Nodding off: Investigating global relationships in wild Fabeae-rhizobia symbioses

Sarah Roberts S1888329

20th August 2020

Thesis submitted in partial fulfilment for the
MSc in the Biodiversity and Taxonomy of Plants

Abstract

Wild plants from the large tribe Fabeae are distributed almost globally. They and their bacterial symbionts are often especially well adapted to harsh conditions. Few studies have examined the influence of biogeography on the rhizobia.

A large new dataset of whole draft genome sequences of wild *Vicia* and *Lathyrus* symbionts from Eurasia, Japan and North and South America was used to investigate whether evolutionary relatedness in the plant hosts or their biogeography might exert a stronger influence on their distribution. Average Nucleotide Identity was calculated to analyse the species boundaries of the group. Separate phylogenies using 120 core chromosomal genes, *nod*A, *nod*C and *nod*D nodulation genes, and 16S rRNA were also constructed to further explore phylogenetic and symbiotic relationships.

The results showed trends for geographical grouping over large areas with robust separation between genospecies in different locations, suggesting some local specialization but that overall biogeography might have stronger influence on the rhizobia. Host plants appeared generally promiscuous and there was no apparent correlation with the evolutionary phylogeny of the hosts and their choice of rhizobial partners. The 16S sequence is highly conserved across the accessions. Phylogenies for the nodulation genes reflected the selection by Fabeae plants of biovar *viciae* genes.

Acknowledgements

Massive thanks to my supervisor Greg Kenicer for his tremendous knowledge, patience, wit, wisdom, inspiration and conspiration throughout this project.

My sincere thanks to Peter Young and Euan James for the benefit of their very special rhizobial expertise.

Thanks to Greg Kenicer, Matias Morales, Euan James and Louis Ronse de Craene for collecting the microbes for my magical dataset, and thanks to Catherine Kidner, Mark Hughes and Flávia Pezzini for their timely advice on getting it to work.

My gratitude to all the MSc tutors for their wonderful instruction in the planty wonder of the world.

My thanks to all my MSc classmates, it was a privilege to share the journey with.

Thanks to Jacqui Pestell, for making working at the same time possible.

Thanks to Thomas and Jo Green for their huge support throughout the MSc.

And finally thanks to my beloved Owen, for his unwavering support and wizardly help with python-wrangling, without whom this project would have been risible.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Project overview

The symbiosis that occurs in root nodules between rhizobial bacteria and legumes enables to plants to obtain nitrogen from the atmosphere rather than relying on its presence in the soil. Biological nitrogen fixation (BNF) by plants represents the most cost-effective and least polluting and way of enriching soils, especially in regions which are subject to climate stresses, and can play a key part in mitigating the effects of climate change (Zahran, 1999). The most important BNF contribution to ecosystems is made by the symbioses between *Rhizobium leguminosarum* and legumes, which include the large tribe Fabeae (Herridge et al., 2008). The symbiosis allows rapid colonisation of new terrains and endurance in harsh environments for the plants (Sprent et al., 2017)

Most studies of the Fabeae-rhizobia relationship focus on crops. These will not be considered here except where appropriate. The role of biogeography of the host plants in the partnership has not been extensively studied, as rhizobia may frequently be consciously or unconsciously distributed by humans along with the legumes. It is possible that the rhizobia of wild legumes may give a more resolved biogeographical picture.

The aim of this study was to use the explanatory power of whole rhizobial genomes to look for the influence of evolutionary relatedness of the plant hosts and the influence of biogeography on the symbiosis. The dataset consisted of 68 whole draft genome sequences extracted from bacteria cultured from wild Fabeae species from UK, Crete, Spain, North America, Canada, Japan, Chile and Argentina. The dataset includes samples from various distances, habitats and environmental conditions, in accordance with recommendations to tackle the complex mechanisms generating microbial biogeographic patterns (Martiny et al., 2006).

In order to do this, an Average Nucleotide Identity (ANI) table was constructed to analyse the species boundaries of the group and this was related to the geographical area of collection. A phylogeny using 120 vertically transmitted core genes was constructed as proxy for the species. Phylogenies of 16S rRNA sequence and of nodulation genes nodA, nodC and nodD were also constructed. The results of these were analysed against the phylogeny and biogeography of the host plants. A review of the literature on host-rhizobia relationships in wild Fabeae was conducted to interpret the results.

1.2. Tribe Fabeae

The tribe Fabeae are Inverted Repeat Lacking Clade legumes comprising around 380 species of legumes in large genera *Lathyrus* L. (160 species) and *Vicia L*. (about 140 species) (Schaefer et al., 2012) in family Leguminosae (Fabaceae). Until recently, *Lens* Mill. was nested in *Vicia*, *Pisum* L. and *Vavilovia* Fed. were nested in *Lathyrus*, and the genus *Vicia* was paraphyletic with L*athyrus* nested within. This thesis is written in a time of transition for the taxonomy of Fabeae, see Table 1. for and overview of the names in transition and the terms used in this project. Recent molecular work has resolved some of the relationships within the Tribe. Genera *Ervum* and *Ervilia* Link are being established to resolve the paraphyly of *Vicia*, (Coulot and Rabaute, 2016). Pea, *Pisum sativum* L. (three species) has been moved into *Lathyrus* as *Lathyrus oleraceus* Lam. Monospecific genus *Vavilovia formosa* (Steven) Fed was found to be nested within *Lathyrus* and sister to pea, and has been recombined as *Lathyrus formosus* (Steven) Kenicer (Kenicer and Parsons, 2021). The four former species of lentil *Lens* have been transferred to *Vicia* as *Vicia lens* (Miller) De Candolle. It should be noted that *Vicia nigricans* here refers to the large New World species (our samples were South American) *V. nigricans* Hooker & Arnott, rather than the newly renamed *Vicia nigricans* (M. Bieberstein) Cosson & Germain. This latter name should take precedence as the basionym (*Ervum nigricans* Bieberstein (1808) takes precedence. The South American species is awaiting an updated name.

Former name	**New name**
Vavilovia formosa (Steven) Fed.	*Lathyrus formosus* (Steven) Kenicer
Pisum sativum L.	*Lathyrus oleraceus* Lam.
Vicia tetrasperma (L.) Schreb.	*Ervum tetrasperma* L.
Vicia hirtsuta (L.) Gray	*Vicilla hirsuta* (L.) Opiz
Lens nigricans (M.Bieb.) Webb & Berthel.	*Vicia nigricans* (M. Bieberstein) Cosson & Germain
Vicia nigricans Hooker & Arnott	?

Table 1. Fabeae species with names in transition.
The names used in this study are highlighted in red.

All other scientific plant names in this study are in accordance with their listing in *International Plant Names Index* (IPNI, 2021)

Tribe Fabeae is thought to have evolved in the mid Miocene period, in the Eastern Mediterranean. There have been many long-distance dispersal events into Eurasia, Africa and at least seven times to the Americas (Schaefer et al., 2012).

Tribe Fabeae have an almost worldwide distribution (Schaefer et al., 2012). This includes important crops pea, broad bean and lentil, and their relationships with symbionts are well studied. The tribe also contains some important ornamentals (Kenicer and Parsons, 2021), as well as forage and fodder and green manure crops (Drouin et al., 1996), many of which are of increasing potential importance to agriculture (Cevheri, 2014; Gritli et al., 2020; Kumar et al., 2013; Mohammed et al., 2020). Fabeae are also important to pollinators and honey production (Sui et al., 2009). Tribe Fabeae contains some of the earliest crops. There is evidence for the cultivation of pea, lentil, broad bean, and bitter vetch *(V. ervilia)* from Neolithic times, becoming staples of Bronze Age diets in the Near East and Europe (Zohary and Hopf, 1973). In Lathyrus there are some economically and horticulturally important ornamentals: *L. odoratus, L. vernus, L. grandiflorus, L. tuberosus, L. latifolius*, and there are potentially many more suitable species (Güneş, 2019; Kenicer, 2008; Kenicer and Parsons, 2021).

Many plants in the large genera *Lathyrus* and *Vicia* occupy interesting niches and are adapted to harsh conditions and poor, degraded, polluted, saline, alkaline, acid or drought-prone soils, and special adaptations to these conditions are equally important in the nodules as they are in the plants (Abdelkrim et al., 2019; Gritli et al., 2020; Li et al., 2016; Mohammed et al., 2020). Many Fabeae plants enjoy ruderal or weedy habitats with wide distributions and many can tolerate harsh or stressful conditions (Ampomah et al., 2012). Study of their symbionts is necessary in order to help their conservation (Atallah et al., 2008; Smýkal et al., 2017).

1.3. Rhizobia

Rhizobia are free-living soil organisms that can reduce nitrogen from the air using the enzyme nitrogenase. This process is known as biological nitrogen fixation. Rhizobia in dense populations can contribute over 100 kilograms of nitrogen per hectare per year (Vitousek et al., 2002), and up to 330 kg per hectare per year in symbiosis with a legume such as *Vicia faba* (Peoples et al., 1995).

The primary symbiotic rhizobial partners of Fabeae are *Rhizobium leguminosarum* species complex (Rlc) (Young, John M et al., 2001; J. Peter W. Young et al., 2021). A review of the literature on legume-rhizobia symbioses in field soils was recently published, finding twenty-eight recent studies reporting *Rhizobium* as the only Fabeae symbionts, and reporting a high degree of specificity with *Rhizobium leguminosarum* for the tribe Fabeae, although this was not exclusive and other rhizobia within various genera of Alphaproteobacteria can nodulate with the plants (Andrews and Andrews, 2017).

1.4. Rhizobium leguminosarum

Rhizobium leguminosarum (Frank 1879) Frank 1889 AL was first described by Frank in 1879, at that point the name was used to denote any bacterium that was found nodulating with plants. (Rivas et al., 2009). More than one thousand studies have been published on the species (Young et al., 2006).

Relationships within Alphaproteobacteria have been recently revised. Under current circumscription, the genus *Rhizobium* is not monophyletic. A revision of R*hizobium* Frank was published in 2001, amalgamating *Agrobacterium*, *Allorhizobium* and *Rhizobium* into the monophyletic species *Rhizobium* (Young, John M et al., 2001). The taxonomic status of the species Rhizobium leguminosarum was revised in 2008 (Ramírez-Bahena et al., 2008a) and again in 2020 (Hördt et al., 2020). However, a sister for the genus *Rhizobium* has not been identified. *Rhizobium anhuiense*, characterized by Zhang et al. (2015), was found to be sister to *R. leguminosarum* (J. Peter W. Young et al., 2021).

1.5. Root nodules

Co-evolution between the legumes and rhizobia is directed by symbiosis-specific genes in each partner (Mutch and Young, 2004). The symbiosis takes place within root nodules are specialized cells and tissues formed as a result of symbiogenesis between the bacteria and the legume hosts. Nodules were described by Chapman and Margulis (1998) as "Our best example of symbiospecific morphogenesis"

Fabeae produce indeterminate nodules with bacteroids which usually occupy symbiosomes singly, often become very enlarged and terminally differentiated. IRLC legumes exert stringent controls over their symbionts which may help them to be particularly effective at nitrogen fixation (Sprent et al., 2017). Nodules generally contain one strain of nitrogen-fixing rhizobia but other microbes live inside the nodules and the surrounding rhizosphere (Muresu et al., 2008). The symbiotic partnership of Fabeae and the rhizobia is mediated by symbiosis genes of biovar *viciae* (Mutch and Young, 2004).

Host plants are generally promiscuous, although in particular habitats strain-specific mutualisms are found to occur (Andrews and Andrews, 2017). A comparative study of wild and cultivated legumes found that plants from tribe Fabeae nodulate fairly indiscriminately with *Rhizobium leguminosarum* biovar *viceae* and the wild legumes were more promiscuous than the crop plants (Mutch and Young, 2004).

1.6. Climate change, changing soils, and the need for food security

Climate change, changing soils, the need for food security and the need to reduce industrially produced N fertilizers give urgency to the study of wild relatives of crop plants and their bacterial partners.

Despite the importance to ecosystems of the large number of wild Fabeae plants, their diversity and widespread inhabitation of temperate, Mediterranean and semi-arid parts of the world, symbiotic partners of wild Fabeae plants are little studied.

Biogeographical studies of microbes are necessary in order to understand their dispersal limitation, population composition, diversity, and the relationships they have with other organisms and their ecosystems (Martiny et al., 2006). Nodulated legumes have a key role to play in land-based climate change mitigation approaches (Roe et al., 2019). Crop rotation using legumes can reduce reliance on synthetic fertilizers and legumes in grasslands promote carbon sequestration (Soussana et al., 2004). Shifting to mixed agricultural systems can reduce nitrogen runoff (Chatterjee, 2009) and reduce CO_2 emissions by from crop systems by 28%- 60% (Jensen et al., 2012).

