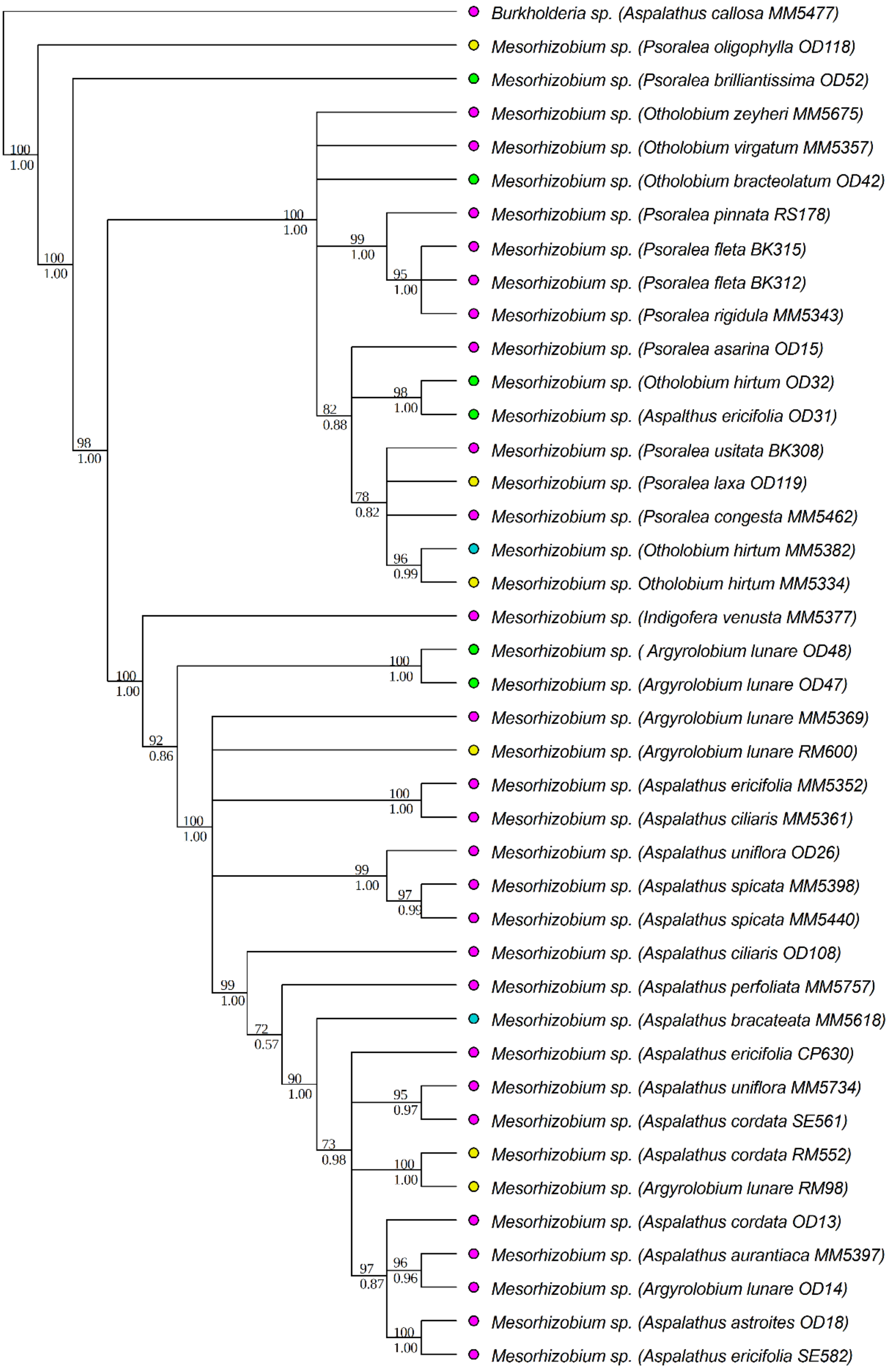


**Figure 3.4.** Phylogenetic relationships of *Mesorhizobium* strains based on *nodA* data. Names of the legume hosts and rhizobial strain numbers are in parentheses. Strain numbers with the prefixes OD- and MM- are from the study by Lemaire et al. (2015a). The rest were newly generated in this study. ML bootstrap (%) and BI posterior probabilities are shown above and below nodes, respectively. Colored circles indicate the soil types of the sites where the legumes and their symbionts were collected, blue: granite; green: limestone; pink: sandstone, yellow: shale.

*Analyses of phylogenetic signals*

For *Burkholderia*, 72 % of the strains occurred on sandstone, 24 % on granite, 4 % on shale and none were on limestone-derived soils, whereas for *Mesorhizobium*, 54 % of the strains were on sandstone, 20 % on granite, 17 % on shale and 9 % on limestone-derived soils (Appendix B3). From the chromosomal gene tree, a comparison of the phylogenetic D statistic with the random shuffling of parameter values along the tips of the phylogeny showed a significant phylogenetic signal for sandstone (D = 0.133; p = 0.00) and a strong phylogenetic signal for granite (D = -0.22; p = 0.00) for *Burkholderia*. *Mesorhizobium* had significant phylogenetic signals for sandstone (D = 0.433; p = 0.0009) and granite (D = 0.252; p = 0.0006) and a strong phylogenetic signal for limestone-derived (D = -0.359; p = 0.0006) soils (Table 3.1). On the other hand, when the D statistic was compared to the Brownian threshold model all, but the *Mesorhizobium* on shale-derived soils, were as clumped on the phylogeny as expected under a Brownian motion model (Table 3.1).

For the *nodA* phylogeny, *Burkholderia* showed similar patterns of phylogenetic signal as observed for the chromosomal genes (Table 3.1). Despite having similar patterns to those of the chromosomal genes for sandstone and limestone-derived soils, *nodA* results for *Mesorhizobium* showed evidence of random dispersion of parameter values for granite-derived soils (D = 1.752; p = 0.833), with a significant departure (p = 0.032) from a Brownian threshold model (Table 3.1), indicating a lack of phylogenetic structure for this parameter.

Analyses based on the chromosomal gene trees showed significant phylogenetic signals for pH for both *Burkholderia* (Pagel's $\lambda = 0.642$; p = 0.019) and *Mesorhizobium* (Pagel's $\lambda = 0.508$; p = 0.027). On the other hand, altitude showed no significant phylogenetic signal on either *Burkholderia* (Pagel's $\lambda = 0.402$; p = 0.081) or *Mesorhizobium* (Pagel's $\lambda = 0.217$; p = 0.097). Similar patterns were observed for analyses based on the *nodA* trees of both rhizobial genera (Table 3.2).

**Table 3.1.** Results of the tests of phylogenetic signals on soil types using the D statistic on the combined chromosomal gene (16S rRNA and *recA*), and the *nodA* trees of *Burkholderia* and *Mesorhizobium*.

| DNA region | Genus | Soil Type | *D* | p-value Random Shuffle | p-value Brownian Motion |
|---|---|---|---|---|---|
| 16S rRNA and *recA* | *Burkholderia* | Granite | −0.220 | 0.00 | 0.717 |
| | | Sandstone | 0.133 | 0.00 | 0.384 |
| | | Shale | 0.956 | 0.375 | 0.144 |
| | *Mesorhizobium* | Granite | 0.252 | 0.0006 | 0.235 |
| | | Limestone | −0.359 | 0.0006 | 0.744 |
| | | Sandstone | 0.133 | 0.0009 | 0.056 |
| | | Shale | 0.975 | 0.419 | 0.001 |
| *nodA* | *Burkholderia* | Granite | −0.208 | 0.00 | 0.715 |
| | | Sandstone | 0.082 | 0.00 | 0.424 |
| | | Shale | 0.617 | 0.160 | 0.291 |
| | *Mesorhizobium* | Granite | 1.752 | 0.833 | 0.032 |
| | | Limestone | −0.871 | 0.0005 | 0.914 |
| | | Sandstone | 0.162 | 0.002 | 0.423 |
| | | shale | 0.492 | 0.118 | 0.218 |

**Table 3.2.** Results of the tests of phylogenetic signals for altitude and pH using Pagel's λ on the chromosomal (16S rRNA and *recA*) and *nodA* trees of *Burkholderia* and *Mesorhizobium*.

| Genus | Gene Type | Variable | Pagel's λ | *p*-values |
|---|---|---|---|---|
| *Burkholderia* | Chromosomal | Altitude | 0.402 | 0.081 |
| | | pH | 0.643 | 0.019 |
| | *nodA* | Altitude | 0.093 | 0.389 |
| | | pH | 0.840 | 0.0008 |
| *Mesorhizobium* | Chromosomal | Altitude | 0.217 | 0.097 |
| | | pH | 0.508 | 0.027 |
| | *nodA* | Altitude | 0.767 | 0.999 |
| | | pH | 0.912 | 0.016 |

*Diversity of rhizobial symbionts nodulating Indigofera superba*

In total, 87 and 86 strains were isolated and successfully sequenced for *recA* and *nodC*, respectively, from *Indigofera superba* and its sympatric legumes in Vogelgat Nature Reserve. All strains that were isolated from the root nodules of *Indigofera superba* were identified (based on BLASTn searches on GenBank) as *Burkholderia,* and the highest matches (≥ 95 % similarity) were known *Burkholderia* species from South Africa, i.e. *B. dilworthii*, *B. kirstenboschensis*, *B. rhynchosiae*, *B. sprentiae* and *B. tuberum*. The highest matching (% similarity) GenBank sequences for the various strains are provided as part of the supplementary materials (Appendix B1). On the other hand, all strains isolated from *Psoralea pullata*, which occurs in sympatry with *I. superba* were *Mesorhizobium*. The BI and ML analyses of the *recA* and *nodC* matrices produced some fairly resolved and supported topologies (Figure 3.5 and Figure 3.6). *Indigofera superba* symbionts were part of multiple distinct clades, most of which included strains isolated from other legume species of the CCR in both the *recA* and the *nodC* trees (Figure 3.5 and Figure 3.6).

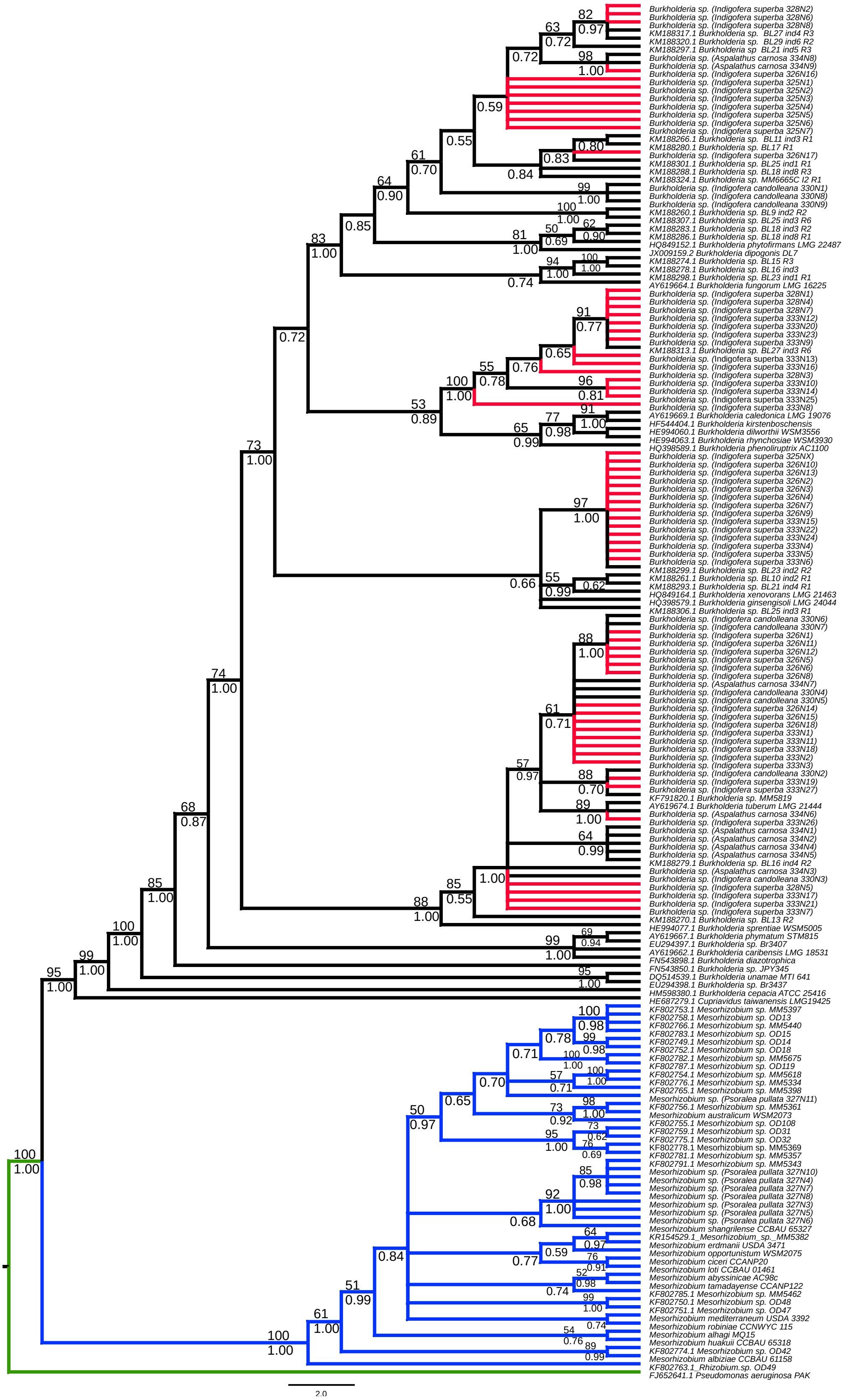


**Figure 3.5.** Phylogenetic relationships of rhizobial strains based on *recA* data, showing the phylogenetic position of strains isolated from *Indigofera superba* (red nodes) in relation to rhizobial strains of other legumes in the CCR, blue: *Mesorhizobium* strains; black: *Burkholderia* strains, green: outgroup. Maximum Likelihood (ML) Bootstrap support values (%) are shown above the nodes and Bayesian posterior probabilities, below the branches.

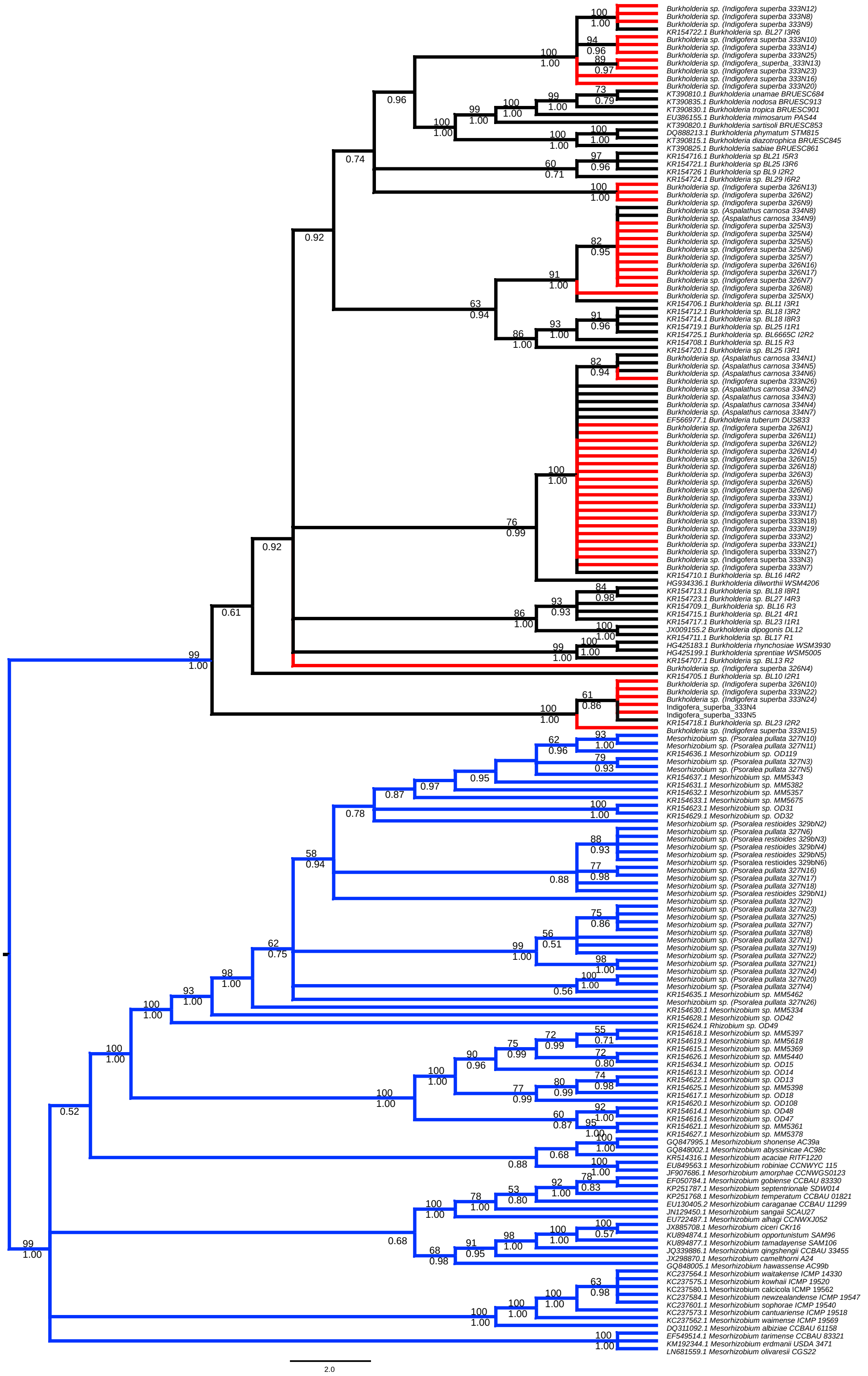


**Figure 3.6.** Phylogenetic relationships of rhizobial strains based on *nodC* data, showing the phylogenetic position of strains isolated from *I. superba* (red nodes) in relation to rhizobial strains of other legumes in the CCR, blue: *Mesorhizobium* strains; and black: *Burkholderia* strains. Maximum Likelihood Bootstrap support values (%) are shown above branches and Bayesian posterior probabilities below branches.

## 3.4 Discussion

The main objective of this study was to determine if the three ecological parameters: altitude, pH and soil type show phylogenetic structuring for the two predominant rhizobial genera, *Burkholderia* and *Mesorhizobium* (Lemaire et al., 2015a), in the Core Cape Subregion of South Africa. For both genera, the results showed significant phylogenetic signals for soil type and pH, but not for altitude. Soil type can be viewed as an indicator of the nutrient status of the habitats based on literature (Ojeda et al., 2010; Thwaites and Cowling, 1988) and on the results of Chapter 2, which showed that sandstone habitats are the most nutrient-impoverished relative to the granite and shale substrates. Limestone soils are generally more fertile than the granite, sandstone and shale substrates (Cowling et al., 1994; Witkowski and Mitchell, 1987). Soil type is also related to pH, with the following general ranges, sandstone pH: 3-4.5, granite pH: 4.5-5.5, shale pH: 5.5-6.5 and limestone pH: >6.5 (Cowling, 1990; Thwaites and Cowling, 1988; Witkowski and Mitchell, 1987). Therefore, it is unsurprising that the results show similar patterns for pH and soil type. Having established that edaphic factors influence the biogeography of legumes in the landscape (Chapter 2), the present chapter sought to determine whether edaphic factors also influence the distribution of the legume symbionts (rhizobia) in the landscape.

Consistent with previous studies (Bontemps et al., 2010; Pires et al., 2018; Stopnisek et al., 2014), *Burkholderia* strains showed a preference for acidic soils, as indicated by the large proportion (72 %) of its strains that were collected from the highly acidic sandstone habitats and its complete absence in the limestone habitats, which have alkaline conditions (Appendix B3). The findings of significant phylogenetic signals on the acidic sandstone and granite habitats indicate that in addition to the genus-wide preference for acidic conditions, *Burkholderia* strains are not randomly distributed within these soil types, but closely related strains tend to occupy similar habitats with respect to soil type and pH. Thus, these ecological parameters have a significant influence on *Burkholderia*'s distribution within the CCR landscape. On the other hand, the results indicated that *Mesorhizobium* tolerates a wider range of soil types and pH conditions because it had nearly equal proportions of its strains isolated from the highly acidic and infertile sandstones and the higher pH and nutrient-richer substrates (Appendix B3). The finding of significant phylogenetic signals for granite, limestone and sandstone, and for pH indicates that despite the wider tolerance range of *Mesorhizobium* as a genus, the distribution of various strains is phylogenetically structured. Thus, for each of the different soil types and pH conditions of the CCR there are particular strains of *Mesorhizobium* that are adapted to them. This is consistent with observations from other biomes showing that *Mesorhizobium* species exhibit high diversity in their tolerance to various pH conditions (Brígido and Oliveira, 2013; Brígido et al., 2017; Laranjo and Oliveira, 2011). This could explain the predominance of *Mesorhizobium* (in addition to *Burkholderia*) in the CCR (Lemaire et al., 2015a). Overall, the results suggest that *Mesorhizobium* has a wider soil type and pH tolerance range than *Burkholderia*, but strains of both genera exhibit phylogenetic clustering within their distribution ranges.

The finding of a significant phylogenetic signal for granite-derived soils (for *Mesorhizobium*) based on the chromosomal gene tree, versus a lack of phylogenetic signal for the same parameter on the *nodA* tree suggests that the chromosomal and nodulation genes have different evolutionary histories, possibly due to horizontal inheritance of the nodulation genes. This would be unsurprising as Chapter 4 of this thesis and other studies (Beukes et al., 2013; Lemaire et al., 2015b) show that horizontal gene transfer (HGT) is a common phenomenon among CCR rhizobia, resulting in conflicting phylogenetic signals between chromosomal and nodulation gene trees.

The observed variation in the biogeographical structuring of the different rhizobia with respect to soil type and pH has implications for the biogeography of legumes in the CCR. This is particularly the case considering the findings of Chapter 2 of this thesis, that discrete edaphic habitats are characterised by discrete legume assemblages in the Cape Peninsula, which points to an important role of edaphic factors in driving legume biogeography. Considering that soil nutrients are a limiting factor to plants in the CCR (Witkowski and Mitchell, 1987), the ecological advantage that nitrogen fixation confers on legumes must be key to their success in such an environment. Therefore, if edaphic factors also limit the distribution of rhizobia, legumes that exhibit high rhizobial specificity, e.g. species of the tribe Podalyrieae, which are only nodulated by *Burkholderia* (Lemaire et al., 2015a), might fail to establish in habitats that are unsuitable for their rhizobial symbionts (Parker, 2001; Simonsen et al., 2017). In such a case, the biogeography of such legumes would also be driven by the distribution of their specific symbionts. This could explain the sparse representation of the tribe Podalyrieae in the limestone habitats (three out of 104 species) whereas most of its species occur in sandstone habitats (i.e. where *Burkholderia* are the predominant symbionts) in the CCR (Schutte, 2012). On the contrary, promiscuous legume lineages, e.g. *Aspalathus* and *Indigofera*, or those that are nodulated by *Mesorhizobium*, e.g. *Psoralea* and *Otholobium* (Lemaire et al., 2015a) are widespread in diverse soil types of the CCR (Schutte, 2012). These patterns suggest that rhizobia play a significant role in the distribution of legumes in the CCR. Furthermore, in a glasshouse experiment where legumes from the Fynbos and Grassland biomes were grown in soils from both biomes, Fynbos legumes were only able to nodulate in Fynbos soil (Lemaire et al., unpublished). This indicates a potential role of rhizobia specificity in driving the distribution of legumes in the various biomes of Southern Africa. This, therefore, opens up avenues for further research. For example, can rhizobia specificity explain why some Cape clades that occur outside the Fynbos are restricted to sandstone habitats? Also, why are Genistoid legumes that occur in the CCR nodulated by *Mesorhizobium*

(Lemaire et al., 2015a), whereas those of the Great Escarpment are nodulated by *Bradyrhizobium* (Beukes et al., 2016)?

Although altitude is highly heterogeneous and it has been found to play a significant role in driving plant diversification in the CCR (Verboom et al., 2015), the results of the present study showed no evidence of phylogenetic structuring of the two predominant rhizobial genera, for this ecological parameter. Similar results were obtained in the study conducted by Lemaire et al. (2015a) for both the chromosomal (16S rRNA) and *nodA* genes of *Mesorhizobium*. Likewise, the *nodA* genetic diversity in *Burkholderia* was not significantly correlated with altitude (Lemaire et al., 2015a). The only disparity is that they found a positive correlation between altitude and genetic diversity of 16S rRNA for *Burkholderia*. The disparity is likely because the current study considered overall phylogenetic signals based on a combination of both 16S rRNA and *recA*, whereas the previous study only considered genetic distances of a single chromosomal marker: 16S rRNA. These findings are contrary to the observed biogeographic structuring of *Burkholderia* communities along altitudinal gradients in Brazil (Bontemps et al., 2010). Such conflicting results have also been observed in other rhizobial genera e.g. *Ensifer* in Northern China (Zhao et al., 2014) versus central China (Chen et al., 2004). However, the sampling of the present study spanned an altitude range of 10 – 1000 m above sea level, whereas the highest altitude in the CCR is 2249 m (Verboom et al., 2015). Notably, in the case of Brazil, the *Mimosa* L. species which are hosts to *Burkholderia* (Bontemps et al., 2010) largely occur in high-altitude habitats (Simon and Proença, 2000). This suggests that the observed variation in *Burkholderia* diversity with changing altitude (Bontemps et al., 2010) is a feature of high-altitude areas. Thus, in the case of the CCR, the current data may not be sufficient to allow for conclusive inferences on the role of altitude in rhizobial biogeography for the region. Hence, future studies, sampling high altitude areas could allow for further investigation of the effect of altitude on rhizobial diversity and turnover in the CCR.

The finding that all strains isolated from the root nodules of the rare *Indigofera superba* belong to the genus *Burkholderia* (despite the availability of *Mesorhizobium*, which was isolated from its sympatric species, *Psoralea pullata*) suggests a potential symbiotic specificity at the generic level. However, the clustering of the different strains in several distinct clades points towards association with multiple divergent lineages within the genus *Burkholderia*. Notably, the nodulation of *Indigofera* species by *Burkholderia* has only been reported from the CCR (Lemaire et al., 2015a), whereas in other areas they are nodulated by alpha-rhizobia (Gehlot et al., 2012). Moreover, the specificity (at the rhizobial generic level) observed in *I. superba* is a common phenomenon among CCR legume genera that are nodulated by multiple rhizobia genera, e.g. *Aspalathus*. If these strains that cluster with divergent lineages are capable of nodulating *I. superba*, it would be a similar scenario to that of the widespread

*Podalyria calyptrata*, which is nodulated by strains from several distinct lineages within *Burkholderia* (Lemaire et al., 2016b), i.e. no symbiotic specificity at the intrageneric level. The results also suggest that the strains isolated from *I. superba* are not genetically distinct since they were part of various clades that included strains isolated from legumes that occur outside its distribution range. Overall, these results lead to the hypothesis that *I. superba* does not exhibit rhizobia specificity at the intrageneric level. More studies are required to test this hypothesis and this could involve testing if the various strains are able to induce nodulation on *I. superba* and determining if *I. superba* is able to form nodules in soils from outside its distribution range. A lack of nodulation from these soils would indicate that the restricted distribution of *I. superba* is due to rhizobia specificity.

### 3.5 Conclusions

The study of the legume-rhizobia relationship in Southern Africa is still at its infancy and although the CCR has received much attention relative to the rest of the sub-continent, only a small proportion of its legume diversity has been studied. Nevertheless, the patterns that are emerging from these few studies suggest that these below-ground mutualists of the legumes might be significant drivers of the distribution of legumes in the CCR. The findings of the present study suggest that while *Burkholderia* has an affinity for the acidic and nutrient-poor soils of the CCR, *Mesorhizobium* has a wider soil type and pH tolerance range, allowing various strains to thrive in habitats of varying edaphic stress. The presence of such ecologically diverse symbionts, coupled with the edaphic heterogeneity of the CCR landscape provide opportunities for the legumes to diversify, and this might explain the high species richness of the family. With the finding that rhizobia contribute towards the structuring of legume assemblages in the CCR, it is plausible that rhizobia also have a strong influence on the structuring of legume assemblages within and across the different biomes of Southern Africa, and this is one possible future direction for research.

# 4. Horizontal gene transfer among rhizobia of the Core Cape Subregion of southern Africa

## 4.1 Introduction

Southern Africa has a high diversity of Fabaceae (≈1600 species) and distinct legume assemblages characterise its various biomes and vegetation types, culminating in what has been termed "Leguminochoria" of Southern Africa (Trytsman et al., 2016). The Core Cape Subregion (CCR) of South Africa is one such Leguminochorion, comprising over 760 legume species mostly belonging to four tribes of the subfamily Papilionoideae, i.e. Crotalarieae, Indigofereae, Podalyrieae and Psoraleeae (Manning and Goldblatt, 2012). The CCR is known for its extremely oligotrophic soils, in which nitrogen (N) and phosphorus (P) are limiting to plants (Stock and Lewis, 1986; Witkowski and Mitchell, 1987). These conditions are particularly unsuitable for most legumes as they require high leaf N and P contents (McKey, 1994; Werner et al., 2015). Therefore, it is unsurprising that besides retaining fundamental adaptations for the acquirement and efficient use of P (Lambers et al., 2011), most legumes of the CCR engage in a symbiotic relationship with bacteria (rhizobia) which fix atmospheric N in return for photosynthates from the legumes (Sprent et al., 2017). This process occurs within specialised structures in the roots of the legumes, called nodules (Sprent et al., 2010). Although the rhizobia nodulating legumes in southern Africa are generally understudied, the body of literature documenting the legume-rhizobia relationships in the CCR is accumulating steadily (Beukes et al., 2013; de Faria et al., 1989; Elliott et al., 2007; Garau et al., 2009; Gerding et al., 2012; Hassen et al., 2012; Lemaire et al., 2016a, 2016b, 2015a, 2015b).

One of the patterns emerging from the studies of rhizobia in the CCR is that while the legumes are nodulated by most of the known genera of N-fixing bacteria (*Bradyrhizobium*, *Burkholderia*, *Ensifer*, *Mesorhizobium* and *Rhizobium*), the genera *Burkholderia* and *Mesorhizobium* are the predominant symbionts (Beukes et al., 2013; Gerding et al., 2012; Lemaire et al., 2015a). Secondly, some legume-rhizobia associations are highly specialised (e.g. *Mesorhizobium* symbionts associating exclusively with the Psoraleeae, and the Podalyrieae associating chiefly with *Burkholderia*), while others are generalist/promiscuous associations, such as in the Crotalarieae and Indigofereae which associate with symbionts of multiple rhizobial genera (Lemaire et al., 2015a). Finally, several new species of rhizobia have been described from the CCR and these include *Burkholderia aspalathi*, *B. dilworthii*, *B. kirstenboschensis*, *B. rhynchosiae* and *B. sprentiae* (De Meyer et al., 2014, 2013a, 2013b; Mavengere et al., 2014; Steenkamp et al., 2015) and more are yet to be described as several strains, distinct from known species from elsewhere in the world, have been isolated (Beukes et al., 2013;

Lemaire et al., 2015a). Therefore, considering the existence of both specificity and promiscuity among the legumes and the fact that only a small proportion of the legume species has been studied thus far (≈12 %), there may still be more rhizobial diversity within the CCR waiting to be discovered.

The legume-nodulating species of the genus *Burkholderia* have two major centres of diversity, the CCR of South Africa and the Brazilian Caatinga and Cerrado biomes in South America (Bournaud et al., 2013; Lemaire et al., 2015a). Notably, the symbionts from these two regions differ markedly in their legume host species and molecular composition. In South America, *Burkholderia* species associate chiefly with Mimosoid legumes (Bournaud et al., 2013; Elliott et al., 2009), while in South Africa, they associate mostly with Papilionoid legumes (Beukes et al., 2013; Garau et al., 2009; Lemaire et al., 2015a). Only one *Burkholderia* species, *B. tuberum*, is common to both regions, but each region has a genetically distinct symbiovar of the species (Gyaneshwar et al., 2011; Mishra et al., 2012). Thus, the symbionts of the two regions belong to different species and current evidence indicates that they do not nodulate each other's legume hosts (Elliott et al., 2007; Lemaire et al., 2016a; Liu et al., 2014). They also differ in the composition and molecular arrangement of their symbiosis (nodulation and nitrogen fixation) genes (De Meyer et al., 2016). Consequently, it has been suggested that while the *Burkholderia* of both regions probably share a common ancestor that occurred in both continents during the Cambrian period, when Africa and South America were adjoined, the symbionts must have followed different evolutionary trajectories after the continents separated (Zheng et al., 2017). The validity of the hypothesis on the vicariance of *Burkholderia* needs investigation as an origin in any one of the two areas followed by dispersal into the other is also plausible.

On the other hand, *Mesorhizobium* nodulates a wide diversity of legumes in various parts of the globe. Some examples include *Ademisia* spp. in Chile (Gerding et al., 2017), *Astragalus* and *Oxytropis* in northern Europe (Ampomah et al., 2017), various Papilionoid legumes in the CCR (Gerding et al., 2012; Lemaire et al., 2015a) and *Sophora* spp. in New Zealand (Tan et al., 2015). In a study of rhizobial symbionts of the genus *Lessertia* in the CCR, Gerding et al. (2012) found that *Mesorhizobium* was the primary symbiont of the species and although the strains were clustered with known *Mesorhizobium* species from other parts of the world in the phylogeny of chromosomal genes (*dnaK* and 16S rRNA), the CCR strains formed a distinct clade in their nodulation gene (*nodA*) phylogeny. Similar results were reported by Lemaire et al. (2015a) for *Mesorhizobium* strains isolated from 28 (out of 65) legume species that occur in the CCR. The findings of these two studies indicate that the *nodA* gene of the studied *Mesorhizobium* strains of the CCR has a distinct evolutionary history from that of the chromosomal genes. Therefore, the *nodA* genes of both primary symbionts of the

CCR legumes (*Burkholderia* and *Mesorhizobium*) appear to have evolved independently of their counterparts in other parts of the globe.

DNA regions commonly sequenced for phylogenetic studies of rhizobia include housekeeping genes (located on chromosomes) and nodulation (*nod*) and nitrogen fixation (*nif*) genes, which reside in mobile extra-chromosomal plasmids or symbiotic islands (Peix et al., 2015; Zahran, 2017). Nodulation genes are involved in the formation of nodules and determination of host specificity (Perret et al., 2000; Wang et al., 2012), while nitrogen fixation genes encode enzymes responsible for the conversion of atmospheric nitrogen into ammonia and for various proteins that regulate the process (Kneip et al., 2007). The mobile elements that carry *nod* and *nif* genes can be transferred, laterally, from one organism to another (including disparate evolutionary lineages) in a process called horizontal gene transfer (HGT), unlike the usual parent-to-offspring transfer (vertical inheritance) of genetic material across succeeding generations (Juhas et al., 2009; MacLean et al., 2007). The process of HGT can occur through various mechanisms including transformation, transduction and conjugation (Davison, 1999; Paul, 1999). The resultant organism (recipient of the transferred genetic material) is a chimeric species, i.e. different parts of its genome have different evolutionary histories (Eisen, 2000) and phylogenetic analyses based on these disparate portions of its genome often yield incongruent results (Huang and Gogarten, 2006; Paul, 1999). Horizontal gene transfer is important in the study of microbial evolution because not only does it lead to the formation of new lineages (Gogarten et al., 2002), but also to the acquisition of adaptations for extreme environments (Andrews and Andrews, 2017; Ochman et al., 2000; Remigi et al., 2016; Wiedenbeck and Cohan, 2011).

Incongruences between the phylogenies of symbiosis (*nod* and *nif*) genes and those of chromosomal genes have been reported in several studies on rhizobia and this has been inferred as an indication of horizontal inheritance of the symbiosis genes (Aoki et al., 2013; Chen et al., 2003; Huang and Gogarten, 2006; Liu et al., 2012; Moulin et al., 2004; Silva et al., 2018). Notably, HGT does not preclude vertical inheritance of symbiosis genes (Bournaud et al., 2013; Moulin et al., 2004; Zheng et al., 2017) and evidence suggests that both processes have played a significant role in the evolution and spread of symbioses genes in bacteria (De Meyer et al., 2016; Zheng et al., 2017). Although HGT is widely studied, most of the studies reporting on its prevalence in rhizobia tend to focus on closely related species (mostly within the same genus), leaving the prospect of HGT between different rhizobial genera e.g. HGT from alpha to beta-rhizobia and *vice versa*, poorly studied. This also applies to studies reporting on HGT in *Burkholderia* and *Mesorhizobium* in the CCR (Beukes et al., 2013; Gerding et al., 2012; Lemaire et al., 2015b). Furthermore, evidence of HGT among and within the other existing rhizobia genera (i.e. *Bradyrhizobium*, *Ensifer* and *Rhizobium*) in the CCR, has not been tested. Determining the occurrence of HGT across various genera of rhizobia is important as recent studies have shown that an understanding of the extent, direction and frequency of HGT across diverse rhizobial lineages is key to unravelling the origin of nodulation and nitrogen fixation genes (Aoki et al., 2013; De Meyer et al., 2016; Zheng et al., 2017). Considering the unique evolutionary history of the CCR rhizobia, it has been suggested that frequent events of HGT within and between rhizobial genera may be connected with the diversification of the legume flora of the region (Beukes et al., 2013).