Strains of bacterial symbionts of wild plants adapted to conditions such as salinity, drought, high temperatures and acid soils enjoyed wide host ranges, and such strains around the world could help tolerate changing climates and conditions (Moschetti et al., 2005).

Improved understanding of the phytomicrobiome will help us understand how better to inoculate legumes for crops and conservation (Martínez-Hidalgo and Hirsch, 2017) . The study of bacterial strain genomes present in wild species can uncover which bacterial strains and recombinants have benefitted from selection pressures of particular hosts, soils and climates. (Zahran, 2001)Wild plants and their nodulating partners can potentially provide valuable genetic diversity as germplasm as a source of alleles and traits that can be used to improve productivity and resistance to stress in their crop relatives (Smýkal, 2011; Smýkal et al., 2015), , and potential inoculants to assist with conservation of rare species and revegetation of degraded soils and preservation of biodiversity (Atallah et al., 2008). . As local bacteria that are less efficient nitrogen fixers can sometimes outcompete inoculant strains to form nodules with plants (Boivin et al., 2020), this could allow us to select strains best adapted for inoculation (Baymiev et al., 2011).

2. Literature review

2.1. Introduction: aim of the review

This chapter presents a comprehensive review of the current state of research on wild members of tribe Fabeae and their rhizobial partners. Relatively few papers have been published on the wild partnership, in contrast to the large body of work on nodulation in widespread agricultural Fabeae crops–pea, broad bean, and lentil, which will not be reviewed here. However many studies look at both crop and wild systems together, particularly in research on orphan crops, neglected and underutilized semi-wild fodder and forage crops in which the wild symbioses are harnessed for cultivation, these will be discussed here as appropriate.

Some reviews which are important to this study have looked at work done on both crop and wild systems together and a large part of current understanding of the Fabeae-rhizobia relationship comes from studies on cultivated species. Work from global studies on legumes that had been carried out in the field on the specificity in legume-rhizobia symbioses is reviewed in Andrews and Andrews (2017). They looked at studies across all families in Leguminosae and found seventeen (mostly crop-based) studies in tribe Fabeae that reported a highly specific relationship with *Rhizobium leguminosarum* sv. viciae as the preferred symbiont in nodules. Other studies showed that effective nodulation was also reported with other rhizobia but similarity in *nif*H and *nod*C genes suggested that the nodulation genes were transferred laterally to bacteria adapted to particular soils, (Andrews and Andrews, 2017). It is expected that the symbionts of wild legumes will follow the same patterns, but that studies of wild mutualists may be able to help clarify host-bacteria relationships in ways that may not be possible by looking at legumes that have a history of use by humans, as it is more likely that the rhizobia associated with the crop plants have been unconsciously distributed along with the crops (Mutch and Young, 2004).

A global review of the biogeography of nodulated plants and their symbionts is equally valuable for an overview of the general literature on the migrations of the systems (Sprent et al., 2017), although Fabeae are not a chief focus of their study. However a major study on the systematics, biogeography and character evolution of tribe Fabeae (Schaefer et al., 2012) and a highly resolved study of the *Rhizobium leguminosarum* species complex (J. Peter W. Young et al., 2021) provide invaluable background to this study and will be referred to extensively.

The wild Fabeae plants harbour more botanical diversity and might be expected to harbour more symbiont diversity than crop species. A review of this literature is worthwhile as the studies of the wild systems give special perspectives which can help us to understand the operation and coevolution of the partners and of the partnership.

In order to find published studies of the wild Fabeae-rhizobia symbiosis, I conducted searches of Pubmed, Google Scholar and Web of Science. Initial searches using keywords "Fabeae"OR "Lathyrus" OR"Vicia"OR "Vavilovia" OR"Lens"OR "Pisum" AND "Rhizobium"OR "rhizobia"OR "nodulation" returned over 60,000 results. After irrelevant subjects and exclusively crop-focused papers were filtered from the results around 490 papers remained, which were further filtered to leave only those with a major focus on symbioses with wild host plants, around 80 articles. Of these, 36 *** were found to be particularly informative about the research techniques, geography and taxonomic groups of the wild partnership and these are summarised in Table 2. **p.?

The following review summarises these and other studies of wild Fabeae-rhizobial systems. To some extent the purpose of this study is to find studies that look at the diversity of Fabeae plants and their rhizobial symbionts in an evolutionary and biogeographical context, of which there are relatively few. The majority of the work on the wild partnership is concerned with the diversity of plants and bacteria, and it is often a major focus of studies, so this will be discussed first in section 2.2. The promiscuity and specificity of the partnership is related to the diversity of the species but will be discussed here separately in the following section2.3. Studies with a biogeographical focus are reviewed in 2.4, and the final section 2.5 looks at studies on purposes and techniques

2.2. Studies of plant and bacterial diversity

The diversity of rhizobia nodulating Fabeae plants is a topic in several recent publications (Adiguzel et al., 2010; Ampomah et al., 2012; Ampomah and Huss-Danell, 2016; Aoki et al., 2010; Atallah et al., 2008; De Meyer et al., 2011; Dello Jacovo et al., 2019; I. Guefrachi et al., 2013; Han et al., 2008; Kan et al., 2007; Kurchak et al., 2011; Lei et al., 2008; Mutch and Young, 2004; Safronova et al., 2014; Smýkal, 2011; J Peter W Young et al., 2021; Zahran, 2001; Zézé et al., 2001)

Lathyrus and *Vicia* share similar distributions and habitats. They are mostly mesophytes in open woodland and forest margin habitats, with alpine and littoral species and some drought tolerant species (Kenicer et al., 2005). James et al. (2011) found that effective nodulation is more pronounced in poorer soils, such as coastal and seashores, and that the same plants growing in more fertile soils, such as woodlands, may obtain their nitrogen from the soil rather than symbionts. The diversity of rhizobial symbionts of wild plants of tribe Fabeae is a key topic of recent study, including the identification of species and species diversity associated with particular plants or in particular locations. Some new bacterial species have recently been identified and described. (Kenicer and Parsons, 2021)

In publications of *Minutes of the meetings of International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Rhizobia and Agrobacteria* since 2004, nine new rhizobial species nodulating Fabeae plants are noted to have been validly published. Of these, six were crop and three found nodulating wild host species. The wild species were *Rhizobium multihospitium* nodulating *Lathyrus odoratus*, *Vicia hirsuta* and other legumes (Han et al., 2008), *Bosea lathyri* LMG 26379 from host plant *Lathyrus latifolius* (De Meyer and Willems, 2012), and *Bosea vaviloviae* LMG 28367 from host plant *Vavilovia formosa* (Safronova et al., 2015).

This may reflect where work is concentrated more than the actual abundance of uncharacterized species in soils and rhizospheres. Other records of rhizobial species in this period include, for instance, several species found with rice and maize roots, and species associated with deep sea sediments and polluted soils found in metagenomic studies (de Lajudie et al., 2021; de Lajudie and Martinez-Romero, 2017; de Lajudie and Young, 2019; Lindström and Martínez-Romero, 2005). More studies turn up potential new species, particularly as metagenomic methods are used to explore soil populations (Parks et al., 2017). The process of formal description is complex (de Lajudie et al., 2019) and there may be many rhizobial genome sequences in the public datasets which are misattributed.(Young, 2021)

2.2.1. Studies of particular host plants and their symbionts

Studies of particular host plants are revealing about their symbiotic partnerships and the role that this plays within ecosystems. The symbiotic relationship of the common and widespread host plant *Lathyrus japonicus* is well studied. Seashore plants *Lathyrus japonicus* are generally dispersed by sea and colonise shingle soils throughout the northern temperate and Arctic coastal regions (Kenicer and Parsons, 2021). The plant is a halophyte, cold tolerant, and found in dune systems where it can act as a stabilizer. In one study authors found nodC sequences in *Rhizobium anhuiense* in China identical to those from reference *Rhizobium anhuiense* from Japan, suggesting these genes evolved specifically with their host plant. The authors suggested that low diversity in symbiosis genes was found, possibly due to a combination of selection pressure by saline and alkaline soil and by the host plants (Li et al., 2016). However (Aoki et al., 2010) found the plant to be nodulated by diverse rhizobia with genetic variation in the nod genes.

Further understanding of the environmental adaptation and phenotypic variation in nodulation of *Lathyrus japonicus* is achieved with studies that look at the physiology and ecology of the bacteria. In studies in northern Quebec, strains *of Rhizobium leguminosarum* bv. *vicieae* isolated from *L. japonicus,* and *L. pratensis* were studied. Strains of the two species were closely related but distinguishable by experimental growth at 5˚ C, reflecting environmental adaption to cold conditions (Drouin et al., 1996). Further studies looking at cold adaptation were done in Northern Quebec with the Rlv symbionts of *Lathyrus japonicus* and *Lathyrus pratensis.* Cold conditions induce the production of cold shock proteins, unsaturated fatty acids, and cold acclimation proteins - cold-adapted strains produce more though the efficiency of nodulation at cold temperatures depends also on the plants and their interactions. (Drouin et al., 2000) This was taken further in studies of seasonal changes in perennial root nodules of *L. japonicus*: starch grains and reducing sugar concentrations get depleted in the winter, but soluble sugar and non-reducing sugar is at a peak in midwinter, which is thought likely to act as an antifreeze to increase cold tolerance during the winter dormant period (Chinnasamy and Bal, 2003). There are also seasonal changes in oleosomic lipids and fatty acids: nodules fatten up for winter (Chinnasamy et al., 2003).

Lathyrus formosus (Steven) Kenicer formerly (*Vavilovia formosa* (Steven)Fed.) has been the subject of study by various authors using multiple approaches Combining studies of plants and rhizobia with ecological studies could be invaluable for conservation work on vulnerable and threatened plants such as this. *L. formosus* is found in isolated ecological habitats in mountain refugia. It is not found on the ICUN databases but is understood to be in danger of extinction (Mikić et al., 2014) . It was suggested by (Smýkal et al., 2017) as a Biotically Sustained Glacial Relict species as per (Hampe and Jump, 2011) in that the legume is dependent on its cold-adapted bacterial symbionts, and both partners are limited by climate. Studies have been carried out on the plant and its symbionts (V. Safronova et al., 2017; Safronova et al., 2015, 2014) rhizobial isolates form a distinct group inside Rlc. The sym genes are particularly distant from those of other Rhizobium leguminosarum bv. *viciae* and *trifolii* and include *nodX* gene, the diversity could be elicited by host plants, but housekeeping genes could be caused by local adaptation to soils or genetic drift (Kimeklis et al., 2019). (Smýkal et al., 2017). Whole genome sequencing and characterization of the nodules with particular attention to sym genes was carried out by Chirak et al., (2019) , although this was interpreted with an assumption that the host plant occupies an ancestral position in Fabeae that is not borne out in other phylogenies (Mikić et al., 2014; Schaefer et al., 2012; Smýkal, 2011).

2.2.2. Diversity of species within regions.

In a comprehensive study in in Flanders, Belgium, 3810 strains were isolated from 43 indigenous legumes from different genera in different ecoregions. Of the four *Lathyrus* and *Vicia* species sampled they found that with the exception of some plants of *Vicia sativa* which were found to be nodulated by *Bradyrhizobium japonicum,* all nodulated almost exclusively with *Rhizobium leguminosarum*. A large number of *R. leguminosarum* isolates were produced by the study, which show a large diversity of 156 distinct clusters when analysed using rep-PCR (De Meyer et al., 2011).

A multilayered approach was undertaken by (Ampomah et al., 2012) in a study of nodulation of wild plants in natural ecosystems in Scotland and Sweden across diverse habitats and altitudes and local climates. Nodules were taken from the same species indifferent locations in both countries, The authors sampled from different locations soil carbon and nitrogen content in these locations were also assessed. Further studies were undertaken to examine the diversity and phylogeny of bacterial symbionts of *Vicia* species in Sweden, Norway and Siberia. 14 genotypes of *Rhizobium leguminosarum* bv. *vicieae* were found In nodulation tests the isolates were able to cross-nodulate suggesting that hosts may be quite promiscuous but that soil conditions could be selecting for different genotypes. (Ampomah and Huss-Danell, 2016)

2.2.3. Nodules contain other bacteria with unknown function

Phyllobacterium was found in root nodules of *Lathyrus numidicus* (Mantelin et al., 2006)

Diverse genera were cultured from wild *Lathyrus* species in Bashkortostan, Russia. However no nodulation genes were looked at and nodulation tests were not carried out and the authors point out some of these may simply be occupying the nodules rather than fixing nitrogen. (Baymiev et al., 2012)

2.3. Research on promiscuity and specificity

Research on the biodiversity of rhizobia with the plants tends to uncover specificity and or promiscuity in the host-bacteria relationship and this is reported in some studies (Andrews and Andrews, 2017; Atallah et al., 2008; Baymiev et al., 2012; Chen et al., 2010; De Meyer and Willems, 2012; Han et al., 2008; Kimeklis et al., 2019; Lei et al., 2008; Li et al., 2016; Smýkal et al., 2017; Zahran, 1999).

2.3.1. Specificity

For legumes to colonise new ground successfully, compatible symbiotic partners need to be present in soils. (Aoki et al., 2010; Mutch and Young, 2004). However the mobile symbiosis genes that direct the bacterial interactions with the host plants can be shared between bacteria by conjugation (Cavassim et al., 2020) so these can potentially be donated by incoming bacteria not so well adapted to local conditions.

Nodulation genes, *nod*ABC are considered major determinants of host specificity (Sprent et al., 2017) Legume specificity is encoded in accessory genes that can be laterally transferred between species and genera (Rogel et al., 2011) sometimes quite distantly related bacterial species (Baymiev et al., 2011; Broughton and Perret, 1999).

*Nod*D is found only in rhizobia and as the nearest homologues of the genes in other bacteria have been shown to be very dissimilar, direct amplification of these genes from soil samples has been shown to be informative to give sensitive characterization of the diversity of rhizobia in the soil community. (Zézé et al., 2001)

It has been suggested that specificity can arise from geographical isolation. *Microvirga ossetica,* from the family Methylobacteriaceae was found to be effectively nodulating the high altitude and geographically isolated species *Vicia alpestris*. It was suggested that the species had diverged and are maintained by genetic isolation (V. I. Safronova et al., 2017)

A bacterial isolate from a native *Vicia* species near Tromsø in Northern Norway was found to have a similar genetic profile to an isolate from Haparanda in northern Sweden, 475 km apart (Ampomah and Huss-Danell, 2016). In a study in Sierra Nevada, Spain, diverse Rlc rhizobia were isolated from wild *Lathyrus* and *Vicia* species, some which clustered closely in analyses with reference strains isolated from other continents (Villadas et al., 2017)

2.3.2. Promiscuity of Fabeae plants

Plants in tribe Fabeae are found prefer *Rhizobium leguminosarum* but to nodulate promiscuously with different genospecies and the partnership varies in degree of specificity.

In a study designed to examine the relationships between symbionts of crop symbionts and indigenous bacteria, chromosomal and nodulation genes of Rhizobium leguminosarum bv. vicieae were studied with wild plants *Vicia cracca*, *V. hirsuta*, V. *sativa, Lathyrus pratensis, L. aphaca, L. nissolia*, and crops (pea, *Pisum sativum* cv. Kelvedon Wonder and broad bean, *V. faba* cv. The Sutton). Wild legumes were found to be more promiscuous than crops with their rhizobia. The authors described "a population of symbionts that is diversified by plasmid transfer and shared fairly indiscriminately by local wild legume hosts". (Mutch and Young, 2004)

2.3.3. Diverse rhizobia also nodulate with Fabeae plants

Reports of Fabeae plants found nodulating with other rhizobia altogether are not uncommon. In Bashkortostan, Russia, diverse symbionts were found with wild Fabeae: *Rhizobium tropici* nodulating with *Lathyrus vernus* L. Bernh. and *Lathyrus sylvestris*, *Agrobacterium* spp. nodulating with *Lathyrus palustris* L. and *Phyllobacterium myrsinacearum* with *Lathyrus gmelinii* Fritsch (Baymiev et al., 2012) .

In a study in China, wild *Vicia* species were frequently nodulated by Rhizobium leguminosarum; however *Rhizobium gallicum*, *Mesorhizobium huakuii*, *Ensifer meliloti* and *Bradyrhizobium* spp, were also isolated (Lei et al., 2008).

2.3.4. Bacteria are also promiscuous

Rhizobium multihospitum presented as the primary symbiont of diverse legumes in Xinjiang, China A new species of *Rhizobium hospitium* was found nodulating quite disparate host species growing wild, including *Lathyrus odoratus, Astragalus* spp. and tree species *Halimodendron halodendron* Northwestern Xinjiang region of China. Although the host plants harboured different strains of the bacterium, the nodD and nifH genes were found to be identical. (Han et al., 2008) The role of the soils here is interesting. Studies of nodulation in the arid, saline desert soils of this part of Northwestern China have uncovered unusual rhizobia (Chen et al., 1995). A survey of wild diverse legumes in XInjiang found that *Vicia cracca* grown in the poor soil could be nodulated by a newly discovered bacterium *Mesorhizobium alhag*i which had been isolated from *Alhagi sparsifolia*, despite having dissimilar *nod*A and *nod*C genes to other rhizobia found nodulating these local plants (Chen et al., 2010).

2.4. Biogeographical research

1n 1934 Lourens Bass-Becking framed a hypothesis for the biogeography of microbes that "Everything is everywhere; the environment selects." (Baas-Becking, 1934)

It has been further proposed that dispersal of prokaryotes is random, and that because of the very large numbers of individuals, spatial distributions should therefore be random (Finlay, 2002) But a review of biogeography of microorganisms finds recent studies suggest that microbial communities adapt to their environments, diverge, and can be maintained by genetic isolation. Rates of biogeographic processes seem to be more varied than for larger organisms and size of organism body and community means that considerations of scale are important in studies of population distribution and dispersal. (Martiny et al., 2006)

A review of the biogeography of legumes and bacterial symbionts was recently published. Here, the symbionts are examined both separately and together, treating the partnership rather than focusing one or the other. Distinct longitudinal and latitudinal variations are found in the distribution of nodulated legumes and their symbionts. Members of tribe Fabeae generally nodulate with *Rhizobium* which is found on all continents (Sprent et al., 2017). in large part, this is the inspiration for the current study.

A few primary research papers on the Fabeae-rhizobium symbiosis were concerned with biogeography. A large study of different regions of Flanders (De Meyer et al., 2011), and a study of Sweden and Scotland (Ampomah et al., 2012). (Sui et al., 2009) found that the distribution of rhizobia in various locations in temperate Northern China were determined by geography of the host plant as well as the environment. Though these studies are valuable, a greater geographic spread of data would be useful to understand the biogeography of the partnership.

2.5. Techniques used in wild *Rhizobium* studies

Some of the most fundamental studies simply survey the presence of nodulation in systems. For some purposes, presence, morphology, and pink colour of nodules indicating likely high levels of leghaemoglobin are adequate indication of effective nodulation remain useful for determining probable contribution to nitrogen resources. Indeed, simple *in situ* surveys of nodulated plants remain useful to determine this. Visual inspection is important for ecological work and as preliminary work for further investigations (Ampomah et al., 2012; James et al., 2011) and light and scanning electron microscopy is still valuable (Mutch and Young, 2004). Many studies on rhizobial systems have used morphological and culture approaches, but few were in wild host strains. One such morphological study of nodulation in wild plants in agricultural pastures of Turkey was published by Küçük and Cevheri, (2014). This kind of study is important, but probably not reported as much as it should be. These quick, simple and inexpensive studies are perhaps often carried out as undergraduate projects, preliminary work collecting samples for further study, or for local agricultural or commercial purposes. Unfortunately most of these studies probably go unpublished. As publication remits focus on novelty and molecular work is prioritised, it is possible that much valuable ecological work may be missed from the literature.

Questions of diversity and specificity are increasingly better resolved as molecular methods have changed rapidly over recent times and this has led to new approaches and innovations in methodology. In recent work 16S rRNA sequences are used to distinguish genera and species outside *Rhizobium leguminosarum*, with additional 'housekeeping' genes to gain finer resolution of the relationships within Rlc. A summary of some of these methods and the genes used in recent studies is included in Table 2 and a general review of materials and methods used in recent work is in greater detail here in Chapter 3 (page *.).

Most recently, whole genome and plasmid studies have become inexpensive and support for bioinformatic data processing is ever more available, which has greatly opened up the range of potential work that can be done to examine the functionality of genes, create large and very robust phylogenies. Relationships can be resolved to a fine degree as is shown by Young et al., (2021) and there is great scope for increasing the breadth of knowledge about the wild systems.

2.6. Purposes of studies of wild plant-rhizobia systems

The study of nodulation in wild Fabeae is most commonly related to issues which directly affect humans, in some cases as part of an increasingly urgent need for food security in changing climates. The bacterial symbiosis can potentially contribute to environmental maintenance and improvement and conservation.

Many studies looked directly at orphan crops (neglected and underutilized legumes). Crop-wild relative studies of Fabeae symbionts are important both for the large agricultural crops and orphan, fodder and forage crops. Several studies explore the role the partnership can play in phytoremediation and conservation both of plants and ecosystems is a topic of interest

2.6.1. Studies of Orphan crops

The role of nodulation in supporting orphan crops is currently being investigated for agricultural potential for withstanding some of the effects of climate change as the strongly tolerant of drought, heat stress and water stagnation. Grass pea, *Lathyrus sativus* L., is considered one such crop although it is an important crop in Ethiopia, Tunisia and India. The rhizobial partnership of *L. sativus* has been studied by several authors recently (Abdelkrim et al., 2019, 2020; I. Guefrachi et al., 2013; Jiao et al., 2011; Kumar et al., 2013; Mahdavi, 2007; Mohammed et al., 2020; Rathi et al., 2021; Sui et al., 2009; Xiong et al., 2014).

A study of wild rhizobia in forage and underutilized grain legume species found that the wild rhizobia showed greater variation in stress tolerance than reference strains and potentially additional fitness for the crop plants (Moschetti et al., 2005). *Lathyrus cicera*, red pea is grown as a forage crop (Gritli et al., 2020) Other potential grain and forage crops include *L. hirsutus, L. cicera*, *L. ochrus*, *L. tingitanus*, *L. latifolius* and *L. sylvestris* (Pandey, 2000)

2.6.2. Improving food crops: Crop-wild relatives

An important study of the diversity and specificity of *Rhizobium leguminosarum* bv. *vicieae* on crop and wild rhizobia was published to give an overview of the host range evolution of the bacteria with tribe Fabeae (Mutch and Young, 2004), see section 2.5.2 below.

However some studies of wild plant symbioses are intended to be used directly in crop improvement. In a comparison of crop and wild rhizobia, the rhizobia associated with crop plants were found to be more adapted to anthropogenic stress factors, whereas the rhizobia of wild *Vicia* were found to have higher species diversity of bacteria, higher polymorphism of traits adapted to local soils, edaphic stress factors and rhizosphere; and be more capable of efficient nodulation. (Kurchak et al., 2011)

An approach in older studies was to recover bacterial isolates which are suited to particular habitat types, such as a study to find near-desert adapted rhizobia in for potential inoculation on introduced legumes in Australia(Hely and Ofer, 1972). Current practice is to find the best adapted autochthonous legumes and rhizobia where possible (Villadas et al., 2017). One such study was done in Northern Turkey. Nodules from wild *Vicia cracca* plants growing at high altitudes were investigated as suitable symbionts of *V. cracca* plants grown as forage crops at lower altitudes. The potential inoculants were better adapted for stress and drought and were recruited to help reduce reliance on artificial fertilizer. (Adiguzel et al., 2010)

2.6.3. Phytoremediation and conservation

Nodulated legumes can contribute to the sustainable management of polluted soils. This has been studied with various Fabeae hosts (Guefrachi et al., 2013; Ruiz-Díez et al., 2012). and more particularly with *Lathyrus sativus.* A community of lead-resistant plant growth promoting rhizobacteria which was used to inoculate L. sativus increased the growth of the plant and increases lead uptake. This decreased the bioavailability of lead in the soils (Abdelkrim et al., 2018). In another study, the impact of field inoculation with a bacterial cocktail including potential nodulating rhizobia along with growing *L. sativus* gave significant improvement in soil quality and fertility. (Abdelkrim et al., 2020)

A survey in degraded landscapes in Lebanon established the nodulation status and potential use for legumes in revegetation projects. 11 Fabeae species were found (listed in Table 2) and two species, *Vicia hybrida* L. and *Vicia palaestina* Boiss. were selected for revegetation (Atallah et al., 2008).Salt tolerant rhizobia have been suggested to enable *L. maritimus* have a role in dune stabilization (Li et al., 2016).

Vicia, Lathyrus and *Trifolium* species are sown as forage legumes for goats and sheep to create areas of clear land that can act as firebreaks in a fire-prone area of high ecological significance and species endemism in the Sierra Nevada mountains, Spain. Inoculation of seeds or legumes that associate with fast-growing rhizobia of rhizobial endosymbionts already known to inhabit these soils helps the plants to get established quickly. (Villadas et al., 2017)

2.7. Conclusion

Studies of the rhizobial symbiosis with wild Fabeae are few compared to those with their crop plant relatives, though some work has compared crop and wild rhizobial symbioses.