In an effort to determine the origins of nitrogen fixation in legumes, researchers have sought to establish the origin of symbiosis genes and the ancestral symbionts of legumes. The search for answers to these questions is ongoing. One hypothesis is that nodulation genes originated in alpha rhizobia, particularly *Bradyrhizobium*, followed by a transfer into other genera including beta rhizobia (Bontemps et al., 2010; Chen et al., 2003; Moulin et al., 2001; Rogel et al., 2011). Some of the observations on which this hypothesis is premised include that *Bradyrhizobium* is the most diverse rhizobial genus that associates with a wide diversity of legumes in various habitats (Parker, 2015), it is the primary symbiont of basally branching Caesalpinioid legumes (dos Santos et al., 2017; Yao et al., 2017) and it has high variation in its *nodA* sequences compared to other rhizobial genera (Martínez-Romero et al., 2010). However, a study of some complete rhizobial genome sequences, conducted by Aoki et al. (2013), showed that only the beta rhizobial genomes contained *nodIJ* genes and their nonsymbiotic paralogues (DRA-ATPase/permease genes). On this basis, it was concluded that the *nodIJ* genes originated in the beta proteobacteria, followed by a transfer to alpha proteobacteria (Aoki et al., 2013). Therefore, *Burkholderia* are also likely to be the ancestral symbionts of legumes.

A recent genome study by Zheng et al. (2017) supports the hypothesis that *Burkholderia* is the ancestral symbiont of legumes, but postulates that their nodulation genes originated from a common ancestor with alpha rhizobia. The genome study by De Meyer et al. (2016), which reported distinct composition and molecular arrangement of symbiosis genes between Mimosoid-nodulating *Burkholderia* (i.e. from South America) and Papilionoid-nodulating *Burkholderia* (i.e. from Southern Africa-CCR), suggests that these symbionts have different sources of symbioses genes. De Meyer et al. (2016) also showed that the CCR *Burkholderia* has a complex history of acquisition of nodulation genes, with *nodABC* being more similar to alpha rhizobia, while *nodIJ* genes are more similar to beta rhizobia. These results suggest that while the *nodIJ* genes originated from beta rhizobia, the *nodABC* genes of CCR rhizobia may have originated from alpha rhizobia, followed by transfer into *Burkholderia.* Testing for the occurrence and direction of HGT among the CCR rhizobial symbionts and reconstructing the ancestral symbionts on a phylogeny of the CCR legumes might shed more light on this subject, and these are the core objectives of the present study.

In this study, it is hypothesized that horizontal gene transfer has not been limited to closely related rhizobial species, but it has also occurred between distantly related species (i.e. different genera of alpha and beta rhizobia). Thus, HGT might have played a significant role in the evolution of rhizobial symbionts in the CCR. The hypothesis is tested by determining whether a phylogeny of the rhizobial symbionts that is based on two chromosomal genes (16S rRNA and *recA*) is congruent to that of their nodulation gene, *nodA*. Considering that the CCR is a centre of diversity of Papilionoid-nodulating *Burkholderia* (Gyaneshwar et al., 2011; Lemaire et al., 2016a), and given observations that *Burkholderia* are adapted to the prevalent acidic, nutrient-poor soils of the region (Elliott et al., 2009; Stopnisek et al., 2014), it is hypothesized that *Burkholderia* are the ancestral symbionts of CCR legumes. This hypothesis is tested by reconstructing the ancestral rhizobial genus of CCR legumes on a phylogeny of the legumes for which N-fixing symbionts are currently known.

## 4.2 Materials and Methods

### *Selection of rhizobia DNA sequences*

The study was based on DNA sequences from previous studies (Beukes et al., 2013; Gerding et al., 2012; Howieson et al., 2013; Lemaire et al., 2015a) where rhizobia had been isolated from indigenous CCR legume root nodules and some newly generated sequences. Two chromosomal (16S rRNA and *recA*) and one nodulation gene (*nodA*) were selected because they are the most sequenced markers used in most studies of CCR rhizobia, which allows for combining data from multiple studies. Newly generated sequences represent strains isolated from indigenous legumes occurring in the CCR (mostly the Fynbos biome). The strains represented four genera of α-proteobacteria, i.e. *Bradyrhizobium*, *Ensifer*, *Mesorhizobium* and *Rhizobium* and one genus of β-proteobacteria, *Burkholderia*. Some reference strains for these genera were downloaded from Genbank and incorporated into the data matrices. A list of all accessions used in the study is provided as part of the supplementary materials (Appendix C1). Strains for the α-rhizobia were isolated from root nodules of the following legume genera: *Argyrolobium* Eckl. & Zeyh., *Aspalathus* L., *Indigofera* L., *Lessertia* DC., *Psoralea* L., *Otholobium* C.H.Stirt., *Tephrosia* Pers., and *Virgilia* Poir. The β-rhizobial strains were symbionts of *Amphithalea* Eckl. & Zeyh., *Aspalathus* L., *Bolusafra* Kuntze, *Cyclopia* Vent., *Dipogon* Leibm., *Hypocalyptus* Thunb., *Indigofera* L., *Lebeckia* Thunb., *Podalyria* Willd., *Rafnia* Thunb. and *Virgilia* Poir.

### *Nodule collection and isolation of rhizobia*

For the sequences generated in this study, field nodules were collected from diverse indigenous legume species occurring within the Cape Peninsula. Details of the sampling sites and how they were chosen are as described in Chapter 2 of this thesis. Nodule sampling followed standard protocols

(Howieson et al., 2016). Isolation of rhizobia followed the methods described in Lemaire et al. (2015a) with one modification on the surface sterilisation of the nodules, where a 4% solution of sodium hypochlorite (NaOCl) was used instead of 0.1% acidified mercuric chloride.

### *DNA extraction, PCR amplification and sequencing*

DNA was extracted using a modified version of the CTAB protocol (Doyle and Doyle, 1987) as previously described (Dludlu, 2010). Polymerase chain reactions (PCR) were conducted to amplify 16S rRNA, *recA* and *nodA* using an Applied Biosystems GeneAmp 2700 thermal cycler (Applied Biosystems, Foster City, CA, USA). Primer pairs used were 16S-f27 and 16S-r1485 (Lane, 1991) for 16S rRNA; recA-63F and recA-504R (Gaunt et al., 2001) for *recA*; and nodA-1F and nodA-2R (Haukka et al., 1998) for *nodA*. Each PCR reaction had a total volume of 25 µl: comprising 19.92 µl of water, 2 µl of 10 × DNA polymerase buffer with $Mg^{2+}$, 0.4 µl of 10 mM dNTP, 0.8 µl each of forward and reverse primers (10 µM), 0.08 µl *Taq* Polymerase (Kapa Biosystems) and 1 µl template DNA. With the exception of 16S rRNA, which was amplified using the reaction conditions of Weisburg et al. (1991), *recA* and *nodA* were amplified according to the conditions described by the authors of the respective primers. PCR products were loaded onto ethidium bromide-stained agarose gels (1 %) and subjected to electrophoresis using 0.5 × TBE. The gels were observed under UV light to identify successfully amplified samples. Amplified products were enzymatically purified using the Exo/SAP protocol (Werle et al., 1994) and sent to Macrogen (Macrogen, Netherlands) for sequencing with the same primers used for PCR amplifications.

### *Data analysis*

The forward and reverse DNA sequence contigs were assembled using the Staden package version 2.0.0 (Staden et al., 2000) and aligned using the online version of MAFFT (Katoh, 2002). The alignments were viewed in Bioedit version 7.1.9 (Hall, 1999) and equivocally aligned fragments were adjusted manually. Phylogenetic analyses of the aligned matrices were performed on the CIPRES web portal (https://www.phylo.org), through a maximum likelihood (ML) approach, using RAxML version 8.2.10 (Stamatakis, 2014) and Bayesian inference (BI), as implemented in MrBayes version 3.2.6 (Ronquist and Huelsenbeck, 2003). The ML analyses employed the GTR + GAMMA substitution model and statistical support on nodes was evaluated using the nonparametric rapid bootstrapping technique (Stamatakis et al., 2008), with 1000 replicates. For the BI analysis, the best model of nucleotide substitution was determined using jModelTest2 version 2.1.6 (Posada, 2008), employing the Bayesian Information Criterion. The BI analyses were run for as many generations as necessary to achieve chain convergence (5-10 million generations). A conservative burnin of 25% was applied to all BI analyses and convergence of the chains was assessed using Tracer version 1.6 (Rambaut et al., 2014).

Determination of the combinability of the different DNA data partitions followed the method used by Pirie et al. (2009, 2008). Individual data partitions were analysed separately using the ML procedures described above and the output topologies were visually examined for presence of conflicting nodes that had ≥ 70 % bootstrap support. Nodes that had < 70 % bootstrap support were considered unsupported and, therefore, when no supported conflict was observed, the partitions were considered combinable (Pirie et al., 2009, 2008). This approach was chosen over the incongruence length difference (ILD) test (Mickevich and Farris, 1981) because the latter only tests for overall incongruence between data partitions without detecting local conflict that is due to specific taxa or clades (Pirie et al., 2009). There was no conflict observed between 16S rRNA and *recA* and, therefore, these partitions were combined in subsequent analyses. However, the *nodA* partition had significantly supported conflict with both chromosomal markers and, therefore, it was analysed separately.

The study used a threshold of 97 % sequence similarity in the 16S rRNA to delimit operational taxonomic units (OTUs). These were used to analyse for the extent of incongruence between the chromosomal and nodulation gene trees as an indicator of horizontal gene transfer within and between the various rhizobial genera. Phylogenetic trees of the strains were reconstructed by ML and BI techniques as described above. The extent of parallel evolution between the chromosomal and nodulation gene trees was assessed through a reconciliation analysis performed on Jane version 4 (Conow et al., 2010). This technique seeks to reconcile pairs of phylogenies, a dependent and an independent phylogeny, subject to biologically plausible events (i.e. co-speciation, duplication, duplication with host switching, loss and failure to diverge) and the costs of such events (Conow et al., 2010). The extent of congruence between the phylogenies was assessed by maximising the number of co-speciation (i.e. vertical gene transfer) events and minimising the possible number of non-codiversification events (horizontal gene transfer) using the default settings of event costs. Permutation tests were used to determine if the two phylogenies were randomly related or that the number of co-speciation events was not larger than expected due to mere chance. Reconciliation analyses, in which the scores of the optimal and initial reconstructions were compared with those of randomly generated topologies were performed using the best trees obtained from the ML analyses as input. Randomization tests were conducted with 1000 randomly permuted trees and a population size of 100. The cost distributions of random sample solutions and statistical significance was calculated in a cost histogram in Jane.

A phylogeny of the CCR legumes for which rhizobial symbionts are known was reconstructed using two chloroplast markers, *rbcL* and *matK*. These were sequenced as part of the DNA barcoding project for Cape legumes and for some species sequences were downloaded from Genbank (see supplementary materials, Appendix C2). *Polygala subtilis* Kunth (Polygalaceae) was used as an outgroup. Combinability of the two data partitions was assessed as described above for the rhizobial data sets. Phylogeny reconstruction by BI and ML was conducted as described above. A character matrix scoring the genera of rhizobia that nodulate the various legume species was created based on the strain identities as inferred from the chromosomal markers. The genera were coded as follows: A (*Burkholderia*), B (*Mesorhizobium*), C (*Ensifer*), D (*Bradyrhizobium*) and E (*Rhizobium*). The best scoring ML tree was used for the reconstruction of ancestral rhizobial genera, which was performed on RASP version 4.0 beta (Yu et al., 2015) using the Bayesian Binary MCMC method. The analysis was run for 5 million generations, sampling every 1000 generations. A fixed JC (Jukes Canto)) model and Equal among-site rate variation were used, employing a null root distribution. The maximum number of genera was set at one because none of the sampled taxa were nodulated by more than one genus of rhizobia. Additionally, a cophylogeny analysis (involving the legume phylogeny vs. rhizobia phylogeny) was conducted to test for evidence of co-diversification between the legumes and their rhizobial symbionts. This was done for both the chromosomal and the *nodA* trees. The analyses were performed in Jane using similar settings as those used for the test of congruence between the chromosomal and *nodA* genes.

## 4.3 Results

*Phylogenetic relationships*

Newly isolated strains were identified to genus level by conducting BLASTn searches of their 16S rRNA and *recA* sequences on the GenBank database. Results of the searches placed the various strains into the genera *Bradyrhizobium*, *Burkholderia*, *Ensifer*, *Mesorhizobium* and *Rhizobium*. A majority of the *Burkholderia* strains had at least 97 % similarity to known South African rhizobial strains i.e. *B. dilworthii*, *B. kirstenboschensis*, *B. rhynchosiae*, *B. sprentiae* and *B. tuberum*, whereas the strains of the alpha rhizobial genera were similar to strains from various parts of the world, such as *Bradyrhizobium elkanii*, *Mesorhizobium erdmanii*, *Rhizobium leguminosarum* and *Ensifer medicae*.

The aligned 16S rRNA dataset comprised 139 taxa and 1763 characters, while the *recA* alignment comprised 151 taxa and 1107 characters. The best model of nucleotide substitution for both datasets was GTR+I+G. Phylogenetic analysis of the individual chromosomal regions (16S rRNA and *recA*) yielded trees of similar topologies, with no significantly supported conflict ($\geq$ 70 % bootstrap support), and for most branches, the *recA* tree had higher support values than the 16S rRNA tree. The ML and BI analyses produced similar topologies, but clade support on the BI trees was mostly higher than the corresponding bootstrap support values on the ML trees. The concatenated chromosomal regions (16S rRNA and *recA*) produced a more resolved and moderately supported topology that was congruent with the individual gene trees (Figure 4.1; Appendix C3, supplementary materials). Most of the strains clustered within clades that correspond to previously established rhizobia-legume associations. For example, legumes of the tribes Psoraleeae (*Psoralea* and *Otholobium*) and Genisteae (*Argyrolobium*) were associated with *Mesorhizobium* symbionts. Likewise, Podalyrieae species (except *Virgilia divaricata*), were associated with *Burkholderia* symbionts, while *Indigofera* and *Aspalathus* species were associated with several genera of rhizobia (Appendix C3). The reference strains of all the studied genera of rhizobia, including those from elsewhere in the world (i.e. non-CCR) were intermixed with CCR strains in the chromosomal phylogeny (Figure 4.1; Appendix C3). Some strains from the *Mesorhizobium* and *Burkholderia* clades did not cluster with any of the known rhizobia reference strains (Appendix C3), suggesting that these strains represent undescribed species that are new to science.

The aligned matrix for *nodA* consisted of 112 taxa and 768 characters. The *nodA* tree is presented as an unrooted tree since the gene does not occur on non-nodulating bacterial genera, which could have otherwise been used as outgroups (Figure 4.2; Appendix C4, supplementary materials). The topology of the *nodA* tree was largely incongruent with that of the chromosomal genes (Figure 4.2). First, the genus *Burkholderia* was not monophyletic. The non-CCR *Burkholderia* reference strains formed a separate clade from the CCR *Burkholderia* and the former was closely related to *Ensifer*, *Mesorhizobium* and *Rhizobium*, whereas the latter was closely related to *Bradyrhizobium* (Figure 4.2; Appendix C4). Likewise, CCR *Mesorhizobium* strains were not intermixed with the *Mesorhizobium* reference strains i.e. non-CCR *Mesorhizobium* (Figure 4.2; Appendix C4). Strains of the genus *Rhizobium* also followed the same pattern (Figure 4.2; Appendix C4). On the other hand, reference strains for *Bradyrhizobium* and *Ensifer* were intermixed with their CCR counterparts (Figure 4.2; Appendix C4). Two strains of alpha rhizobia (i.e. *Rhizobium* sp. symbiont of *Virgilia divaricata* and *Mesorhizobium* sp. symbiont of *Psoralea oligophylla*) were embedded within the CCR *Burkholderia* clade and one strain of CCR *Burkholderia* (symbiont of *Rafnia triflora*) was nested within the CCR *Mesorhizobium* clade (Figure 4.2; Appendix C4). In addition, one strain of CCR *Mesorhizobium* (symbiont of *Psoralea brilliantissima*) was nested within *Ensifer*, while two CCR *Rhizobium* strains (symbionts of *Aspalathus* sp.) were nested within CCR *Mesorhizobium* (Figure 4.2; Appendix C4). Infrageneric relationships were also not congruent between the chromosomal and *nodA* trees, especially for the CCR *Burkholderia* and *Mesorhizobium*.

**Figure 4.1.** Phylogenetic relationships of rhizobial strains based on 16S rRNA and *recA* data. Clades are coloured to reflect the different rhizobial genera and the phylogenetic positions of non-CCR strains are marked with asterisks (*). A detailed version of this figure with ML bootstrap (%), BI posterior probabilities and tip labels is provided as part of the supplementary materials (Appendix C3).

The conflicting phylogenetic relationships between the *nodA* and the 16S rRNA and *recA* gene trees suggest different evolutionary histories of the chromosomal and extra-chromosomal genes, possibly due to horizontal transfer of nodulation genes within and among the different genera. Additionally, the separation of non-CCR *Burkholderia, Mesorhizobium* and *Rhizobium* strains from their CCR counterparts indicates a unique evolutionary history of their *nodA* genes. This validates using only CCR strains to determine the extent of HGT of *nodA* within and between the various rhizobial genera of the region. Due to the large number of closely related strains in the main dataset, which could be representing the same species, strains with < 97 % sequence similarity for the chromosomal genes were selected and used for an in-depth analysis of the extent of incongruence between the chromosomal and the *nodA* gene trees for the CCR rhizobia.

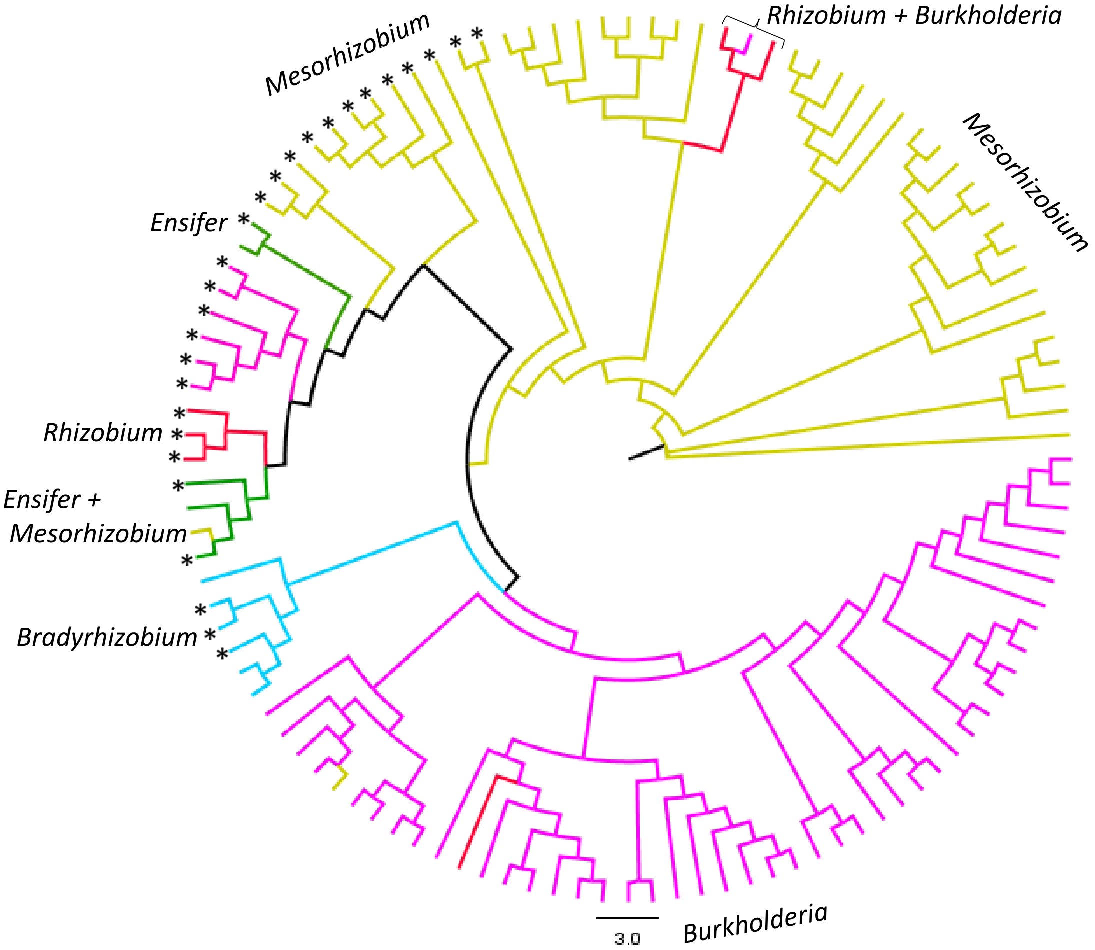


**Figure 4.2.** Phylogenetic relationships of rhizobial strains based on *nodA* data. Clades are coloured to reflect the different rhizobial genera and the phylogenetic positions of non-CCR strains are marked with asterisks (*). A detailed version of this figure with ML bootstrap (%), BI posterior probabilities and tip labels is provided as part of the supplementary materials (Appendix C4).

### *Degree of incongruence between the chromosomal and nodulation genes*

The dataset that was used for analysing the extent of incongruence between the chromosomal and nodulation genes comprised 46 sequences, representing strains of all five genera of rhizobia that were studied in the main dataset. The chromosomal gene tree was largely incongruent with the nodulation gene tree even on visual inspection (Figure 4.3). Both trees were fully resolved and a majority of clades were strongly supported (Figure 4.3), allowing for inferences of the evolutionary history of the constituent gene regions. This was achieved through the cophylogeny approach (reconciliation) in which the nodulation gene tree was mapped onto the chromosomal gene tree. The reconciliation analysis produced 6790 solutions of equal cost (68), of which the number of co-speciation events (i.e. vertical transmission of genes) ranged between 12 and 15, while the number of duplication and host switching events (i.e. lateral gene transfer) was between 30 and 33 (Cost = 68). Nodes associated with co-speciation events are marked with some asterisks in Figure 4.3. A statistical test of the significance of the congruence between the two trees, based on randomization tests, rejected the null hypothesis that the chromosomal and the nodulation gene trees are randomly related ($p < 0.01$). This indicates that although there is a high degree of horizontal transfer of symbiotic genes, the two trees share a significant amount of co-speciation events and thus the frequency of parallel evolution is greater than would be expected by mere chance. A plot of the frequency distribution of the event costs is shown in Appendix C5 (supplementary materials).

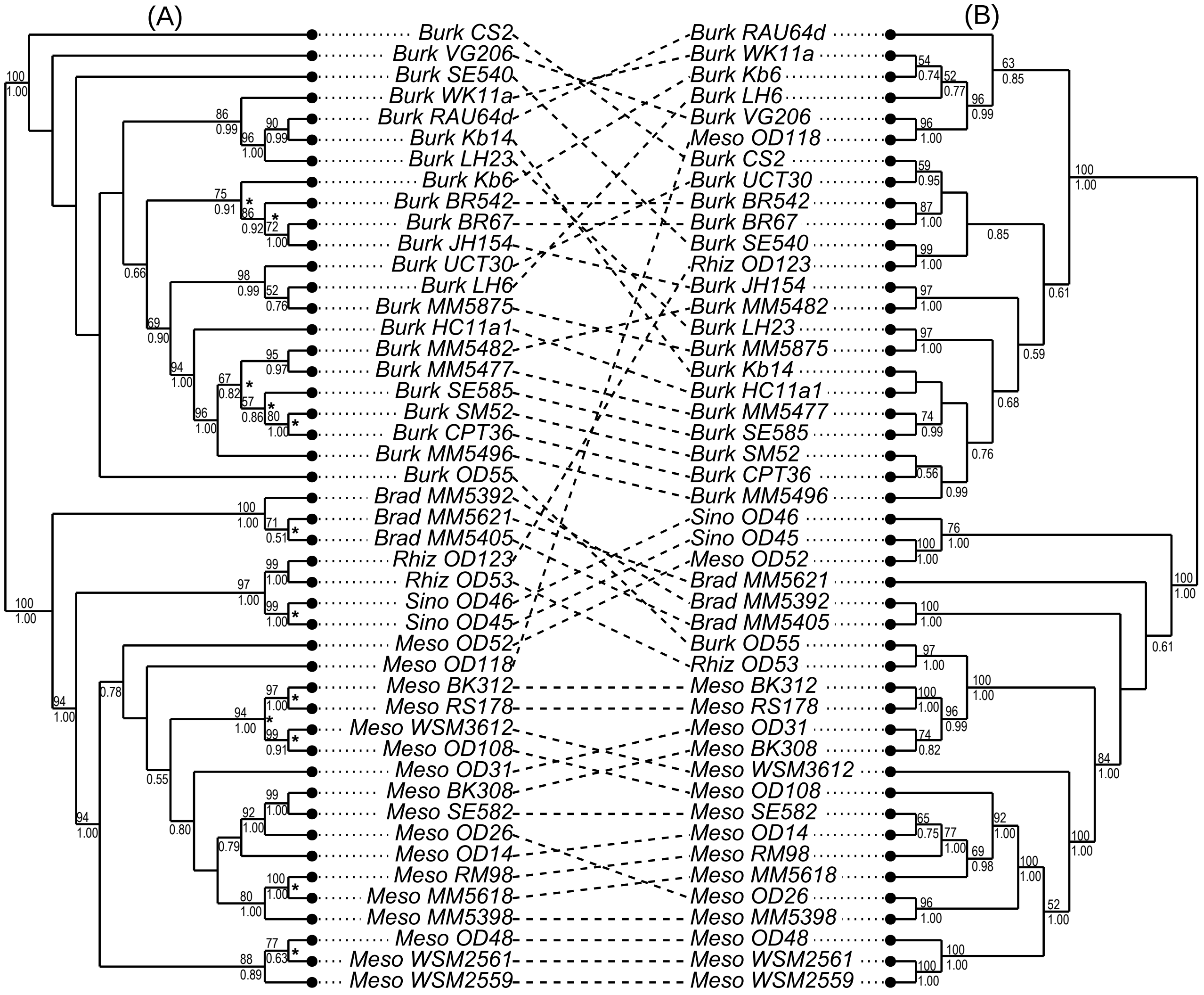


**Figure 4.3.** Phylogenetic incongruences between chromosomal (16S rRNA and *recA*) and nodulation (*nod*A) gene sequence data for CCR rhizobial strains. This is a comparison of the best ML trees of (**A**) chromosomal and (**B**) *nodA* genes. Support values for the ML (bootstrap %) and BI (posterior probabilities) analyses are shown above and below the nodes, respectively. Dashed lines indicate the species associations between the chromosomal and nodulation gene trees. Nodes highlighted with an asterisk (*) represent events of co-speciation as revealed by the reconciliation analysis.

### *Legume phylogeny and reconstruction of ancestral rhizobial state*

The *matK* matrix of the legumes comprised 72 taxa and had an aligned length of 1592 characters, while that of *rbcL* had 68 taxa and an aligned length of 1457 characters. The individual partitions produced trees of similar topologies, with no supported conflicts and were thus concatenated. The concatenated analysis produced a well-resolved tree and a majority of the branches were strongly supported (Figure 4.4). A majority of the genera were monophyletic and their phylogenetic relationships were consistent with established infrafamilial relationships (LPWG, 2017). The only exception was *Rafnia*, which was embedded within *Lebeckia* (Figure 4.4).

Reconstruction of the ancestral rhizobia genera that nodulate the legumes indicated that *Burkholderia* are the ancestral symbionts for the CCR legumes (probability = 0.98), while *Mesorhizobium* and other α-rhizobial genera are derived symbionts of the legumes (Figure 4.4). The most recent common ancestor (MRCA) to the Genistoids, which include the Crotalarieae (i.e. *Aspalathus*, *Calobota*, *Lebeckia* and *Rafnia*), Genisteae (i.e. *Argyrolobium*) and the Podalyrieae (i.e. *Amphithalea*, *Cyclopia*, *Podalyria* and *Virgilia*) was nodulated by *Burkholderia* (probability = 0.96). The MRCA of the Psoraleeae was nodulated by *Mesorhizobium* (probability = 0.97). There were four events of symbiont switching from *Burkholderia* to *Mesorhizobium*, two events from *Burkholderia* to *Ensifer*, and one event each from *Burkholderia* to *Bradyrhizobium* and to *Rhizobium* (Figure 4.4). There were two events of reversal from *Mesorhizobium* to *Burkholderia* and one event each of switching from *Mesorhizobium* to *Ensifer* and to *Bradyrhizobium* (Figure 4.4).

### *Co-diversification of the legumes and their rhizobia*

The cophylogenetic analysis of the legume phylogeny against the chromosomal phylogeny of the rhizobia produced 6301 solutions of equal cost (111), with 15 cospeciation events and 53 duplication and host switching events. Randomization tests rejected the null hypothesis that the two trees are randomly related ($p < 0.01$). A histogram showing the frequency distribution of the event costs is shown in Appendix C6 (supplementary materials). On the other hand, the analysis of co-diversification between the *nodA* tree and the legume tree yielded 1000 solutions of equal cost (97), of which there were 13 events of cospeciation and 46 events of duplication and host switching. Significance testing based on randomization tests did not reject the null hypothesis that the two trees are randomly related ($p > 0.05$). A plot of the frequency distribution of the event costs is provided in the supplementary materials (Appendix C7).

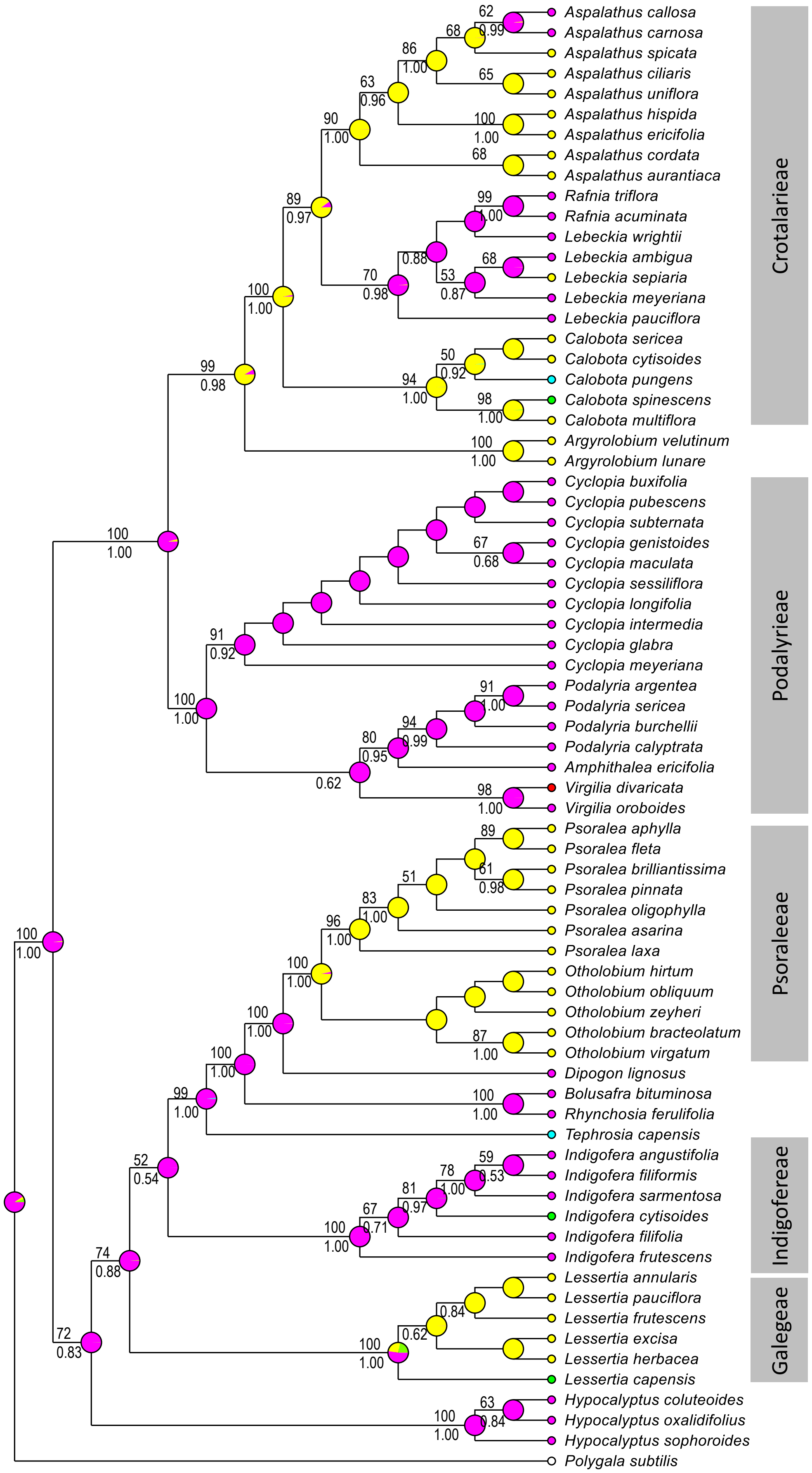

Aspalathus callosa
Aspalathus carnosa
Aspalathus spicata
Aspalathus ciliaris
Aspalathus uniflora
Aspalathus hispida
Aspalathus ericifolia
Aspalathus cordata
Aspalathus aurantiaca
Rafnia triflora
Rafnia acuminata
Lebeckia wrightii
Lebeckia ambigua
Lebeckia sepiaria
Lebeckia meyeriana
Lebeckia pauciflora
Calobota sericea
Calobota cytisoides
Calobota pungens
Calobota spinescens
Calobota multiflora
Argyrolobium velutinum
Argyrolobium lunare
Cyclopia buxifolia
Cyclopia pubescens
Cyclopia subternata
Cyclopia genistoides
Cyclopia maculata
Cyclopia sessiliflora
Cyclopia longifolia
Cyclopia intermedia
Cyclopia glabra
Cyclopia meyeriana
Podalyria argentea
Podalyria sericea
Podalyria burchellii
Podalyria calyptrata
Amphithalea ericifolia
Virgilia divaricata
Virgilia oroboides
Psoralea aphylla
Psoralea fleta
Psoralea brilliantissima
Psoralea pinnata
Psoralea oligophylla
Psoralea asarina
Psoralea laxa
Otholobium hirtum
Otholobium obliquum
Otholobium zeyheri
Otholobium bracteolatum
Otholobium virgatum
Dipogon lignosus
Bolusafra bituminosa
Rhynchosia ferulifolia
Tephrosia capensis
Indigofera angustifolia
Indigofera filiformis
Indigofera sarmentosa
Indigofera cytisoides
Indigofera filifolia
Indigofera frutescens
Lessertia annularis
Lessertia pauciflora
Lessertia frutescens
Lessertia excisa
Lessertia herbacea
Lessertia capensis
Hypocalyptus coluteoides
Hypocalyptus oxalidifolius
Hypocalyptus sophoroides
Polygala subtilis
Crotalarieae
Podalyrieae
Psoraleeae
Indigofereae
Galegeae

**Figure 4.4.** Phylogenetic relationships of legumes with ancestral state reconstruction (by Bayesian Binary MCMC) of the genera of rhizobia nodulating them. This is the best ML tree based on *matK* and *rbcL* data. Support values for the ML (bootstrap %) and BI (posterior probabilities) analyses are shown above and below the nodes, respectively. Rhizobia genera are colour-coded as follows: Pink: *Burkholderia*, Yellow: *Mesorhizobium*, Blue: *Bradyrhizobium*, Red: *Rhizobium*, Green: *Ensifer*. Pies at the nodes indicate the reconstructed probabilities of the various states.

## 4.4 Discussion

The diversity of rhizobial genera encountered in this study is consistent with recent studies showing that *Burkholderia* and *Mesorhizobium* are the predominant symbionts of the CCR legumes (Beukes et al., 2013; Gerding et al., 2012; Lemaire et al., 2015a). Likewise, the types of legume-rhizobia associations reported in previous studies were upheld, i.e. Podalyrieae chiefly associating with *Burkholderia*, while the Psoraleeae and Genisteae associate with *Mesorhizobium.* Similarly, the Crotalarieae and Indigofereae’s association with both alpha and beta rhizobia was maintained. This suggests that despite the small proportion of the legume flora whose rhizobial associations have been studied thus far (≈ 15 %), the broad patterns of legume tribe and rhizobia genera associations are consistent, which allows for generalisations to be made at these taxonomic levels.