Motivations for study include conservation of endangered species and sustainable management of ecosystems, and improvement of legumes used for agricultural purposes.

Some studies are ecological, some look at specific plants and their symbionts. Key topics of study include diversity of strains and species, promiscuity and specificity of hosts and rhizobia, environmental adaptation to particular niches and geographic regions.

Some plants and their symbionts have been studied in detail, such as *Lathyrus japonicus* and *Lathyrus formosus*, and some regions such as UK, Sweden, North Africa and parts of Northwestern China have been more studied.

Biogeography is of interest in some studies, although it is not a central focus of many studies. There has been no recent systematic study of the biogeography of wild rhizobial symbionts of tribe Fabeae. Few high resolution studies of biogeography have been done and there is very little information on wild legumes in certain regions, especially in both North and South America.

2.8. Possible Further study

All comparative studies reviewed so far depend on a small number of housekeeping and nodulation genes. As whole genome sequencing becomes more affordable and processing becomes less costly of time and computing power, higher resolution of interrelationships between the symbionts and insights into their relationships with host plants could be obtained.

A clear picture of the extent of regional coverage by certain bacteria is not emerging. The limits of dispersal of the species are not established and the role of microbial communities in the maintaining the distribution of rhizobia and the functioning of the symbiosis is very little known. Tribe Fabeae is interesting biogeographically, with many long-distance dispersal events. Further biogeographical studies of their symbionts in North and South America would be of particular interest.

In the following chapter I present my work on a novel dataset of whole genome sequences from rhizobia nodulating with wild Fabeae. Following the precedent of Young et al. (J Peter W Young et al., 2021)]] 120 core chromosomal genes will be used, along with plasmid-borne nodulation genes to reconstruct phylogenetic relationships, the possible influence of host phylogeny or biogeography that is acting on the rhizobial symbionts is examined.

Table 2. Overview of important recent studies of wild Fabeae and their rhizobia
With notes on the region of study and methods of analysis

Host	symbiont	Region	Analysis	author
		Ethiopia, Southern		(Asfaw et al., 2020)
L. sativus		Tunisia, Northern and Central		(Abdelkrim et al., 2020)
Lathyrus pratensis L., *L. japonicus* Willd. , *L. linifolius* (Reichard) Bässler, *L. palustris* L. *L. Vernus* (L. Bernh) *Vicia cracca* L., *Vicia hirsuta* L., *Vicia sativa* L. *Vicia sepium* L. *V. sylvatica* L., *V. tetrasperma* (L.) Schreb. V. cracca, V. hirsuta, V. sepium, V. tetrasperma and V. sylvatica V. cracca V. *multicaulis*	14 genotypes of *Rhizobium leguminosarum* bv. *viciae*	Sweden and Scotland Sweden, Norway, Siberia	atpD, recA, nodA	(Ampomah et al., 2012) (Ampomah and Huss-Danell, 2016)
		Vyritsa, Leningrad		(Andronov et al., 2015)
L. Japonicus	*R. leguminosarum* bv. *viciae*	Japan, Canada	61 rhizobial strains: placed into 3 genomic groups (I, II, III) RAPD fingerprinting, RFLP and 16–23S ITS region; 16S rRNA (*rrs), recA, nodB,* and *nodC* genes.	(Aoki et al., 2010)

Lathyrus ciliolatus Sam ex. Hech. f. *Lathyrus aphaca* L. *Vicia palaestina* Boiss., *Lathyrus annuus* L, *Lathyrus digitatus* (M Bieb. *Vicia hybrida* L., *Vicia peregrina* L., *Vicia lutea* L.*, Lathyrus gorgonii Parl.* *Vicia sativa* L. ssp *cordata* Hoppe.	Presence of effective nodules	Mediterranean Lebanon	Survey of occurrence of plants and in-situ survey of nodulation of plants	(Atallah et al., 2008)
L. *vernus*	Bv. *viceae* nodulation genes are introduced into *Rhizobium leguminosarum*, R*hizobium tropici*, A*grobacterium* sp., and P*hyllobacterium* sp. by horizontal transfer from *R. leguminosarum* bv. *viceae* strains, conferring the ability to form symbioses with *Lathyrus* spp		Comparative analysis of *nifH, nifD*, and *nodA* opposed to 16S rRNA, showing that nodulation genes are transferred to bacteria other than *R. leguminosarum*	(Baymiev et al., 2011)
L. vernus L. Bernh., *L. gmelinii* Fritsch, *L. tuberosus L.*, *L. pratensis* L., *L. litvinovi*i Iljin, *L. sylvestris* L.*, L. pisiformis* L., *L. palustris* L., *L. palescens* (Bieb.) C. Koch*)*	The majority *R. leguminosarum*, but also *R. tropici*, *Agrobacterium* sp., *Phyllobacterium myrsinacearum*	Bashkortostan, Russia	PCR, 16S only	(Baymiev et al., 2012)
		Algeria, Eastern		(Belhadi et al., 2018)
Vicia sativa L. subsp. *sativa* L., *Vicia narbonensis* L. var. *narbonensis* L., *Vicia palaestina* Boiss., *Vicia hybrida* L., *Vicia lutea* L. var. *lutea* Boiss. ET Ball., *Pisum sativum* L. subsp. *sativum*				(Cevheri, 2014)

L. var. *sativum* L, *Lathyrus cassius* Boiss				
Lathyrus japonicus		Newfoundland, Canada		(Chinnasamy and Bal, 2003)
		Turkey		(ÇİLDİR, 2011)
L. vernus, L. palustris, and L. gmelinii		North Caucasus, Russian Federation		(Chubukova et al., 2011) - plant symbiosis genes only.
Lathyrus latifolius, L. pratensis, L. sylvestris, L. tuberosus, Vicia cracca, V. hirsuta, V. thyroids, V. sativa, V. sepium, V. tetrasperma, V. villosa	Almost exclusively *Rhizobium leguminosarum*, but *Bradyrhizobium japonicum* isolated from some *V. sativa*	Flanders	16S rRNA, RecA, nifH, nodC 3810 strains 43 plant species.	(De Meyer et al., 2011)
Lathyrus linifolius L.			16S rRNA, nodA and nodD used to understand rhizobial diversity.	(Dello Jacovo et al., 2019)
L. japonicus L. pratensis		Quebec		(Drouin et al., 1996)(Drouin et al., 2000)
L. cicera		Tunisia, Northern and Central		(Gritli et al., 2020)
Lathyrus japonicus		Newfoundland, Canada		(Chinnasamy et al., 2003; Gurusamy et al., 2000)
Many diverse legumes	*Rhizobium multihospitum*	China, Xinjiang		(Han et al., 2008)
Vicia and *Lathyrus* spp		East Scotland		(James et al., 2011)

Vicia	*Rhizobium leguminosarum*	Qinghai-Tibet	ARDRA, AFLP, DNA–DNA hybridization and 16S rDNA sequencing,	(Kan et al., 2007)
Lathyrus oleraceus		Norther Ossetia	hkg: 16S rRNA, *gln*II, *glt*A, and *dna*K; and sym: *nod*A, *nod*C, *nod*D, and *nif*H) of *Vavilovia* compared with those from other Rlv and Rlt. they assume *Vavilovia* to be ancestral -contrast with other authors	(Kimeklis et al., 2019)
Vicia sativa L. subsp. *sativa* L., *Vicia narbonensis* L. var. *narbonensis* L., *Vicia palaestina* Boiss., *Vicia hybrida* L., *Vicia lut*ea L. var. lutea Boiss. ET Ball., *Pisum sativum* L. subsp. *sativum* L. var. *sativum* L, *Lens culinaris* Medik., *Lathyrus cassius* Boiss.,		Turkey		(Küçük and Cevheri, 2014)
L. japonicus		Shandong Peninsula, China vs. Japanese strains	Phylogenetic relationships: concatenated sequences of recA; also, 16S rRNA atpD, glnII - genospecies	(Li et al., 2016)

			MLSA 97% similarity: also, nodC and nifH *R. anhuiensis* has the same recA genotype as Japanese coastal samples - might have the same origin.	
Wild Vicia	*Rhizobium leguminosarum, Rhizobium gallicum, Mesorhizobium huakuii, Ensifer meliloti and Bradyrhizobium spp*	Temperate China	Amplified 16S rDNA restriction analysis (ARDRA) and restriction fragment length polymorphism (RFLP) of 16S–23S internally transcribed spacer (ITS) region	(Lei et al., 2008)
Lathyrus numidicus	*Phyllobacterium ifriqiyense* sp. nov. LMG 22831T * 16S and atpD genes only - no nodulation genes/nodulation test	Southern Tunisia		(Mantelin et al., 2006)
wild *Vicia hybrida, V. villosa, V. villosa var. pseudocracca, V. sativa, Lathyrus ochrus, L. clymenum, L. annus* and *Galega officinalis and crop* legumes *Vicia faba, P. sativum.*		Italy		(Moschetti et al., 2005)

. *L. aphaca* and *V. hirsuta*.				
Rlv and wild *V. cracca, V. hirsuta, V. sativa, L. pratensis L. aphaca, L. nissolia;* crops (pea, *Pisum sativum* cv. Kelvedon Wonder and broad bean, *V. faba* cv. The Sutton				(Mutch and Young, 2004)
Vavilovia formosa (Stev.) Fed.	*Bosea vaviloviae* sp. Nov.	North Ossetia, Dagestan, and Armenia	Multiple housekeeping genes	(V. Safronova et al., 2017; Safronova et al., 2015, 2014)
L. japonicus *L. davidii, L. quinquenervius, L. odoratus, L. pratensis, L. sativus*		China	. Use of ARDRA, IGS-RFLP, TP-RAPD and MLE.E for capturing detailed information. Biogeography of rhizobia.	(Sui et al., 2009)
Lathyrus cicera, Lathyrus intricatus (Lathyrus hygrophilus), Vicia sativa subsp. angustifolia (Vicia amphicarpa), Vicia disperma, and Trifolium spp.	High diversity - all strains were phylogenetically close to the species from the Rhizobium leguminosarum group, although they were not identified as any of them.	Mountains of Sierra Nevada National Park (South Spain).	16S rRNA, *atpD, recA, glnII* and nodC used:	(Villadas et al., 2017)
Lathyrus -A survey of a very diverse bacteria nodulating wild legumes, some novel associations - strains	Slow growing Bradyrhizobium strains found only in China and USA prior to this study.	Tunisia	Polyphasic approach - DNA Restriction Analysis (ARDRA) of the 16S rDNA and SDS-PAGE of total all	(Zakhia et al., 2004)

which might be new genospecies -, including *Lathyrus numidicus, Ebenus*.			proteins and 16S rDNA sequence comparison.	
Vicia faba, *Pisum sativum*	*Rhizobium anhuiense* sp. nov.			(Zhang et al., 2015)
L. japonicus L. pratensis		Quebec		Drouin

3. Methods and Results

This section provides an expanded methods section for the purposes of the MSc thesis and background on the technique

3.1. Introduction

A typical *Rhizobium leguminosarum* bacterium has a genome of 7.7 Mb on a circular chromosome, with three to eight but more normally six circular plasmids. 61% of total composition is G + C. All three rRNA operons ad 52 tRnA genes are on the chromosome.

Most essential genes for encoding proteins are vertically inherited and situated on the chromosome, though most functional classes also occur on the plasmids. Nodulation genes are situated on the plasmids and are frequently exchanged between bacteria by conjugation (Bañuelos-Vazquez et al., 2020; Ding and Hynes, 2009; Young et al., 2006).

Many genes have orthologs in close relatives *Agrobacterium*, *Ensifer,* (*Sinorrhizobium)* and *Mesorhizobium*, and rhizobia which acquire suitable nodulation genes can form symbioses with Fabeae plants (Young et al., 2006).

Rhizobium leguminosarum is a gram-negative bacterium with a core circular chromosome and three to eight circular plasmids, usually six (Young et al., 2021). Most of the vertically inherited essential genes for encoding proteins are on the chromosome. Many functional components occur on the plasmids. The mobile and often horizontally transferred symbiosis genes are also on the plasmids. (Bañuelos-Vazquez et al., 2020). As distantly related bacteria can share genes, the circumscription of species can be thought of in terms of genotypic cohesion with clusters of genetic similarity and minimal homologous recombination within distant species (Kumar et al., 2015).

3.1.1. Current Methods.

Current methods of investigating species delimitation, relationships and activity of bacteria tend to rely heavily on molecular methods. DNA-DNA hybridization (DDH) has been used for species delimitation for around 25 years. The genomes of strains are hybridized in a lab to enable pair-wise comparisons of the similarity between genomes and therefore their relatedness, and the complex procedure is highly accurate. (Wayne et al., 1987) This approach was used in describing new species by (Ampomah and Huss-Danell, 2016; Drouin et al., 1996; Han et al., 2008; Kan et al., 2007; V. I. Safronova et al., 2017; Xiong et al., 2014) Further characterization of species can include phenotypic studies such as stress-tolerance, temperature responses (Moschetti et al., 2005) In many of the recent studies on wild legumes and their microbial symbionts. amplified ribosomal DNA restriction analysis (ARDRA), which is amplification by PCR of the 16s region followed by restriction using enzymes and analysis has also been recently used (Han et al., 2008; Lei et al., 2008; Safronova et al., 2015; Sánchez-Cañizares et al., 2018; Sui et al., 2009; Zakhia et al., 2004).

In general bacteriological work a whole 16S sequence may be considered adequate to identify a strain to species (Johnson et al., 2019), *Rhizobium leguminosarum* species complex is not well resolved by this highly conserved sequence alone (J Peter W Young et al., 2021), additional protein coding genes add greater reliability of identifications (Chun et al., 2018; Young et al., 2006). 'Housekeeping' genes such as recA, DNAk, glnII, rpoB are stable and less likely to be horizontally transferred by other species are used for phylogenetic analysis. Combinations of genes used in analyses in recent studies are included in Table 2.

Comparatively small genome size and recent developments mean that whole bacterial genomes can now be sequenced quickly and relatively cheaply using high-throughput processes. Computational advances and capacity make it possible for the assembly and analysis of whole genome sequences, with rapid gains in the granularity of understanding we can achieve. As in this study, large amounts of gene sequence data can be retrieved from draft genomes for analysis, even without full assembly of the genome.

New methods for minimal standards for the description of new genera of prokaryotes have been recommended by (de Lajudie et al., 2019) Minimal standards for the use of genome data for the taxonomy of prokaryotes have been proposed (Chun et al., 2018)

Quality control of large amounts of data is important and programmes such as CheckM (Parks et al., 2015) and Quast(Gurevich et al., 2013) allow for assessment of the quality of the genome data.

Average Nucleotide Identity (ANI) is a whole-genome similarity metric which is one of the Overall Genome Relatedness Index measures that can be calculated digitally and is accepted as appropriate for the delineation of bacterial species (Chun et al., 2018). ANI does not calculate evolutionary relatedness, but shows robust resolution between strains. Due to the high volume of genetic material available in the genome the ANI method has been shown to give highly resolved analyses of species boundaries for bacteria from diverse lineages and distinct genetic discontinuity has been consistently reported for bacterial genomes: conspecifics (on the basis of genospecies concepts) are found at >95% ANI and intergeneric boundaries at <83%. (Jain et al., 2018) Results from this method closely reflect results from the traditional method of DNA-DNA hybridization (Goris et al., 2007; Richter and Rosselló-Móra, 2009) .

Current guidance for producing the best single tree for a phylogenetic analysis indicates that a maximum likelihood analysis can be performed on a concatenated alignment of conserved, vertically inherited, single copy genes. (Lang et al., 2013; Parks et al., 2018)

120 ubiquitous single-copy genes had been evaluated to be phylogenetically informative and known as the *bac120 marker set.* This set includes the commonly used housekeeping genes *rpo*B dnaK glnII gyrB (Parks et al., 2017)

In the intitial stages of this project a multilocus sequence analysis (MLSA) approach was planned using a suite of seven housekeeping genes. Sequences of atpD, DNAK, glnII, gltA, gyrB, recA, and rpoB were extracted from the contigs using the extraction method above and aligned in Aliview (Larsson, 2014). The MLSA approach was discontinued in favour of the potentially higher resolution offered by analysis following the protocols of (J Peter W Young et al., 2021). This approach involves the use of Python programming language to automate operations with the larger volumes of data involved with the concatenated genes of the dataset.

3.1.2. Aims of the analysis

The aim of this study was to use a large new dataset of whole genomes extracted from bacteria cultured from wild Fabeae species form UK, Crete, Spain, USA, Canada, Chile and Argentina to examine the biogeographical and phylogenetic relationships of the bacteria with reference to what is known about host plant phylogeny and biogeography.

The objectives of the study were:

1. to construct an ANI table to analyse the species boundaries of the group and
2. to relate this to the geographical area of collection and
3. to construct a phylogeny using 120 genes as proxy for the species and
4. to construct another phylogeny using nodulation genes
5. to analyse at the results of these analyses against phylogeny and biogeography of the host plants

3.2. Materials and Methods

3.2.1. The Study Dataset

The study dataset consists of 67 draft bacterial whole genome sequences from bacteria cultured from root nodule samples from Fabeae species. These were mostly wild or cultivated wild species. The dataset spreadsheet *Accessions_References.xlsx* (Appendix 5) shows these strains along with host plant and collection details where obtainable.

Genome files were inspected with Artemis software (Rutherford et al., 2000) for an overview of completeness and the presence of annotated genes.

Fabeae are known to nodulate with *Rhizobium leguminosarum* and close relatives (J Peter W Young et al., 2021). Other species are sometimes found in nodules and there are new investigations into nodule community assembly, however the focus of the study was on nitrogen-fixing symbionts of the plant hosts so unlikely candidates were removed from the dataset.

3.2.2. Methods used before the start of the project

This study was performed on draft whole genomes which had already been assembled into contigs and annotated. Prior to the start of the project the materials had been treated according to the following methods:

The nodules were collected from wild plants by Greg Kenicer (RBGE, Edinburgh, UK), Louis Ronse de Craene (RBGE, Edinburgh, UK), Matias Morales (CONICET, Buenos Aires Argentina) and Euan James (Hutton Institute, Invergowrie, UK) between 2018 and 2021 and stored in silica gel.

Genomes were sequenced by MicrobesNG (http://www.microbesng.com) according to their protocols available at https://microbesng.com/microbesng-faq/, briefly outlined below. Each accession sample was grown on in pure culture before being harvested and resuspended with cryoperservative for DNA extraction. The suspension was lysed with a buffer solution incubated for 25 min at 37°C. Proteinase K and SDS (were added and incubated for 5 min at 65°C. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer. DNA was quantified with the Quant-iT dsDNA HS kit (ThermoFisher Scientific)   assay in an Eppendorf plate reader.

For Illumina Sequencing, genomic DNA libraries were prepared using the Nextera XT Library Prep Kit following the manufacturer’s protocol with the following modifications: input DNA was increased 2-fold, and PCR elongation time was increased to 45s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Libraries were quantified using the Kapa Biosystems Library Quantification Kit for

Illumina then sequenced using Illumina sequencers using a 250bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 (Bolger et al., 2014). De novo assembly was performed on samples using SPAdes version 3.7 (Bankevich et al., 2012), and contigs were annotated using Prokka 1.11 (Seemann, 2014).

3.2.3. Reference genomes

The full genome of *Rhizobium anhuiense*, sister to the *Rhizobium leguminsarum* complex (Rlc) and the genomes of 18 type strains of genospecies of the Rlc identified by Young et al., (2021) were downloaded from the Genbank database (Clark et al., 2016). Further type strain reference genomes for species outside this group were later downloaded. All reference genomes listed in the spreadsheet along with their accession numbers, along with further reference strains, in spreadsheet Accessions_references.xlsx (Appendix 6.)

Reference genomes were added to the analyses, which were performed on 89 taxa.

Full assembly of genomes was not possible within the time for the project. However preliminary assessments of data quality were made to follow the initial steps recommended for minimum standards for working with whole bacterial genome material in the taxonomy of prokaryotes (Chun et al., 2018).

3.2.4. Quality assessment with QUAST

To evaluate the size and quality of the draft genomes, analyses were performed on all the accessions using QUAST quality assessment tool for genome assemblies. (Gurevich et al., 2013). The length of the genomes and contigs and Nx scores for all accessions was calulated. The mean of the total length of the genomes in numbers of bases and the was N50 score by position of the base within the genome were calculated in Microsoft Excel spreadsheets over all taxa Results are displayed in a spreadsheet, Quast_report_averages.xlsx (Appendix 7)

3.2.5. 16s for genus level identification.

To determine the identity of accessions to genus level, 16S rRNA gene was analysed. 16SrRNA, housekeeping and nodulation genes were extracted from the genome files using a BEDTools (Quinlan and Hall, 2010). Reference sequences for 16S rRNA were downloaded from The Ribosomal Database Project (Cole et al., 2014) The 16S sequences were aligned along with the reference strains using MUSCLE (Edgar, 2004) in Aliview (Larsson, 2014) (table) and adjusted and inspected by hand.

The 16S sequences were queried against the Genbank database (Clark et al., 2016) using BLAST blastn (Camacho et al., 2009). Clear outliers at this stage were accessions from host plants *Lathyrus crassipes* Gillies ex Hook. & Arn. and *Vicea lutea* Pall. ex M.Bieb. The accessions isolated from *L. crassipes* aligned most nearly with other rhizobial species known to nodulate with plants, however the *V. lutea* accession was identified as *Pseudomonas* sp. and eliminated from the dataset.

To run the following analyses languages Python scripts written by (J Peter W Young et al., 2021) were used and adapted where necessary. The scripts used here are in the folder P_Young-Python-scripts-adapted (Appendix 8)

3.2.6. Average Nucleotide Identity

To identify groupings of species boundaries between the study strains, pairwise ANI was calculated using FastANI 1.31, a high-throughput alignment-free approximate sequence-mapping method (Jain et al., 2018).

FastANI was run from terminal using the default settings were used (kmer = 16, fragment length = 3000, minimum shared fraction = 0.2) using a reference text file. The output was written to a tab separated file. The number of fragments per genome ranged from 1974 to 3800, and the number of matches in pairwise comparisons ranged from 1151 to 3788.

A Python script adapted from (J Peter W Young et al., 2021) was run to plot a square matrix of all-by-all ANI values and to display the results using the Seaborn library (ANI_squareplot,

Appendix X). At Young's suggestion (Young, 2021) we amended find_genes.py script to allow the Seaborn clustermap to organize the data. This is a hierarchical clustering algorithm which uses a clustering implementation from Scipy.

Calculations were later repeated to include the potential reference strains for taxa which presented as possible new species.

To extract full-length gene sequences for each of the accessions for analysis, a Python script was run to find matches to protein or DNA sequences in fasta files of the contigs, using standards files for reference. (Reference sequences are at Appendix 9)

A blastn search or tblastn search of the whole genomes was run using the standards files as query. The best hits from the results were indexed and written to a tab-delimited file sorted by scaffold and query (Camacho et al., 2009). The gene sequence for each top hit was extracted and aligned with *CrustalO* Clustal Omega programme (Sievers et al., 2011). Gene sequences were aligned and concatenated using *CrustalO.* This process was completed for the core *Bac120* set of genes, nodulation genes *nod*A, *nod*C and *nod*D, and 16S rRNA genes. For the core 120 genes, a single file was produced with single gene alignments assembled into longer strings of concatenated sequences for each accession.

3.2.7. Phylogenetic analysis

The protocols published in (J Peter W Young et al., 2021) were followed to produce phylogenetic analyses. To further understand the evolutionary relationships between the accessions a phylogeny was prepared from the dataset using the *bac120 marker set* (Parks et al., 2017).DNA nucleotide sequences coding for the proteins were used here rather than the protein sequences in order to gain a higher resolution for the intraspecific relationships of our dataset.

Approximately Maximum-Likelihood trees were created for concatenated core and nodulation gene sequences using FastTree Version 2.1.11 (Price et al., 2010) The default settings were used. Heuristic neighbour-joining to find the topology of tree space; nearest-neighbour interchanges and subtree pruning to correct for branch-length distortions;

maximum likelihood nucleotide distances were initially calculated using Jukes-Cantor model of nucleotide evolution. As a model test for the breadth of data in the different functional core genes of the *bac120* dataset was not obtainable the analysis was repeated using generalized time-reversal (GTR) model of nucleotide evolution for comparison.

To estimate support for branches Shimodaira-Hasegawa test was used with 1000 replicates.

Nodulation genes *nod*A, *nod*C and *nod*D were extracted from the dataset. The nodulation genes of all accessions were blasted against reference genes from biovars *viceae*, *trifolii* and phaseolii. A report for all accessions showing similarity matches is in spreadsheet NodACD_symbiovars.xlsx. (Appendix 6)

Phylogenetic analyses were conducted on the nodulation genes and on 16S sequences of all accessions using the protocols above. Maximum likelihood gene trees were constructed in F displayed in Figtree and saved in different formats.