The observation of conflicting phylogenetic relationships between the chromosomal gene tree and the *nodA* gene tree indicates that the chromosomal and extra-chromosomal genes have disparate evolutionary histories, likely due to horizontal inheritance of the nodulation gene (*nodA*). The separation of the CCR strains from the non-CCR strains in the *nodA* tree (for *Burkholderia, Mesorhizobium* and *Rhizobium*) supports previous studies which showed that nodulation genes of the CCR rhizobia have a different evolutionary history from those found in other parts of the world (De Meyer et al., 2016; Gerding et al., 2012; Lemaire et al., 2016a; Zheng et al., 2017). On the other hand, the intermixing of CCR rhizobia with the reference strains of non-CCR rhizobia in the chromosomal tree (Figure 4.1) indicates that the symbionts share a common ancestor in deep evolutionary time. Molecular dating analyses on rhizobia could shed some light on the origins and diversification of the various rhizobial genera.

*Bradyrhizobium* and *Ensifer*, on the other hand, showed similar infrageneric relationships in both the chromosomal and *nodA* gene trees, with an intermixing of CCR and non-CCR strains on both trees (Figure 4.1 and Figure 4.2). This suggests a similar evolutionary history between the chromosomal and *nodA* genes, which could be due to vertical inheritence of both chromosomal and nodulation genes. However, a study of *Bradyrhizobium* strains that nodulate some South African legumes of the Great Escarpment showed that their *nodA* genes were distinct from those occurring elsewhere in the world and their phylogeny was incongruent with their chromosomal gene tree (Beukes et al., 2016). This suggests that the *nodA* gene of the studied *Bradyrhizobium* strains of the CCR legumes has a different evolutionary history from that of the Great Escarpment. Considering that the number of *Bradyrhizobium* and *Ensifer* strains encountered in CCR rhizobia studies of indigenous legumes is relatively small compared to *Burkholderia* and *Mesorhizobium*, the strength of this pattern remains to be determined. Overall, the results indicate a complex evolutionary history of the rhizobia genera of the CCR, characterised by both vertical and horizontal inheritence of nodulation genes.

The observation of some strains from different genera being embedded within other genera among the CCR strains, e.g. the *Mesorhizobium* and *Rhizobium* strains that were nested within *Burkholderia*, the *Mesorhizobium* strain that was part of an *Ensifer* clade or the *Burkholderia* strain that was embedded within *Mesorhizobium* (Figure 4.2; Figure 4.3), suggests possible events of HGT of *nodA* between the different genera. This is in addition to the incongruence between the chromosomal and *nodA* gene trees within the genera (Figure 4.3). Such high levels of HGT are incomparable to other centres of diversity of these genera of rhizobia. For instance, in the Brazillian Cerrado and Caatinga biomes, where *Burkholderia* are the primary symbionts of *Mimosa* species, Bontemps et al. (2010) found no significant levels of HGT. Pires et al. (2018) also found no evidence of HGT between *Burkholderia* and *Rhizobium* strains that nodulate different species of the genus *Mimosa*. Likewise, HGT is rare in *Rhizobium*, which is also one of the predominant symbionts of *Mimosa* species in Mexico (Bontemps et al., 2016). These findings indicate that the CCR rhizobia have a unique evolutionary history characterised by frequent events of horizontal gene transfer (hybridisation) within and possibly across diverse genera.

The reconstruction of the ancestral rhizobia on the phylogeny of the legumes supported the hypothesis that *Burkholderia* are the ancestral symbionts of the CCR legumes (Figure 4.4). Although the reconstruction was based on a small subset of the legume flora due to the small number of documented legume-rhizobia relationships in the region, it is consistent with other lines of evidence. For instance, the reconstruction suggested that the ancestors of older legume lineages such as that of the Genistoid legumes (i.e. Crotalarieae, Genisteae and the Podalyrieae) were nodulated by *Burkholderia*, while the switch to *Mesorhizobium* symbionts is a derived condition (e.g. the younger Psoraleeae clade, i.e. *Psoralea* and *Otholobium*). Indeed, the Genistoids are an older lineage that originated 44-46 million years ago (Edwards and Hawkins, 2007; Schnitzler et al., 2011), while the Psoraleeae are a Miocene lineage that originated 8-12 million years ago (Bello, 2016). It is worth noting that the most speciose legume taxa (i.e. *Aspalathus* and the Psoraleeae) are predominantly nodulated by *Mesorhizobium*, which has been shown to have a wider tolerance to various ecological factors (e.g. soil pH) (Laranjo et al., 2014; Laranjo and Oliveira, 2011) than *Burkholderia*, which has a predilection for acidic conditions (Stopnisek et al., 2014). This indicates a potential influence of the heterogeneous edaphic landscape of the CCR (Cowling et al., 2009, 1996; Richards et al., 1997) on the speciation of legumes that are nodulated by *Mesorhizobium.*

The results of the cophylogeny analysis between the legumes' and the rhizobia's chromosomal phylogeny showed that the two trees share a significant amount of co-speciation events, indicating that the frequency of parallel evolution is greater than would be expected due to chance alone. This indicates that the two partners may have influenced each other's evolution. For instance, an increase in the diversity of rhizobial symbionts which are adapted to diverse ecological conditions in the landscape could have triggered diversification in the legumes. This is consistent with the observations from the ancestral rhizobia reconstructions that the speciose legume clades are mostly taxa that are nodulated by *Mesorhizobium*, which has a wider ecological tolerance range (Laranjo et al., 2014; Laranjo and Oliveira, 2011). Given the small proportion of the CCR legumes for which rhizobial partners are presently known, this hypothesis requires further investigation with a more representative phylogeny of the legumes and their symbionts. The finding of no significant co-diversification between the legume tree and the *nodA* tree (which is contrary to the chromosomal tree) accentuates the observation that the two genomes have disparate evolutionary histories due to the prevalence of HGT.

Interestingly, while the CCR's Fynbos biome has similar climatic and edaphic conditions to those of the Kwongan in southwestern Australia (Yates et al., 2010), the two floristic regions have distinct legume floras. This is despite evidence of transoceanic dispersal of plant taxa among the various landmasses of the southern hemisphere during the Cenozoic (Crisp et al., 2009). The only closely related legumes shared between the two regions are the tribe Hypocalypteae, an endemic to the CCR Fynbos biome (Schutte and Van Wyk, 1998) which is sister to the Australian Mirbelieae and Bossiaeeae tribes (Wojciechowski et al., 2004). The Hypocalypteae are exclusively nodulated by *Burkholderia* (Beukes et al., 2013), while the Australian Mirbeloids and Bossiaeeae are predominantly nodulated by *Bradyrhizobium* (Stępkowski et al., 2012). Recent attempts to introduce some South African Fynbos legumes (*Lessertia* species) as pasture crops into Western Australia have been unsuccessful and rhizobial analysis indicated that Australian rhizobia were outcompeting that of the introduced legumes, albeit not fixing any nitrogen (Gerding et al., 2014, 2013). On the other hand, *Dipogon lignosus* L., a South African Fynbos legume that has become invasive in Australia and New Zealand, is nodulated by *Burkholderia* strains that are phylogenetically related to strains of South African *Burkholderia* (Liu et al., 2014). This suggests that the rhizobia were co-introduced along with the legume host and they have managed to break through the competitive barriers in their new environment. Likewise, two other South African legumes, *Psoralea pinnata* and *P. arborea*, seem to have overcome this barrier as they have become invasive in the Mediterranean and Temperate regions of Australia (Stirton et al., 2015). This accentuates the observation that the establishment of introduced rhizobial strains in new environments depends on various environmental factors e.g. pH, soil nutrients and temperature, and the introduced strain's ability to compete with the local rhizobial community which is already well-adapted to the environment (Toro, 1996). Therefore, it is conceivable that the establishment of legumes dispersing out of the CCR may be limited by the availability of compatible rhizobia and their ability to compete with the native rhizobial strains in the new habitats. Hence, the lack of shared legumes between the Fynbos and Kwongan biomes could be related to rhizobia.

### 4.5 Conclusions

The study has shown that for the majority of the rhizobial symbionts of the CCR legumes, *Burkholderia*, *Mesorhizobium* and *Rhizobium*, the *nodA* genes have different evolutionary histories from their counterparts in other parts of the world. Secondly, high levels of hybridisation among the symbionts through horizontal transfer of the *nodA* gene have led to a complex evolutionary history of the CCR rhizobia. The study has also shown that *Burkholderia* is the ancestral symbiont of the CCR legumes and there is evidence of co-diversification between the legumes and their symbionts, which highlights a potential contribution of the legume-rhizobia interaction to the diversification of legumes in the CCR. This hypothesis requires further confirmation with a larger phylogeny of the legumes and their symbionts.

# 5. General Discussion and Conclusion

The distributions of living organisms are influenced by several factors and the theoretical framework of Pearson and Dawson (2003) hypothesizes that the strength of the various factors is governed by spatial scale. The framework posits that climate has the strongest effects at the regional, continental and global scales (> 200 km), whereas edaphic factors impact more at local and site scales (< 10 km), with biotic interactions strongest at local to micro scales (Pearson and Dawson, 2003). Although several studies support this theory (Eiserhardt et al., 2011; Luoto et al., 2006; Pearson et al., 2004; Thuiller et al., 2004), others show departures from it (e.g. Heikkinen et al., 2007; Austin and van Niel, 2011). This points toward some exceptions to the generality of the framework.

The species-rich Core Cape Subregion of southern Africa (CCR), with its heterogeneous landscapes and steep ecological gradients in climate, edaphic factors, topography, biotic factors (pollinator availability, dispersal agents, rhizobial and mycorrhizal associations), etc. (Campbell, 1986; Lambers et al., 2010; Lemaire et al., 2015a; Manning and Goldblatt, 2012) provides an opportunity to test the applicability of this theory. The CCR also has Fabaceae as the second most speciose plant family (Schutte, 2012), a group of plants that has colonised nearly all kinds of ecosystems throughout the globe (LPWG, 2017; Schrire et al., 2005; Sprent et al., 2017). However, the legumes of the CCR are neither uniformly nor randomly distributed, but tend to occupy particular habitats in discrete legume assemblages within the landscape. Thus, the legumes provide a useful model assemblage for studying the drivers of species distributions in the CCR, which was the overall aim of the study.

Using the Cape Peninsula as a microcosm of the CCR (based on its legume species composition and ecological heterogeneity (Cowling et al., 1996; Trinder-Smith et al., 1996)), the study determined whether edaphic factors have influenced the legume species composition at local and site scales as predicted by the theoretical framework of Pearson and Dawson (2003). The findings were consistent with the theory as discrete legume assemblages were associated with distinct edaphic habitats within the Cape Peninsula (Chapter 2). In addition, legume species that are unique to particular edaphic habitats (indicator species) were identified, which suggests a strong effect of edaphic factors on their biogeography. The observed indicator species of the various edaphic habitats were also species that occupy similar edaphic habitats in the broader CCR, outside the Cape Peninsula, which suggests a broader applicability of the findings at a wider scale. The study can act as a prototype for future studies seeking to broaden the search for indicator species including non-legume taxa both in the CCR and other similarly heterogeneous landscapes. Since the CCR has steep ecological gradients, the contributions of other ecological factors e.g. altitude, aspect and biotic interactions also need to be determined in future studies. This would allow for an assessment of the relative influences of the various factors on species distributions in the CCR and it would strengthen the identification of indicator species for the various habitats. The identification of indicator species could provide an ability to predict the prevailing ecological conditions of an ecosystem based on its species composition, which could aid in conservation planning, such as delineating protection-worthy areas that maximise ecological heterogeneity. For example, the study identified *Psoralea repens* and *Otholobium bracteolatum* as indicator species for the Dune sands of the CCR, a coastal habitat that is under threat from habitat transformation.

Apart from edaphic factors, biotic interactions are also expected to strongly influence species distributions at local and micro scales (Pearson and Dawson, 2003). In the case of legumes, one such interaction involves their symbiotic association with rhizobia that fix nitrogen in the legume root nodules. This interaction is particularly important for legumes occurring in the CCR because most of its soils are oligotrophic (Stock and Lewis, 1986; Witkowski and Mitchell, 1987), whereas legumes generally have a high nitrogen demanding lifestyle (McKey, 1994; Reed, 2017; Werner et al., 2015). The legume-rhizobia relationship can be highly specific, e.g. CCR legumes of the tribe Podalyrieae associate only with rhizobia of the genus *Burkholderia* and those of the tribe Psoraleeae associate only with *Mesorhizobium* (Lemaire et al., 2015a). Therefore, availability of suitable symbionts may limit the legume distributions in the CCR. However, the Baas Becking hypothesis regarding the biogeography of microorganisms states that “everything is everywhere, but the environment selects”. Based on this hypothesis, there should be no biogeographic structuring of rhizobia, and hence, legumes should not be limited by the availability of suitable rhizobial symbionts. Thus, the present study sought to test whether the Baas Becking hypothesis holds for rhizobia in the CCR and to investigate the role of rhizobia in the distributions of legumes in the CCR.

Having established the role of edaphic factors in the distribution of legumes in the CCR (Chapter 2), their influence on the distribution of the legumes’ symbionts was investigated (Chapter 3). Results indicated strong phylogenetic structuring of *Burkholderia* and *Mesorhizobium* (the primary symbionts of CCR legumes) with respect to pH and soil types. This indicates the presence of biogeographic structuring of rhizobia at the local scale, whereby closely related strains occupy similar edaphic habitats and vice versa. In addition, *Burkholderia* was found to be restricted to acidic edaphic habitats, while *Mesorhizobium* had a wider distribution range spanning acidic to alkaline habitats. In turn, an analysis of the legume species distribution patterns showed that those that are only nodulated by *Burkholderia* are restricted to acidic soil types, mostly of sandstone parent rock materials, while those that are nodulated by *Mesorhizobium* are more widespread, occupying all the different soil types within the CCR. These results add to the growing body of evidence that refutes the Baas Becking hypothesis on microbial biogeography. They further indicate that the interaction of legumes and rhizobia (biotic interactions) contributes towards the legume biogeography in local spatial scales, in accordance with the framework of Pearson and Dawson (2003). The observation of differential tolerance of both major rhizobial genera to soil acidity and soil types, coupled with the generally heterogeneous ecological landscape of the CCR, exposes avenues for further research into their contribution towards diversification of the legumes (Chapter 4). An understanding of the drivers of legume diversification in the CCR would unravel the key ecological processes that generated such a high diversity of legumes, which implies that conservation action that facilitates and maintains such processes is key to conserving their diversity.

It is puzzling why some legumes species are narrowly distributed. The study on the diversity of rhizobia that associates with the rare and narrowly distributed *Indigofera superba* (Chapter 3) revealed that this legume species exhibits rhizobial specificity, at the generic level, as all the strains associated with it belonged to the genus *Burkholderia*. However, the strains were diverse as they formed several distinct clades, clustering together with strains from various habitats within the CCR. This indicates a potential lack of rhizobial specificity at the infrageneric level. The results were similar to what Lemaire et al. (2016b) found with rhizobia associated with the widespread *Podalyria calyptrata*, which indicates that the narrowly distributed *I. superba* is perhaps limited by other factors. Testing for nodulation capability of *I. superba* on soils from outside its distribution range could shed light on whether rhizobia limits its distribution.

Evidence (from literature) of biogeographic structuring of rhizobia at regional, continental and global scales (De Meyer et al., 2016; Bontemps et al., 2016; Sprent et al., 2017; Zheng et al., 2017) counters both the Baas Becking hypothesis and the hierarchical framework of Pearson and Dawson (2003). This is because in addition to demonstrating the existence of a biogeographic structure for rhizobia it shows that the influence of the biotic interaction of legumes and rhizobia extends beyond the local and micro scales. Of particular relevance in the case of the CCR is that its legume flora mostly comprises Papilionoid legumes (Schutte, 2012), which are predominantly nodulated by *Burkholderia* and *Mesorhizobium* (Lemaire et al., 2015a), while the other subfamilies e.g. Caesalpinioideae, which are largely nodulated by *Bradyrhizobium* (Sprent et al., 2017) are generally lacking in the CCR (Manning and Goldblatt, 2012). Moreover, the Papilionoid counterparts of the CCR legumes e.g. the Crotalarieae and Genisteae, which occur in the Grassland and Savanna biomes of South Africa are nodulated by *Bradyrhizobium* (Beukes et al., 2016). In a glasshouse experiment in which some CCR legumes were grown on soils from both the Savanna and Fynbos biomes, the CCR legumes only nodulated in their native soils, suggesting a lack of compatible rhizobia in the Savanna soils (Lemaire et al. unpublished). At the global scale, while the CCR and the Kwongan of Southwestern Australia have similar climatic and edaphic conditions (Yates et al., 2010) and share a number of genera for other major plant families (Crisp et al., 2004), e.g. the Proteaceae and Restionaceae, this is not the case for Fabaceae. Recent attempts to introduce some CCR legumes into Australia were unsuccessful and rhizobial analyses indicated that the rhizobial strains of the introduced legumes were outcompeted by local strains, which did not fix any nitrogen (Gerding et al., 2014; 2013). These patterns suggest a strong influence of rhizobia on the establishment of CCR legumes outside their native range. They also reveal that the influence of biotic interactions on species distributions extends beyond the local and micro scales. Therefore, studies on the distributions of organisms such as the legumes need to consider the effect of biotic interactions at all spatial scales.

In order to unearth the role of the legume-rhizobia relationship in driving the diversification of legumes in the CCR, an understanding of the evolutionary history of both symbiotic partners is crucial. Hence, the evolutionary history of chromosomal (16S rRNA and *recA*) and the extra-chromosomal nodulation gene (*nodA*) of the rhizobia was investigated and a phylogeny of the legumes was used to reconstruct the ancestral rhizobial symbionts of the CCR legumes (Chapter 4). The results showed evidence of horizontal inheritance of the *nodA* gene within and between rhizobial genera and that the *nodA* genes of CCR rhizobia (*Burkholderia*, *Mesorhizobium,* and *Rhizobium)* are phylogenetically distinct from those of their counterparts from other parts of the globe. This indicates a unique evolutionary history of the *nodA* genes of these CCR rhizobial symbionts, which is consistent with previous studies (Beukes et al., 2013; De Meyer et al., 2016; Lemaire et al., 2016a). On the other hand, the lack of segregation of CCR and non-CCR symbionts for *Bradyrhizobium* and *Ensifer* on the *nodA* gene tree suggested a common origin of *nodA* genes for both CCR and non-CCR symbionts of these genera. This indicates a complex evolutionary history of the CCR symbionts. Although several studies based on whole genome sequences of some *Burkholderia* strains (Aoki et al., 2013; De Meyer et al., 2016; Zheng et al., 2017) have attempted to unravel the origins of nodulation genes, they have only been able to reveal the complexity of their evolutionary history. Thus, the origin of nodulation genes remains an unanswered question that requires further investigation. This might require incorporating nitrogen fixing bacteria that associate with non-legume plants analysed in the context of a dated phylogeny of the plants.

The results (of Chapter 4) also indicated that *Burkholderia* are the most likely ancestral symbionts of the CCR legumes, while the switch to *Mesorhizobium* and the other alpha-rhizobial genera is a derived condition. This is consistent with the hypothesis that *Burkholderia* are ancient symbionts of legumes (Bontemps et al., 2010). However, more work is needed to confirm this hypothesis at both the CCR and global levels using a comprehensive and dated phylogeny of the legumes in order to pinpoint the exact times when the various associations arose. Besides, at the global level, the hypothesis that *Bradyrhizobium* is the ancestral symbiont of legumes (Sprent et al., 2017) has not yet been disproved. Cophylogenetic analyses conducted as part of this study also showed evidence of coevolution between the legumes and their symbionts based on the chromosomal gene trees (Chapter 4). This indicates a potential role of the legume-rhizobia symbiosis in driving the diversification of the CCR legumes, a hypothesis that needs further confirmation with a more comprehensive phylogeny of both the legumes and their symbionts. This would allow for testing of whether shifts in rhizobial symbionts e.g. from *Burkholderia* (which is restricted to acidic habitats) to *Mesorhizobium* (with its wider tolerance range to pH and soil type as shown in Chapter 3) were associated with increases in diversification rates. Increased diversification rates would explain the high species richness in some legume taxa that are nodulated by *Mesorhizobium*, such as the genus *Aspalathus* and the tribe Psoraleeae.

Overall, the study has shown that edaphic factors and biotic interactions between legumes and rhizobia have played a significant role in the evolution and distribution of legumes in the Core Cape Subregion of southern Africa. Although the results are consistent with the hierarchical framework of Pearson and Dawson (2003) for the study area involved, when larger scale distribution patterns of legumes and rhizobia are considered, it appears that the strength of the legume-rhizobia interaction in driving legume distributions is not limited to the local and micro scales as predicted by the framework. Thus, species distribution studies on taxa that are associated with strong biotic interactions need to consider the effect of these interactions at all scales. The study has also shown evidence of biogeographic structuring of rhizobia in the CCR, which is contrary to the Baas Becking hypothesis of no biogeographic structure for microbes. The potential contribution of rhizobia to the diversification of legumes in the CCR has been shown and this requires further investigation in a larger sample of both symbiotic partners.  Therefore, the study should form a basis for future studies on the biogeography of legumes and non-legume plants, in the context of biotic interactions and steep ecological gradients, in the CCR.

# References

Abdelmoumen, H., Filali-Maltouf, A., Neyra, M., Belabed, A., El Idrissi, M.M., 1999. Effect of high salts concentrations on the growth of rhizobia and responses to added osmotica. Journal of Applied Microbiology 86, 889-898.

Allsopp, N., Stock, W.D., 1993. Mycorrhizal status of plants in the Cape Floristic Region, South Africa. Bothalia 23, 91-104.

Ampomah, O.Y., Mousavi, S.A., Lindström, K., Huss-Danell, K., 2017. Diverse *Mesorhizobium* bacteria nodulate native *Astragalus* and *Oxytropis* in arctic and subarctic areas in Eurasia. Systematic and Applied Microbiology 40, 51-58.

Andrews, M., Andrews, M.E., 2017. Specificity in legume-rhizobia symbioses. International Journal of Molecular Sciences 18, 705.

Aoki, S., Ito, M., Iwasaki, W., 2013. From β-to α-proteobacteria: the origin and evolution of rhizobial nodulation genes *nodIJ*. Molecular Biology and Evolution 30, 2494-2508.

Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16, 743-753.

Ardley, J., 2017. Legumes of the Thar Desert and their nitrogen fixing *Ensifer* symbionts. Plant and Soil 410, 517-520.

Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography 38, 1-8.

Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-revisited. American Journal of Botany 97, 1296-1303.

Bell, D.T., 2001. Ecological response syndromes in the flora of southwestern Western Australia: fire resprouters versus reseeders. The Botanical Review 67, 417-440.

Bello, A., 2016. Taxonomy and Evolutionary Studies on the Genus *Psoralea* L. (Psoraleeae, Fabaceae) Doctoral Thesis, University of Cape Town.

Bello, A., Daru, B.H., Stirton, C.H., Chimphango, S.B., van der Bank, M., Maurin, O., Muasya, A.M., 2015. DNA barcodes reveal microevolutionary signals in fire response trait in two legume genera. AoB Plants 7, plv124.

Béna, G., Lyet, A., Huguet, T., Olivieri, I., 2005. *Medicago–Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. Journal of Evolutionary Biology 18, 1547-1558.

Bertrand, R., Perez, V., Gégout, J., 2012. Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of *Quercus pubescens* in France. Global Change Biology 18, 2648-2660.

Beukes, C.W., Palmer, M., Manyaka, P., Chan, W.Y., Avontuur, J.R., Van Zyl, E., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., Varghese, N., Mikhailova, N., Stamatis, D., Reddy, T.B.K., Daum, C., Shapiro, N., Markowitz, V., Ivanova, N., Kyrpides, N., Woyke, T., Blom, J., Whitman, W.B., Venter, S.N., Steenkamp, E.T., 2017. Genome data provides high support fro generic boundaries in *Burkholderia sensu lato*. Frontiers in Microbiology 8, 1154. Doi: 10.3389/fmicb.2017.01154

Beukes, C.W., Stępkowski, T., Venter, S.N., Cłapa, T., Phalane, F.L., le Roux, M.M., Steenkamp, E.T., 2016. Crotalarieae and Genisteae of the South African Great Escarpment are nodulated by novel *Bradyrhizobium* species with unique and diverse symbiotic loci. Molecular Phylogenetics and Evolution 100, 206-218.

Beukes, C.W., Venter, S.N., Law, I.J., Phalane, F.L., Steenkamp, E.T., 2013. South African Papilionoid legumes are nodulated by diverse *Burkholderia* with unique nodulation and nitrogen-fixation loci. PLoS One 8, e68406.

Blomberg, S.P., Garland, T., 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology 15, 899-910.

Bontemps, C., Elliott, G.N., Simon, M.F., Dos Reis Junior, F B, Gross, E., Lawton, R.C., Neto, N.E., Louriero, D.M., De Faria, S.M., Sprent, J.I., 2010. *Burkholderia* species are ancient symbionts of legumes. Molecular Ecology 19, 44-52.

Bontemps, C., Rogel, M.A., Wiechmann, A., Mussabekova, A., Moody, S., Simon, M.F., Moulin, L., Elliott, G.N., Lacercat-Didier, L., Dasilva, C., 2016. Endemic *Mimosa* species from Mexico prefer alphaproteobacterial rhizobial symbionts. New Phytologist 209, 319-333.

Bordeleau, L.M., Prévost, D., 1994. Nodulation and nitrogen fixation in extreme environments. Plant and Soil 161, 115-125.

Born, J., Linder, H.P., Desmet, P., 2007. The Greater Cape Floristic Region. Journal of Biogeography 34, 147-162.

Boulangeat, I., Gravel, D., Thuiller, W., 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecology Letters 15, 584-593.

Bournaud, C., de Faria, S.M., dos Santos, J M F, Tisseyre, P., Silva, M., Chaintreuil, C., Gross, E., James, E.K., Prin, Y., Moulin, L., 2013. *Burkholderia* species are the most common and preferred nodulating symbionts of the *Piptadenia* group (tribe Mimoseae). PLoS One 8, e63478.

Brígido, C., Alexandre, A., Laranjo, M., Oliveira, S., 2007. Moderately acidophilic mesorhizobia isolated from chickpea. Letters in Applied Microbiology 44, 168-174.

Brígido, C., Glick, B.R., Oliveira, S., 2017. Survey of Plant Growth-Promoting Mechanisms in Native Portuguese Chickpea *Mesorhizobium* Isolates. Microbial Ecology 73, 900-915.

Brígido, C., Oliveira, S., 2013. Most acid-tolerant chickpea mesorhizobia show induction of major chaperone genes upon acid shock. Microbial Ecology 65, 145-153.

Broadley, M.R., White, P.J., Hammond, J.P., Zelko, I., Lux, A., 2007. Zinc in plants. New Phytologist 173, 677-702.

Brown, G., Mitchell, D.T., 1986. Influence of fire on the soil phosphorus status in sand plain lowland fynbos, south-western Cape. South African Journal of Botany 52, 67-72.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J., Green, J.L., 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences 105, 11505-11511.

Campbell, B.M., 1986. Vegetation classification in a floristically complex area: The Cape Floristic Region. South African Journal of Botany 52, 129-140.

Campbell, B.M., Moll, E.J., 1977. The forest communities of Table Mountain, South Africa. Plant Ecology 34, 105-115.

Cardoso, E., Nogueira, M.A., Zangaro, W., 2017. Importance of Mycorrhizae in Tropical Soils, in: de Azevedo J.L., Quecine M.C. (Eds.), Diversity and Benefits of Microorganisms from the Tropics. Springer, pp. 245-267.

Certini, G., 2005. Effects of fire on properties of forest soils: a review. Oecologia 143, 1-10.

Chen, W.X., Wang, E.T., David Kuykendall, L., 2005. *Mesorhizobium*. Bergey's Manual of Systematics of Archaea and Bacteria.

Chen, W., Moulin, L., Bontemps, C., Vandamme, P., Béna, G., Boivin-Masson, C., 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. J. Bacteriol. 185, 7266–72. doi:10.1128/JB.185.24.7266-7272.2003

Chen, W., Huang, Q., Xiong, X., 2004. Distribution and biodiversity of soybean rhizobia in the soils of Shennongjia forest reserve, China. Biology and Fertility of Soils 40, 306-312.

Chen, Z., Zha, T., Jia, X., Wu, Y., Wu, B., Zhang, Y., Guo, J., Qin, S., Chen, G., Peltola, H., 2015. Leaf nitrogen is closely coupled to phenophases in a desert shrub ecosystem in China. Journal of Arid Environments 122, 124-131.

Chimphango, S.B., Potgieter, G., Cramer, M.D., 2015. Differentiation of the biogeochemical niches of legumes and non-legumes in the Cape Floristic Region of South Africa. Plant Ecology 216, 1583-1595.

Clark, D.B., Palmer, M.W., Clark, D.A., 1999. Edaphic factors and the landscape-scale distributions of Tropical Rain Forest trees. Ecology 80, 2662-2675.

Conow, C., Fielder, D., Ovadia, Y., Libeskind-Hadas, R., 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms for Molecular Biology 5, 16.

Cowling, R.M., 1990. Diversity components in a species-rich area of the Cape Floristic Region. Journal of Vegetation Science, 699-710.

Cowling, R.M., Witkowski, E., Milewski, A.V., Newbey, K.R., 1994. Taxonomic, edaphic and biological aspects of narrow plant endemism on matched sites in Mediterranean South Africa and Australia. Journal of Biogeography, 651-664.

Cowling, R.M., Procheş, Ş, Partridge, T.C., 2009. Explaining the uniqueness of the Cape flora: incorporating geomorphic evolution as a factor for explaining its diversification. Molecular Phylogenetics and Evolution 51, 64-74.

Cowling, R.M., Rundel, P.W., Lamont, B.B., Arroyo, M.K., Arianoutsou, M., 1996. Plant diversity in Mediterranean-climate regions. Trends in Ecology and Evolution 11, 362-366.

Cox, C.B., 2001. The biogeographic regions reconsidered. Journal of Biogeography 28, 511-523.

Cramer, M.D., 2010. Phosphate as a limiting resource: introduction. Plant and Soil 334, 1-10.

Crews, T.E., 1999. The presence of nitrogen fixing legumes in terrestrial communities: Evolutionary vs ecological considerations, in: Townsend A.R. (Ed.), New Perspectives on Nitrogen Cycling in the Temperate and Tropical Americas. Springer, pp. 233-246.

Crisp, M.D., Arroyo, M.T., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S., Weston, P.H., Westoby, M., Wilf, P., Linder, H.P., 2009. Phylogenetic biome conservatism on a global scale. Nature 458, 754.

Crisp, M.D., Cook, L.G., 2013. How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. Annual Review of Ecology, Evolution, and Systematics 44, 303-324.

Crisp, M., Cook, L., Steane, D., 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? Philosophical Transactions of the Royal Society of London B: Biological Sciences 359, 1551-1571.

Davison, J., 1999. Genetic exchange between bacteria in the environment. Plasmid 42, 73-91.

De Cáceres, M., Legendre, P., Moretti, M., 2010. Improving indicator species analysis by combining groups of sites. Oikos 119, 1674-1684.

De Faria, S.M., Lewis, G.P., Sprent, J.I., Sutherland, J.M., 1989. Occurrence of nodulation in the Leguminosae. New Phytologist 111, 607-619.

De Meyer, S.E., Cnockaert, M., Ardley, J.K., Maker, G., Yates, R., Howieson, J.G., Vandamme, P., 2013a. *Burkholderia sprentiae sp. nov*., isolated from *Lebeckia ambigua* root nodules. International Journal of Systematic and Evolutionary Microbiology 63, 3950.

De Meyer, S.E., Cnockaert, M., Ardley, J.K., Trengove, R.D., Garau, G., Howieson, J.G., Vandamme, P., 2013b. *Burkholderia rhynchosiae sp. nov*., isolated from *Rhynchosia ferulifolia* root nodules. International Journal of Systematic and Evolutionary Microbiology 63, 3944.

De Meyer, S.E., Briscoe, L., Martínez-Hidalgo, P., Agapakis, C.M., de-los Santos, P.E., Seshadri, R., Reeve, W., Weinstock, G., O'Hara, G., Howieson, J.G., 2016. Symbiotic *Burkholderia* species show diverse arrangements of *nif/fix* and *nod* genes and lack typical high-affinity cytochrome *cbb3* oxidase genes. Molecular Plant Microbe Interactions 29, 609-619.

De Meyer, S.E., Cnockaert, M., Ardley, J.K., Van Wyk, B., Vandamme, P.A., Howieson, J.G., 2014. *Burkholderia dilworthii sp. nov.*, isolated from *Lebeckia ambigua* root nodules. International Journal of Systematic and Evolutionary Microbiology 64, 1090.

Dean, S., Farrer, E.C., Menges, E.S., 2015. Fire Effects on Soil Biogeochemistry in Florida Scrubby Flatwoods. The American Midland Naturalist 174, 49-64.

Degefu, T., Wolde-meskel, E., Woliy, K., Frostegård, Å, 2017. Phylogenetically diverse groups of *Bradyrhizobium* isolated from nodules of tree and annual legume species growing in Ethiopia. Systematic and Applied Microbiology 40, 205-214.

Delamuta, J.R.M., Menna, P., Ribeiro, R.A., Hungria, M., 2017. Phylogenies of symbiotic genes of *Bradyrhizobium* symbionts of legumes of economic and environmental importance in Brazil support the definition of the new symbiovars *pachyrhizi* and *sojae*. Systematic and Applied Microbiology.

Depoorter, E., Bull, M.J., Peeters, C., Coenye, T., Vandamme, P., Mahenthiralingam, E., 2016. *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. Appllied Microbiology and Biotechnology 100, 5215-5229.

Diamantis, V., Pagorogon, L., Gazani, E., Gkiougkis, I., Pechtelidis, A., Pliakas, F., Elsen, E., Doerr, S.H., Ritsema, C.J., 2017. Use of Clay Dispersed in Water for Decreasing Soil Water Repellency. Land Degradation and Development 28, 328-334.

Diniz-Filho, J.A.F., Santos, T., Rangel, T.F., Bini, L.M., 2012. A comparison of metrics for estimating phylogenetic signal under alternative evolutionary models. Genetics and Molecular Biology 35, 673-679.

Dludlu, M., 2010. Systematic studies of the southern African Psoraleoid legumes Master's Dissertation, University of Cape Town.

Dobritsa, A.P., Samadpour, M., 2016. Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia gen. nov*. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. International Journal of Systematic and Evolutionary Microbiology 66, 2836-2846.

Dos Santos, J.M.F., Alves, P.A.C., Silva, V.C., Rhem, M.F.K., James, E.K., Gross, E., 2017. Diverse genotypes of *Bradyrhizobium* nodulate herbaceous *Chamaecrista* (Moench) (Fabaceae, Caesalpinioideae) species in Brazil. Syst. Appl. Microbiol. 40, 69–79. doi:10.1016/j.syapm.2016.12.004

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19, 11-15.

Doyle, J.J., 2016. Chasing unicorns: Nodulation origins and the paradox of novelty. American Journal of Botany 103, 1865-1868.

Doyle, J.J., 2011. Phylogenetic perspectives on the origins of nodulation. Molecular Plant-Microbe Interactions 24, 1289-1295.

Doyle, J.J., Luckow, M.A., 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology 131, 900-910.

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., Guisan, A., 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. Journal of Vegetation Science 24, 593-606.

Duke, J., 2012. Handbook of Legumes of World Economic Importance Springer Science & Business Media.

Edwards, D., Hawkins, J.A., 2007. Are Cape floral clades the same age? Contemporaneous origins of two lineages in the genistoids *s.l.* (Fabaceae). Molecular Phylogenetics and Evolution 45, 952-970.

Eisen, J.A., 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Current Opinion in Genetics and Development 10, 606-611.

Eiserhardt, W.L., Svenning, J., Kissling, W.D., Balslev, H., 2011. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. Annals of Botany 108, 1391-1416.

Elliott, G.N., Chen, W.M.W.-M., Bontemps, C., Chou, J.H.J.-H., Young, J.P.W., Sprent, J.I., James, E.K., 2007. Nodulation of *Cyclopia* spp. (Leguminosae, Papilionoideae) by *Burkholderia tuberum*. Ann. Bot. 100, 1403–1411. doi:10.1093/aob/mcm227

Elliott, G.N., Chou, J.H., Chen, W.M., Bloemberg, G.V., Bontemps, C., Martínez-Romero, E., Velázquez, E., Young, J.P.W., Sprent, J.I., James, E.K., 2009. *Burkholderia* spp. are the most competitive symbionts of *Mimosa,* particularly under N-limited conditions. Environmental Microbiology 11, 762-778.