3.2.8. TYGS

To determine closely related strains, genome sequence data from taxa not found to be identified as known rhizobial species are submitted to Type Strain Genome Server (TYGS) at https://tygs.dsmz.de . This is a free high-throughput web-server for whole genome based prokaryote taxonomic analysis. (Meier-Kolthoff and Göker, 2019) Submitted genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm (Ondov et al., 2016) to give an approximation of intergenomic relatedness to find the ten type strains with the smallest MASH distances per genome. 16S rDNA gene sequences were then extracted from submitted genomes using RNAmmer (Lagesen et al., 2007) and each sequence was compared with the 14883 bacterial type strains available in the TYGS database using BLAST (Camacho et al., 2009). Distances were calculated using Genome Blast Distance Phylogeny (Meier-Kolthoff et al., 2013) along with the 50 closest type strains to finally determine the 10 closest to the query genomes.

The results were provided on 15-06-2021

3.3. Results

3.3.1. Quality of data: Quast

66 of the 67 study accessions were calculated. (One was dropped from the analysis- *Lathyrus latifolius* USA could not be read. However all core and nodulation genes were later found in the genome, so this was possibly a syntax error in the file.)

Results were written to Icarus and Quast contig browsers in HTML see Appendix X

The mean of the total length of the genomes in numbers of bases and the was N50 score by position of the base within the genome were calculated in Microsoft Excel spreadsheets over all taxa. Average total length of the genome sequences (number of bases) was 7713445 which is consistent with the expected size of a whole genome in *Rhizobium leguminosarum* (Young et al., 2006). Standard deviation from the mean was 794998.072. Taxa with significantly different reports 1.5 standard deviations above or below were noted for future reference as there may be missing data, duplications or other errors in these sequences which may add uncertainty to later analyses or interpretation

The mean N50 for all taxa was 225329.591 with a standard deviation of 107362.72.

Taxon	**Total Length**	**N50**
Lathyrus_crassipes_GK448_Argentina	6312775	213868
Vicia_nigricans_GK460_Argentina	10615091	8266
Lathyrus_nigrivalvis_GK438_Argentina	9298075	7337
Lathyrus_japonicus_GK373_Canada	8128171	42309
Lathyrus_polyphyllus_GK12_12_UK	11796400	150999

Table 3. Taxa with >= 1.5 standard deviation from mean N50 (above in green, below in red)

3.3.2. Results of the phylogenetic analyses

All 120 core genes were present in each strain with high coverage. . The results were displayed with Figtree 1.4.4 and saved to files as .svg files which were exported to Adobe Illustrator and annotated with taxonomic interpretation (Figure 3) and geographical interpretation (figure 4). Full size, scalable (readable) pdfs are included as Appendix 3 and 4 respectively.

*Nod*A gene was missing from *Lathyrus crassipes* accessions and some of the reference strains which were omitted from the *nod*A analysis.

Matches to bv. *viciae* for *nodA, nodC* and *nod*D were found for all study taxa with high percentage similarities: above 90% for nodC and nodD, and above 88% for nodA: and very low E values in all cases. The exception was the strain from host plant *Lathyrus crassipes*: nodC gene was weakly similar (24%) to bv. *trifolii*, nodD weakly similar to bv. *phaseoli* (34%) and the *nod*A missing altogether. Further examination of this accession would be needed to find out why. The trees for these genes and the 16S tree are collated in Figure 5.

3.3.3. For 16S

The 16s tree shows very close similarity for all the study accessions except for accessions from *Vicia bithynica* and *Lathyrus littoralis* which are closely grouped with *Rhizobium leguminosaru*m 248 (gsQ) and Rhizobium leguminosarum FA23 gsK, all of which have the same sequence variation between positions 5 - 89 in the alignments. (Figure 5, as above.)

3.3.4. For ANI

The results if the ANI calculations were exported as jpegs and annotated using Adobe Photoshop. The table is displayed with to show hresholds indicating genospecies boundaries (Figure 1) and with annotations for geography of host plants (Figure 2). Clearer versions of these figures are uploaded into Appendix 1 and 2 respectively.

The results of the ANI table bear out the clear boundary between genospecies within the RLc group and between Rlc and other *Rhizobium* species as identified by (J Peter W Young et al., 2021)

Vicia and *Lathyrus* symbionts seem distributed randomly and fairly equally through the chart. Accessions from Japan and South America are clear outliers in these tables.

Five accessions from hosts *Lathyrus magellanicus* Lam. from Chile and Argentina are the same or closely related species, along with *Vicia nigricans* Hook. & Arn. collected in Argentina., These are weakly similar to strains nodulating *Vicia graminea* Sm. and *Lathyrus nigrivalvis* Burkart also in Argentina which have a similarity of 94% to each other: these seven accessions form a distinct cluster well separated from the other accessions in the study.

Lathyrus crassipes accession from Argentina presents as very dissimilar from all others.

Accessions *Vicia bifolia* Nakai and strains of *Lathyrus japonicus* Willd., *Lathyrus quinquenervius* (Miq.) Litv., *Lathyrus davidii* Hance, and *Vicia unijuga* A. Braun collected in Japan form a group together with higher than 95% similarity between all their pairwise identities, though with little similarity to the rest of the group.

3.3.5. for TYGS

Accessions cultured from Japan are *Rhizobium pisi* (formerly *R. leguminosarum* DSM 30132 type strain, DSM 30132(T)=NCIMB 11478(T)), also clustered with *Rhizobium fabeae, e*xcept for the accession hosted by *Vicia bifolia* which is returned as potential sp. nov. However is close to R. fabae and R. vallis. All the South American accessions including *Lathyrus crassipes* also present as potential new species. A phylogram of these matches was produced (Figure 6).

JH142 *Vicia lutea* and JH124 *V. tetrasperma* came out as *R. ruizarguesonis*, and *V. sativa* JHI54 and the two *L. linifolius* JHI1093 and JHI1096 as potential spp. nov., although these are placed as RLc genospecies K, L and L repectively in the core genes phylogeny.

3.4. Limitations of this study

In order to satisfy Koch's postulates the rhizobia cultured and sequenced to provide the genetic material used in this study would need to be introduced to sterile plant roots which would then be grown up and assessed for effective nodulation. Restrictions to access to laboratories and suitable research glasshouses meant that this was not possible within the timeframe of the study. This adds a layer of uncertainty to any results as we can not say for sure that the strains sampled were effectively fixing nitrogen for their host plants. In addition, for those taxa which were found to have absent nod genes (*Lathyrus crassipes*) we do not know whether the strain always lacked these genes or if they had been lost in the process of sequencing, which is a common occurrence. (Young, 2021)

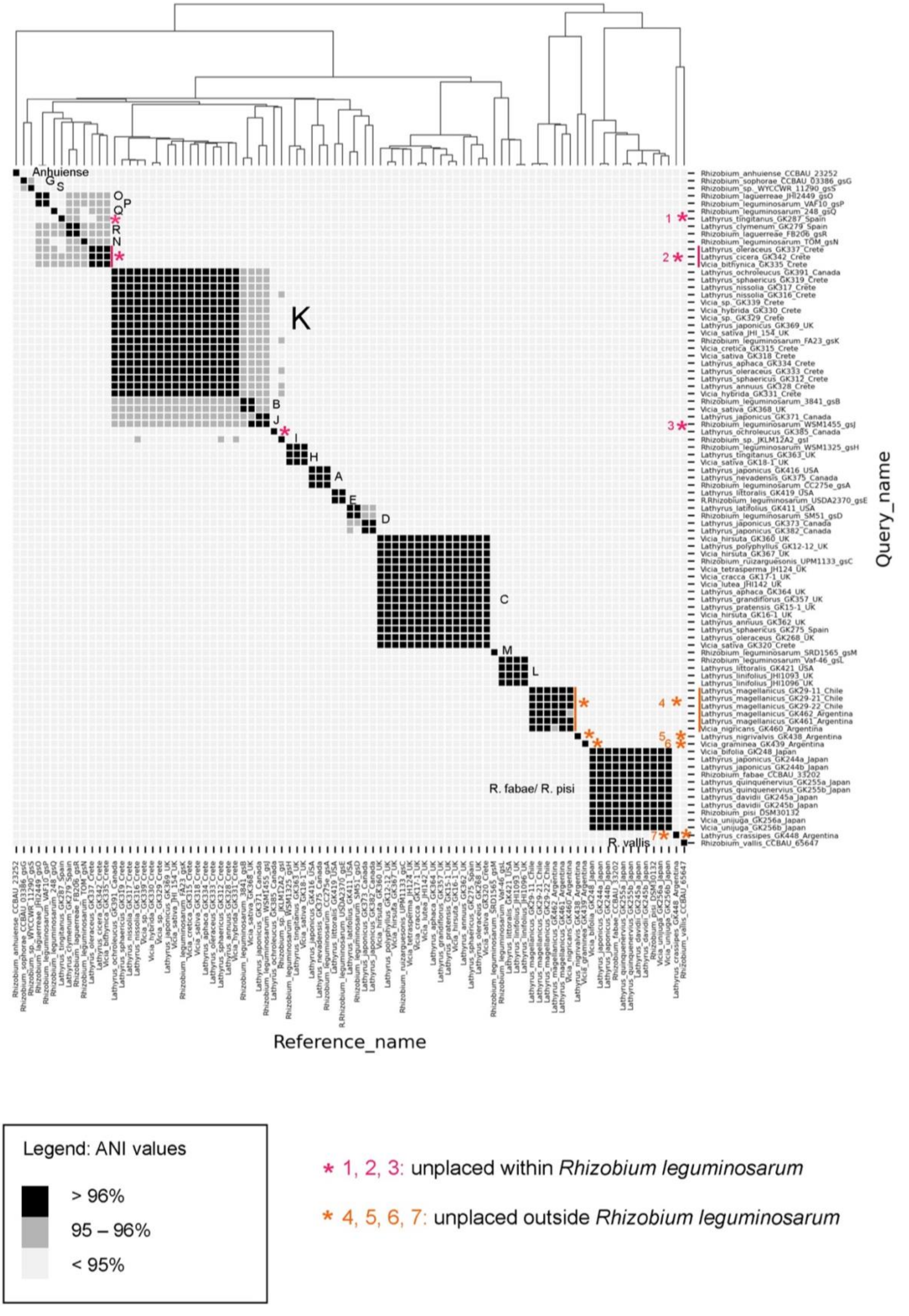
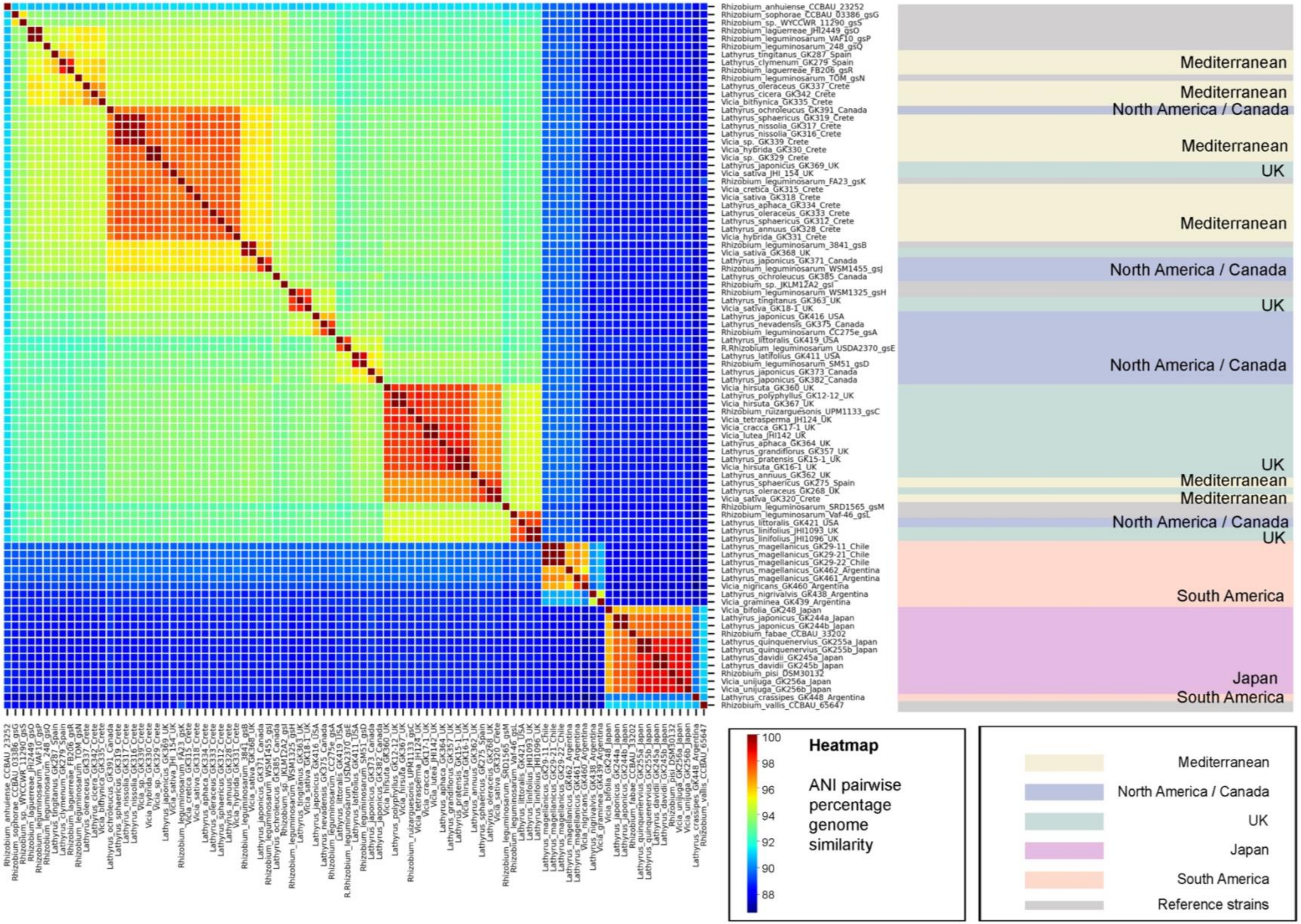


Figure 1. ANI table showing clustering of genospecies of Rhizobium leguminosarum species complex as defined by (J Peter W Young et al., 2021) Thresholds indicate genospecies boundaries with similarity over 95% (grey) and over 96% (black). The dendrogram at the top shows distance values and should not be interpreted as showing ancestral relationships.

For a full scale , zoomable PDF please see Appendix 1.

Figure 2. ANI Table showing pairwise percentage genome similarity for all accessions, annotated on the right to show the geographic location of collection. PDF: Appendix 2.



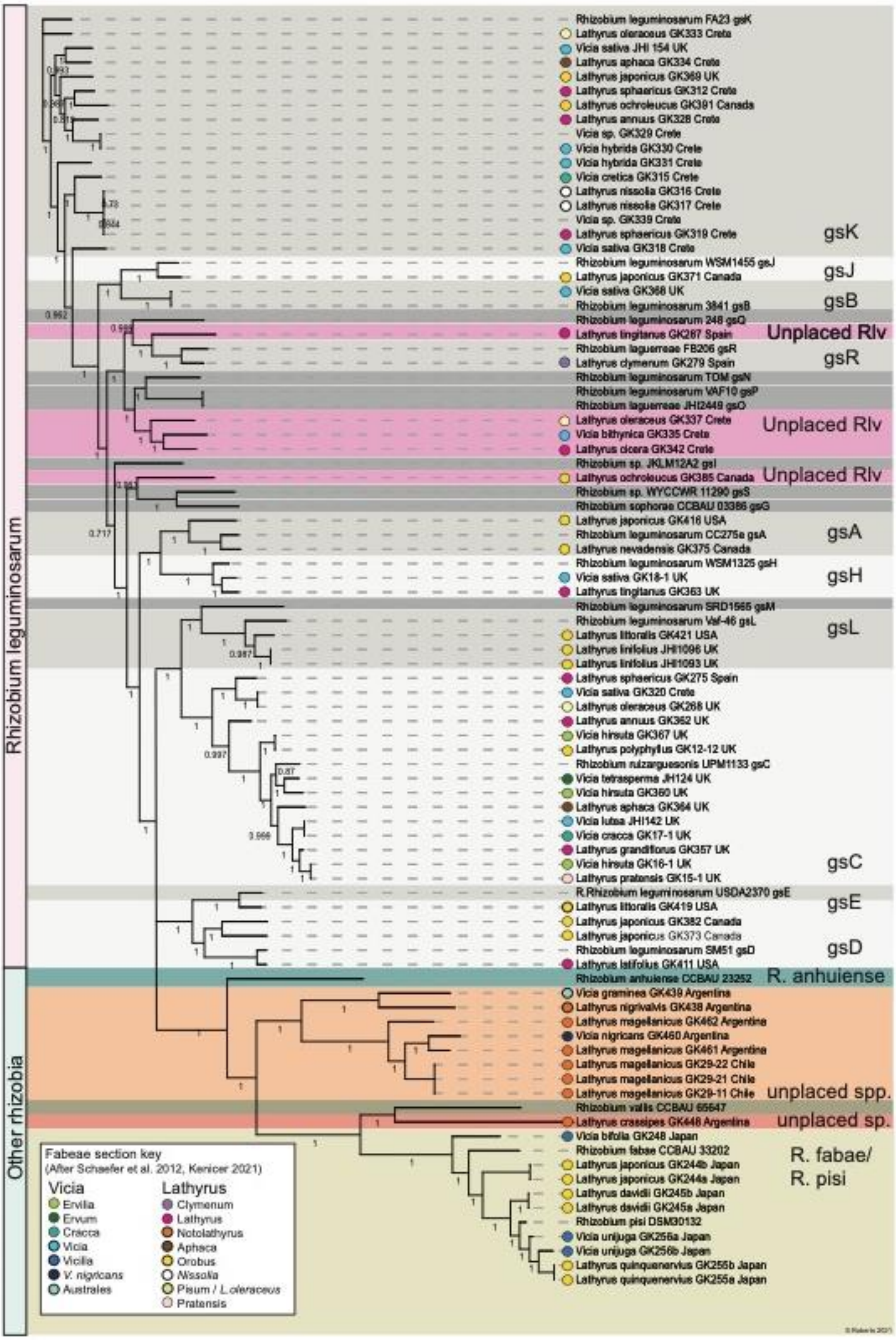


Figure 3. Phylogenetic analysis of bac120 set of core chromosomal genes for all accessions, with taxonomic annotations

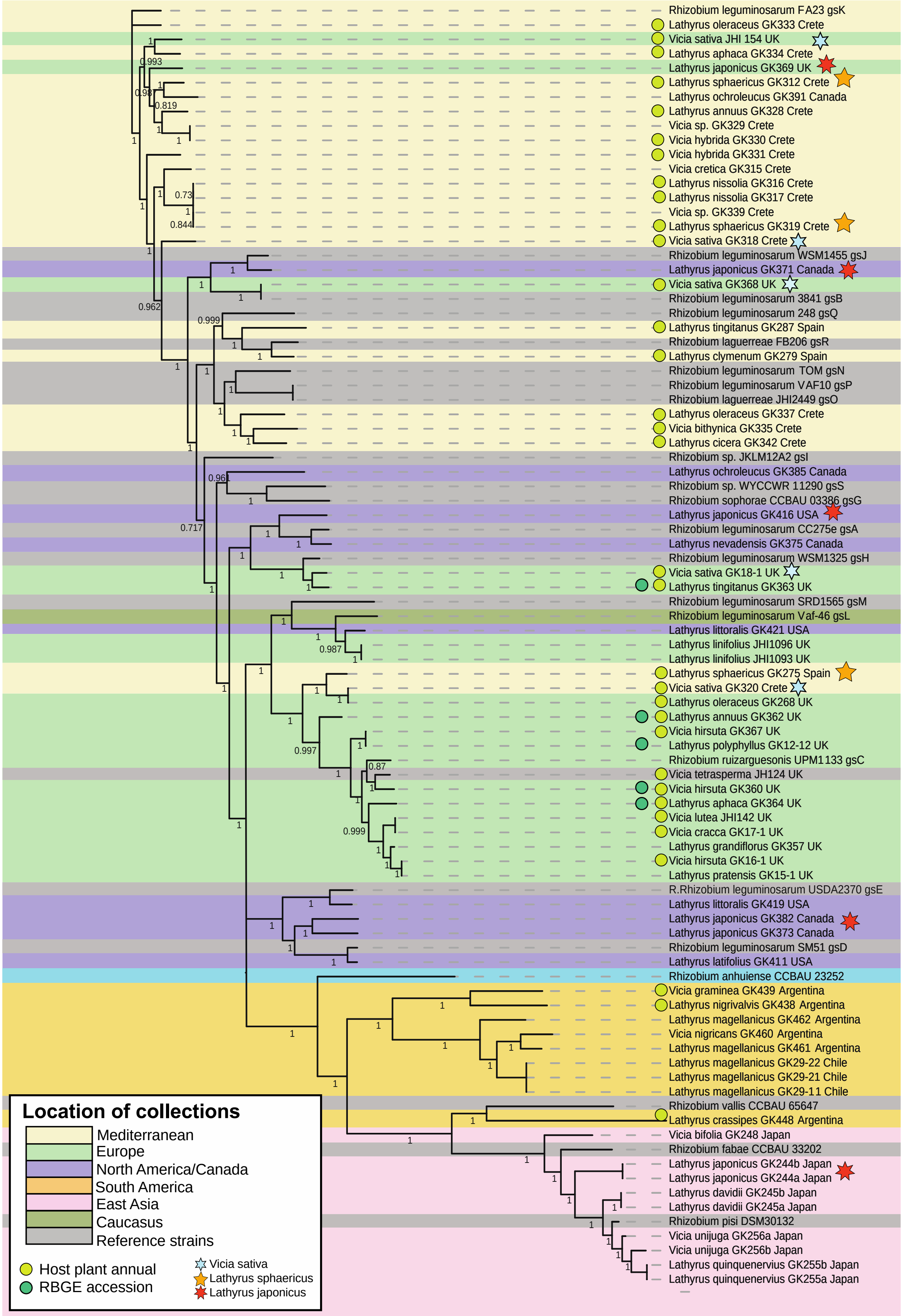


Figure 4. Phylogenetic analysis of bac120 set of core chromosomal genes for all accessions; geographical annotations

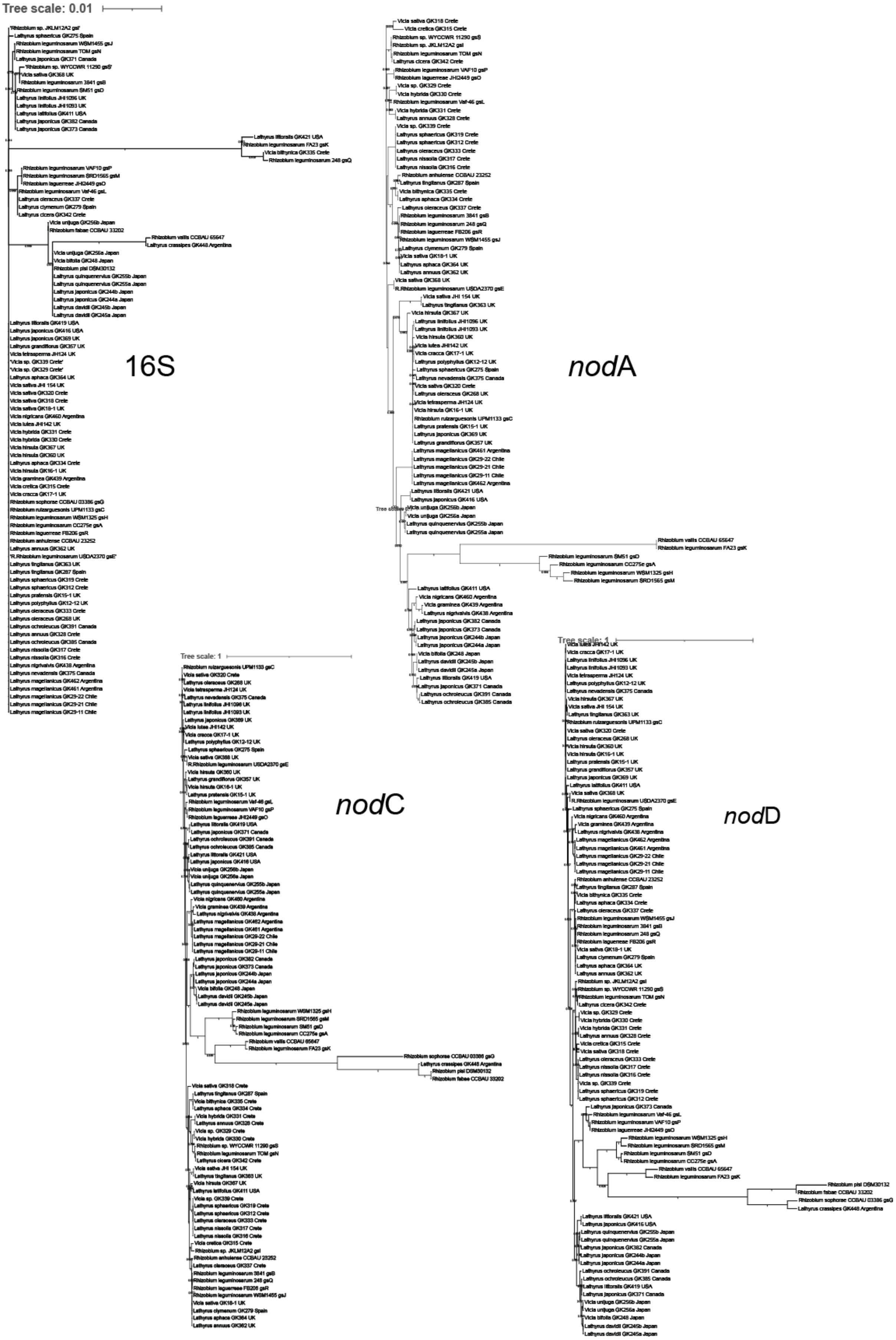


Figure 5 Maximum-likelihood trees of nodACD genes and the highly conserved 16S sequence.