Ellis, A.G., Weis, A.E., 2006. Coexistence and differentiation of ‘flowering stones’: the role of local adaptation to soil microenvironment. Journal of Ecology 94, 322-335.

Ellis, A.G., Weis, A.E., Gaut, B.S., 2006. Evolutionary radiation of “stone plants” in the genus *Argyroderma* (Aizoaceae): unraveling the effects of landscape, habitat, and flowering time. Evolution 60, 39-55.

Esler, K.J., Cowling, R.M., 1993. Edaphic factors and competition as determinants of pattern in South African Karoo vegetation. South African Journal of Botany 59, 287-295.

Estrada-de Los Santos, P., Rojas-Rojas, F.U., Tapia-García, E.Y., Vásquez-Murrieta, M.S., Hirsch, A.M., 2016. To split or not to split: an opinion on dividing the genus *Burkholderia*. Annals of Microbiology 66, 1303-1314.

Estrada-de Los Santos, P., Vacaseydel-Aceves, N.B., Martnez-Aguilar, L., Cruz-Hernndez, M.A., Mendoza-Herrera, A., Caballero-Mellado, J., 2011. *Cupriavidus* and *Burkholderia* species associated with agricultural plants that grow in alkaline soils. The Journal of Microbiology 49, 867-876.

Ferguson, B., Lin, M., Gresshoff, P.M., 2013. Regulation of legume nodulation by acidic growth conditions. Plant signaling & behavior 8, e23426.

Flores, M., Morales, L., Avila, A., Gonzlez, V., Bustos, P., Garca, D., Mora, Y., Guo, X., Collado-Vides, J., Pinero, D., 2005. Diversification of DNA sequences in the symbiotic genome of *Rhizobium etli*. Journal of Bacteriology 187, 7185-7192.

Fontaneto, D., Barraclough, T.G., Chen, K., Ricci, C., Herniou, E.A., 2008. Molecular evidence for broad-scale distributions in bdelloid rotifers: everything is not everywhere but most things are very widespread. Molecular Ecology 17, 3136-3146.

Fontaneto, D., Hortal, J., 2012. Microbial biogeography: is everything small everywhere, in: Ogilvie L.A., Hirsch P.R. (Eds.), Microbial Ecological Theory: Current Perspectives. Caister Academic Press, Norfolk, pp. 87-98.

Francisco-Ortega, J., Santiago-Valentín, E., Acevedo-Rodrguez, P., Lewis, C., Pipoly III, J., Meerow, A.W., Maunder, M., 2007. Seed plant genera endemic to the Caribbean Island biodiversity hotspot: a review and a molecular phylogenetic perspective. The Botanical Review 73, 183-234.

Freckleton, R.P., Harvey, P.H., Pagel, M., 2002. Phylogenetic analysis and comparative data: a test and review of evidence. The American Naturalist 160, 712-726.

Fritz, S.A., Purvis, A., 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. Conservation Biology 24, 1042-1051.

Garau, G., Yates, R.J., Deiana, P., Howieson, J.G., 2009. Novel strains of nodulating *Burkholderia* have a role in nitrogen fixation with papilionoid herbaceous legumes adapted to acid, infertile soils. Soil Biology and Biochemistry 41, 125-134.

Gaunt, M.W., Turner, S.L., Rigottier-Gois, L., Lloyd-Macgilp, S.A., Young, J., 2001. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. International Journal of Systematic and Evolutionary Microbiology 51, 2037-2048.

Gehlot, H.S., Panwar, D., Tak, N., Tak, A., Sankhla, I.S., Poonar, N., Parihar, R., Shekhawat, N.S., Kumar, M., Tiwari, R., 2012. Nodulation of legumes from the Thar Desert of India and molecular characterization of their rhizobia. Plant and Soil 357, 227-243.

Geml, J., 2017. Altitudinal Gradients in Mycorrhizal Symbioses, in: Tedersoo L. (Ed.), Biogeography of Mycorrhizal Symbiosis. Springer, pp. 107-123.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., Young, N.D., 2005. Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiology 137, 1228-1235.

Gerding, M., Howieson, J.G., O’Hara, G.W., Real, D., Bräu, L., 2013. Establishment and survival of the South African legume *Lessertia* spp. and rhizobia in Western Australian agricultural systems. Plant and Soil 370, 235-249.

Gerding, M., O’Hara, G.W., Bräu, L., Nandasena, K., Howieson, J.G., 2012. Diverse *Mesorhizobium* spp. with unique *nodA* nodulating the South African legume species of the genus *Lessertia*. Plant and Soil 358, 385-401.

Gerding, M., O’Hara, G.W., Howieson, J.G., Bräu, L., 2014. Overcoming non-selective nodulation of *Lessertia* by soil-borne rhizobium in the presence of inoculant *Mesorhizobium*. Plant and Soil 380, 117-132.

Gerding, M., Oyarzúa, P., Garca, L., Fischer, S., Norambuena, C., Barahona, V., Del Pozo, A., Ovalle, C., 2017. Diversity and symbiotic effectiveness of *Adesmia* spp. root nodule bacteria in central and southern Chile. Symbiosis 72, 61-72.

Giraldo, R., Caballero, W., Camacho-Tamayo, J., 2016. Mantel test for spatial functional data. AStA Advances in Statistical Analysis, 1-19.

Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. Molecular Biology and Evolution 19, 2226-2238.

Goldblatt, P., Manning, J., 2000. Cape Plants: A Conspectus of the Cape Flora of South Africa. National Botanical Institute.

Goldblatt, P., Manning, J.C., 2002. Plant diversity of the Cape region of southern Africa. Annals of the Missouri Botanical Garden, 281-302.

Goslee, S.C., Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22, 1-19.

Graham, P.H., Viteri, S.E., Mackie, F., Vargas, A.T., Palacios, A., 1982. Variation in acid soil tolerance among strains of *Rhizobium phaseoli*. Field Crops Research 5, 121-128.

Graham, P.H., Vance, C.P., 2003. Legumes: importance and constraints to greater use. Plant Physiology 131, 872-877.

Graziano, M., Lamattina, L., 2005. Nitric oxide and iron in plants: an emerging and converging story. Trends in Plant Science 10, 4-8.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8, 993-1009.

Gyaneshwar, P., Hirsch, A.M., Moulin, L., Chen, W., Elliott, G.N., Bontemps, C., Estrada-de los Santos, P., Gross, E., dos Reis Jr, F B, Sprent, J.I., 2011. Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. Molecular Plant-Microbe Interactions 24, 1276-1288.

Hall, J.S., McKenna, J.J., Ashton, P.M.S., Gregoire, T.G., 2004. Habitat characterizations underestimate the role of edaphic factors controlling the distribution of *Entandrophragma*. Ecology 85, 2171-2183.

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

Hänsch, R., Mendel, R.R., 2009. Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). Current Opinion in Plant Biology 12, 259-266.

Hassen, A., Bopape, F., Habig, J., Lamprecht, S., 2012. Nodulation of rooibos (*Aspalathus linearis* Burm. f.), an indigenous South African legume, by members of both the α-Proteobacteria and β-Proteobacteria. Biology and Fertility of Soils 48, 295-303.

Haukka, K., Lindström, K., Young, J.P., 1998. Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. Applied and Environmental Microbiology 64, 419-426.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Körber, J., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecology and Biogeography 16, 754-763.

Helme, N.A., Trinder-Smith, T.H., 2006. The endemic flora of the Cape Peninsula, South Africa. South African Journal of Botany 72, 205-210.

Herendeen, P.S., Crepet, W.L., Dilcher, D.L., 1992. The fossil history of the Leguminosae: phylogenetic and biogeographical implications, in: Herendeen P.S., Dilher D.L. (Eds.), Advances in Legume Systematics 4, pp. 303-316.

Higgins, S.I., Richardson, D.M., Cowling, R.M., Trinder-Smith, T.H., 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. Conservation Biology 13, 303-313.

Hopper, S.D., Gioia, P., 2004. The southwest Australian floristic region: evolution and conservation of a global hot spot of biodiversity. Annual Review of Ecology, Evolution and Systematics 35, 623-650.

Howieson, J.G., De Meyer, S.E., Vivas-Marfisi, A., Ratnayake, S., Ardley, J.K., Yates, R.J., 2013. Novel *Burkholderia* bacteria isolated from *Lebeckia ambigua* - a perennial suffrutescent legume of the fynbos. Soil Biology and Biochemistry 60, 55.

Howieson, J.G., Yates, R.J., Bala, A., Hungria, M., 2016. Collecting nodules for isolation of rhizobia, in: Howieson, J.G., Dilworth, M.J. (Eds.), Working with Rhizobia. Australian Centre for International Agricultural Research, Canberra, pp. 25–38.

Huang, J., Gogarten, J.P., 2006. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. Trends in Genetics 22, 361-366.

Jensen, E.S., Hauggaard-Nielsen, H., 2003. How can increased use of biological $N_2$ fixation in agriculture benefit the environment? Plant and Soil 252, 177-186.

Jeong, J., Guerinot, M.L., 2009. Homing in on iron homeostasis in plants. Trends in Plant Science 14, 280-285.

John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarrete, H., Vallejo, M., 2007. Soil nutrients influence spatial distributions of tropical tree species. Proceedings of the National Academy of Sciences 104, 864-869.

Juhas, M., van der Meer, J R, Gaillard, M., Harding, R.M., Hood, D.W., Crook, D.W., 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiology Reviews 33, 376-393.

Kampfenkel, K., Van Montagu, M., Inzé, D., 1995. Effects of iron excess on *Nicotiana plumbaginifolia* plants (implications to oxidative stress). Plant Physiology 107, 725-735.

Kanu, S.A., Dakora, F.D., 2012. Symbiotic nitrogen contribution and biodiversity of root-nodule bacteria nodulating *Psoralea* species in the Cape Fynbos, South Africa. Soil Biology and Biochemistry 54, 68-76.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059-3066.

Keet, J.H., Ellis, A.G., Hui, C., Le Roux, J.J., 2017. Legume-rhizobium symbiotic promiscuity and effectiveness do not affect plant invasiveness. Annals of Botany 119, 1319-1331.

Klock, M.M., Barrett, L.G., Thrall, P.H., Harms, K.E., 2016. Differential plant invasiveness is not always driven by host promiscuity with bacterial symbionts. AoB Plants 8, plw060.

Klopper, R.R., Gautier, L., Chatelain, C., Smith, G.F., Spichiger, R., 2007. Floristics of the angiosperm flora of sub-Saharan Africa: an analysis of the African Plant Checklist and Database. Taxon 56, 201-208.

Kneip, C., Lockhart, P., Voß, C., Maier, U., 2007. Nitrogen fixation in eukaryotes – New models for symbiosis 12, 1–12. doi: 10.1186/1471-2148-7-55

Körner, C., 2007. The use of ‘altitude’in ecological research. Trends in Ecology and Evolution 22, 569-574.

Kruger, F.J., 1983. Plant community diversity and dynamics in relation to fire, in: Kruger F.J., Mitchell D.T., Jarvus J.U.M. (Eds.), Mediterranean-Type Ecosystems. Springer, Berlin, pp. 446-472.

Kulkarni, S., Surange, S., Nautiyal, C.S., 2000. Crossing the limits of *Rhizobium* existence in extreme conditions. Current Microbiology 41, 402-409.

Lafay, B., Burdon, J.J., 1998. Molecular diversity of rhizobia occurring on native shrubby legumes in southeastern Australia. Applied and Environmental Microbiology 64, 3989-3997.

Lambers, H., Brundrett, M.C., Raven, J.A., Hopper, S.D., 2011. Plant mineral nutrition in ancient landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. Plant and Soil 348, 7.

Lambers, H., Raven, J.A., Shaver, G.R., Smith, S.E., 2008. Plant nutrient-acquisition strategies change with soil age. Trends in Ecology and Evolution 23, 95-103.

Lamont, B., 1982. Mechanisms for enhancing nutrient uptake in plants, with particular reference to Mediterranean South Africa and Western Australia. The Botanical Review 48, 597-689.

Lane, D.J., 1991. 16S/23S rRNA Sequencing, in: Stackebrandt E.G., M (Ed.), Nucleic Acid Techniques in Bacterial Systematics. John Wiley and Sons, New York, pp. 115-148.

Laranjo, M., Oliveira, S., 2011. Tolerance of *Mesorhizobium* type strains to different environmental stresses. Antonie van Leeuwenhoek 99, 651-662.

Laranjo, M., Alexandre, A., Oliveira, S., 2014. Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus. Microbiol. Res. 169, 2–17. doi:10.1016/j.micres.2013.09.012

Larimer, A.L., Clay, K., Bever, J.D., 2014. Synergism and context dependency of interactions between arbuscular mycorrhizal fungi and rhizobia with a prairie legume. Ecology 95, 1045-1054.

Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Systematic Biology 54, 575-594.

Lavin, M., Matos, A.B., 2008. The impact of ecology and biogeography on legume diversity, endemism, and phylogeny in the Caribbean region: a new direction in historical biogeography. The Botanical Review 74, 178-196.

Lawrence, J.G., Ochman, H., 1997. Amelioration of bacterial genomes: rates of change and exchange. Journal of Molecular Evolution 44, 383-397.

Le Maitre, D.C., Midgley, J.J., 1992. Plant reproductive ecology, in: Cowling R.M. (Ed.), The Ecology of Fynbos: Nutrients, Fire and Diversity. Oxford University Press, Cape Town, pp. 135-175.

Le Maitre, D.C., Gaertner, M., Marchante, E., Ens, E., Holmes, P.M., Pauchard, A., O’Farrell, P.J., Rogers, A.M., Blanchard, R., Blignaut, J., Richardson, D.M., 2011. Impacts of invasive Australian acacias: implications for management and restoration. Diversity and Distributions 17, 1015-1029.

Le Roux, J.J., Mavengere, N.R., Ellis, A.G., 2016. The structure of legume-rhizobium interaction networks and their response to tree invasions. AoB Plants 8, plw038.

Lemaire, B., Chimphango, S.B., Stirton, C., Rafudeen, S., Honnay, O., Smets, E., Chen, W.M., Sprent, J., James, E.K., Muasya, A.M., 2016a. Biogeographical Patterns of Legume-Nodulating *Burkholderia* spp.: from African Fynbos to Continental Scales. Applied and Environmental Microbiology 82, 5099-5115.

Lemaire, B., Dlodlo, O., Chimphango, S., Stirton, C., Schrire, B., Boatwright, J.S., Honnay, O., Smets, E., Sprent, J., James, E.K., 2015a. Symbiotic diversity, specificity and distribution of rhizobia in native legumes of the Core Cape Subregion (South Africa). FEMS Microbiology Ecology 91, 1-17.

Lemaire, B., Van Cauwenberghe, J., Chimphango, S., Stirton, C., Honnay, O., Smets, E., Muasya, A.M., 2015b. Recombination and horizontal transfer of nodulation and ACC deaminase (*acdS*) genes within Alpha-and Betaproteobacteria nodulating legumes of the Cape Fynbos biome. FEMS Microbiology Ecology 91.

Lemaire, B., Van Cauwenberghe, J., Verstraete, B., Chimphango, S., Stirton, C., Honnay, O., Smets, E., Sprent, J., James, E.K., Muasya, A.M., 2016b. Characterization of the papilionoid–*Burkholderia* interaction in the Fynbos biome: The diversity and distribution of beta-rhizobia nodulating *Podalyria calyptrata* (Fabaceae, Podalyrieae). Systematic and Applied Microbiology 39, 41-48.

Lewis, G., Schrire, B., MacKinder, B., Lock, M., 2005. Legumes of the World Royal Botanic Gardens, Kew.

Li, H., Wang, W., Mortimer, P.E., Li, R., Li, D., Hyde, K.D., Xu, J., Soltis, D.E., Chen, Z., 2015. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. Scientific Reports 5.

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

Liu, X., Wei, S., Wang, F., James, E.K., Guo, X., Zagar, C., Xia, L.G., Dong, X., Wang, Y.P., 2012. *Burkholderia* and *Cupriavidu*s spp. are the preferred symbionts of *Mimosa* spp. in Southern China. FEMS Microbiology Ecology 80, 417-426.

Liu, W.Y., Ridgway, H.J., James, T.K., James, E.K., Chen, W., Sprent, J.I., Young, J.P.W., Andrews, M., 2014. *Burkholderia* sp. induces functional nodules on the South African invasive legume *Dipogon lignosus* (Phaseoleae) in New Zealand soils. Microbial Ecology 68, 542-555.

Lopes, A.S., Cox, F.R., 1977. A survey of the fertility status of surface soils under “Cerrado” vegetation in Brazil. Soil Science Society of America Journal 41, 742-747.

Lopes-Santos, L., Castro, D.B.A., Ferreira-Tonin, M., Corrêa, D.B.A., Weir, B.S., Park, D., Ottoboni, L.M.M., Neto, J.R., Destéfano, S.A.L., 2017. Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis gen. nov.*, *comb. nov*. Antonie van Leeuwenhoek 110, 727-736.

LPWG, 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66, 44-77.

LPWG, 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. Taxon 62, 217-248.

Luoto, M., Heikkinen, R.K., Pöyry, J., Saarinen, K., 2006. Determinants of the biogeographical distribution of butterflies in boreal regions. Journal of Biogeography 33, 1764-1778.

Luoto, M., Heikkinen, R.K., 2008. Disregarding topographical heterogeneity biases species turnover assessments based on bioclimatic models. Global Change Biology 14, 483-494.

Maathuis, F.J., 2009. Physiological functions of mineral macronutrients. Current Opinion in Plant Biology 12, 250-258.

MacLean, A.M., Finan, T.M., Sadowsky, M.J., 2007. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiology 144, 615-622.

Maistry, P.M., Cramer, M.D., Chimphango, S.B., 2013. N and P colimitation of $N_2$-fixing and N-supplied fynbos legumes from the Cape Floristic Region. Plant and Soil 373, 217-228.

Maistry, P.M., Muasya, A.M., Valentine, A.J., Chimphango, S.B., 2015. Increasing nitrogen supply stimulates phosphorus acquisition mechanisms in the fynbos species *Aspalathus linearis*. Functional Plant Biology 42, 52-62.

Manning, J., Goldblatt, P., 2012. Plants of the Greater Cape Floristic Region 1: The Core Cape Flora. Strelitzia 29, South African National Biodiversity Institute, Pretoria.

Martínez-Romero, J.C., Ormeno-Orrillo, E., Rogel, M.A., López-López, A., Martínez-Romero, E., 2010. Trends in rhizobial evolution and some taxonomic remarks, in: Pontarotti, P. (Ed.), Evolutionary biology–concepts, molecular and morphological evolution. Springer Berlin Heidelberg, pp. 301–315.

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews: Microbiology 4, 102.

Martiny, J.B., Eisen, J.A., Penn, K., Allison, S.D., Horner-Devine, M.C., 2011. Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108, 7850-7854.

Masson-Boivin, C., Giraud, E., Perret, X., Batut, J., 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends in Microbiology 17, 458-466.

Mavengere, N.R., Ellis, A.G., Le Roux, J.J., 2014. *Burkholderia aspalathi sp. nov*., isolated from root nodules of the South African legume *Aspalathus abietina* Thunb. International Journal of Systematic and Evolutionary Microbiology 64, 1906.

McKey, D., 1994. Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle, in: Sprent J.I., McKey D. (Eds.), Advances in Legume Systematics 5, pp. 211-228.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature Communications 8, 14363.

Mickevich, M.F., Farris, J.S., 1981. The implications of congruence in *Menidia*. Systematic Zoology 30, 351-370.

Mikola, P.U., 1986. Relationship between nitrogen fixation and mycorrhiza. World Journal of Microbiology and Biotechnology 2, 275-282.

Mishra, R.P., Tisseyre, P., Melkonian, R., Chaintreuil, C., Mich, L., Klonowska, A., Gonzalez, S., Beba, G., Laguerre, G., Moulin, L., 2012. Genetic diversity of *Mimosa pudica* rhizobial symbionts in soils of French Guiana: investigating the origin and diversity of *Burkholderia phymatum* and other beta-rhizobia. FEMS Microbiol. Ecol. 79, 487–503.

Mitchell, D.T., Brown, G., Jongens-Roberts, S.M., 1984. Variation of forms of phosphorus in the sandy soils of coastal fynbos, south-western Cape. The Journal of Ecology, 575-584.

Molina-Venegas, R., Rodríguez, M., 2017. Revisiting phylogenetic signal; strong or negligible impacts of polytomies and branch length information? BMC Evolutionary Biology 17, 53.

Moulin, L., Béna, G., Boivin-Masson, C., Stępkowski, T., 2004. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. Mol. Phylogenet. Evol. 30, 720–732.

Moulin, L., James, E.K., Klonowska, A., Faria, S.M., Simon, M.F., 2015. Phylogeny, diversity, geographical distribution, and host range of legume-nodulating Betaproteobacteria: what is the role of plant taxonomy? in: de Bruijn F.J. (Ed.), Biological Nitrogen Fixation. John Wiley & Sons Inc, Chichester, UK, pp. 177-190.

Moulin, L., Munive, A., Dreyfus, B., Boivin-Masson, C., 2001. Nodulation of legumes by members of the *beta*-subclass of Proteobacteria. Nature 411, 948.

Mucina, L., Rutherford, M.C., 2006. The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19, South African National Biodiversity Institute, Pretoria.

Nagajyoti, P.C., Lee, K.D., Sreekanth, T., 2010. Heavy metals, occurrence and toxicity for plants: a review. Environmental Chemistry Letters 8, 199-216.

Ndlovu, J., Richardson, D.M., Wilson, J.R.U., Le Roux, J.J., Ladiges, P., 2013. Co-invasion of South African ecosystems by an Australian legume and its rhizobial symbionts. Journal of Biogeography 40, 1240-1251.

Niemi, G.J., McDonald, M.E., 2004. Application of ecological indicators. Annual Review of Ecology, Evolution and Systematics 35, 89-111.

O'Malley, M.A., 2008. 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. Studies in History and Philosophy of Biological and Biomedical Sciences 39, 314-325.

O'Malley, M.A., 2007. The nineteenth century roots of 'everything is everywhere'. Nature Reviews: Microbiology 5, 647.

Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299.

Ojeda, F., Pausas, J.G., Verd, M., 2010. Soil shapes community structure through fire. Oecologia 163, 729-735.

Ojha, A., Tak, N., Rathi, S., Chouhan, B., Rao, S.R., Barik, S.K., Joshi, S.R., Sprent, J.I., James, E.K., Gehlot, H.S., 2017. Molecular characterization of novel *Bradyrhizobium* strains nodulating *Eriosema chinense* and *Flemingia vestita*, important unexplored native legumes of the sub-Himalayan region (Meghalaya) of India. Systematic and Applied Microbiology.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., Pearse, W., 2012. Caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5 283, 7.

Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401, 877.

Palmer, M.W., 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74, 2215-2230.

Panwar, J., Tarafdar, J.C., 2006. Distribution of three endangered medicinal plant species and their colonization with arbuscular mycorrhizal fungi. Journal of Arid Environments 65, 337-350.

Parker, M.A., 2015. The spread of *Bradyrhizobium* lineages across host legume clades: from *Abarema* to *Zygia*. Microbial Ecology 69, 630-640.

Parker, M.A., 2012. Legumes select symbiosis island sequence variants in *Bradyrhizobium*. Molecular Ecology 21, 1769-1778.

Parker, M.A., 2001. Mutualism as a constraint on invasion success for legumes and rhizobia. Diversity and Distributions 7, 125-136.

Parker, M.A., Malek, W., Parker, I.M., 2006. Growth of an invasive legume is symbiont limited in newly occupied habitats. Diversity and Distributions 12, 563-571.

Paul, J.H., 1999. Microbial gene transfer: an ecological perspective. Journal of Molecular Microbiology and Biotechnology 1, 45-50.

Payn, T.W., Clough, M.E., 1989. Factors affecting soil phosphate distribution and reactions in Cape forest soils. South African Forestry Journal 150, 24-31.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12, 361-371.

Pearson, R.G., Dawson, T.P., Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27, 285-298.

Peix, A., Ramrez-Bahena, M.H., Velzquez, E., Bedmar, E.J., 2015. Bacterial associations with legumes. Critical Reviews in Plant Sciences 34, 17-42.

Perret, X., Staehelin, C., Broughton, W.J., 2000. Molecular basis of symbiotic promiscuity. Microbiol. Mol. Biol. Rev. 64, 180–201.

Pilon-Smits, E.A., Quinn, C.F., Tapken, W., Malagoli, M., Schiavon, M., 2009. Physiological functions of beneficial elements. Current Opinion in Plant Biology 12, 267-274.

Pires, R. de C., dos Reis Junior, F.B., Zilli, J.E., Fischer, D., Hoffman, A., James, E.K., Simon, M.F., 2018. Soil characteristics determine the rhizobia in association with different species of *Mimosa* in central Brazil. Plant and Soil 423, 411-428.

Pirie, M.D., Humphreys, A.M., Barker, N.P., Linder, H.P., 2009. Reticulation, data combination and inferring evolutionary history: an example from Danthonioideae (Poaceae). Systematic Biology 58, 612-628.

Pirie, M.D., Humpreys, A.M., Galley, C., Barker, N.P., Verboom, G.A., Orlovich, D., Draffin, S.J., Lloyd, K., Baeza, C.M., Negritto, M., 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. Molecular Phylogenetics and Evolution 48, 1106-1119.

Platero, R., James, E.K., Rios, C., Iriarte, A., Sandes, L., Zabaleta, M., Battistoni, F., Fabiano, E., 2016. Novel *Cupriavidus* strains isolated from root nodules of native Uruguayan *Mimosa* species. Applied and Environmental Microbiology 82, 3150-3164.

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

Poulsen, Z.C., Hoffman, M.T., 2015. Changes in the distribution of indigenous forest in Table Mountain National Park during the 20th Century. South African Journal of Botany 101, 49-56.

Power, S.C., Cramer, M.D., Verboom, G.A., Chimphango, S.B., 2011. Legume seeders of the Cape Floristic Region inhabit more fertile soils than congeneric resprouters-sometimes. Plant Ecology 212, 1979-1989.

Power, S.C., Cramer, M.D., Verboom, G.A., Chimphango, S.B., 2010. Does phosphate acquisition constrain legume persistence in the fynbos of the Cape Floristic Region? Plant and Soil 334, 33-46.

Pregitzer, C.C., Bailey, J.K., Hart, S.C., Schweitzer, J.A., 2010. Soils as agents of selection: feedbacks between plants and soils alter seedling survival and performance. Evolutionary Ecology 24, 1045-1059.

Queloz, V., Sieber, T.N., Holdenrieder, O., McDonald, B.A., Grünig, C.R., 2011. No biogeographical pattern for a root-associated fungal species complex. Global Ecology and Biogeography 20, 160-169.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahbek, C., 2005. The role of spatial scale and the perception of large-scale species-richness patterns. Ecology Letters 8, 224-239.

Raimondo, D., Staden, L.v., Foden, W., Victor, J.E., Helme, N.A., Turner, R.C., Kamundi, D.A., Manyama, P.A., 2009. Red List of South African Plants 2009. South African National Biodiversity Institute.

Rambaut, A., Suchard, M., Drummond, A., 2014. Tracer, version 1.6. 0. http://tree.bio.ed.ac.uk/software/tracer/ [accessed 15 May 2017].

Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G., Guisan, A., 2009. Land use improves spatial predictions of mountain plant abundance but not presence-absence. Journal of Vegetation Science 20, 996-1008.

Rausch, T., Wachter, A., 2005. Sulfur metabolism: a versatile platform for launching defence operations. Trends in Plant Science 10, 503-509.

Raven, P.H., Polhill, R.M., 1981. Biogeography of the Leguminosae, in: Polhill R.M., Raven P.H. (Eds.), Advances in Legume Systematics, Part 1. Royal Botanic Gardens, Kew, UK, pp. 27-34.

Reed, S.C., 2017. Disentangling the complexities of how legumes and their symbionts regulate plant nitrogen access and storage. New Phytologist 213, 478-480.

Remigi, P., Zhu, J., Young, J.P.W., Masson-Boivin, C., 2016. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. Trends in Microbiology 24, 63-75.

Requena, N., Perez-Solis, E., Azcon-Aguilar, C., Jeffries, P., Barea, J., 2001. Management of indigenous plant-microbe symbioses aids restoration of desertified ecosystems. Applied and Environmental Microbiology 67, 495-498.

Revell, L.J., 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217-223.

Richards, M.B., Cowling, R.M., Stock, W.D., 1997a. Soil factors and competition as determinants of the distribution of six fynbos Proteaceae species. Oikos, 394-406.

Richards, M.B., Cowling, R.M., Stock, W.D., 1997b. Soil nutrient dynamics and community boundaries in the fynbos vegetation of South Africa. Plant Ecology 130, 143-153.

Rieseberg, L.H., 1997. Hybrid Origins of Plant Species. Annual Review of Ecology and Systematics 28, 359-389.

Rios, N.E., Bart, H.L., 2010. GEOLocate 3.2.2.

Rodríguez-Echeverría, S., Le Roux, J.J., Crisóstomo, J.A., Ndlovu, J., 2011. Jack-of-all-trades and master of many? How does associated rhizobial diversity influence the colonization success of Australian *Acacia* species? Diversity and Distributions 17, 946-957.

Rogel, M.A., Ormeno-Orrillo, E., Martinez-Romero, E., 2011. Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst. Appl. Microbiol. 34, 96–104.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Rotaru, V., Sinclair, T.R., 2009. Interactive influence of phosphorus and iron on nitrogen fixation by soybean. Environmental and Experimental Botany 66, 94-99.

Rout, G.R., Sahoo, S., 2015. Role of iron in plant growth and metabolism. Reviews in Agricultural Science 3, 1-24.

Rout, M.E., Callaway, R.M., 2012. Interactions between exotic invasive plants and soil microbes in the rhizosphere suggest that ‘everything is not everywhere’. Annals of Botany 110, 213-222.

Sánchez-Ramírez, S., Wilson, A.W., Ryberg, M., 2017. Overview of Phylogenetic Approaches to Mycorrhizal Biogeography, Diversity and Evolution, in: Tedersoo L. (Ed.), Biogeography of Mycorrhizal Symbiosis. Springer, pp. 1-37.

Sankhla, I.S., Tak, N., Meghwal, R.R., Choudhary, S., Tak, A., Rathi, S., Sprent, J.I., James, E.K., Gehlot, H.S., 2017. Molecular characterization of nitrogen fixing microsymbionts from root nodules of *Vachellia* (*Acacia*) *jacquemontii*, a native legume from the Thar Desert of India. Plant and Soil 410, 21-40.

Sawana, A., Adeolu, M., Gupta, R.S., 2014. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia gen. nov*. harboring environmental species. Frontiers of Genetics 5, 429.

Schaefer, H., Hechenleitner, P., Santos-Guerra, A., de Sequeira, M.M., Pennington, R.T., Kenicer, G., Carine, M.A., 2012. Systematics, biogeography, and character evolution of the legume tribe

Fabeae with special focus on the middle-Atlantic island lineages. BMC Evolutionary Biology 12, 250.

Schettino, A., Scotese, C.R., 2005. Apparent polar wander paths for the major continents (200 Ma to the present day): a palaeomagnetic reference frame for global plate tectonic reconstructions. Geophysical Journal International 163, 727-759.

Schnitzler, J., Barraclough, T.G., Boatwright, J.S., Goldblatt, P., Manning, J.C., Powell, M.P., Rebelo, T., Savolainen, V., 2011. Causes of Plant Diversification in the Cape Biodiversity Hotspot of South Africa. Systematic Biology 60, 343-357.

Schrire, B.D., Lewis, G.P., Lavin, M., 2005. Biogeography of the Leguminosae, in: Lewis G., Schrire B., MacKinder B., Lock M. (Eds.), Legumes of the World. Royal Botanic Gardens, Kew, pp. 21-54.

Schutte, A.L., 2012. Fabaceae, in: Manning J.C., Goldblatt P. (Eds.), Plants of the Greater Cape Floristic Region 1: The Core Cape Flora. Strelitzia 29, South African National Biodiversity Institute, Pretoria, pp. 518-582.

Schutte, A.L., Vlok, J., Van Wyk, B., 1995. Fire-survival strategy - a character of taxonomic, ecological and evolutionary importance in fynbos legumes. Plant Systematics and Evolution 195, 243-259.

Schutte, A.L., Van Wyk, B., 1998. The tribal position of *Hypocalyptus* Thunberg (Fabaceae). Novon, 178-182.

Shane, M.W., Cramer, M.D., Lambers, H., 2008. Root of edaphically controlled Proteaceae turnover on the Agulhas Plain, South Africa: phosphate uptake regulation and growth. Plant, Cell & Environment 31, 1825-1833.

Shen, C., Xiong, J., Zhang, H., Feng, Y., Lin, X., Li, X., Liang, W., Chu, H., 2013. Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. Soil Biology and Biochemistry 57, 204-211.

Siddig, A.A., Ellison, A.M., Ochs, A., Villar-Leeman, C., Lau, M.K., 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. Ecological Indicators 60, 223-230.

Silva, V.C., Casaes, P.A., Rhem, M.F.K., dos Santos, J.F.M., James, E.K., Gross, E., 2018. Brazilian species of *Calliandra* Benth. (tribe Ingeae) are nodulated by diverse strains of *Paraburkholderia*. Systematic and Applied Microbiology 41, 241-250.

Silva, C., Vinuesa, P., Eguiarte, L.E., Souza, V., Martínez-Romero, E., 2005. Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato,* a widely distributed bacterial symbiont of diverse legumes. Molecular Ecology 14, 4033-4050.

Simmons, M.T., Cowling, R.M., 1996. Why is the Cape Peninsula so rich in plant species? An analysis of the independent diversity components. Biodiversity and Conservation 5, 551-573.

Simon, L., Kiss, F., Bak, A., Hajdu, F., Hrcsik, T.Z., Balogh, Pais, I., 1989. Effect of gallium on photosynthetic pigments and peroxidase activity of *Chlorella pyrenoidosa*. Journal of Plant Nutrition 12, 1123-1140.

Simon, M.F., Proença, C., 2000. Phytogeographic patterns of *Mimosa* (Mimosoideae, Leguminosae) in the Cerrado biome in Brazil: an indicator genus of high-altitude centers of endemism? Biological Conservation 96, 279-296.

Simonsen, A.K., Dinnage, R., Barrett, L.G., Prober, S.M., Thrall, P.H., 2017. Symbiosis limits establishment of legumes outside their native range at a global scale. Nature Communications 8.

Slabbert, E., Jacobs, S.M., Jacobs, K., 2014. The soil bacterial communities of South African fynbos riparian ecosystems invaded by Australian Acacia species. PloS one 9, e86560.

Smith, D.S., Schweitzer, J.A., Turk, P., Bailey, J.K., Hart, S.C., Shuster, S.M., Whitham, T.G., 2012. Soil-mediated local adaptation alters seedling survival and performance. Plant and Soil 352, 243-251.

Soares, M.P., Reys, P., Pifano, D.S., S, J.L.d., Silva, P.O.d., Santos, T.M., Silva, F.G., 2015. Relationship between edaphic factors and vegetation in savannas of the Brazilian Midwest region. Revista Brasileira de Cincia do Solo 39, 821-829.

Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters 10, 1115-1123.

Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G., 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proceedings of the National Academy of Sciences 92, 2647-2651.

Sprent, J.I., Ardley, J.K., James, E.K., 2013. From North to South: a latitudinal look at legume nodulation processes. South African Journal of Botany 89, 31-41.

Sprent, J.I., Geoghegan, I.E., Whitty, P.W., James, E.K., 1996. Natural abundance of 15 N and 13 C in nodulated legumes and other plants in the cerrado and neighbouring regions of Brazil. Oecologia 105, 440-446.

Sprent, J.I., 2009. Legume Nodulation John Wiley & Sons, Chichester.

Sprent, J.I., 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytologist 174, 11-25.

Sprent, J.I., 2001. Nodulation in Legumes Royal Botanic Gardens, Kew.

Sprent, J.I., Ardley, J., James, E.K., 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytologist 215, 40-56.

Sprent, J.I., Gehlot, H.S., 2010. Nodulated legumes in arid and semi-arid environments: are they important? Plant Ecology and Diversity 3, 211-219.

Sprent, J.I., Odee, D.W., Dakora, F.D., 2009. African legumes: a vital but under-utilized resource. Journal of Experimental Botany 61, 1257-1265.

Staden, R., Beal, K., Bonfield, J.K., 1998. The Staden package. Methods in Molecular Biology 132, 115-130.

Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stamatakis A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAXML webservers. Systematic Biology 57, 758-771.

Steenkamp, E.T., van Zyl, E., Beukes, C.W., Avontuur, J.R., Chan, W.Y., Palmer, M., Mthombeni, L.S., Phalane, F.L., Sereme, T.K., Venter, S.N., 2015. *Burkholderia kirstenboschensis sp. nov.*

nodulates papilionoid legumes indigenous to South Africa. Systematic and Applied Microbiology 38, 545-554.

Stępkowski, T., Watkin, E., McInnes, A., Gurda, D., Gracz, J., Steenkamp, E.T., 2012. Distinct *Bradyrhizbium* communities nodulate legumes native to temperate and tropical monsoon Australia. Molecular Phylogenetics and Evolution 63, 265-277.

Stirton, C.H., 1988. The naturalized species of *Lupinus* (Fabaceae) in southern Africa. Bothalia 18, 25-29.

Stirton, C.H., 1982a. Fabaceae. Bothalia 14, 69-77.

Stirton, C.H., 1982b. The genus *Medicago* (Leguminosae) in southern Africa. Bothalia 14, 27-35.

Stirton, C.H., Stajsic, V., Bello, A., 2015. Naturalised species of *Psoralea* (Fabaceae: Psoraleeae) in Australia. Muelleria 33, 97-107.

Stock, W.D., Lewis, O., 1986. Soil nitrogen and the role of fire as a mineralizing agent in a South African coastal fynbos ecosystem. The Journal of Ecology, 317-328.

Stock, W.D., Verboom, G.A., 2012. Phylogenetic ecology of foliar N and P concentrations and N: P ratios across mediterranean-type ecosystems. Global Ecology and Biogeography 21, 1147-1156.

Stopnisek, N., Bodenhausen, N., Frey, B., Fierer, N., Eberl, L., Weisskopf, L., 2014. Genus-wide acid tolerance accounts for the biogeographical distribution of soil *Burkholderia* populations. Environmental Microbiology 16, 1503-1512.

Stropp, J., Ladle, R.J., Malhado, M., Ana, C., Hortal, J., Gaffuri, J., H Temperley, W., Olav Skien, J., Mayaux, P., 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. Global Ecology and Biogeography 25, 1085-1096.

Takhtajan, A., 1986. Floristic Regions of the World (Translated by TJ. Crovello, Edited by A. Cronquist) University of California Press, Berkeley, California.

Talbot, J.M., Bruns, T.D., Taylor, J.W., Smith, D.P., Branco, S., Glassman, S.I., Erlandson, S., Vilgalys, R., Liao, H., Smith, M.E., 2014. Endemism and functional convergence across the

North American soil mycobiome. Proceedings of the National Academy of Sciences 111, 6341-6346.

Tan, H.W., Heenan, P.B., De Meyer, S.E., Willems, A., Andrews, M., 2015. Diverse novel mesorhizobia nodulate New Zealand native *Sophora* species. Systematic and Applied Microbiology 38, 91-98.

Tang, C., Robson, A.D., Dilworth, M.J., 1990. The role of iron in nodulation and nitrogen fixation in *Lupinus angustifolius* L. New Phytologist 114, 173-182.

Telford, R.J., Vandvik, V., Birks, H.J.B., 2006. Dispersal limitations matter for microbial morphospecies. Science 312, 1015.

Ter Braak, C.J., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167-1179.

Ter Braak, C., 1987. CANOCO, a FORTRAN Program for Canonical Community Ordination by Partial Detrended Canonical Correspondence Analysis, Principal Components Analysis and Redundancy Analysis (Version 2.1) TNO Institute of Applied Computer Science, Wageningen.

Thilakarathna, M.S., Raizada, M.N., 2017. A meta-analysis of the effectiveness of diverse rhizobia inoculants on soybean traits under field conditions. Soil Biology and Biochemistry 105, 177-196.

Thuiller, W., Araujo, M.B., Lavorel, S., 2004. Do we need land-cover data to model species distributions in Europe? Journal of Biogeography 31, 353-361.

Thwaites, R.N., Cowling, R.M., 1988. Soil-vegetation relationships on the Agulhas plain, South Africa. Catena 15, 333-345.

Tiffney, B.H., 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. Journal of the Arnold Arboretum 66, 243-273.

Toro, A., 1996. Nodulation competitiveness in the Rhizobium-legume symbiosis. World Journal of Microbiology and Biotechnology 12, 157-162.

Trinder-Smith, H., Cowling, R.M., Linder, H.P., 1996. Profiling a besieged flora: endemic and threatened plants of the Cape Peninsula, South Africa. Biodiversity and Conservation 5, 575-589.

Trinder-Smith, T.H., Kidd, M.M., Anderson, F., 2006. Wild Flowers of the Table Mountain National Park Botanical Society of South Africa, Cape Town.

Trytsman, M., Westfall, R.H., Breytenbach, P.J., Calitz, F.J., van Wyk, A.E., 2016. Diversity and biogeographical patterns of legumes (Leguminosae) indigenous to southern Africa. PhytoKeys 70, 53-96.

Tuohy, J.M., Prior, J.A., Stewart, G.R., 1991. Photosynthesis in relation to leaf nitrogen and phosphorus content in Zimbabwean trees. Oecologia 88, 378-382.

Van der Niet, T., Johnson, S.D., Linder, H.P., 2006. Macroevolutionary data suggest a role for reinforcement in pollination system shifts. Evolution 60, 1596-1601.

Verboom, G.A., Bergh, N.G., Haiden, S.A., Hoffmann, V., Britton, M.N., 2015. Topography as a driver of diversification in the Cape Floristic Region of South Africa. New Phytologist 207, 368-376.

Verboom, G.A., Linder, H.P., Stock, W.D., 2004. Testing the adaptive nature of radiation: growth form and life history divergence in the African grass genus *Ehrharta* (Poaceae: Ehrhartoideae). American Journal of Botany 91, 1364-1370.

Verdú, M., 2000. Ecological and evolutionary differences between Mediterranean seeders and resprouters. Journal of Vegetation Science 11, 265-268.

Vincent, J.M., 1970. A Manual for the Practical Study of the Root-Nodule Bacteria. Blackwell Scientific Publications, Oxford and Edinburgh.

Vitousek, P.M., Porder, S., Houlton, B.Z., Chadwick, O.A., 2010. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. Ecological Applications 20, 5-15.

Wandrag, E., Sheppard, A., Duncan, R., Hulme, P., 2013. Reduced availability of rhizobia limits the performance but not invasiveness of introduced *Acacia*. Journal of Ecology 101, 1103-1113.

Wang, J., Meier, S., Soininen, J., Casamayor, E.O., Pan, F., Tang, X., Yang, X., Zhang, Y., Wu, Q., Zhou, J., 2017. Regional and global elevational patterns of microbial species richness and evenness. Ecography 40, 393-402.

Wang, J., Soininen, J., Zhang, Y., Wang, B., Yang, X., Shen, J., 2011. Contrasting patterns in elevational diversity between microorganisms and macroorganisms. Journal of Biogeography 38, 595-603.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. Journal of Bacteriology 173, 697-703.

Wells, M.J., Balsinhas, A.A., Joffe, H., Engelbrecht, V.M., Harding, G., Stirton, C.H., 1986. A catalogue of problem plants in southern Africa. Memoirs of the Botanical Survey of South Africa 53, 1-658.

Werle, E., Schneider, C., Renner, M., Vlker, M., Fiehn, W., 1994. Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Research 22, 4354.

Werner, G.D., Cornwell, W.K., Cornelissen, J.H., Kiers, E.T., 2015. Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. Proceedings of the National Academy of Sciences 112, 10262-10269.

Werner, G.D., Cornwell, W.K., Sprent, J.I., Kattge, J., Kiers, E.T., 2014. A single evolutionary innovation drives the deep evolution of symbiotic $N_2$-fixation in angiosperms. Nature Communications 5, 4087.

Whitaker, R.J., Grogan, D.W., Taylor, J.W., 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301, 976-978.

Whittaker, R.J., Willis, K.J., Field, R., 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28, 453-470.

Wiedenbeck, J., Cohan, F.M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiology Reviews 35, 957-976.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J., Guisan, A., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews 88, 15-30.

Witkowski, E., Mitchell, D.T., 1987. Variations in soil phosphorus in the Fynbos biome, South Africa. The Journal of Ecology, 1159-1171.

Wojciechowski, M.F., Lavin, M., Sanderson, M.J., 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. American Journal of Botany 91, 1846-1862.

Yang, J., Smith, H.G., Sherratt, T.N., Wilkinson, D.M., 2010. Is there a size limit for cosmopolitan distribution in free-living microorganisms? A biogeographical analysis of testate amoebae from polar areas. Microbial Ecology 59, 635-645.

Yao, Y., Sui, X.H., Zhang, X.X., Wang, E.T., Chen, W.X., 2017. *Bradyrhizobium ferriligni sp. nov*., isolated from effective nodules of *Erythrophleum fordii*. Int. J. Syst. Evol. Microbiol. 65, 1831–1837.

Yates, C.J., Elith, J., Latimer, A.M., Le Maitre, D., Midgley, G.F., Schurr, F.M., West, A.G., 2010. Projecting climate change impacts on species distributions in megadiverse South African Cape and Southwest Australian Floristic Regions: opportunities and challenges. Austral Ecology 35, 374-391.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Molecular Phylogenetics and Evolution 87, 46-49.

Zahran, H.H., 2017. Plasmids impact on rhizobia-legumes symbiosis in diverse environments. Symbiosis, 1-17.

Zahran, H.H., 1999. *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiology and Molecular Biology Reviews 63, 968-989.

Zavala, L.M., García-Moreno, J., Gordillo-Rivero, Á.J., Jordn, A., Mataix-Solera, J., 2014. Natural soil water repellency in different types of Mediterranean woodlands. Geoderma 226, 170-178.

Zhang, J.J., Lou, K., Jin, X., Mao, P.H., Wang, E.T., Tian, C.F., Sui, X.H., Chen, W.F., Chen, W.X., 2012. Distinctive *Mesorhizobium* populations associated with *Cicer arietinum* L. in alkaline soils of Xinjiang, China. Plant and Soil 353, 123-134.

Zhao, L., Fan, M., Zhang, D., Yang, R., Zhang, F., Xu, L., Wei, X., Shen, Y., Wei, G., 2014. Distribution and diversity of rhizobia associated with wild soybean (*Glycine soja* Sieb. & Zucc.) in Northwest China. Systematic and Applied Microbiology 37, 449-456.

Zheng, J.Z., Wang, R., Liu, R.R., Chen, J.J., Wei, Q., Wu, X.Y., Pang, X.W., James, E.K., Liu, X.Y., 2017. The structure and evolution of beta-rhizobial symbiotic genes deduced from their complete genomes. Immunome Research 13 (2).

# Supplementary materials

## Appendix A

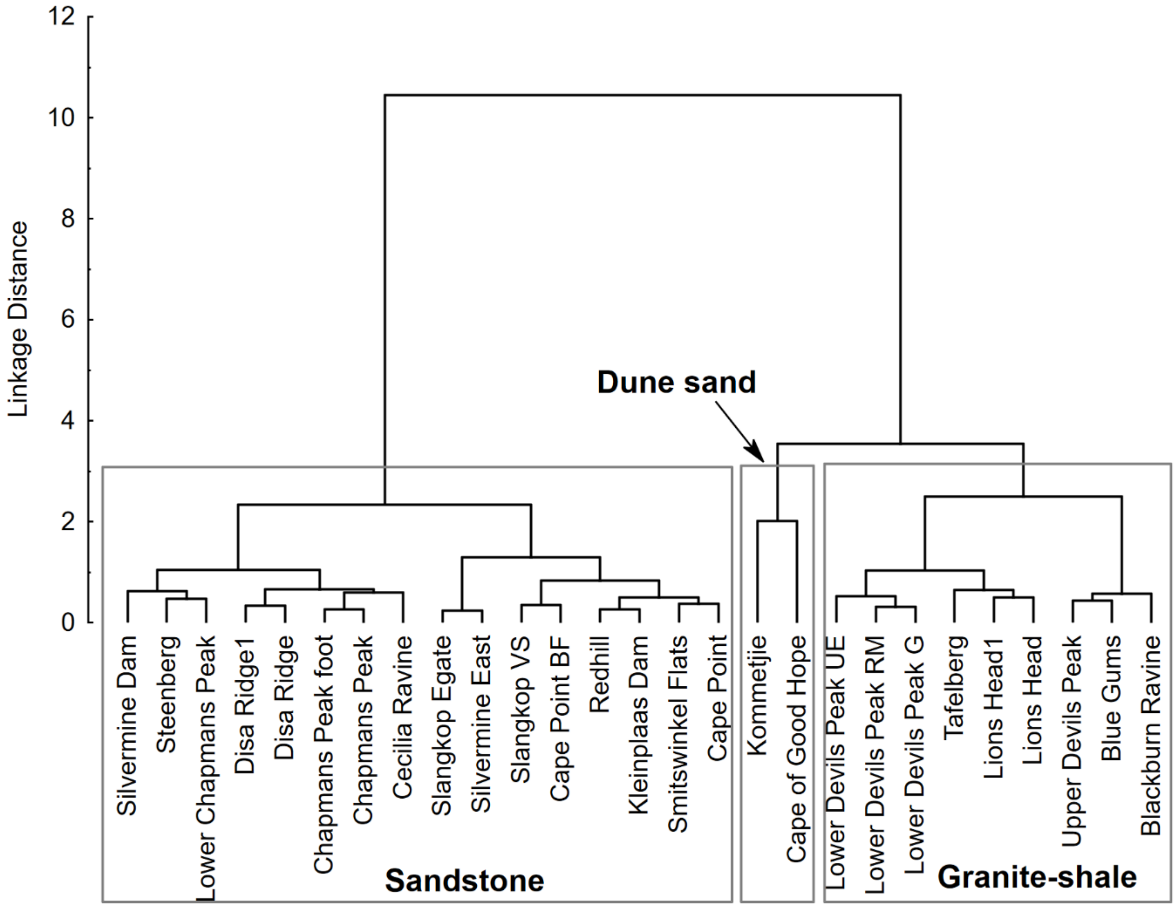
**Appendix A1.** List of legume species collected from the 27 sites studied. Geological types (hereafter referred to as soil types) of the sites are indicated in parentheses.

| Site | Species | Tribe |
|---|---|---|
| Lions Head<br>(Granite) | *Aspalathus chenopoda* L.<br>*Aspalathus cordata* (L.) R.Dahlgren<br>*Aspalathus ericifolia* P.J.Bergius<br>*Aspalathus laricifolia* Lam.<br>*Indigofera candolleana* Meisn.<br>*Indigofera incana* Thunb.<br>*Indigofera psoraloides* L.<br>*Lebeckia plukenetiana* E.Mey.<br>*Lessertia capensis* Druce<br>*Lotus angustissimus* L.<br>*Otholobium hirtum* (L.) C.H.Stirt.<br>*Psoralea asarina* (P.J.Bergius) T.M.Salter<br>*Podalyria sericea* R.Br. | Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Indigofereae<br>Indigofereae<br>Indigofereae<br>Crotalarieae<br>Galegeae<br>Loteae<br>Psoraleeae<br>Psoraleeae<br>Podalyrieae |
| Lions Head 1<br>(Granite) | *Aspalathus chenopoda* L.<br>*Aspalathus arida* E.Mey.<br>*Aspalathus cordata* (L.) R.Dahlgren<br>*Dipogon lignosus* (L.) Verdc.<br>*Indigofera psoraloides* L.<br>*Otholobium hirtum* (L.) C.H.Stirt.<br>*Otholobium virgatum* (Burm.f) C.H.Stirt.<br>*Psoralea pinnata* L.<br>*Psoralea asarina* (P.J.Bergius) T.M.Salter<br>*Rafnia triflora* Thunb.<br>*Virgilia oroboides* (P.J.Bergius) T.M.Salter | Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Phaseoleae<br>Indigofereae<br>Psoraleeae<br>Psoraleeae<br>Psoraleeae<br>Psoraleeae<br>Crotalarieae<br>Podalyrieae |
| Silvermine Dam<br>(Sandstone) | *Bolusafra bituminosa* Kuntze<br>*Cyclopia genistoides* Vent.<br>*Liparia splendens* (Burm.f.) Bos & de Wit<br>*Podalyria calyptrata* Willd.<br>*Psoralea pinnata* L.<br>*Psoralea aculeata* Thunb. *ex* Harv. & Sond.<br>*Virgilia oroboides* (P.J.Bergius) T.M.Salter | Phaseoleae<br>Podalyrieae<br>Podalyrieae<br>Podalyrieae<br>Psoraleeae<br>Psoraleeae<br>Podalyrieae |
| Disa Ridge<br>(Sandstone) | *Cyclopia genistoides* Vent.<br>*Indigofera glomerata* E.Mey.<br>*Indigofera sarmentosa* Herb.Holm. *ex* Harv. & Sond.<br>*Indigofera candolleana* Meisn.<br>*Liparia splendens* (Burm.f.) Bos & de Wit<br>*Psoralea aculeata* Thunb. *ex* Harv. & Sond.<br>*Psoralea pinnata* L.<br>*Psoralea congesta* C.H.Stirt. & Muasya<br>*Xiphotheca fruticosa* (L.) A.L.Schutte & B.-E.van Wyk | Podalyrieae<br>Indigofereae<br>Indigofereae<br>Indigofereae<br>Podalyrieae<br>Psoraleeae<br>Psoraleeae<br>Psoraleeae<br>Podalyrieae |
| Disa Ridge 1<br>(Sandstone) | *Aspalathus capitata* L.<br>*Cyclopia genistoides* Vent.<br>*Indigofera candolleana* Meisn.<br>*Indigofera sarmentosa* Herb.Holm. *ex* Harv. & Sond.<br>*Liparia splendens* (Burm.f.) Bos & de Wit<br>*Psoralea pinnata* L.<br>*Psoralea aculeata* Thunb. *ex* Harv. & Sond.<br>*Xiphotheca fruticosa* (L.) A.L.Schutte & B.-E.van Wyk | Crotalarieae<br>Podalyrieae<br>Indigofereae<br>Indigofereae<br>Podalyrieae<br>Psoraleeae<br>Psoraleeae<br>Podalyrieae |
| Upper Devils Peak<br>(Shale) | *Argyrolobium lunare* Druce<br>*Bolusafra bituminosa* Kuntze | Genisteae<br>Phaseoleae |

| Site | Species | Tribe |
|---|---|---|
| | *Indigofera filiformis* Thunb. | Indigofereae |
| | *Lotononis umbellata* Benth. | Crotalarieae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Psoralea pinnata* L. | Psoraleeae |
| | *Psoralea imbricata* (L.) T.M.Salter | Psoraleeae |
| | *Rhynchosia capensis* Schinz | Phaseoleae |
| | *Virgilia oroboides* (P.J.Bergius) T.M.Salter | Podalyrieae |
| Tafelberg<br>(Shale) | *Aspalathus chenopoda* L. | Crotalarieae |
| | *Aspalathus cordata* (L.) R.Dahlgren | Crotalarieae |
| | *Aspalathus ericifolia* P.J.Bergius | Crotalarieae |
| | *Aspalathus hispida* ssp. *albiflora* (Eckl. & Zeyh.) R.Dahlgren | Crotalarieae |
| | *Indigofera psoraloides* L. | Indigofereae |
| | *Otholobium hirtum* (L) C.H.Stirt. | Psoraleeae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Otholobium uncinatum* (Eckl. & Zeyh.) C.H.Stirt. | Psoraleeae |
| | *Psoralea asarina* (P.J.Bergius) T.M.Salter | Psoraleeae |
| | *Psoralea imbricata* (L.) T.M.Salter | Psoraleeae |
| | *Psoralea aphylla* L. | Psoraleeae |
| | *Sutherlandia frutescens* (L.) R.Br. ex W.T.Aiton | Galegeae |
| | *Virgilia oroboides* (P.J.Bergius) T.M.Salter | Podalyrieae |
| Chapmans Peak<br>(Sandstone) | *Aspalathus barbata* (Lam.) R.Dahlgren | Crotalarieae |
| | *Indigofera filiformis* Thunb. | Indigofereae |
| | *Indigofera mauritanica* Thunb. | Indigofereae |
| | *Otholobium fruticans* (L.) C.H.Stirt. | Psoraleeae |
| | *Rhynchosia capensis* Schinz | Phaseoleae |
| | *Xiphotheca reflexa* (Thunb.) A.L.Schutte & B.-E.van Wyk | Podalyrieae |
| Lower Chapmans Peak<br>(Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus divaricata* Thunb. | Crotalarieae |
| | *Aspalathus arida* E.Mey. | Crotalarieae |
| | *Aspalathus ericifolia* P.J.Bergius | Crotalarieae |
| | *Indigofera mauritanica* Thunb. | Indigofereae |
| | *Indigofera filiformis* Thunb. | Indigofereae |
| | *Otholobium fruticans* (L.) C.H.Stirt. | Psoraleeae |
| | *Rhynchosia capensis* Schinz | Phaseoleae |
| Chapmans Peak foot<br>(Sandstone) | *Indigofera filiformis* Thunb. | Indigofereae |
| | *Indigofera mauritanica* Thunb. | Indigofereae |
| | *Liparia splendens* (Burm.f.) Bos & de Wit | Podalyrieae |
| | *Otholobium fruticans* (L.) C.H.Stirt. | Psoraleeae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| Lower Devils Peak G<br>(Shale) | *Aspalathus cephalotes* Thunb. | Crotalarieae |
| | *Aspalathus chenopoda* L. | Crotalarieae |
| | *Aspalathus hispida* ssp. *albiflora* (Eckl. & Zeyh.) R.Dahlgren | Crotalarieae |
| | *Aspalathus ericifolia* P.J.Bergius | Crotalarieae |
| | *Aspalathus variegata* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus cordata* (L.) R.Dahlgren | Crotalarieae |
| | *Indigofera psoraloides* L. | Indigofereae |
| | *Otholobium uncinatum* (Eckl. & Zeyh.) C.H.Stirt. | Psoraleeae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Otholobium hirtum* (L.) C.H.Stirt. | Psoraleeae |
| | *Psoralea asarina* (P.J.Bergius) T.M.Salter | Psoraleeae |
| Lower Devils Peak UE<br>(Shale) | *Aspalathus cordata* (L.) R.Dahlgren | Crotalarieae |
| | *Aspalathus biflora* E.Mey. | Crotalarieae |
| | *Aspalathus cephalotes* Thunb. | Crotalarieae |
| | *Aspalathus chenopoda* L. | Crotalarieae |
| | *Aspalathus crenata* ( L. ) R.Dahlgren | Crotalarieae |
| | *Aspalathus cymbiformis* DC. | Crotalarieae |
| | *Aspalathus hispida* ssp. *albiflora* (Eckl. & Zeyh.) R.Dahlgren | Crotalarieae |
| | *Indigofera psoraloides* L. | Indigofereae |
| | *Otholobium hirtum* (L.) C.H.Stirt. | Psoraleeae |
| | *Otholobium uncinatum* (Eckl. & Zeyh.) C.H.Stirt. | Psoraleeae |
| | *Psoralea asarina* (P.J.Bergius) T.M.Salter | Psoraleeae |

| Site | Species | Tribe |
|---|---|---|
| | *Psoralea pinnata* L. | Psoraleeae |
| | *Tephrosia capensis* Pers. | Milletieae |
| Lower Devils Peak RM (Shale) | *Argyrolobium lunare* Druce | Genisteae |
| | *Aspalathus hispida* ssp. *albiflora* (Eckl. & Zeyh.) R.Dahlgren | Crotalarieae |
| | *Aspalathus chenopoda* L. | Crotalarieae |
| | *Aspalathus cordata* (L.) R.Dahlgren | Crotalarieae |
| | *Aspalathus ericifolia* P.J.Bergius | Crotalarieae |
| | *Aspalathus cephalotes* Thunb. | Crotalarieae |
| | *Aspalathus macrantha* Harv. | Crotalarieae |
| | *Bolusafra bituminosa* Kuntze | Phaseoleae |
| | *Indigofera psoraloides* L. | Indigofereae |
| | *Lotus angustissimus* L. | Loteae |
| | *Otholobium hirtum* (L.) C.H.Stirt. | Psoraleeae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Psoralea asarina* (P.J.Bergius) T.M.Salter | Psoraleeae |
| Silvermine East (Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus carnosa* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus capensis* (Walp.) R.Dahlgren | Crotalarieae |
| | *Aspalathus ericifolia* P.J.Bergius | Crotalarieae |
| | *Aspalathus callosa* L. | Crotalarieae |
| | *Cyclopia galioides* DC. | Podalyrieae |
| | *Indigofera glomerata* E.Mey. | Indigofereae |
| | *Liparia parva* Vogel ex Walp. | Podalyrieae |
| | *Liparia splendens* (Burm.f.) Bos & de Wit | Podalyrieae |
| | *Podalyria calyptrata* Willd. | Podalyrieae |
| | *Psoralea pinnata* L. | Psoraleeae |
| Steenberg (Sandstone) | *Aspalathus juniperina* Thunb. | Crotalarieae |
| | *Aspalathus cymbiformis* DC. | Crotalarieae |
| | *Aspalathus callosa* L. | Crotalarieae |
| | *Indigofera glomerata* E.Mey. | Indigofereae |
| | *Indigofera candolleana* Meisn. | Indigofereae |
| | *Rafnia capensis* (L.) Schinz | Crotalarieae |
| Redhill (Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus carnosa* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus capensis* (Walp.) R.Dahlgren | Crotalarieae |
| | *Indigofera glomerata* E.Mey. | Indigofereae |
| | *Liparia parva* Vogel ex Walp. | Podalyrieae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Otholobium fruticans* (L.) C.H.Stirt. | Psoraleeae |
| | *Podalyria biflora* Sims | Podalyrieae |
| | *Psoralea pinnata* L. | Psoraleeae |
| Kleinplaas Dam (Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus capensis* (Walp.) R.Dahlgren | Crotalarieae |
| | *Aspalathus cymbiformis* DC. | Crotalarieae |
| | *Aspalathus carnosa* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus chenopoda* L. | Crotalarieae |
| | *Indigofera glomerata* E.Mey. | Indigofereae |
| | *Liparia parva* Vogel ex Walp. | Podalyrieae |
| | *Otholobium virgatum* (Burm.f) C.H.Stirt. | Psoraleeae |
| | *Podalyria biflora* Sims | Podalyrieae |
| | *Psoralea pinnata* L. | Psoraleeae |
| Cape Point (Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus juniperina* Thunb. | Crotalarieae |
| | *Aspalathus carnosa* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus capensis* (Walp.) R.Dahlgren | Crotalarieae |
| | *Aspalathus callosa* L. | Crotalarieae |
| | *Bolusafra bituminosa* Kuntze | Phaseoleae |
| Cape Point BF (Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh. | Podalyrieae |
| | *Aspalathus capensis* (Walp.) R.Dahlgren | Crotalarieae |
| | *Aspalathus carnosa* Eckl. & Zeyh. | Crotalarieae |
| | *Aspalathus hispida* Thunb. subsp. *hispida* | Crotalarieae |

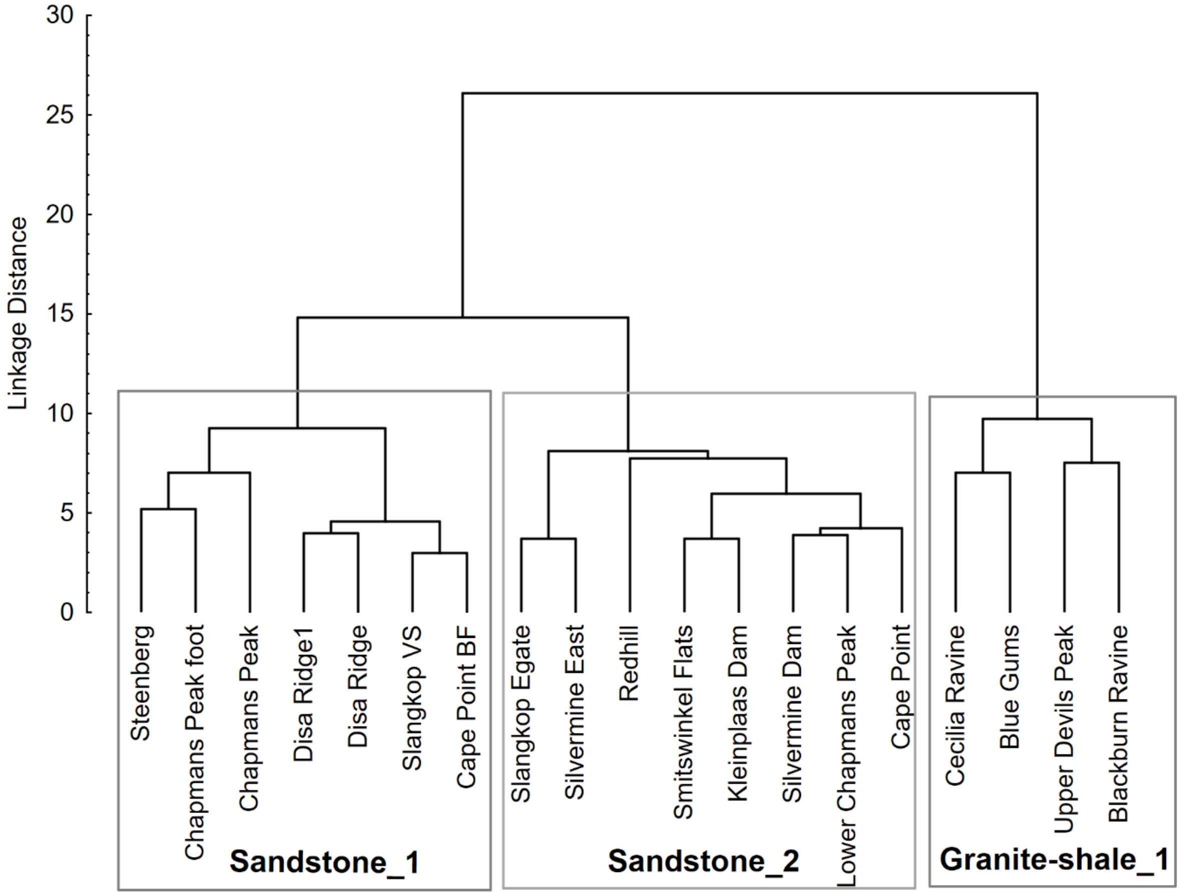
| **Site** | **Species** | **Tribe** |
|---|---|---|
| | *Psoralea pinnata* L.<br>*Psoralea aphylla* L. | Psoraleeae<br>Psoraleeae |
| Cape of Good Hope<br>(Dune sand) | *Otholobium bracteolatum* (Eckl. & Zeyh.) C.H.Stirt.<br>*Psoralea repens* P.J.Bergius | Psoraleeae<br>Psoraleeae |
| Blackburn Ravine<br>(Granite) | *Argyrolobium lunare* Druce<br>*Aspalathus divaricata* Thunb.<br>*Aspalathus ericifolia* P.J.Bergius<br>*Bolusafra bituminosa* Kuntze<br>*Indigofera candolleana* Meisn.<br>*Indigofera filiformis* Thunb.<br>*Indigofera frutescens* L.f.<br>*Podalyria calyptrata* Willd.<br>*Xiphotheca fruticosa* (L.) A.L.Schutte & B.-E.van Wyk | Genisteae<br>Crotalarieae<br>Crotalarieae<br>Phaseoleae<br>Indigofereae<br>Indigofereae<br>Indigofereae<br>Podalyrieae<br>Podalyrieae |
| Cecilia Ravine<br>(Sandstone) | *Aspalathus divaricata* Thunb.<br>*Dipogon lignosus* (L.) Verdc.<br>*Indigofera frutescens* L.f.<br>*Liparia splendens* (Burm.f.) Bos & de Wit | Crotalarieae<br>Phaseoleae<br>Indigofereae<br>Podalyrieae |
| Blue Gums<br>(Granite) | *Argyrolobium lunare* Druce<br>*Aspalathus macrantha* Harv.<br>*Bolusafra bituminosa* Kuntze<br>*Indigofera filiformis* Thunb.<br>*Indigofera frutescens* L.f.<br>*Otholobium virgatum* (Burm.f) C.H.Stirt.<br>*Otholobium spicatum* (L.) C.H.Stirt.<br>*Psoralea monophylla* (L.) C.H.Stirt.<br>*Psoralea pinnata* L.<br>*Rhynchosia capensis* Schinz | Genisteae<br>Crotalarieae<br>Phaseoleae<br>Indigofereae<br>Indigofereae<br>Psoraleeae<br>Psoraleeae<br>Psoraleeae<br>Psoraleeae<br>Phaseoleae |
| Slangkop E Gate<br>(Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh.<br>*Aspalathus carnosa* Eckl. & Zeyh.<br>*Aspalathus capensis* (Walp.) R.Dahlgren<br>*Aspalathus hispida* Thunb. subsp. *hispida*<br>*Aspalathus juniperina* Thunb.<br>*Aspalathus chenopoda* L.<br>*Bolusafra bituminosa* Kuntze<br>*Otholobium virgatum* (Burm.f) C.H.Stirt.<br>*Podalyria sericea* R.Br. | Podalyrieae<br>Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Crotalarieae<br>Phaseoleae<br>Psoraleeae<br>Podalyrieae |
| Slangkop VS<br>(Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh.<br>*Aspalathus carnosa* Eckl. & Zeyh.<br>*Aspalathus capensis* (Walp.) R.Dahlgren<br>*Indigofera brachystachya* (DC.) E.Mey.<br>*Lebeckia plukenetiana* E.Mey. | Podalyrieae<br>Crotalarieae<br>Crotalarieae<br>Indigofereae<br>Crotalarieae |
| Kommetjie<br>(Dune sand) | *Wiborgia obcordata* Thunb. | Crotalarieae |
| Smitswinkel Flats<br>(Sandstone) | *Amphithalea ericifolia* (L.) Eckl. & Zeyh.<br>*Aspalathus capensis* (Walp.) R.Dahlgren<br>*Aspalathus carnosa* Eckl. & Zeyh.<br>*Otholobium virgatum* (Burm.f) C.H.Stirt.<br>*Psoralea pinnata* L. | Podalyrieae<br>Crotalarieae<br>Crotalarieae<br>Psoraleeae<br>Psoraleeae |



**Appendix A2.** Dendrogram based on a cluster analysis of the overall soil chemical and physical characteristics of the 27 sites. Groups are demarcated by the rectangles and labelled according to their corresponding geology.

**Appendix A3.** Standardised coefficients of the soil variables on the first and second roots of the DFA performed for the site groups recovered from cluster analysis of all 27 sites.

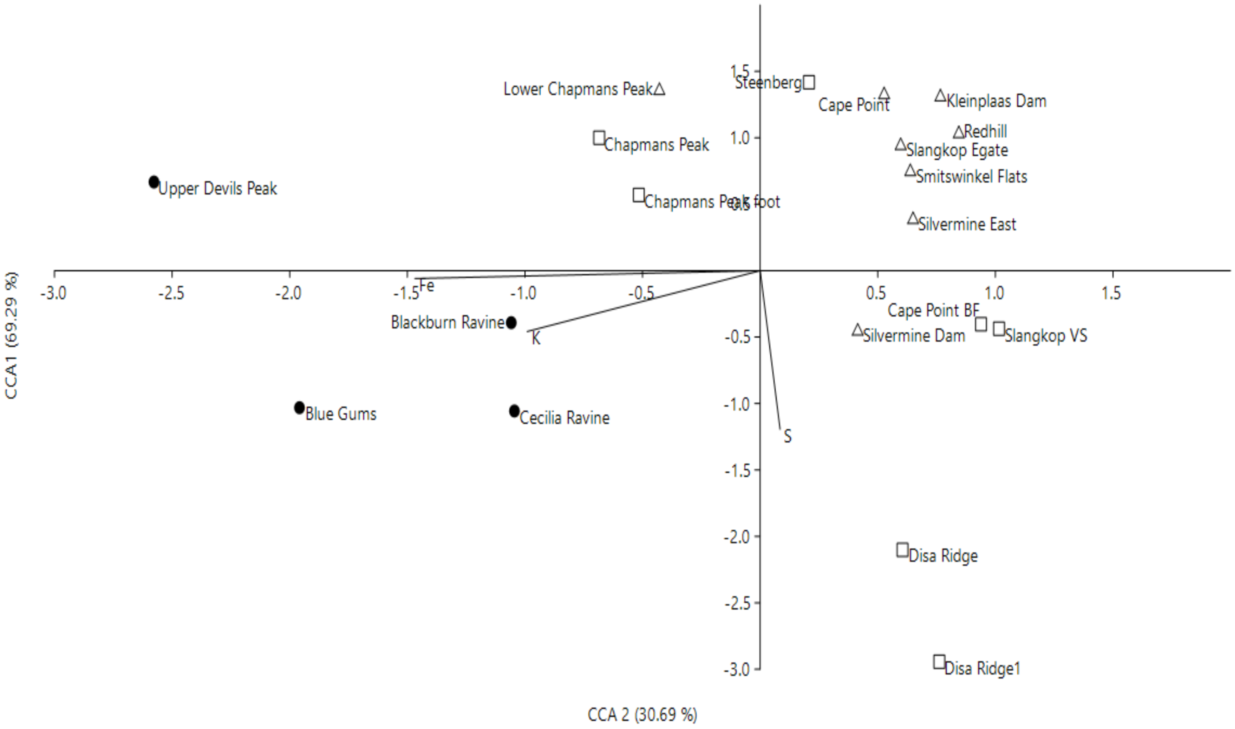
| **Variable** | **Root 1** | **Root 2** |
|---|---|---|
| Al | -0.86261 | 0.393315 |
| P | -0.32960 | -0.638200 |
| Mn | 0.47302 | -0.295012 |
| Fe | 1.07594 | -0.435907 |
| P Bray II | -0.65352 | -0.429609 |
| Na | -1.10757 | -0.976477 |
| K | 0.81209 | -0.309421 |
| N | 0.53305 | 0.257540 |
| C | -0.24291 | 0.648931 |
| Sand | -0.68139 | -0.041753 |
| Eigenvalue | 12.63005 | 5.650721 |
| Cumulative Probability | 0.69089 | 1.000000 |



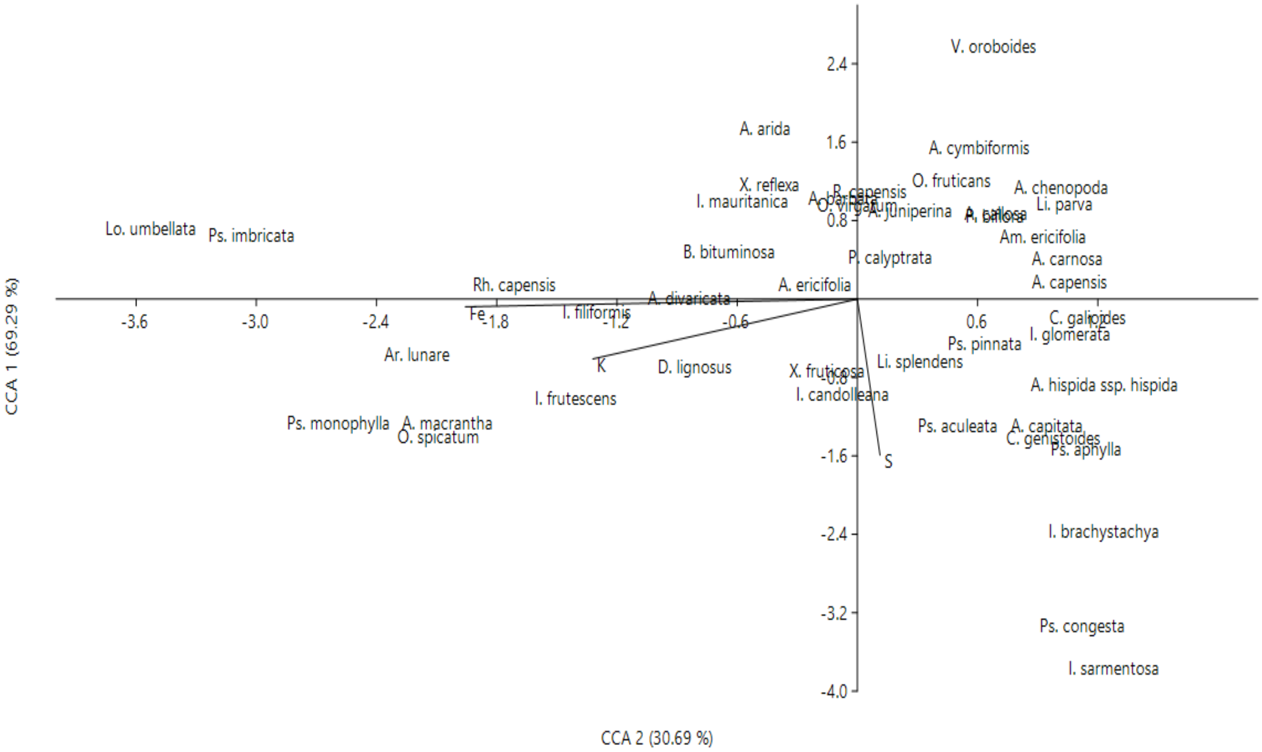
**Appendix A4.** Dendrogram based on a cluster analysis of the overall soil chemical and physical characteristics of the subset of the data comprising three granite-shale sites and all sandstone sites. Groups are demarcated by the rectangles and labelled according to their corresponding geology.

**Appendix A5.** Standardised coefficients of the soil variables on the first and second roots of the DFA performed on sandstone and three granite-shale sites only.

| **Variable** | **Root 1** | **Root 2** |
|---|---|---|
| Fe | 1.393437 | 0.356737 |
| P Bray II | 0.617727 | 0.381330 |
| pH | -0.546128 | -0.115199 |
| N | 0.089955 | 0.612179 |
| C | 0.775673 | -0.868729 |
| K | -0.627372 | -0.694054 |
| Eigenvalue | 4.204756 | 0.768651 |
| Cumulative Probability | 0.845448 | 1.000000 |



**Appendix A6.** (a) Species-conditional biplot based on canonical correspondence analysis of soil characteristics and legume species composition data (comprising the three granite-shale sites and all sandstone sites), showing the ordination of legume sites along the first two CCA axes. Open squares represent Sandstone_1 sites, open triangles represent Sandstone_2 sites, while the closed circles represent Granite-shale_1 sites. Soil characteristics are indicated by lines whose lengths indicate the strength of their correlations with the ordination axes.



**Appendix A6.** (b) Species-conditional biplot based on canonical correspondence analysis of soil characteristics and legume species composition data (comprising the three granite-shale sites and all sandstone sites), showing the ordination of legume species. Soil characteristics are indicated by lines, whose lengths indicate the strength of their correlations with the ordination axes. Genera are abbreviated as in Figure 2.4b.

# Appendix B

**Appendix B1.** BLASTn search results for the various rhizobial strains isolated as part of this study. Percentage similarity to their closest sequences on GenBank are shown.

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| 16S rRNA | RM581 | *Argyrolobium lunare* | *Mesorhizobium sp. CCBAU 33406 (JQ339800.1)* | 95 |
| | RM600 | *Argyrolobium lunare* | *Mesorhizobium sp. strain WC23.1d (MG050062.1)* | 99 |
| | RM98 | *Argyrolobium lunare* | *Mesorhizobium sp. strain WC23.1d (MG050062.1)* | 100 |
| | SE571 | *Aspalathus argyrella* | *Burkholderia sp. WSM4177 (HE862275.1)* | 97 |
| | SE572 | *Aspalathus argyrella* | *Burkholderia sp. HC6.4b (HF674718.1)* | 99 |
| | CPT36 | *Aspalathus capensis* | *Burkholderia sp. strain WC21.1b (MG050053.1)* | 97 |
| | SW5 | *Aspalathus carnosa* | *Burkholderia sp. WK1.1f (HF674688.1)* | 99 |
| | BR621 | *Aspalathus chenopoda* | *Mesorhizobium sp. strain WC23.1d (MG050062.1)* | 98 |
| | BR624 | *Aspalathus chenopoda* | *Mesorhizobium sp. strain WC23.1d (MG050062.1)* | 96 |
| | RM552 | *Aspalathus cordata* | *Mesorhizobium loti NZP2037 (CP016079.1)* | 99 |
| | SE551 | *Aspalathus cordata* | *Burkholderia sp. BL25 (KR154605.1)* | 97 |
| | SE563 | *Aspalathus cordata* | *Mesorhizobium sp. CCBAU 33404 (JQ339778.1)* | 99 |
| | BR604 | *Aspalathus ericifolia* | *Mesorhizobium sp. CCBAU 11214 ( EF149002.1)* | 100 |
| | BR605 | *Aspalathus ericifolia* | *Mesorhizobium loti NZP2037 (CP016079.1)* | 99 |
| | CP627 | *Aspalathus ericifolia* | *Mesorhizobium sp. RPJ3 (FJ985026.1)* | 97 |
| | CP630 | *Aspalathus ericifolia* | *Mesorhizobium sp. strain WC23.1d (MG050062.1)* | 99 |
| | LH6 | *Dipogon lignosus* | *Burkholderia sp. WSM4178 (HE862279.1)* | 99 |
| | SE555 | *Indigofera filifolia* | *Burkholderia sp. CS13775 (KF791624.1)* | 97 |
| | SE557 | *Indigofera filifolia* | *Burkholderia sp. CS 2 (AY178065.1)* | 99 |
| | VG182 | *Indigofera filifolia* | *Burkholderia sp. BL16 R3 (KR154595.1)* | 99 |
| | VG185 | *Indigofera filifolia* | *Burkholderia sp. BL16 R3 (KR154595.1)* | 99 |
| | BR62 | *Indigofera filiformis* | *Burkholderia sp. RAU2i (HF674680.1)* | 97 |
| | BR542 | *Indigofera frutescens* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | BR543 | *Indigofera frutescens* | *Burkholderia sp. BL6665C (KR154611.1)* | 98 |
| | BR544 | *Indigofera frutescens* | *Burkholderia sp. BL6665C (KR154611.1)* | 98 |
| | BR545 | *Indigofera frutescens* | *Burkholderia sp. BL25 (KR154605.1)* | 97 |

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| | BR67 | *Indigofera frutescens* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | BR68 | *Indigofera frutescens* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | BR69 | *Indigofera frutescens* | *Burkholderia sp. RAU2i (HF674680.1)* | 98 |
| | CP589 | *Indigofera frutescens* | *Burkholderia sp. WSM4177 (HE862275.1)* | 98 |
| | CP591 | *Indigofera frutescens* | *Burkholderia sp. UCT 31 (AY178074.1)* | 98 |
| | BR537 | *Indigofera mauritanica* | *Burkholderia sp. RAU2j (HF674681.1)* | 98 |
| | SE573 | *Indigofera sarmentosa* | *Burkholderia sp. Cpub 6 (AY178071.1)* | 99 |
| | SM52 | *Indigofera sarmentosa* | *Burkholderia sp. WSM4177 (HE862275.1)* | 98 |
| | BR103 | *Indigofera sp.* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | PK221 | *lebeckia ambigua* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | PK225 | *lebeckia ambigua* | *Burkholderia sp. BL25 (KR154605.1)* | 100 |
| | RH601 | *Lebeckia sp.* | *Burkholderia sp. HC6.4b (HF674718.1)* | 99 |
| | RH602 | *Lebeckia sp.* | *Burkholderia sp. Cpub 6 (AY178071.1)* | 99 |
| | RH625 | *Lebeckia sp.* | *Burkholderia sp. WSM4177 (HE862275.1)* | 99 |
| | SE574 | *Liparia splendens* | *Burkholderia sp. HC6.4b (HF674718.1)* | 99 |
| | BR619 | *Otholobium fruticans* | *Mesorhizobium sp. MM5377 ( KF802585.1)* | 99 |
| | LH610 | *Otholobium hirtum* | *Mesorhizobium cantuariense strain ICMP 19515 (NR_137373.1)* | 99 |
| | LH611 | *Otholobium hirtum* | *Mesorhizobium ciceri strain VAR5.8 ( KY515337.1)* | 100 |
| | LH612 | *Otholobium hirtum* | *Mesorhizobium ciceri strain VAR5.8 ( KY515337.1)* | 100 |
| | RM579 | *Otholobium hirtum* | *Mesorhizobium sp. strain WC33b* | 98 |
| | RH616 | *Podalyria argentea* | *Burkholderia sp. WSM4177 (HE862275.1)* | 99 |
| | JH160 | *Podalyria calyptrata* | *Burkholderia sp. BL16 R3 (KR154595.1)* | 99 |
| | KM195 | *Podalyria calyptrata* | *Burkholderia sp. RAU2j (HF674681.1)* | 99 |
| | RM173 | *Podalyria calyptrata* | *Burkholderia sp. BL6665C (KR154611.1)* | 99 |
| | RM175 | *Podalyria calyptrata* | *Burkholderia sp. WK1.1f (HF674688.1)* | 99 |
| | VG193 | *Podalyria calyptrata* | *Burkholderia sp. BL16 R3 (KR154595.1)* | 99 |
| | VG204 | *Podalyria calyptrata* | *Burkholderia sp. RAU2i (HF674680.1)* | 99 |
| | VG206 | *Podalyria calyptrata* | *Burkholderia sp. BL16 R3 (KR154595.1)* | 97 |
| | WB164 | *Podalyria calyptrata* | *Burkholderia sp. RAU2f (HF674677.2)* | 99 |
| | WB168 | *Podalyria calyptrata* | *Burkholderia sp. BL29 I6R2 (KR154610.1)* | 99 |
| | SE584 | *Podalyria sericea* | *Burkholderia sp. HC6.4b (HF674718.1)* | 99 |

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| | SE596 | *Psoralea aphylla* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 99 |
| | SE597 | *Psoralea aphylla* | *Mesorhizobium sp. ORS1080 (AJ295082.1)* | 99 |
| | RM548 | *Psoralea asarina* | *Mesorhizobium sp. CCBAU 41174 (KP116972.1)* | 99 |
| | BK312 | *Psoralea fleta* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 99 |
| | BK315 | *Psoralea fleta* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 99 |
| | RH607 | *Psoralea pinnata* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 98 |
| | RM569 | *Psoralea pinnata* | *Mesorhizobium sp. CCBAU 41174 (KP116972.1)* | 99 |
| | RS176 | *Psoralea pinnata* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 100 |
| | RS178 | *Psoralea pinnata* | *Mesorhizobium sp. strain WSM4692 (MF949006.1)* | 100 |
| | SE592 | *Psoralea pinnata* | *Mesorhizobium sp. ORS1080 (AJ295082.1)* | 100 |
| | BK308 | *Psoralea usitata* | *Mesorhizobium loti NZP2037 (CP016079.1)* | 100 |
| | LH22 | *Virgilia oroboides* | *Burkholderia kirstenboschensis HC1.1bc (HF674714.1)* | 99 |
| | LH23 | *Virgilia oroboides* | *Burkholderia kirstenboschensis HC1.1bc (HF674714.1)* | 99 |
| | SE539 | *Virgilia oroboides* | *Burkholderia sp. CB2 (AY178059.1)* | 99 |
| | SE540 | *Virgilia oroboides* | *Burkholderia sp. CB2 (AY178059.1)* | 98 |
| | | | | |
| *nodA* | SE571 | *Aspalathus argyrella* | *Burkholderia tuberum strain DUS833 (EF566976.1)* | 99 |
| | CPT36 | *Aspalathus capensis* | *Burkholderia tuberum strain DUS833 (EF566976.1)* | 99 |
| | LH6 | *Dipogon lignosus* | *Burkholderia rhynchosiae strain WSM3930 (EU219866.1)* | 100 |
| | LH4 | *Indigofera candolleana* | *Paraburkholderia sprentiae WSM5005 (CP017562.1)* | 98 |
| | SE555 | *Indigofera filifolia* | *Burkholderia sp. BL21-ind2-R1 (KM188382.1)* | 100 |
| | SE556 | *Indigofera filifolia* | *Burkholderia sp. BL21-ind4-R3 (KM188386.1)* | 99 |
| | SE557 | *Indigofera filifolia* | *Burkholderia sp. BL21-ind2-R1 (KM188382.1)* | 100 |
| | VG182 | *Indigofera filifolia* | *Burkholderia sp. BL16-ind1-R2 (KM188366.1)* | 100 |
| | VG185 | *Indigofera filifolia* | *Burkholderia sp. BL16-ind1-R2 (KM188366.1)* | 99 |
| | BR62 | *Indigofera filiformis* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | BR542 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 99 |
| | BR544 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | BR545 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | BR67 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | BR68 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | BR69 | *Indigofera frutescens* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | CP588 | *Indigofera frutescens* | *Burkholderia sp. MM5482-R2 ( KF791743.1)* | 100 |
| | CP589 | *Indigofera frutescens* | *Burkholderia sp. MM5482-R2 ( KF791743.1)* | 100 |
| | CP590 | *Indigofera frutescens* | *Burkholderia sp. MM5482-R2 ( KF791743.1)* | 100 |
| | CP591 | *Indigofera frutescens* | *Burkholderia sp. MM5482-R2 ( KF791743.1)* | 100 |
| | BR534 | *Indigofera mauritanica* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 99 |
| | BR535 | *Indigofera mauritanica* | *Burkholderia sp. BL55-R2 (KM188420.1)* | 100 |
| | BR536 | *Indigofera mauritanica* | *Burkholderia sp. BL55-R2 (KM188420.1)* | 100 |
| | BR537 | *Indigofera mauritanica* | *Burkholderia sp. BL55-R2 (KM188420.1)* | 100 |
| | SM52 | *Indigofera sarmentosa* | *Burkholderia sp. MM5496-R1 (KF791745.1)* | 99 |
| | BR103 | *Indigofera sp.* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | BR152 | *Indigofera sp.* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | PK221 | *Lebeckia ambigua* | *Burkholderia sp. BL21-ind5-R2 (KM188387.1)* | 100 |
| | PK225 | *Lebeckia ambigua* | *Burkholderia sp. BL21-ind5-R2 (KM188387.1)* | 100 |
| | RH602 | *Lebeckia sp.* | *Burkholderia sp. MM5496-R1 (KF791745.1)* | 100 |
| | RH625 | *Lebeckia sp.* | *Burkholderia sp. MM5496-R1 (KF791745.1)* | 100 |
| | RH614 | *Podalyria argentea* | *Burkholderia sp. BL29-ind5-R2 (KM188409.1)* | 99 |
| | RH615 | *Podalyria argentea* | *Burkholderia sp. BL29-ind5-R2 (KM188409.1)* | 99 |
| | RH616 | *Podalyria argentea* | *Burkholderia sp. BL29-ind5-R2 (KM188409.1)* | 99 |
| | JH154 | *Podalyria calyptrata* | *Burkholderia sp. BL18-ind8-R3 (KM188379.1)* | 99 |
| | JH160 | *Podalyria calyptrata* | *Burkholderia sp. BL16-ind1-R2 (KM188366.1)* | 99 |
| | KM195 | *Podalyria calyptrata* | *Burkholderia sp. BL29-ind6-R1 ( KM188410.1)* | 99 |
| | KM198 | *Podalyria calyptrata* | *Paraburkholderia sprentiae WSM5005 (CP017562.1)* | 99 |
| | LH1 | *Podalyria calyptrata* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 100 |
| | RM173 | *Podalyria calyptrata* | *Burkholderia sp. BL18-ind8-R2 (KM188378.1)* | 99 |
| | RM175 | *Podalyria calyptrata* | *Burkholderia sp. BL18-ind8-R2 (KM188378.1)* | 100 |
| | VG188 | *Podalyria calyptrata* | *Burkholderia sp. BL21-ind4-R1 (KM188384.1)* | 100 |
| | VG193 | *Podalyria calyptrata* | *Burkholderia sp. BL21-ind4-R1 (KM188384.1)* | 100 |
| | VG204 | *Podalyria calyptrata* | *Burkholderia sp. BL16-ind1-R2 (KM188366.1)* | 99 |
| | VG206 | *Podalyria calyptrata* | *Burkholderia sp. BL9-ind2-R2 (KM188351.1)* | 100 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | WB164 | *Podalyria calyptrata* | *Burkholderia sp. BL21-ind4-R1 (KM188384.1)* | 99 |
| | WB168 | *Podalyria calyptrata* | *Burkholderia sp. BL28-ind1-R3 (KM188428.1)* | 99 |
| | WB229 | *Podalyria calyptrata* | *Burkholderia sp. BL29-ind6-R1 ( KM188410.1)* | 99 |
| | WB232 | *Podalyria calyptrata* | *Burkholderia sp. BL21-ind5-R2 (KM188387.1)* | 100 |
| | SE584 | *Podalyria sericea* | *Burkholderia sp. BL13-R2 (KM188361.1)* | 99 |
| | LH7 | *Podalyria sericea* | *Burkholderia sp. MM6669-R2 (KM188348.1)* | 99 |
| | SE585 | *Podalyria sericea* | *Burkholderia sp. BL13-R2 (KM188361.1)* | 99 |
| | SE586 | *Podalyria sericea* | *Burkholderia sp. BL13-R2 (KM188361.1)* | 99 |
| | SE587 | *Podalyria sericea* | *Burkholderia sp. MM5482-R2 ( KF791743.1)* | 100 |
| | LH22 | *Virgilia oroboides* | *Burkholderia sp. MM5878 (KF840398.1)* | 100 |
| | LH23 | *Virgilia oroboides* | *Burkholderia sp. MM5878 (KF840398.1)* | 100 |
| | SE538 | *Virgilia oroboides* | *Burkholderia sp. BL17-R1 (KM188371.1)* | 100 |
| | SE539 | *Virgilia oroboides* | *Burkholderia sp. BL17-R1 (KM188371.1)* | 100 |
| | SE540 | *Virgilia oroboides* | *Burkholderia sp. BL17-R1 (KM188371.1)* | 100 |
| | SE541 | *Virgilia oroboides* | *Burkholderia sp. BL17-R1 (KM188371.1)* | 100 |
| | TM140 | *Virgilia oroboides* | *Burkholderia sp. BL17-R1 (KM188371.1)* | 100 |
| | RM600 | *Argyrolobium lunare* | *Mesorhizobium sp. MM5369 (KF802693.1)* | 98 |
| | RM98 | *Argyrolobium lunare* | *Mesorhizobium sp. OD13 (KF802674.1)* | 99 |
| | RM552 | *Aspalathus cordata* | *Mesorhizobium sp. OD13 (KF802674.1)* | 99 |
| | SE561 | *Aspalathus cordata* | *Mesorhizobium sp. OD13 (KF802674.1)* | 99 |
| | CP630 | *Aspalathus ericifolia* | *Mesorhizobium sp. MM5734 (KF802683.1)* | 99 |
| | SE582 | *Aspalathus ericifolia* | *Mesorhizobium sp. OD18 (KF802669.1)* | 100 |
| | BK312 | *Psoralea fleta* | *Mesorhizobium sp. MM5343 (KF802700.1)* | 99 |
| | BK315 | *Psoralea fleta* | *Mesorhizobium sp. MM5343 (KF802700.1)* | 98 |
| | RS178 | *Psoralea pinnata* | *Mesorhizobium sp. MM5343 (KF802700.1)* | 99 |
| | BK308 | *Psoralea usitata* | *Mesorhizobium sp. BL637 (KP013169.1)* | 99 |
| *recA* | 334N1 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |
| | 334N2 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |
| | 334N3 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |
| | 334N4 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| | 334N5 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |
| | 334N6 | *Aspalathus carnosa* | *Burkholderia tuberum STM4252 (HE864382.1)* | 99 |
| | 334N7 | *Aspalathus carnosa* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 334N8 | *Aspalathus carnosa* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 334N9 | *Aspalathus carnosa* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 330N1 | *Indigofera candolleana* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 330N2 | *Indigofera candolleana* | *Burkholderia sp. MM5819 (KF791820.1)* | 99 |
| | 330N3 | *Indigofera candolleana* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 330N4 | *Indigofera candolleana* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 330N5 | *Indigofera candolleana* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 330N6 | *Indigofera candolleana* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 330N7 | *Indigofera candolleana* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 330N8 | *Indigofera candolleana* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 330N9 | *Indigofera candolleana* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N1 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N2 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N3 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N4 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N5 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N6 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325N7 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 325NX | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N1 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N1 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N10 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N11 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N12 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N13 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N14 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | 326N15 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N16 | *Indigofera superba* | *Burkholderia sp. MM6511-R2 (KF791816.1)* | 99 |
| | 326N17 | *Indigofera superba* | *Burkholderia sp. OD25-R1 (KF791830.1)* | 99 |
| | 326N18 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N2 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 326N2 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N3 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N3 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N4 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N4 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 326N5 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N5 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N6 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N6 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 326N7 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 326N7 | *Indigofera superba* | *Burkholderia sp. OD25-R1 (KF791830.1)* | 99 |
| | 326N8 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N8 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 326N9 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 328N1 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 328N2 | *Indigofera superba* | *Burkholderia sp. OD120 (KF791814.1)* | 99 |
| | 328N3 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 328N4 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 328N5 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 328N6 | *Indigofera superba* | *Burkholderia sp. OD120 (KF791814.1)* | 99 |
| | 328N7 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 328N8 | *Indigofera superba* | *Burkholderia sp. OD120 (KF791814.1)* | 98 |
| | 333N1 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 333N10 | *Indigofera superba* | *Burkholderia sp. 973N3 (KT718932.1)* | 99 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | 333N11 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 333N12 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 333N13 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 333N14 | *Indigofera superba* | *Burkholderia sp. 973N3 (KT718932.1)* | 99 |
| | 333N15 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N16 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 333N17 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 333N18 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 333N19 | *Indigofera superba* | *Burkholderia sp. MM5819 (KF791820.1)* | 99 |
| | 333N2 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 333N20 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 333N21 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 333N22 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N23 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 333N24 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N25 | *Indigofera superba* | *Burkholderia sp. 973N3 (KT718932.1)* | 99 |
| | 333N26 | *Indigofera superba* | *Burkholderia sp. 988N8 (KT718943.1)* | 99 |
| | 333N27 | *Indigofera superba* | *Burkholderia sp. MM5819 (KF791820.1)* | 99 |
| | 333N3 | *Indigofera superba* | *Burkholderia sp. WSM4175 (HE994066.1)* | 99 |
| | 333N4 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N5 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N6 | *Indigofera superba* | *Burkholderia sp. MM6463B (KF791848.1)* | 99 |
| | 333N7 | *Indigofera superba* | *Burkholderia sp. MM5477-R2 (KF791799.1)* | 99 |
| | 333N8 | *Indigofera superba* | *Burkholderia sp. 989N1 (KT718902.1)* | 99 |
| | 333N9 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718935.1)* | 99 |
| | 327N10 | *Psoralea pullata* | *Mesorhizobium sp. MM5343 (KF802791.1)* | 99 |
| | 327N11 | *Psoralea pullata* | *Mesorhizobium sp. MM5378 (KF802772.1)* | 99 |
| | 327N3 | *Psoralea pullata* | *Mesorhizobium sp. MM5413 (KF802790.1)* | 99 |
| | 3237N4 | *Psoralea pullata* | *Mesorhizobium sp. MM5343 (KF802791.1)* | 99 |

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| | 3237N5 | *Psoralea pullata* | *Mesorhizobium sp. MM5413 (KF802790.1)* | 99 |
| | 3237N6 | *Psoralea pullata* | *Mesorhizobium sp. MM5413 (KF802790.1)* | 99 |
| | 3237N7 | *Psoralea pullata* | *Mesorhizobium sp. MM5343 (KF802791.1)* | 100 |
| | 3237N8 | *Psoralea pullata* | *Mesorhizobium sp. MM5343 (KF802791.1)* | 100 |
| | | | | |
| *nodC* | 334N1 | *Aspalathus carnosa* | *Burkholderia sp. BL16 I4R2 (KR154710.1)* | 99 |
| | 334N2 | *Aspalathus carnosa* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 334N3 | *Aspalathus carnosa* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 334N4 | *Aspalathus carnosa* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 334N5 | *Aspalathus carnosa* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 334N6 | *Aspalathus carnosa* | *Burkholderia sp. BL16 I4R2 (KR154710.1)* | 99 |
| | 334N7 | *Aspalathus carnosa* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 334N8 | *Aspalathus carnosa* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 334N9 | *Aspalathus carnosa* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325N3 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325N4 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325N5 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325N6 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325N7 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 325NX | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 326N1 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N1 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N10 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 100 |
| | 326N11 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N12 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N13 | *Indigofera superba* | *Burkholderia sp. OD123 (KP013138.1)* | 98 |
| | 326N14 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N15 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N16 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 326N17 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |

| DNA region | Strain number | Legume host | Closest match (GenBank Accession No.) | % Similarity |
|---|---|---|---|---|
| | 326N18 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N2 | *Indigofera superba* | *Burkholderia sp. OD123 (KP013138.1)* | 99 |
| | 326N3 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 99 |
| | 326N3 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N4 | *Indigofera superba* | *Burkholderia sp. BL13 R2 (KR154707.1)* | 98 |
| | 326N5 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N6 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N7 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 326N8 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 326N8 | *Indigofera superba* | *Paraburkholderia sprentiae WSM5005 (CP017565.1)* | 99 |
| | 326N9 | *Indigofera superba* | *Burkholderia sp. OD123 (KP013138.1)* | 99 |
| | 333N1 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N10 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |
| | 333N11 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N12 | *Indigofera superba* | *Burkholderia sp. BL27 I3R6 (KR154722.1)* | 99 |
| | 333N13 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |
| | 333N14 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |
| | 333N15 | *Indigofera superba* | *Burkholderia sp. BL13 R2 (KR154707.1)* | 99 |
| | 333N16 | *Indigofera superba* | *Burkholderia sp. 973N7 (KT718871.1)* | 99 |
| | 333N17 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N18 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N19 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N2 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N20 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |
| | 333N21 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N22 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 99 |
| | 333N23 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |
| | 333N24 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 99 |
| | 333N25 | *Indigofera superba* | *Burkholderia sp. 986N1 (KT718833.1)* | 99 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | 333N26 | *Indigofera superba* | *Burkholderia sp. BL16 I4R2 (KR154710.1)* | 99 |
| | 333N27 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N3 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N4 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 99 |
| | 333N5 | *Indigofera superba* | *Burkholderia sp. BL23 I2R2 (KR154718.1)* | 99 |
| | 333N7 | *Indigofera superba* | *Burkholderia tuberum STM678 (AJ306730.1)* | 99 |
| | 333N8 | *Indigofera superba* | *Burkholderia sp. BL27 I3R6 (KR154722.1)* | 99 |
| | 333N9 | *Indigofera superba* | *Burkholderia sp. BL27 I3R6 (KR154722.1)* | 99 |
| | 327N1 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N10 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N11 | *Psoralea pullata* | *Mesorhizobium sp. MM5357 (KR154632.1)* | 99 |
| | 327N16 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N17 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N18 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N19 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N2 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N2 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N20 | *Psoralea pullata* | *Mesorhizobium sp. MM5462 (KR154635.1)* | 99 |
| | 327N21 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N22 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N23 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N24 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N25 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N26 | *Psoralea pullata* | *Mesorhizobium sp. ICMP 12638 (KM018083.1)* | 100 |
| | 327N3 | *Psoralea pullata* | *Mesorhizobium sp. MM5357 (KR154632.1)* | 99 |
| | 327N3 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 327N4 | *Psoralea pullata* | *Mesorhizobium sp. MM5462 (KR154635.1)* | 98 |
| | 327N4 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N5 | *Psoralea pullata* | *Mesorhizobium sp. MM5357 (KR154632.1)* | 99 |

| **DNA region** | **Strain number** | **Legume host** | **Closest match (GenBank Accession No.)** | **% Similarity** |
|---|---|---|---|---|
| | 327N6 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N7 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 327N8 | *Psoralea pullata* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 98 |
| | 329bN1 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 329bN2 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 329bN3 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 329bN4 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 329bN5 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |
| | 329bN6 | *Psoralea restioides* | *Mesorhizobium sp. 998N23 (KT719013.1)* | 99 |

**Appendix B2.** List of reference strains used for the *Indigofera superba* study.

| **DNA region** | **Strain** | **Accession number** |
|---|---|---|
| *nodC* | *Burkholderia dilworthii WSM4206* | HG934336.1 |
| | *Burkholderia dipogonis DL12* | JX009155.2 |
| | *Burkholderia mimosarum PAS44* | EU386155.1 |
| | *Burkholderia phymatum STM815* | DQ888213.1 |
| | *Burkholderia rhynchosiae WSM3930* | HG425183.1 |
| | *Burkholderia sprentiae WSM5005* | HG425199.1 |
| | *Burkholderia tuberum DUS833* | EF566977.1 |
| | *Mesorhizobium abyssinicae AC98c* | GQ847995.1 |
| | *Mesorhizobium alhagi CCNWXJ05-2* | EU130405.2 |
| | *Mesorhizobium amorphae CCNWGS0123* | GQ848005.1 |
| | *Mesorhizobium calcicola ICMP19562* | KC237575.1 |
| | *Mesorhizobium camelthorni A24* | JQ339886.1 |
| | *Mesorhizobium cantuariense ICMP19518* | KC237564.1 |
| | *Mesorhizobium caraganae CCBAU11299* | EF549514.1 |
| | *Mesorhizobium ciceri CKr16* | JX298870.1 |
| | *Mesorhizobium erdmanii USDA3471* | KC237601.1 |
| | *Mesorhizobium gobiense CCBAU83330* | DQ311092.1 |
| | *Mesorhizobium hawassense AC99b* | GQ848002.1 |
| | *Mesorhizobium kowhaii ICMP19520* | KC237573.1 |
| | *Mesorhizobium newzealandense ICMP19547* | KC237580.1 |
| | *Mesorhizobium qingshengii CCBAU33455* | JN129450.1 |
| | *Mesorhizobium robiniae CCNWYC115* | EU722487.1 |
| | *Mesorhizobium sangaii SCAU27* | JF907686.1 |
| | *Mesorhizobium septentrionale SDW14* | KP251768.1 |
| | *Mesorhizobium sophorae ICMP19540* | KC237584.1 |
| | *Mesorhizobium tarimense CCBAU83321* | EF050784.1 |
| | *Mesorhizobium temperatum CCBAU1821* | KM192344.1 |
| | *Mesorhizobium waimense ICMP19569* | JX885708.1 |
| | *Mesorhizobium waitakense ICMP14330* | KC237562.1 |
| | | |
| *recA* | *Burkholderia caledonica LMG19076* | AY619669.1 |
| | *Burkholderia caribensis LMG18531* | AY619662.1 |

| DNA region | Strain | Accession number |
|---|---|---|
| | *Burkholderia cepacia ATCC25416* | HM598380.1 |
| | *Burkholderia diazotrophica* | FN543898.1 |
| | *Burkholderia dilworthii WSM3556* | HE994060.1 |
| | *Burkholderia dipogonis DL7* | JX009159.2 |
| | *Burkholderia fungorum LMG16225* | AY619664.1 |
| | *Burkholderia ginsengisoli LMG24044* | HQ398579.1 |
| | *Burkholderia kirstenboschensis* | HF544404.1 |
| | *Burkholderia phenoliruptrix AC1100* | HQ398589.1 |
| | *Burkholderia phymatum STM815* | AY619667.1 |
| | *Burkholderia phytofirmans LMG22487* | HQ849152.1 |
| | *Burkholderia rhynchosiae WSM3930* | HE994063.1 |
| | *Burkholderia sp. Br3407* | EU294397.1 |
| | *Burkholderia sp. Br3437* | EU294398.1 |
| | *Burkholderia sp. JPY345* | FN543850.1 |
| | *Burkholderia sprentiae WSM5005* | HE994077.1 |
| | *Burkholderia tuberum LMG21444* | AY619674.1 |
| | *Burkholderia unamae MTI-641* | DQ514539.1 |
| | *Burkholderia xenovorans LMG21463* | HQ849164.1 |
| | *Cupriavidus taiwanensis LMG19425* | HE687279.1 |
| | *Mesorhizobium abyssinicae AC98c* | GQ848011.1 |
| | *Mesorhizobium albiziae CCBAU61158* | EU249396.1 |
| | *Mesorhizobium alhagi MQ15* | GU121463.1 |
| | *Mesorhizobium australicum WSM2073* | CP003358.1 |
| | *Mesorhizobium ciceri CCANP20* | HG323887.1 |
| | *Mesorhizobium erdmanii USDA3471* | AJ294371.1 |
| | *Mesorhizobium huakuii CCBAU65318* | EU672500.1 |
| | *Mesorhizobium loti CCBAU01461* | EU672493.1 |
| | *Mesorhizobium mediterraneum USDA3392* | AJ294369.1 |
| | *Mesorhizobium opportunistum WSM2075* | CP002279.1 |
| | *Mesorhizobium robiniae CCNWYC115* | GQ856501.1 |
| | *Mesorhizobium shangrilense CCBAU65327* | EU872249.1 |
| | *Mesorhizobium tamadayense CCANP122* | HG323907.1 |

**Appendix B3.** List of rhizobial strains with binary scoring of their soil types. A score of 0 indicates that a strain was not found in that particular soil type, while a score of 1 means that it was.

| **Genus** | **Strain** | **Granite** | **Limestone** | **Sandstone** | **Shale** |
|---|---|---|---|---|---|
| *Burkholderia* | *Burkholderia sp. (Amphithalea ericifolia MM5482)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus argyrella SE571)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus argyrella SE572)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus callosa MM5477)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus capensis CPT36)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus carnosa MM5496)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Aspalathus carnosa SW5)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Bolusafra bituminosa OD29)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Crotalaria sp. OD120)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Dipogon lignosus LH6)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera angustifolia MM5878)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera candolleana LH4)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera cytisoides MM5819)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filifolia SE555)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filifolia SE556)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filifolia SE557)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filifolia VG182)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filifolia VG185)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera filiformis BR62)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR542)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR543)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR544)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR545)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR67)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR68)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens BR69)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera frutescens CP 589)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera frutescens CP588)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera frutescens CP590)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera frutescens CP591)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera mauritanica BR534)* | 1 | 0 | 0 | 0 |

| Genus | Strain | Granite | Limestone | Sandstone | Shale |
|---|---|---|---|---|---|
| | *Burkholderia sp. (Indigofera mauritanica BR535)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera mauritanica BR536)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera mauritanica BR537)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera sarmentosa SE573)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera sarmentosa SM52)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Indigofera sp. BR103)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera sp. BR152)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Indigofera sp. MM5746)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Lebeckia ambigua PK221)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Lebeckia ambigua PK225)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Lebeckia sp. RH601)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Lebeckia sp. RH602)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Lebeckia sp. RH625)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Liparia splendens SE574)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria argentea RH614)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria argentea RH615)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria argentea RH616)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria burchellii MM5875)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata JH154)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata JH160)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata KM195)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata KM198)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata LH1)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata MM5337)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata OD25)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata RM1730* | 0 | 0 | 0 | 1 |
| | *Burkholderia sp. (Podalyria calyptrata RM175)* | 0 | 0 | 0 | 1 |
| | *Burkholderia sp. (Podalyria calyptrata VG188)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata VG193)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata VG204)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata VG205)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata VG206)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata WB164)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata WB168)* | 0 | 0 | 1 | 0 |

| **Genus** | **Strain** | **Granite** | **Limestone** | **Sandstone** | **Shale** |
|---|---|---|---|---|---|
| | *Burkholderia sp. (Podalyria calyptrata WB229)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria calyptrata WB232)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria sericea LH7)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Podalyria sericea MM5384)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Podalyria sericea SE584)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria sericea SE585)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria sericea SE586)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Podalyria sericea SE587)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Rafnia acuminata OD22)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Rafnia sp. OD28)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides LH22)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Virgilia oroboides LH23)* | 1 | 0 | 0 | 0 |
| | *Burkholderia sp. (Virgilia oroboides MM5366)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides SE538)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides SE539)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides SE540)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides SE541)* | 0 | 0 | 1 | 0 |
| | *Burkholderia sp. (Virgilia oroboides TM140)* | 0 | 0 | 1 | 0 |
| | | | | | |
| *Mesorhizobium* | *Mesorhizobium sp. (Argyrolobium lunare MM5369)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Argyrolobium lunare OD14)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Argyrolobium lunare OD48)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Argyrolobium lunare RM5810)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Argyrolobium lunare RM600)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Argyrolobium lunare RM98)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Argyrolobium velutinum OD47)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathuis astroites OD18)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus aurantiaca MM5397)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus bracteata MM5618)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR621)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR623)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR624)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus chenopoda RM559)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Aspalathus ciliaris CS13166)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ciliaris MM5361)* | 0 | 0 | 1 | 0 |

| Genus | Strain | Granite | Limestone | Sandstone | Shale |
|---|---|---|---|---|---|
| | *Mesorhizobium sp. (Aspalathus ciliaris OD108)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus cordata OD13)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus cordata RM552)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Aspalathus cordata SE551)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus cordata SE561)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus cordata SE563)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia BR604)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia BR605)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP627)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP628)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP630)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia MM5352)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia OD31)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE582)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE598)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE599)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus hispida RM564)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Aspalathus spicata MM5398)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus spicata MM5440)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus uniflora MM5734)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Aspalathus uniflora OD26)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Indigofera angustifolia MM5378)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Indigofera venusta MM5377)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Otholobium bracteolatum OD42)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium fruticans BR619)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium fruticans BR620)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum LH609)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum LH610)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum LH611)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum LH612)* | 1 | 0 | 0 | 0 |

| **Genus** | **Strain** | **Granite** | **Limestone** | **Sandstone** | **Shale** |
|---|---|---|---|---|---|
| | *Mesorhizobium sp. (Otholobium hirtum MM5334)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Otholobium hirtum MM5376)* | 1 | 0 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum MM5382)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum OD32)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Otholobium hirtum RM579)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Otholobium obliquum MM5370)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Otholobium virgatum MM5333)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Otholobium virgatum MM5357)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Otholobium virgatum RM567)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Otholobium virgatum SW44)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Otholobium zeyheri MM5675)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea aphylla SE595)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea aphylla SE596)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea aphylla SE597)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea asarina MM5360)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea asarina OD15)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea asarina RM548)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Psoralea brilliantissima OD52)* | 0 | 1 | 0 | 0 |
| | *Mesorhizobium sp. (Psoralea congesta MM5462)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea fleta BK312)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea fleta BK315)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea laxa OD119)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Psoralea oligophylla OD118)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Psoralea pinnata RH607)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea pinnata RH608)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea pinnata RM569)* | 0 | 0 | 0 | 1 |
| | *Mesorhizobium sp. (Psoralea pinnata RS176)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea pinnata RS178)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea pinnata SE592)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea rigidula MM5343)* | 0 | 0 | 1 | 0 |
| | *Mesorhizobium sp. (Psoralea usitata BK308)* | 0 | 0 | 1 | 0 |

**Appendix B4.** List of rhizobial strains used with the altitude and pH data of their sites

| **Genus** | **Strain** | **Altitude** | **pH** |
|---|---|---|---|
| *Burkholderia* | *Burkholderia sp. (Amphithalea ericifolia MM5482)* | 20 | 4.64 |
| | *Burkholderia sp. (Aspalathus argyrella SE571)* | 486 | 4.97 |
| | *Burkholderia sp. (Aspalathus argyrella SE572)* | 487 | 4.97 |
| | *Burkholderia sp. (Aspalathus callosa MM5477)* | 10 | 4.62 |
| | *Burkholderia sp. (Aspalathus capensis CPT36)* | 254 | 4.07 |
| | *Burkholderia sp. (Aspalathus carnosa MM5496)* | 57 | 3.73 |
| | *Burkholderia sp. (Aspalathus carnosa SW5)* | 112 | 3.53 |
| | *Burkholderia sp. (Bolusafra bituminosa OD29)* | 272 | 3.18 |
| | *Burkholderia sp. (Crotalaria sp. OD120)* | 240 | 4.64 |
| | *Burkholderia sp. (Dipogon lignosus LH6)* | 228 | 5.1 |
| | *Burkholderia sp. (Indigofera angustifolia MM5878)* | 40 | 4.5 |
| | *Burkholderia sp. (Indigofera candolleana LH4)* | 228 | 5.1 |
| | *Burkholderia sp. (Indigofera cytisoides MM5819)* | 179 | 4.36 |
| | *Burkholderia sp. (Indigofera filifolia SE555)* | 296 | 4.97 |
| | *Burkholderia sp. (Indigofera filifolia SE556)* | 296 | 4.97 |
| | *Burkholderia sp. (Indigofera filifolia SE557)* | 296 | 4.97 |
| | *Burkholderia sp. (Indigofera filifolia VG182)* | 240 | 3.43 |
| | *Burkholderia sp. (Indigofera filifolia VG185)* | 240 | 3.43 |
| | *Burkholderia sp. (Indigofera filiformis BR62)* | 410 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR542)* | 412 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR543)* | 412 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR544)* | 412 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR545)* | 412 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR67)* | 404 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR68)* | 408 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens BR69)* | 408 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens CP589)* | 413 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens CP588)* | 413 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens CP590)* | 413 | 4.4 |
| | *Burkholderia sp. (Indigofera frutescens CP591)* | 413 | 4.1 |
| | *Burkholderia sp. (Indigofera mauritanica BR534)* | 414 | 4.4 |
| | *Burkholderia sp. (Indigofera mauritanica BR535)* | 414 | 4.4 |
| | *Burkholderia sp. (Indigofera mauritanica BR536)* | 414 | 4.4 |
| | *Burkholderia sp. (Indigofera mauritanica BR537)* | 414 | 4.4 |
| | *Burkholderia sp. (Indigofera sarmentosa SE573)* | 296 | 4.97 |
| | *Burkholderia sp. (Indigofera sarmentosa SM52)* | 112 | 3.53 |
| | *Burkholderia sp. (Indigofera sp. BR103)* | 411 | 4.4 |
| | *Burkholderia sp. (Indigofera sp. BR152)* | 416 | 4.46 |
| | *Burkholderia sp. (Indigofera sp. MM5746)* | 296 | 4.18 |
| | *Burkholderia sp. (Lebeckia ambigua PK221)* | 447 | 3.56 |
| | *Burkholderia sp. (Lebeckia ambigua PK225)* | 447 | 3.56 |
| | *Burkholderia sp. (Lebeckia sp. RH601)* | 369 | 3.83 |
| | *Burkholderia sp. (Lebeckia sp. RH602)* | 369 | 3.83 |

| Genus | Strain | Altitude | pH |
|---|---|---|---|
| | *Burkholderia sp. (Lebeckia sp. RH625)* | 365 | 3.83 |
| | *Burkholderia sp. (Liparia splendens SE574)* | 296 | 4.97 |
| | *Burkholderia sp. (Podalyria argentea RH614)* | 364 | 3.83 |
| | *Burkholderia sp. (Podalyria argentea RH615)* | 364 | 3.83 |
| | *Burkholderia sp. (Podalyria argentea RH616)* | 364 | 3.83 |
| | *Burkholderia sp. (Podalyria burchellii MM5875)* | 28 | 4.5 |
| | *Burkholderia sp. (Podalyria calyptrata JH154)* | 296 | 4.18 |
| | *Burkholderia sp. (Podalyria calyptrata JH160)* | 296 | 4.18 |
| | *Burkholderia sp. (Podalyria calyptrata KM195)* | 128 | 3.76 |
| | *Burkholderia sp. (Podalyria calyptrata KM198)* | 128 | 3.76 |
| | *Burkholderia sp. (Podalyria calyptrata LH1)* | 228 | 5.1 |
| | *Burkholderia sp. (Podalyria calyptrata MM5337)* | 420 | 4.56 |
| | *Burkholderia sp. (Podalyria calyptrata OD25)* | 272 | 3.62 |
| | *Burkholderia sp. (Podalyria calyptrata RM173)* | 166 | 4.4 |
| | *Burkholderia sp. (Podalyria calyptrata RM175)* | 166 | 4.4 |
| | *Burkholderia sp. (Podalyria calyptrata VG188)* | 241 | 3.43 |
| | *Burkholderia sp. (Podalyria calyptrata VG193)* | 241 | 3.43 |
| | *Burkholderia sp. (Podalyria calyptrata VG204)* | 241 | 3.43 |
| | *Burkholderia sp. (Podalyria calyptrata VG205)* | 241 | 3.43 |
| | *Burkholderia sp. (Podalyria calyptrata VG206)* | 241 | 3.43 |
| | *Burkholderia sp. (Podalyria calyptrata WB164)* | 84 | 3.56 |
| | *Burkholderia sp. (Podalyria calyptrata WB168)* | 84 | 3.56 |
| | *Burkholderia sp. (Podalyria calyptrata WB229)* | 87 | 3.59 |
| | *Burkholderia sp. (Podalyria calyptrata WB232)* | 87 | 3.59 |
| | *Burkholderia sp. (Podalyria sericea LH7)* | 228 | 5.1 |
| | *Burkholderia sp. (Podalyria sericea MM5384)* | 91 | 5.26 |
| | *Burkholderia sp. (Podalyria sericea SE584)* | 312 | 4.97 |
| | *Burkholderia sp. (Podalyria sericea SE585)* | 312 | 4.97 |
| | *Burkholderia sp. (Podalyria sericea SE586)* | 312 | 4.97 |
| | *Burkholderia sp. (Podalyria sericea SE587)* | 312 | 4.97 |
| | *Burkholderia sp. (Rafnia acuminata OD22)* | 272 | 3.62 |
| | *Burkholderia sp. (Rafnia sp. OD28)* | 272 | 3.62 |
| | *Burkholderia sp. (Virgilia oroboides LH22)* | 174 | 5.1 |
| | *Burkholderia sp. (Virgilia oroboides LH23)* | 174 | 5.1 |
| | *Burkholderia sp. (Virgilia oroboides MM5366)* | 296 | 4.18 |
| | *Burkholderia sp. (Virgilia oroboides SE538)* | 313 | 4.97 |
| | *Burkholderia sp. (Virgilia oroboides SE539)* | 313 | 4.97 |
| | *Burkholderia sp. (Virgilia oroboides SE540)* | 313 | 4.97 |
| | *Burkholderia sp. (Virgilia oroboides SE541)* | 313 | 4.97 |
| | *Burkholderia sp. (Virgilia oroboides TM140)* | 686 | 4.97 |
| | | | |
| *Mesorhizobium* | *Mesorhizobium sp. (Argyrolobium lunare MM5369)* | 296 | 4.18 |
| | *Mesorhizobium sp. (Argyrolobium lunare OD14)* | 296 | 3.18 |
| | *Mesorhizobium sp. (Argyrolobium lunare OD48)* | 11.4 | 6.39 |
| | *Mesorhizobium sp. (Argyrolobium lunare RM581)* | 212 | 4.4 |

| **Genus** | **Strain** | **Altitude** | **pH** |
|---|---|---|---|
| | *Mesorhizobium sp. (Argyrolobium lunare RM600)* | 219 | 4.4 |
| | *Mesorhizobium sp. (Argyrolobium lunare RM98)* | 218 | 4.4 |
| | *Mesorhizobium sp. (Argyrolobium velutinum OD47)* | 11.4 | 6.71 |
| | *Mesorhizobium sp. (Aspalathuis astroites OD18)* | 296 | 4 |
| | *Mesorhizobium sp. (Aspalathus aurantiaca MM5397)* | 368 | 4.11 |
| | *Mesorhizobium sp. (Aspalathus bracteata MM5618)* | 582 | 5.19 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR621)* | 415 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR623)* | 415 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus chenopoda BR624)* | 415 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus chenopoda RM559)* | 215 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus ciliaris CS13166)* | 40 | 4.1 |
| | *Mesorhizobium sp. (Aspalathus ciliaris MM5361)* | 296 | 4 |
| | *Mesorhizobium sp. (Aspalathus ciliaris OD108)* | 522 | 4.76 |
| | *Mesorhizobium sp. (Aspalathus cordata OD13)* | 296 | 4 |
| | *Mesorhizobium sp. (Aspalathus cordata RM552)* | 217 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus cordata SE551)* | 296 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus cordata SE561)* | 298 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus cordata SE563)* | 292 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus ericifolia BR604)* | 415 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus ericifolia BR605)* | 141 | 4.4 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP627)* | 114 | 4.07 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP628)* | 114 | 4.07 |
| | *Mesorhizobium sp. (Aspalathus ericifolia CP630)* | 114 | 4.07 |
| | *Mesorhizobium sp. (Aspalathus ericifolia MM5352)* | 272 | 4.88 |
| | *Mesorhizobium sp. (Aspalathus ericifolia OD31)* | 581 | 5.19 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE582)* | 312 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE598)* | 311 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus ericifolia SE599)* | 311 | 4.97 |
| | *Mesorhizobium sp. (Aspalathus hispida RM564)* | 205 | 5.47 |
| | *Mesorhizobium sp. (Aspalathus spicata MM5398)* | 368 | 4.11 |
| | *Mesorhizobium sp. (Aspalathus spicata MM5440)* | 229 | 3.9 |
| | *Mesorhizobium sp. (Aspalathus uniflora MM5734)* | 348 | 3.49 |
| | *Mesorhizobium sp. (Aspalathus uniflora OD26)* | 272 | 3.62 |
| | *Mesorhizobium sp. (Indigofera angustifolia MM5378)* | 68 | 4.5 |
| | *Mesorhizobium sp. (Indigofera venusta MM5377)* | 68 | 4.5 |
| | *Mesorhizobium sp. (Otholobium bracteolatum OD42)* | 11.4 | 6.71 |
| | *Mesorhizobium sp. (Otholobium fruticans BR619)* | 636 | 4.4 |
| | *Mesorhizobium sp. (Otholobium fruticans BR620)* | 363 | 4.4 |
| | *Mesorhizobium sp. (Otholobium hirtum LH609)* | 472 | 4.77 |
| | *Mesorhizobium sp. (Otholobium hirtum LH610)* | 472 | 4.77 |
| | *Mesorhizobium sp. (Otholobium hirtum LH611)* | 472 | 4.77 |
| | *Mesorhizobium sp. (Otholobium hirtum LH612)* | 472 | 4.77 |
| | *Mesorhizobium sp. (Otholobium hirtum MM5334)* | 203 | 5.31 |
| | *Mesorhizobium sp. (Otholobium hirtum MM5376)* | 68 | 4.9 |
| | *Mesorhizobium sp. (Otholobium hirtum MM5382)* | 68 | 4.59 |

| **Genus** | **Strain** | **Altitude** | **pH** |
|---|---|---|---|
| | *Mesorhizobium sp. (Otholobium hirtum OD32)* | 581 | 5.19 |
| | *Mesorhizobium sp. (Otholobium hirtum RM579)* | 247 | 4.4 |
| | *Mesorhizobium sp. (Otholobium obliquum MM5370)* | 330 | 4.18 |
| | *Mesorhizobium sp. (Otholobium virgatum MM5333)* | 203 | 5.31 |
| | *Mesorhizobium sp. (Otholobium virgatum MM5357)* | 296 | 4.26 |
| | *Mesorhizobium sp. (Otholobium virgatum RM567)* | 247 | 4.4 |
| | *Mesorhizobium sp. (Otholobium virgatum SW44)* | 124 | 3.86 |
| | *Mesorhizobium sp. (Otholobium zeyheri MM5675)* | 610 | 5.39 |
| | *Mesorhizobium sp. (Psoralea aphylla SE595)* | 317 | 4.97 |
| | *Mesorhizobium sp. (Psoralea aphylla SE596)* | 317 | 4.97 |
| | *Mesorhizobium sp. (Psoralea aphylla SE597)* | 317 | 4.97 |
| | *Mesorhizobium sp. (Psoralea asarina MM5360)* | 296 | 4.26 |
| | *Mesorhizobium sp. (Psoralea asarina OD15)* | 296 | 3.18 |
| | *Mesorhizobium sp. (Psoralea asarina RM548)* | 162 | 4.4 |
| | *Mesorhizobium sp. (Psoralea brilliantissima OD52)* | 107 | 6.5 |
| | *Mesorhizobium sp. (Psoralea congesta MM5462)* | 20 | 4.5 |
| | *Mesorhizobium sp. (Psoralea fleta BK312)* | 358 | 3.29 |
| | *Mesorhizobium sp. (Psoralea fleta BK315)* | 358 | 3.29 |
| | *Mesorhizobium sp. (Psoralea laxa OD119)* | 240 | 4.64 |
| | *Mesorhizobium sp. (Psoralea oligophylla OD118)* | 240 | 4.64 |
| | *Mesorhizobium sp. (Psoralea pinnata RH607)* | 364 | 3.83 |
| | *Mesorhizobium sp. (Psoralea pinnata RH608)* | 364 | 3.83 |
| | *Mesorhizobium sp. (Psoralea pinnata RM569)* | 210 | 4.4 |
| | *Mesorhizobium sp. (Psoralea pinnata RS176)* | 136 | 3.84 |
| | *Mesorhizobium sp. (Psoralea pinnata RS178)* | 136 | 3.84 |
| | *Mesorhizobium sp. (Psoralea pinnata SE592)* | 321 | 4.97 |
| | *Mesorhizobium sp. (Psoralea rigidula MM5343)* | 753 | 3.29 |
| | *Mesorhizobium sp. (Psoralea usitata BK308)* | 356 | 3.29 |

# Appendix C

**Appendix C1.** List of rhizobial strains of CCR legumes from previous studies and reference strains that were used in this study. - = sequence not available.

| Legume | Taxon | Voucher | 16S rRNA | *recA* | *nodA* |
|---|---|---|---|---|---|
| **Astragaleae** | | | | | |
| | *Lessertia annularis* | WSM3270 | JN544905.1 | - | JN375809.1 |
| | *Lessertia diffusa* | WSM2561 | JN544901.1 | - | JN375805.1 |
| | *Lessertia excisa* | WSM3612 | JN544910.1 | - | JN375816.1 |
| | *Lessertia herbacea* | WSM3602 | JN544916.1 | - | JN375837.1 |
| | *Lessertia microphylla* | WSM2559 | JN544900.1 | - | JN375804.1 |
| | *Lessertia* sp. | Dlodlo 46 | KF802586 | KF802773 | KF802688 |
| **Crotalarieae** | | | | | |
| | *Aspalathus astroites* | Dlodlo 18 | KF802565 | KF802752 | KF802669 |
| | *Aspalathus aurantiaca* | Muasya 5397 | KF802566 | KF802753 | KF802670 |
| | *Aspalathus bracteata* | Muasya 5618 | KF802567 | KF802754 | KF802671 |
| | *Aspalathus callosa* | Muasya 5477 | KF791604 | KF791798 | KF791744 |
| | *Aspalathus carnosa* | Muasya 5496 | KF791607 | KF791801 | KF791745 |
| | *Aspalathus ciliaris* | Dlodlo 108 | KF802568 | KF802755 | KF802672 |
| | *Aspalathus ciliaris* | Muasya 5361 | KF802569 | KF802756 | KF802673 |
| | *Aspalathus ciliaris* | Stirton 13166 | KF802570 | KF802757 | - |
| | *Aspalathus cordata* | Dlodlo 13 | KF802571 | KF802758 | KF802674 |
| | *Aspalathus ericifolia* subsp. ericifolia | Dlodlo 31 | KF802572 | KF802759 | KF802675 |
| | *Aspalathus laricifolia* subsp. laricifolia | Muasya 5372 | KF802574 | KF802761 | - |
| | *Aspalathus* sp. | Dlodlo 49 | KF802576 | KF802763 | KF802678 |
| | *Aspalathus* sp. | Dlodlo 53 | KF802577 | KF802764 | KF802679 |
| | *Aspalathus spicata* | Muasya 5398 | KF802578 | KF802765 | KF802680 |
| | *Aspalathus spicata* | Muasya 5440 | KF802579 | KF802766 | KF802681 |

| | | | | | |
|---|---|---|---|---|---|
| | *Aspalathus uniflora* subsp. uniflora | Dlodlo 26 | KF802580 | KF802767 | KF802682 |
| | *Aspalathus uniflora* subsp. uniflora | Muasya 5734 | KF802581 | KF802768 | KF802683 |
| | *Crotalaria* sp. | Dlodlo 120 | KF791620 | KF791814 | KF791755 |
| | *Rafnia acuminata* | Dlodlo 22 | KF791658 | KF791851 | - |
| | *Rafnia triflora* | Dlodlo 55 | KF791666 | - | KF802702 |
| **Genisteae** | | | | | |
| | *Argyrolobium lunare* subsp*. sericeum*. | Dlodlo 14 | KF802562 | KF802749 | KF802666 |
| | *Argyrolobium lunare* | Dlodlo 48 | KF802563 | KF802750 | KF802667 |
| | *Argyrolobium lunare* subsp*. sericeum* | Muasya 5369 | KF802592 | KF802778 | KF802693 |
| | *Argyrolobium velutinum* | Dlodlo 47 | KF802564 | KF802751 | KF802668 |
| **Hypocalypteae** | | | | | |
| | *Hypocalyptus oxalidifolius* | RAU6.4d | HF674671.1 | HF544365.1 | HF674496.1 |
| | *Hypocalyptus sophoroides* | HC1.1a1 | HF674709.1 | HF544405.1 | HF674484.1 |
| | *Hypocalyprus sophoroides* | WK1.1a | HF674684.1 | HF544380.1 | HF674474.1 |
| **Indigofereae** | | | | | |
| | *Indigofera frutescens* | Muasya 5392 | KF802582 | KF802769 | KF802684 |
| | *Indigofera gracilis* | Muasya 5621 | KF802583 | KF802770 | KF802685 |
| | *Indigofera* sp. | Dlodlo 45 | KF802584 | KF802771 | KF802686 |
| | *Indigofera venusta* | Muasya 5377 | KF802585 | KF802772 | KF802687 |
| **Millettieae** | | | | | |
| | *Tephrosia capensis* | Muasya 5405 | KF802609 | KF802795 | KF802703 |
| **Phaseoleae** | | | | | |
| | *Bolusafra bituminosa* | Dlodlo 29 | KF791610 | KF791804 | KF791747 |
| **Podalyrieae** | | | | | |
| | *Amphithalea ericifolia* subsp. ericifolia | Muasya 5482 | KF791603 | KF791797 | KF791743 |
| | *Cyclopia sessiliflora* | UCT30 | AY178067.1 | HF544431.1 | HF674527.1 |
| | *Cyclopia subternata* | CS2 | AY178065.1 | HF544433.1 | HF674529.1 |

| | | | | | |
|---|---|---|---|---|---|
| | *Podalyria  burchellii* | Muasya 5875 | KF791652 | KF791845 | KF791780 |
| | *Podalyria calyptrata* | Dlodlo 25 | KF791637 | KF791830 | KF791769 |
| | *Podalyria calyptrata* | Muasya 5337 | KF791641 | KF791834 | KF791773 |
| | *Virgilia divaricata* | Dlodlo 123 | KF802610 | - | KF802704 |
| | *Virgilia oroboides* | Kb6 | HF674703.1 | HF544399.1 | HF674500.1 |
| **Psoraleeae** | | | | | |
| | *Otholobium bracteolatum* | Dlodlo 42 | KF802587 | KF802774 | KF802689 |
| | *Otholobium hirtum* | Muasya 5334 | KF802589 | KF802776 | KF802691 |
| | *Otholobium hirtum* | Muasya 5376 | KF802590 | KF802777 | - |
| | *Otholobium virgatum* | Muasya 5333 | KF802594 | KF802780 | - |
| | *Psoralea asarina* | Dlodlo 15 | KF802597 | KF802783 | KF802696 |
| | *Psoralea brilliantissoma* | Dlodlo 52 | KF802606.1 | KF802792.1 | KF802701.1 |
| | *Psoralea oligophylla* | Dlodlo 118 | KF802602 | KF802788 | KF802699 |
| | *Psoralea rigidula* | Muasya 5343 | KF802605 | KF802791 | KF802700 |
| ***Bradyrhizobium*** | | | | | |
| | *Bradyrhizobium canariense* | SEMIA 928 | FJ390904.1 | FJ391148.1 | HQ259487.1 |
| | *Bradyrhizobium elkanii* | SEMIA 5011 | FJ390893.1 | FJ391149.1 | HQ259477.1 |
| | *Bradyrhizobium japonicum* | NZP2309 | FM202364.1 | FM202364.1 | HE583307.1 |
| ***Burkholderia*** | | | | | |
| | *Burkholderia caribensis* | TJ182 | - | - | AJ505309.1 |
| | *Burkholderia diazotrophica* | JPY641T | FN543755.1 | FN543898.1 | - |
| | *Burkholderia diazotrophica* | STM4206 | - | - | FN908414.1 |
| | *Burkholderia dilworthii* | WSM3556 | NR_125580.1 | HE994060.1 | HG934319.1 |
| | *Burkholderia fungorum* | LMG 16225 | NR_118060.1 | AY619664.1 | - |
| | *Burkholderia ginsengisoli* | KMY03 | NR_041288.1 | - | - |
| | *Burkholderia mimosarum* | PAS44 | - | - | EU434822.1 |

| | | | | | |
|---|---|---|---|---|---|
| | *Burkholderia nodosa* | Br3437 | NR_043181.1 | EU294398.1 | - |
| | *Burkholderia nodosa* | Br3461 | AY773192.1 | HQ398587.1 | AY533871.1 |
| | *Burkholderia kirstenboschensis* | Kb2 | HF674702.1 | HF544398.1 | HF674499.1 |
| | *Burkholderia kirstenboschensis* | Kb14 | HF674706.1 | HF544402.1 | HF674503.1 |
| | *Burkholderia phenoliruptrix* | AC1100 | - | ADZ13941.1 | - |
| | *Burkholderia phymatum* | STM815 | NR_075568.1 | AY619667.1 | AJ505318.2 |
| | *Burkholderia phytofirmans* | PsJN | NR_102845.1 | - | - |
| | *Burkholderia rhynchosiae* | WSM3930 | EU219864.1 | JX989220.1 | EU219866.1 |
| | *Burkholderia rhynchosiae* | WSM3937 | NR_116248.1 | JX989221.1 | EU219867.1 |
| | *Burkholderia sabiae* | Br3407 | NR_115261.1 | EU294397.1 | AY533872.1 |
| | *Burkholderia symbiotica* | JPY345 | HM357233.1 | FN543850.1 | - |
| | *Burkholderia sprentiae* | WSM5005 | HQ698903 | HE994061 | HG934334.1 |
| | *Burkholderia tuberum* | STM678 | AJ302311.1 | HQ849162.1 | AJ302321.1 |
| | *Burkholderia unamae* | MTI 641 | NR_027569.1 | DQ514539.1 | - |
| | *Burkholderia xenovorans* | LMG21463 | NR_118083.1 | HQ849164.1 | - |
| ***Ensifer*** | | | | | |
| | *Ensifer medicae* | A321 | L39882 | AJ294381 | EU292001.1 |
| | *Ensifer meliloti* | CCNWSX541 | - | FJ619311.1 | FJ619291.1 |
| | *Ensifer mexicanus* | ITTGS4 | - | EF457970.1 | - |
| | *Ensifer xinjiangense* | CCNWNX165 | - | FJ619308.1 | - |
| ***Mesorhizobium*** | | | | | |
| | *Mesorhizobium abyssinicae* | AC98c | GQ847896.1 | GQ848011.1 | GQ847988.1 |
| | *Mesorhizobium albiziae* | CCBAU 61158 | NR_043549.1 | EU249396.1 | GQ167236.1 |
| | *Mesorhizobium alhagi* | MQ15 | GU083834.1 | GU121463.1 | GU111372.1 |

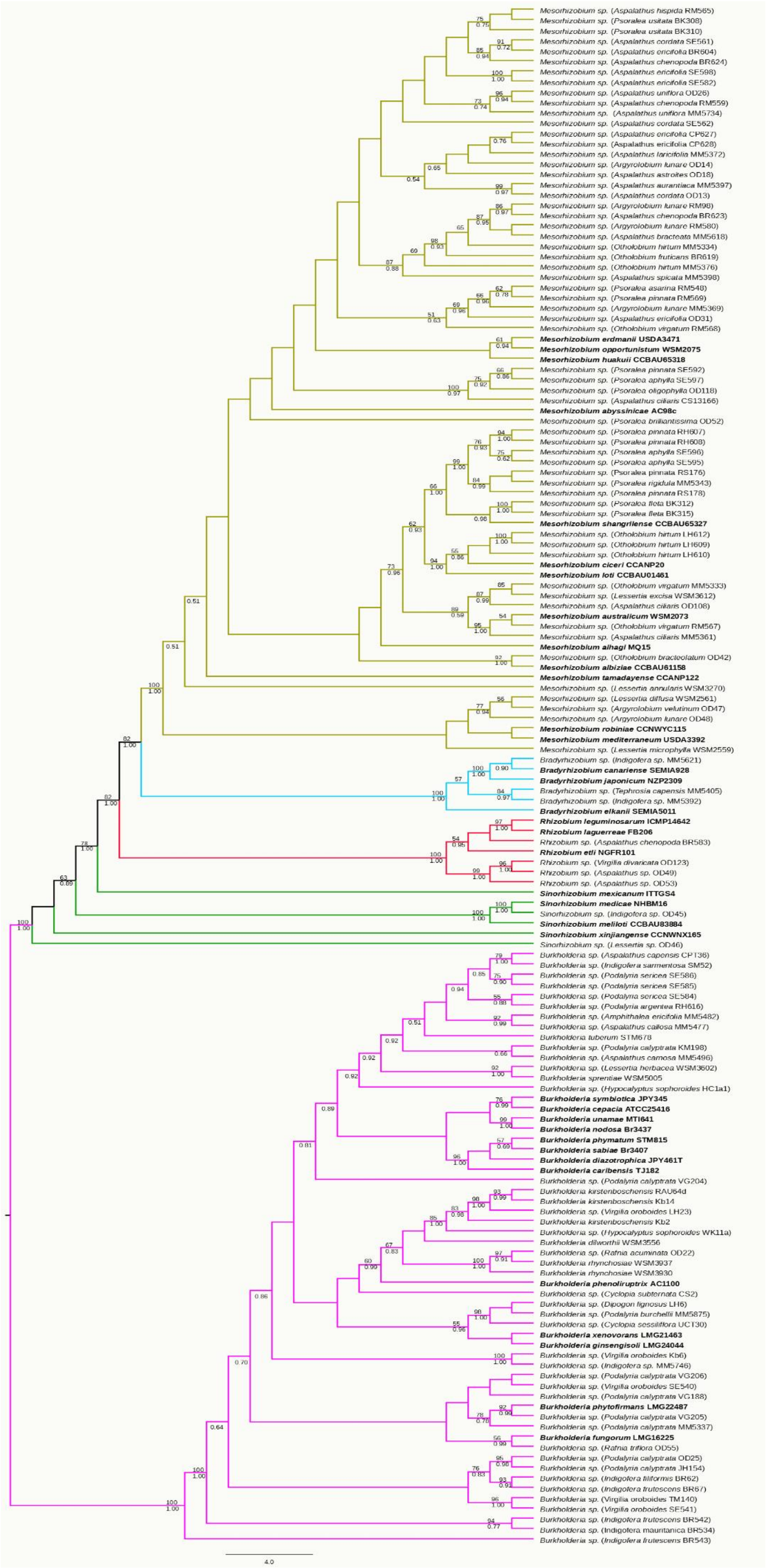
|  |  |  |  |  |  |
|---|---|---|---|---|---|
|  | *Mesorhizobium australicum* | WSM2073 | AY601516.1 | NC_019973.1 | AY601528.1 |
|  | *Mesorhizobium ciceri* | CCANP20 | HF931044.1 | HG323887.1 | LN824188.1 |
|  | *Mesorhizobium erdmanii* | USDA3471 | KM192334.1 | AJ294371.1 | NZ_KI421454.1 |
|  | *Mesorhizobium huakuii* | CCBAU 65318 | EU074202.1 | EU672500.1 | KP251733.1 |
|  | *Mesorhizobium loti* | CCBAU 01461 | - | EU672493.1 | EU687481.1 |
|  | *Mesorhizobium opportunistum* | WSM2075 | NR_074209.1 | NC_015675.1 | AY601530.1 |
|  | *Mesorhizobium robiniae* | CCNWYC 115 | NR_116467.1 | GQ856501.1 | EU849558.1 |
|  | *Mesorhizobium shangrilense* | CCBAU 65318 | NR_116163.1 | EU672501.1 | KP251734.1 |
|  | *Mesorhizobium tamadayense* | CCANP122 | HF931079.1 | HG323907.1 | LN824200.1 |
| ***Rhizobium*** |  |  |  |  |  |
|  | *Rhizobium etli* | NGBFR101 | JQ670251.2 | KP128103.1 | AB917356.1 |
|  | *Rhizobium laguerreae* | FB206 | JN558651.2 | JN558681.2 | JN558701.1 |
|  | *Rhizobium leguminosarum* | ICMP 14642 | AY491062.1 | AY494813.1 | DQ100409.1 |

**Appendix C2.** List of legume species and their Genbank or the Barcode of Life Data Systems (BOLD) accession numbers as used for phylogenetic analyses. - = sequence not available

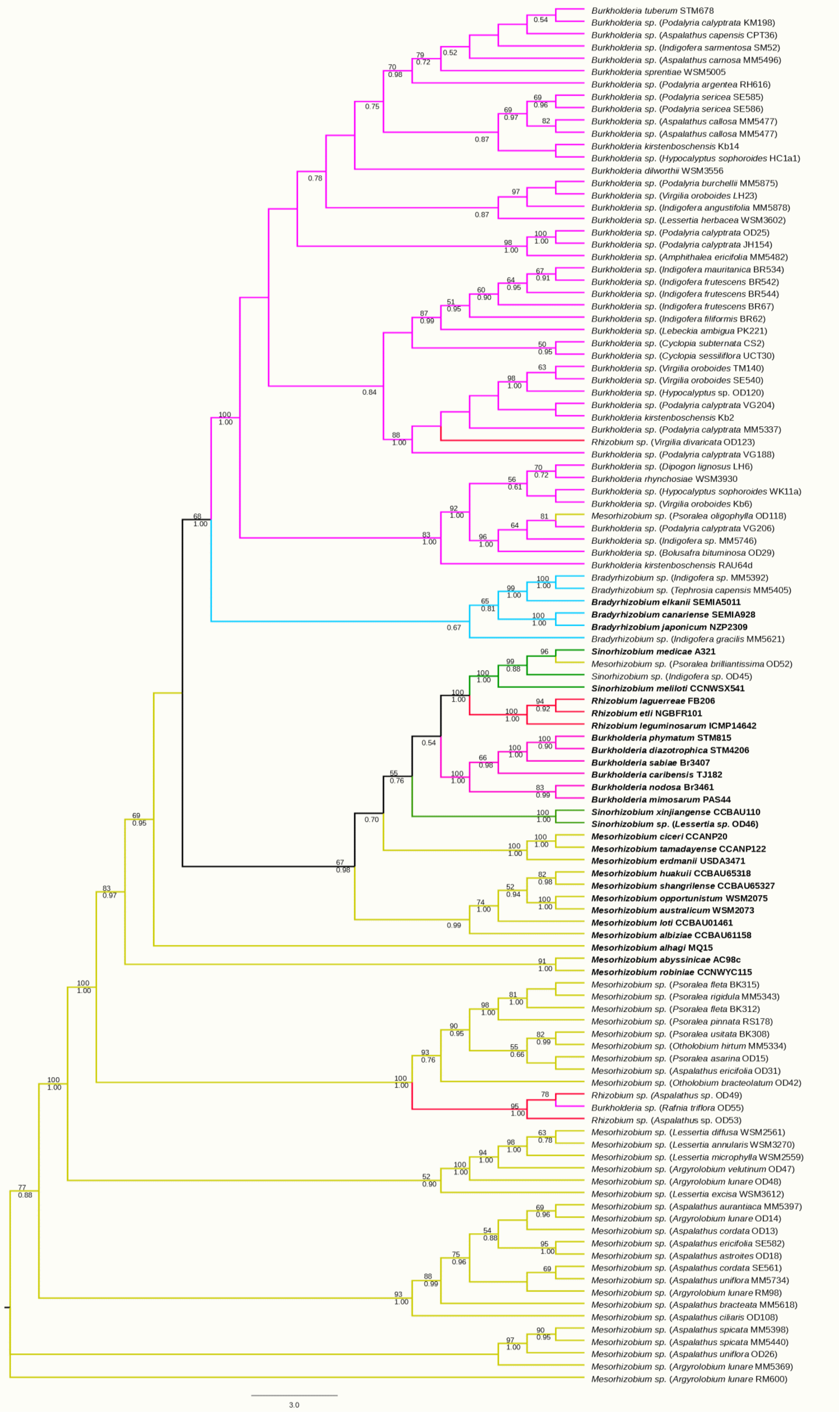
| **Taxa** | **Collector** | **Voucher** | ***rbcL*** | ***matK*** |
|---|---|---|---|---|
| *Aspalathus callosa* | Muasya & Stirton | AMM5477 | FAUCT329-11 | FAUCT329-11 |
| *Aspalathus carnosa* | Muasya & Stirton | AMM5496 | SAFH7264-15 | SAFH7264-15 |
| *Aspalathus spicata* | Muasya & Stirton | AMM5393 | FAUCT324-11 | FAUCT324-11 |
| *Aspalathus uniflora* | Muasya & Stirton | AMM5734 | SAFH7273-15 | SAFH7273-15 |
| *Aspalathus hispida* | Muasya & Stirton | AMM5808 | FAUCT346-11 | FAUCT346-11 |
| *Aspalathus ericifolia* | Muasya & Stirton | AMM4221 | FAUCT263-11 | FAUCT263-11 |
| *Aspalathus aurantiaca* | Muasya & Stirton | AMM5397 | SAFH7263-15 | SAFH7263-15 |
| *Lebeckia ambigua* | Van Wyk | 2900 | EU347917 | - |
| *Lebeckia sepiaria* | Le Roux et al | 10 | EU347936 | - |
| *Lebeckia pauciflora* | Le Roux et al | 7 | EU347935 | - |
| *Lebeckia meyeriana* | Van Wyk | 3551a | EU347904 | - |
| *Rafnia acuminata* | Schutte | 437 | AM931036 | - |
| *Rafnia triflora* | Dlodlo | 55 | SAFH7336-15 | SAFH7336-15 |
| *Lebeckia wrightii* | Muasya & Stirton | AMM5677 | SAFH7300-15 | SAFH7300-15 |
| *Calobota sericea* | J J M van der Meruve | 215 | - | GQ246144 |
| *Calobota cytisodes* | Muasya & Stirton | AMM5379 | SAFH7292-15 | SAFH7292-15 |
| *Calobota pungens* | Boatwright et al | 106 | EU347921 | - |
| *Calobota multiflora* | Boatwright et al | 138 | EU347926 | - |
| *Calobota spinescens* | Boatwright et al | 158 | EU347929.1 | - |
| *Argyrolobium lunare* | Muasya & Stirton | AMM5740 | FAUCT340-11 | FAUCT340-11 |
| *Argyrolobium velutinum* | Dlodlo | 47 | SAFH7261-15 | SAFH7261-15 |
| *Cyclopia pubescens* | Schutte | 685-689 | AM261723 | - |
| *Cyclopia buxifolia* | Muasya & Stirton | AMM3635 | SAFH7279-15 | SAFH7279-15 |
| *Cyclopia longifolia* | Vlok & Schutte | 422 | AM261719 | - |
| *Cyclopia sessiliflora* | Vlok & Schutte | 213 | AM261724 | - |
| *Cyclopia subternata* | Muasya & Stirton | AMM5069 | SAFH7284-15 | SAFH7284-15 |
| *Cyclopia maculata* | Stirton | 13381 | SAFH7283-15 | SAFH7283-15 |
| *Cyclopia genistoides* | JWB | 022 | JX572497 | JX518243.1 |
| *Cyclopia intermedia* | AL | 658 | AM261718.1 | - |
| *Cyclopia meyeriana* | Vlok & Schutte | 251 | AM261721.1 | - |

| **Taxa** | **Collector** | **Voucher** | ***rbcL*** | ***matK*** |
|---|---|---|---|---|
| *Cyclopia glabra* | Schutte | 558 | AM261717.1 | - |
| *Podalyria sericea* | Muasya & Stirton | AMM5384 | SAFH7326-15 | SAFH7326-15 |
| *Podalyria argentea* | Muasya & Stirton | AMM4193 | SAFH7322-15 | SAFH7322-15 |
| *Podalyria burchellii* | Muasya & Stirton | AMM5220 | FAUCT317-11 | FAUCT317-11 |
| *Podalyria calyptrata* | Muasya & Stirton | AMM5428 | FAUCT325-11 | FAUCT325-11 |
| *Amphithalea ericifolia* | Muasya & Stirton | AMM5482 | FAUCT330-11 | FAUCT330-11 |
| *Virgilia oroboides* | Muasya & Stirton | AMM5814 | SAFH7346-15 | SAFH7346-15 |
| *Virgilia divaricata* | Van Wyk | 879-888 | AM260737 | JX517500 |
| *Psoralea oligophylla* | Dlodlo | 118 | SAFH7331-15 | SAFH7331-15 |
| *Psoralea laxa* | Muasya & Stirton | AMM4325 | FAUCT279-11 | FAUCT279-11 |
| *Psoralea asarina* | Muasya & Stirton | AMM5360 | SAFH7328-15 | SAFH7328-15 |
| *Psoralea pinnata* | Muasya & Stirton | AMM3169 | FAUCT005-11 | FAUCT005-11 |
| *Psoralea brilliantissima* | Dlodlo | 52 | SAFH7329-15 | SAFH7329-15 |
| *Psoralea fleta* | Muasya & Stirton | AMM3241 | FAUCT036-11 | FAUCT036-11 |
| *Psoralea aphylla* | Muasya & Stirton | AMM3408 | FAUCT091-11 | FAUCT091-11 |
| *Otholobium virgatum* | Muasya & Stirton | AMM5357 | SAFH7319-15 | SAFH7319-15 |
| *Otholobium bracteolatum* | Muasya & Stirton | AMM3963 | FAUCT229-11 | FAUCT229-11 |
| *Otholobium hirtum* | Muasya & Stirton | AMM3991 | FAUCT232-11 | FAUCT232-11 |
| *Otholobium obliquum* | Muasya & Stirton | AMM3198.1 | FAUCT023-11 | FAUCT023-11 |
| *Otholobium zeyheri* | Muasya & Stirton | AMM5675 | SAFH7320-15 | SAFH7320-15 |
| *Dipogon lignosus* | Muasya & Stirton | AMM4240 | FAUCT268-11 | FAUCT268-11 |
| *Rhynchosia ferulifolia* | Muasya & Stirton | AMM5015 | FAUCT304-11 | FAUCT304-11 |
| *Bolusafra bituminosa* | Dlodlo | 29 | SAFH7274-15 | SAFH7274-15 |
| *Tephrosia capensis* | Muasya & Stirton | AMM5405 | SAFH7274-15 | SAFH7274-15 |
| *Indigofera filiformis* | Muasya & Stirton | AMM3482 | FAUCT108-11\| | FAUCT108-11\| |
| *Indigofera angustifolia* | Muasya & Stirton | AMM4229 | FAUCT265-11 | FAUCT265-11 |
| *Indigofera frutescens* | CS | 01 | JX572692.1 | JX517595 |
| *Lessertia frutescens* | Muasya & Stirton | AMM4150 | FAUCT254-11 | FAUCT254-11 |
| *Lessertia annularis* | H Merxmuller & W Giess | 2897 | - | JQ669621 |
| *Lessertia pauciflora* | Stirton | 13728 | SAFH7305-15 | SAFH7305-15 |
| *Lessertia excisa* | Stirton | 13306 | SAFH7303-15 | SAFH7303-15 |

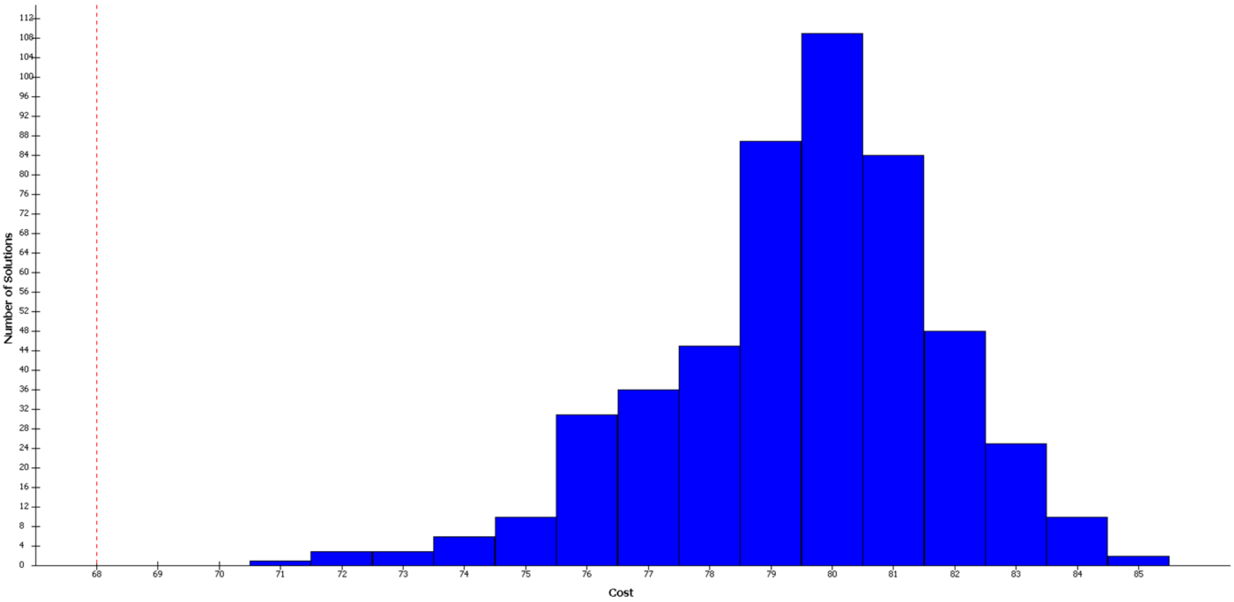
| **Taxa** | **Collector** | **Voucher** | ***rbcL*** | ***matK*** |
|---|---|---|---|---|
| *Lessertia herbacea* | Stirton | 13729 | SAFH7304-15 | SAFH7304-15 |
| *Hypocalyptus oxalidifolius* | Dlodlo | 120 | SAFH7278-15 | SAFH7278-15 |
| *Hypocalyptus oxalidifolius* | Stirton | 13779 | SAFH7286-15 | SAFH7286-15 |
| *Hypocalyptus coluteoides* | C. Burman | 1236 | - | AY386886.1 |
| *Hypocalyptus sophoroides* | Muasya & Stirton | AMM5968 | SAFH7287-15 | SAFH7287-15 |
| *Polygala subtilis* | Bello, M.A. | 719 | EU644683.1 | EU604033.1 |


Mesorhizobium sp. (Aspalathus hispida RM565)
Mesorhizobium sp. (Psoralea usitata BK308)
Mesorhizobium sp. (Psoralea usitata BK310)
Mesorhizobium sp. (Aspalathus cordata SE561)
Mesorhizobium sp. (Aspalathus ericifolia BR604)
Mesorhizobium sp. (Aspalathus chenopoda BR624)
Mesorhizobium sp. (Aspalathus ericifolia SE598)
Mesorhizobium sp. (Aspalathus ericifolia SE582)
Mesorhizobium sp. (Aspalathus uniflora OD26)
Mesorhizobium sp. (Aspalathus chenopoda RM559)
Mesorhizobium sp. (Aspalathus uniflora MM5734)
Mesorhizobium sp. (Aspalathus cordata SE562)
Mesorhizobium sp. (Aspalathus ericifolia CP627)
Mesorhizobium sp. (Aspalathus ericifolia CP628)
Mesorhizobium sp. (Aspalathus laricifolia MM5372)
Mesorhizobium sp. (Argyrolobium lunare OD14)
Mesorhizobium sp. (Aspalathus astroites OD18)
Mesorhizobium sp. (Aspalathus aurantiaca MM5397)
Mesorhizobium sp. (Aspalathus cordata OD13)
Mesorhizobium sp. (Argyrolobium lunare RM98)
Mesorhizobium sp. (Aspalathus chenopoda BR623)
Mesorhizobium sp. (Argyrolobium lunare RM580)
Mesorhizobium sp. (Aspalathus bracteata MM5618)
Mesorhizobium sp. (Otholobium hirtum MM5334)
Mesorhizobium sp. (Otholobium fruticans BR619)
Mesorhizobium sp. (Otholobium hirtum MM5376)
Mesorhizobium sp. (Aspalathus spicata MM5398)
Mesorhizobium sp. (Psoralea asarina RM548)
Mesorhizobium sp. (Psoralea pinnata RM569)
Mesorhizobium sp. (Argyrolobium lunare MM5369)
Mesorhizobium sp. (Aspalathus ericifolia OD31)
Mesorhizobium sp. (Otholobium virgatum RM568)
Mesorhizobium erdmanii USDA3471
Mesorhizobium opportunistum WSM2075
Mesorhizobium huakuii CCBAU65318
Mesorhizobium sp. (Psoralea pinnata SE592)
Mesorhizobium sp. (Psoralea aphylla SE597)
Mesorhizobium sp. (Psoralea oligophylla OD118)
Mesorhizobium sp. (Aspalathus ciliaris CS13166)
Mesorhizobium abyssinicae AC98c
Mesorhizobium sp. (Psoralea brilliantissima OD52)
Mesorhizobium sp. (Psoralea pinnata RH607)
Mesorhizobium sp. (Psoralea pinnata RH608)
Mesorhizobium sp. (Psoralea aphylla SE596)
Mesorhizobium sp. (Psoralea aphylla SE595)
Mesorhizobium sp. (Psoralea pinnata RS176)
Mesorhizobium sp. (Psoralea rigidula MM5343)
Mesorhizobium sp. (Psoralea pinnata RS178)
Mesorhizobium sp. (Psoralea fleta BK312)
Mesorhizobium sp. (Psoralea fleta BK315)
Mesorhizobium shangrilense CCBAU65327
Mesorhizobium sp. (Otholobium hirtum LH612)
Mesorhizobium sp. (Otholobium hirtum LH609)
Mesorhizobium sp. (Otholobium hirtum LH610)
Mesorhizobium ciceri CCANP20
Mesorhizobium loti CCBAU01461
Mesorhizobium sp. (Otholobium virgatum MM5333)
Mesorhizobium sp. (Lessertia excisa WSM3612)
Mesorhizobium sp. (Aspalathus ciliaris OD108)
Mesorhizobium australicum WSM2073
Mesorhizobium sp. (Otholobium virgatum RM567)
Mesorhizobium sp. (Aspalathus ciliaris MM5361)
Mesorhizobium alhagi MQ15
Mesorhizobium sp. (Otholobium bracteolatum OD42)
Mesorhizobium albiziae CCBAU61158
Mesorhizobium tamadayense CCANP122
Mesorhizobium sp. (Lessertia annularis WSM3270)
Mesorhizobium sp. (Lessertia diffusa WSM2561)
Mesorhizobium sp. (Argyrolobium velutinum OD47)
Mesorhizobium sp. (Argyrolobium lunare OD48)
Mesorhizobium robiniae CCNWYC115
Mesorhizobium mediterraneum USDA3392
Mesorhizobium sp. (Lessertia microphylla WSM2559)
Bradyrhizobium sp. (Indigofera sp. MM5621)
Bradyrhizobium canariense SEMIA928
Bradyrhizobium japonicum NZP2309
Bradyrhizobium sp. (Tephrosia capensis MM5405)
Bradyrhizobium sp. (Indigofera sp. MM5392)
Bradyrhizobium elkanii SEMIA5011
Rhizobium leguminosarum ICMP14642
Rhizobium laguerreae FB206
Rhizobium sp. (Aspalathus chenopoda BR583)
Rhizobium etli NGFR101
Rhizobium sp. (Virgilia divaricata OD123)
Rhizobium sp. (Aspalathus sp. OD49)
Rhizobium sp. (Aspalathus sp. OD53)
Sinorhizobium mexicanum ITTGS4
Sinorhizobium medicae NHBM16
Sinorhizobium sp. (Indigofera sp. OD45)
Sinorhizobium meliloti CCBAU83884
Sinorhizobium xinjiangense CCNWNX165
Sinorhizobium sp. (Lessertia sp. OD46)
Burkholderia sp. (Aspalathus capensis CPT36)
Burkholderia sp. (Indigofera sarmentosa SM52)
Burkholderia sp. (Podalyria sericea SE586)
Burkholderia sp. (Podalyria sericea SE585)
Burkholderia sp. (Podalyria sericea SE584)
Burkholderia sp. (Podalyria argentea RH616)
Burkholderia sp. (Amphithalea ericifolia MM5482)
Burkholderia sp. (Aspalathus callosa MM5477)
Burkholderia tuberum STM678
Burkholderia sp. (Podalyria calyptrata KM198)
Burkholderia sp. (Aspalathus carnosa MM5496)
Burkholderia sp. (Lessertia herbacea WSM3602)
Burkholderia sprentiae WSM5005
Burkholderia sp. (Hypocalyptus sophoroides HC1a1)
Burkholderia symbiotica JPY345
Burkholderia cepacia ATCC25416
Burkholderia unamae MTI641
Burkholderia nodosa Br3437
Burkholderia phymatum STM815
Burkholderia sabiae Br3407
Burkholderia diazotrophica JPY461T
Burkholderia caribensis TJ182
Burkholderia sp. (Podalyria calyptrata VG204)
Burkholderia kirstenboschensis RAU64d
Burkholderia kirstenboschensis Kb14
Burkholderia sp. (Virgilia oroboides LH23)
Burkholderia kirstenboschensis Kb2
Burkholderia sp. (Hypocalyptus sophoroides WK11a)
Burkholderia dilworthii WSM3556
Burkholderia sp. (Rafnia acuminata OD22)
Burkholderia rhynchosiae WSM3937
Burkholderia rhynchosiae WSM3930
Burkholderia phenoliruptrix AC1100
Burkholderia sp. (Cyclopia subternata CS2)
Burkholderia sp. (Dipogon lignosus LH6)
Burkholderia sp. (Podalyria burchellii MM5875)
Burkholderia sp. (Cyclopia sessiliflora UCT30)
Burkholderia xenovorans LMG21463
Burkholderia ginsengisoli LMG24044
Burkholderia sp. (Virgilia oroboides Kb6)
Burkholderia sp. (Indigofera sp. MM5746)
Burkholderia sp. (Podalyria calyptrata VG206)
Burkholderia sp. (Virgilia oroboides SE540)
Burkholderia sp. (Podalyria calyptrata VG188)
Burkholderia phytofirmans LMG22487
Burkholderia sp. (Podalyria calyptrata VG205)
Burkholderia sp. (Podalyria calyptrata MM5337)
Burkholderia fungorum LMG16225
Burkholderia sp. (Rafnia triflora OD55)
Burkholderia sp. (Podalyria calyptrata OD25)
Burkholderia sp. (Podalyria calyptrata JH154)
Burkholderia sp. (Indigofera filiformis BR62)
Burkholderia sp. (Indigofera frutescens BR67)
Burkholderia sp. (Virgilia oroboides TM140)
Burkholderia sp. (Virgilia oroboides SE541)
Burkholderia sp. (Indigofera frutescens BR542)
Burkholderia sp. (Indigofera mauritanica BR534)
Burkholderia sp. (Indigofera frutescens BR543)
4.0

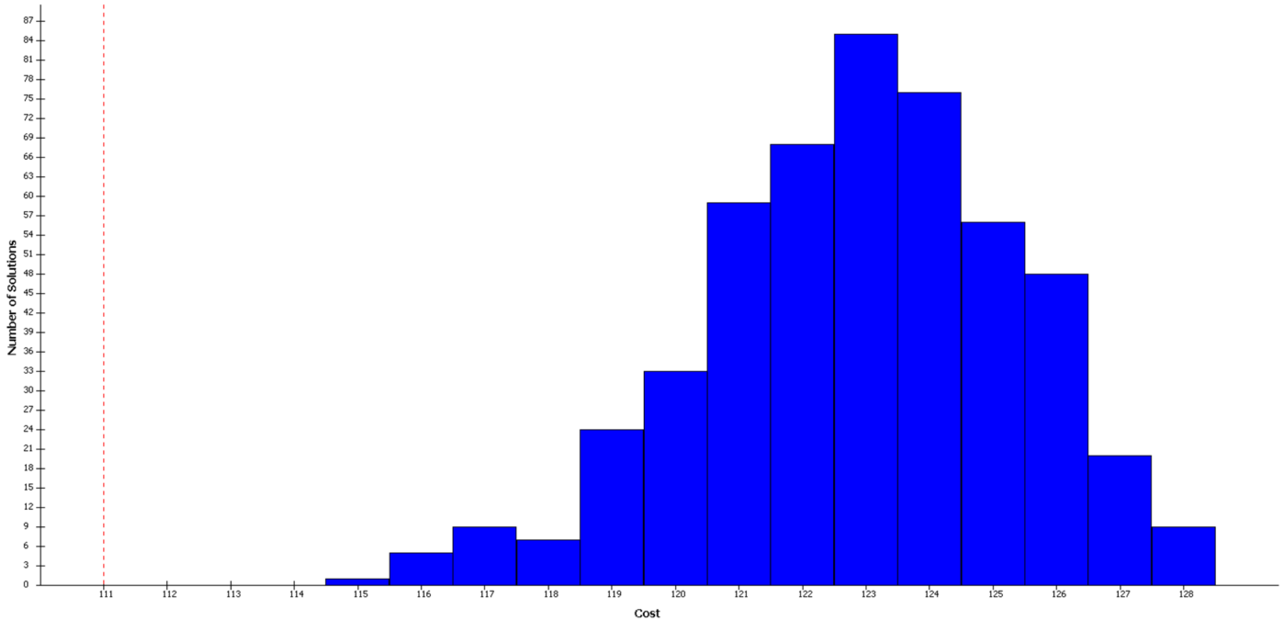
**Appendix C3.** Phylogenetic relationships of rhizobial strains based on 16S rRNA and *recA* data. Clades are coloured to reflect the different rhizobial genera and ML bootstrap (%) and BI posterior probabilities are shown above and below the nodes, respectively. The tip labels indicate strain identities and the legume hosts from which they were isolated are placed in parentheses. Non-CCR strains are labelled in bold.


Burkholderia tuberum STM678
Burkholderia sp. (Podalyria calyptrata KM198)
Burkholderia sp. (Aspalathus capensis CPT36)
Burkholderia sp. (Indigofera sarmentosa SM52)
Burkholderia sp. (Aspalathus carnosa MM5496)
Burkholderia sprentiae WSM5005
Burkholderia sp. (Podalyria argentea RH616)
Burkholderia sp. (Podalyria sericea SE585)
Burkholderia sp. (Podalyria sericea SE586)
Burkholderia sp. (Aspalathus callosa MM5477)
Burkholderia sp. (Aspalathus callosa MM5477)
Burkholderia kirstenboschensis Kb14
Burkholderia sp. (Hypocalyptus sophoroides HC1a1)
Burkholderia dilworthii WSM3556
Burkholderia sp. (Podalyria burchellii MM5875)
Burkholderia sp. (Virgilia oroboides LH23)
Burkholderia sp. (Indigofera angustifolia MM5878)
Burkholderia sp. (Lessertia herbacea WSM3602)
Burkholderia sp. (Podalyria calyptrata OD25)
Burkholderia sp. (Podalyria calyptrata JH154)
Burkholderia sp. (Amphithalea ericifolia MM5482)
Burkholderia sp. (Indigofera mauritanica BR534)
Burkholderia sp. (Indigofera frutescens BR542)
Burkholderia sp. (Indigofera frutescens BR544)
Burkholderia sp. (Indigofera frutescens BR67)
Burkholderia sp. (Indigofera filiformis BR62)
Burkholderia sp. (Lebeckia ambigua PK221)
Burkholderia sp. (Cyclopia subternata CS2)
Burkholderia sp. (Cyclopia sessiliflora UCT30)
Burkholderia sp. (Virgilia oroboides TM140)
Burkholderia sp. (Virgilia oroboides SE540)
Burkholderia sp. (Hypocalyptus sp. OD120)
Burkholderia sp. (Podalyria calyptrata VG204)
Burkholderia kirstenboschensis Kb2
Burkholderia sp. (Podalyria calyptrata MM5337)
Rhizobium sp. (Virgilia divaricata OD123)
Burkholderia sp. (Podalyria calyptrata VG188)
Burkholderia sp. (Dipogon lignosus LH6)
Burkholderia rhynchosiae WSM3930
Burkholderia sp. (Hypocalyptus sophoroides WK11a)
Burkholderia sp. (Virgilia oroboides Kb6)
Mesorhizobium sp. (Psoralea oligophylla OD118)
Burkholderia sp. (Podalyria calyptrata VG206)
Burkholderia sp. (Indigofera sp. MM5746)
Burkholderia sp. (Bolusafra bituminosa OD29)
Burkholderia kirstenboschensis RAU64d
Bradyrhizobium sp. (Indigofera sp. MM5392)
Bradyrhizobium sp. (Tephrosia capensis MM5405)
Bradyrhizobium elkanii SEMIA5011
Bradyrhizobium canariense SEMIA928
Bradyrhizobium japonicum NZP2309
Bradyrhizobium sp. (Indigofera gracilis MM5621)
Sinorhizobium medicae A321
Mesorhizobium sp. (Psoralea brilliantissima OD52)
Sinorhizobium sp. (Indigofera sp. OD45)
Sinorhizobium meliloti CCNWSX541
Rhizobium laguerreae FB206
Rhizobium etli NGBFR101
Rhizobium leguminosarum ICMP14642
Burkholderia phymatum STM815
Burkholderia diazotrophica STM4206
Burkholderia sabiae Br3407
Burkholderia caribensis TJ182
Burkholderia nodosa Br3461
Burkholderia mimosarum PAS44
Sinorhizobium xinjiangense CCBAU110
Sinorhizobium sp. (Lessertia sp. OD46)
Mesorhizobium ciceri CCANP20
Mesorhizobium tamadayense CCANP122
Mesorhizobium erdmanii USDA3471
Mesorhizobium huakuii CCBAU65318
Mesorhizobium shangrilense CCBAU65327
Mesorhizobium opportunistum WSM2075
Mesorhizobium australicum WSM2073
Mesorhizobium loti CCBAU01461
Mesorhizobium albiziae CCBAU61158
Mesorhizobium alhagi MQ15
Mesorhizobium abyssinicae AC98c
Mesorhizobium robiniae CCNWYC115
Mesorhizobium sp. (Psoralea fleta BK315)
Mesorhizobium sp. (Psoralea rigidula MM5343)
Mesorhizobium sp. (Psoralea fleta BK312)
Mesorhizobium sp. (Psoralea pinnata RS178)
Mesorhizobium sp. (Psoralea usitata BK308)
Mesorhizobium sp. (Otholobium hirtum MM5334)
Mesorhizobium sp. (Psoralea asarina OD15)
Mesorhizobium sp. (Aspalathus ericifolia OD31)
Mesorhizobium sp. (Otholobium bracteolatum OD42)
Rhizobium sp. (Aspalathus sp. OD49)
Burkholderia sp. (Rafnia triflora OD55)
Rhizobium sp. (Aspalathus sp. OD53)
Mesorhizobium sp. (Lessertia diffusa WSM2561)
Mesorhizobium sp. (Lessertia annularis WSM3270)
Mesorhizobium sp. (Lessertia microphylla WSM2559)
Mesorhizobium sp. (Argyrolobium velutinum OD47)
Mesorhizobium sp. (Argyrolobium lunare OD48)
Mesorhizobium sp. (Lessertia excisa WSM3612)
Mesorhizobium sp. (Aspalathus aurantiaca MM5397)
Mesorhizobium sp. (Argyrolobium lunare OD14)
Mesorhizobium sp. (Aspalathus cordata OD13)
Mesorhizobium sp. (Aspalathus ericifolia SE582)
Mesorhizobium sp. (Aspalathus astroites OD18)
Mesorhizobium sp. (Aspalathus cordata SE561)
Mesorhizobium sp. (Aspalathus uniflora MM5734)
Mesorhizobium sp. (Argyrolobium lunare RM98)
Mesorhizobium sp. (Aspalathus bracteata MM5618)
Mesorhizobium sp. (Aspalathus ciliaris OD108)
Mesorhizobium sp. (Aspalathus spicata MM5398)
Mesorhizobium sp. (Aspalathus spicata MM5440)
Mesorhizobium sp. (Aspalathus uniflora OD26)
Mesorhizobium sp. (Argyrolobium lunare MM5369)
Mesorhizobium sp. (Argyrolobium lunare RM600)
3.0

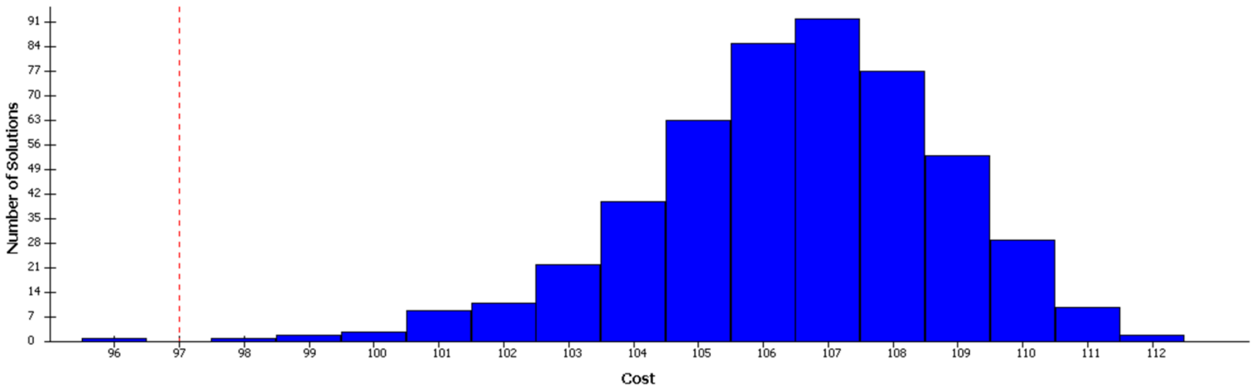
**Appendix C4.** Phylogenetic relationships of rhizobial strains based on *nodA* data. Clades are coloured to reflect the different rhizobial genera and ML bootstrap (%) and BI posterior probabilities are shown above and below the nodes, respectively. The tip labels indicate strain identities and the legume hosts from which they were isolated are placed in parentheses. Non-CCR strains are labelled in bold.



**Appendix C5.** Plot of the frequency distribution of event costs generated by permutation of 1000 random trees in the analysis of congruence between the chromosomal and *nodA* gene trees. The vertical dashed line shows the observed cost of the original reconciliation analysis relative to the event costs for the randomized trees. An overlap of the observed and randomized cost events indicates no significant congruence between the trees.



**Appendix C6.** Plot of the frequency distribution of event costs generated by permutation of 1000 random trees in the analysis of congruence between the legume and the chromosomal gene tree of the rhizobia. The vertical dashed line shows the observed cost of the original reconciliation analysis relative to the event costs for the randomized trees. An overlap of the observed and randomized cost events indicates no significant congruence between the trees.



**Appendix C7.** Plot of the frequency distribution of event costs generated by permutation of 1000 random trees in the analysis of congruence between the legume and the *nodA* gene tree of the rhizobia. The vertical dashed line shows the observed cost of the original reconciliation analysis relative to the event costs for the randomized trees. An overlap of the observed and randomized cost events indicates no significant congruence between the trees.

## Appendix D

Biological Sciences

**Department of Biological Sciences**
University of Cape Town
Private Bag X3, Rhodes Gift, 7701 South Africa
Tel: +27 (0) 21 650 3604 Fax: +27 (0) 21 650 3301
http://www.biologicalsciences.uct.ac.za/

Appendix D1 15 February 2018

**TO WHOM IT MAY CONCERN**

I hereby testify that the first author of the following articles was the lead researcher in the work done towards completing the articles:

1. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2017. Distinct edaphic habitats are occupied by discrete legume assemblages with unique indicator species in the Cape Peninsula of South Africa. *Journal of Plant Ecology* rtx027. doi: 10.1093/jpe/rtx027.
2. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2018. Differential preference of Burkholderia and Mesorhizobium to pH and soil types in the Core Cape Subregion, South Africa. *Genes* 9, 2. doi: 10.3390/genes9010002.
3. Dludlu, M.N., Chimphango, S.B.M., Walker, G., Stirton, C.H., Muasya, A.M., (accepted). Horizontal gene transfer among rhizobia of the Core Cape Subregion of southern Africa. South African Journal of Botany.

Mr Dludlu conceptualised and designed the research and independently conducted the experiments and analysed the data. He wrote the manuscripts and dealt with referees comments. The co-authors contributed by facilitating the work through providing special expertise, funds, facilities, supervision, and comments and suggestions on the manuscripts.

Yours sincerely,

A/Prof A. Muthama Muasya, PhD Supervisor

Biological Sciences

Department of Biological Sciences
University of Cape Town

Private Bag X3, Rhodes Gift, 7701 South Africa
Tel: +27 (0) 21 650 3604 Fax: +27 (0) 21 650 3301
http://www.biologicalsciences.uct.ac.za/

Appendix D2

16 February 2018

**To Whom It May Concern**

I hereby testify that Meshack N Dludlu, the first author of the following published/accepted papers, was the lead researcher in the work done towards completing the papers.

1. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2017. Distinct edaphic habitats are occupied by discrete legume assemblages with unique indicator species in the Cape Peninsula of South Africa. *Journal of Plant Ecology* rtx027. doi: 10.1093/jpe/rtx027.
2. Dludlu, M.N., Chimphango, S.B.M., Stirton, C.H., Muasya, A.M., 2018. Differential preference of Burkholderia and Mesorhizobium to pH and soil types in the Core Cape Subregion, South Africa. *Genes* 9, 2. doi: 10.3390/genes9010002.
3. Dludlu, M.N., Chimphango, S.B.M., Walker, G., Stirton, C.H., Muasya, A.M., (accepted). Horizontal gene transfer among rhizobia of the Core Cape Subregion of southern Africa. South African Journal of Botany.

Mr Dludlu conceptualized and designed the research, and independently conducted the experiments, data collection, and statistical analysis. He independently wrote the manuscripts supported by comments and suggestions from the coauthors and dealt with referees' comments in collaboration with the co-authors. My contribution to the project was through providing specialist expertise, supervision and comments and suggestions on the manuscripts.

Yours sincerely,

Dr Samson Chimphango
PhD Co-supervisor