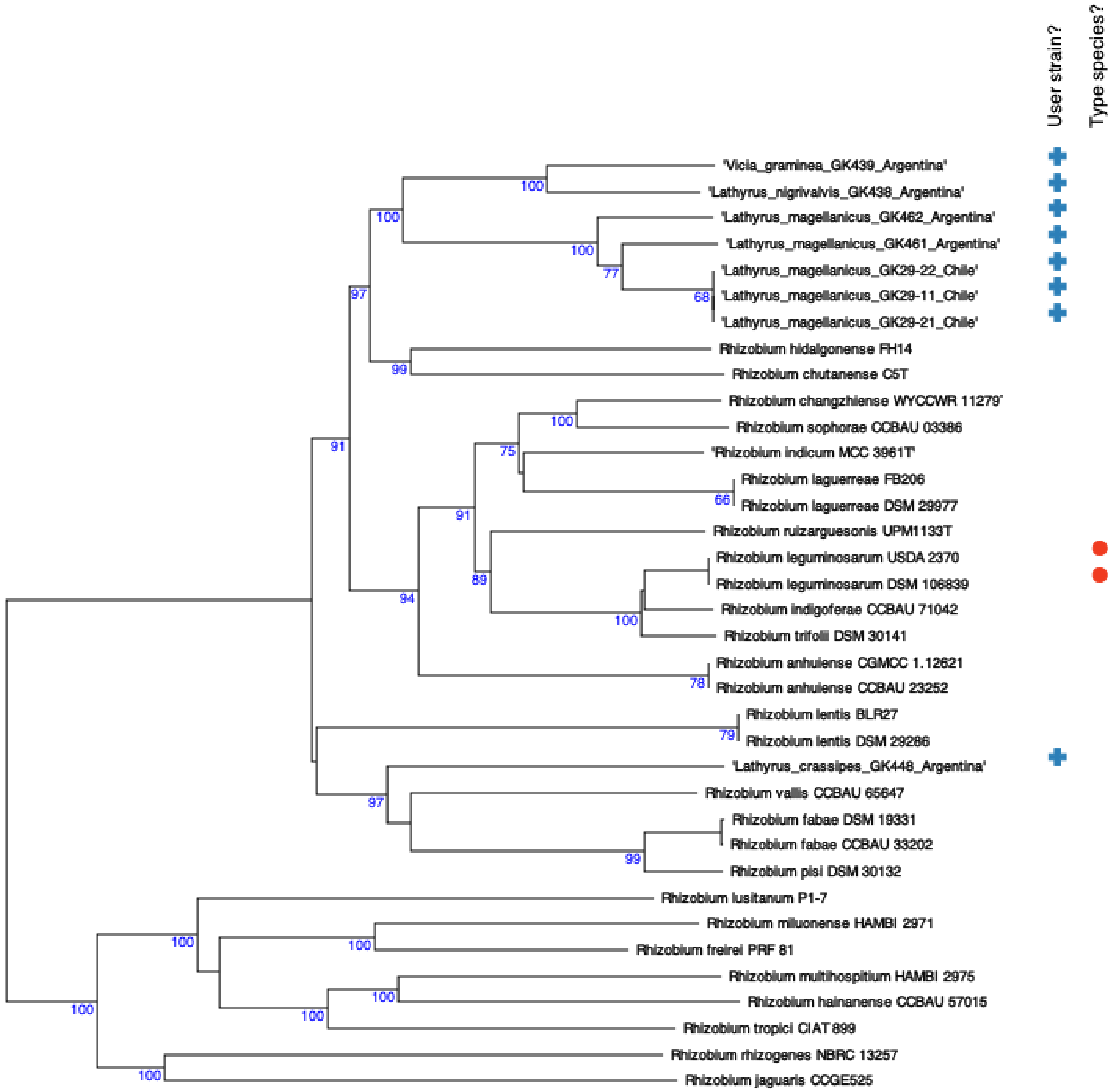


Figure 6. TYGS phylogram showing matches for queries of the whole genomes of the South American accessions.

4. Discussion

Given the lack of a known, alignable sister group to Rhizobium, the trees could not reliably be rooted, so the trees presented do not necessarily show the most basal group as ancestral. However, relationships within the major clades are generally well supported throughout with high SH support for all branches except for the node at the base of gsL which has a value of 0.717.

The results of ANI analysis and phylogenetic analysis of the core 120 rhizobial genes showed general trends for geographical grouping over very large areas, reflecting host plant biogeography and the promiscuity of the plants. The phylogeny of the rhizobia does not seem to be strongly correlated with the phylogeny of the hosts. This is consistent with findings from studies of crop rhizobia (Laguerre Gisèle et al., 2003; Mutch and Young, 2004). Clearly, the influence of geography is stronger than influence of the host plant phylogeny.

The phylogeny for 16S sequences reflected that the region is highly conserved across the accessions. Phylogenies for the nodulation genes reflected the selection by Fabeae plants of symbiovar *viciae* genes.

4.1. Phylogeny of host plants

Lathyrus and *Vicia* have similar biogeographical histories with centres of diversity in the Eastern Mediterranean (Kupicha, 1983). A comprehensive biogeographical and phylogenetic study of Fabeae published in 2012 by (Schaefer et al., 2012) further clarifies work done on Vicia and Lathyrus. For Fabeae there is most likely long-distance dispersal via sea from the Eastern Mediterranean to South America 3-4 Ma (Kenicer, 2007; Kenicer et al., 2005; Schaefer et al., 2012) *Lathyrus* (Notolathyrus group) probably reached South America by long-distance dispersal from Eurasia (Kenicer et al., 2005).

Some important groups of Fabeae are not represented in this study, in particular, *Lens (Vicia lens)* and *Vavilovia (Lathyrus formosus)*. However, there is a cross-section through the tribe, according to relationships determined by molecular analyses (Schaefer et al., 2012).

Genus	**Section**	**Species**	**Notes**
Ervilia		*Vicia hirsuta*	
Ervum		*Vicia tetrasperma*	

	Cracca	*Vicia cretica*	
Vicia	Vicia	*Vicia lutea*, *Vicia sativa, Vicia bithynica*	from different clades within large section Vicia
	Vicilla	*Vicia unijuga, Vicia bifolia*	
	Vicia nigricans	*Vicia nigricans*	
Lathyrus	Clymenum	*Lathyrus clymenum*	
	Lathyrus	*Lathyrus latifolius, Lathyrus grandiflorus*, *Lathyrus annuus, Lathyrus cicera, Lathyrus tingitanus, Lathyrus sphaericus*	
	Notolathyrus	*Lathyrus crassipes, Lathyrus magellanicus, Lathyrus nigrivavlis*	
	Aphaca	*Lathyrus aphaca*	
	Orobus	*Lathyrus vernus, Lathyrus davidii, Lathyrus littoralis, Lathyrus japonicus Lathyrus polyphyllus, Lathyrus ochroleucus* and *Lathyrus nevadensis*	from different clades within large section Orobus
	Lathyrus oleraceus (Pisum sativum)	*Lathyrus oleraceus (Pisum sativum)*	wild pea

Table 4. Host plants of accessions in study by tribe Fabeae, according to relationships determined by molecular analyses by (Schaefer et al., 2012)

The results found that rhizobial genotypes were scattered through the host phylogeny. (Figure 3). Host life cycles and lifestyles appeared to have little bearing on selection of mutualists: European and Mediterranean host species are mostly annuals and often ruderals but the same, or similar bacteria are forming partnerships with both annual and perennial species that are geographically close. Equally, some of the plant hosts seem very promiscuous.

4.2. Promiscuity of host plants

There are host plant species with several accessions from different geographical locations in this study which can give us an idea of the promiscuity of the hosts. These are highlighted in Figure 4.

A distribution through the bacterial phylogeny might be expected from *Lathyrus oleraceus* (*Pisum)*, which is known to be very promiscuous (Laguerre et al., 2007; Ramírez-Bahena et al., 2008b; Zhang et al., 2001) and has a large Mediterranean and Near Eastern native distribution and is naturalised around much of the world. Although the accessions here are from wild-type hosts, the species has been transported by humans as a crop so a global distribution of the symbionts is not unexpected (Shade et al., 2017).

It is interesting that this is also the case for *Lathyrus sphaericus* symbionts in this study. One of the accessions from Mediterranean *Lathyrus sphaericus* host plants, *L. sphaericus* GK 275, was collected from Sierra Morena, Spain, in sandy clay soil disturbed by wild boar and found to be genospecies C; *L. sphaericus.* GK 312 was found to be in gsK.; this was collected from loose sandy clay substrate on the Dikti mountains in Crete. Despite their similar habitats, the accessions are phylogenetically somewhat distant within *Rhizobium leguminosarum.* An accession from nodules of host plant *Vicia sativa* GK 320, which is phylogenetically distant from *L. sphaericus* was also collected from under the Diktaeon Cave of Zeus n Crete in very loose substrate in Crete, was also found to be genospecies C.

Two accessions of *Vicia sativa* in this study from various locations were ascribed to genospecies K, one accession to genospecies B, and one to gsH. These were from hosts GK318 from Crete, no collection data; JHI154 from seaside town Carnoustie, Angus, Scotland and GK368 from along the coast in Monifieth, Angus, Scotland; and GK18-1 collected in the UK. As *Vicia sativa* is a ruderal annual with a widespread wild distribution throughout Eurasia and North Africa (POWO, 2019) and is cultivated for fodder globally it is perhaps unsurprising to find it nodulated promiscuously with partners across the Rhizobium leguminosarum clade.

Lathyrus japonicus is well represented in this study, with accessions from similar habitats in UK, Canada, USA and Japan. *L. japonicus* GK369 Angus, growing behind and among sea defences at Carnoustie, Angus, UK, was assigned to gsK. Accessions GK371 and GK373 from the foreshore of Oak Bay beach Canada, British Columbia were placed in gsJ and gsD. *L. japonicus.* GK382also from British Columbia, collected from sand at Stanley Park Third Beach, Vancouver, was also GSD. *L. japonicus* GK244a and GK244b (different nodules from the same host plant) were collected from very sandy substrate on the coast at the southernmost point of Chiba peninsula in Japan, these were placed in the *Rhizobium pisi/R. fabae* clade. *L. japonicus* GK416, collected from plants growing among driftwood in pure sand on the foreshore at Deception Pass State Park, Washington State, USA was gsA. In order to survive the harsh seashore conditions, rhizobia must also be adapted to cold, salt and drought, and it is interesting that the accessions are distributed widely throughout this rhizobial phylogeny, rather than presenting together as a potential especially adapted strain.

Results here show a strong affinity for *Rhizobium leguminosarum* in much of the Northern Hemisphere but the Japanese, and the South American species form partnerships with other rhizobia. A larger sample size would be desirable to see whether this is a general trend or if the small sample size gives an exaggerated view of this tendency. (Sprent et al., 2017) note a strong North-South difference in nodulation, but it is hard to say what is happening here without sampling more host species. Sampling more deeply in some of the geographical areas, including East Asia, North America and South America are clearly needed.

Lathyrus oleraceus, L. sphaericus, and *Vicia sativa* are common and widespread annuals. *Lathyrus japonicus* is a more habitat-restricted perennial but also has a very large distribution. Good sampling across locations for this study is probably because they are common and easy to find. It is perhaps possible that they are common because they are promiscuous and this is a selective advantage for the plants; an examination of the promiscuity of rare or geographically restricted species would help to understand if there are correlations.

4.3. Promiscuity of bacteria

Reciprocally, it may be possible to make some inferences about the promiscuity of the bacteria. Close relationships in the core 120 genes phylogeny and >96% ANI are found for four accessions in genospecies L. The type strain for this group is *Rhizobium leguminosarum* Vaf-46, a fast-growing rhizobium isolated from root nodules of *Lathyrus formosus* from a montaine habitat in Armenia (Kimeklis et al., 2019). The authors found high nucleotide similarity in the housekeeping genes in strains from this location which was interpreted as adaptation to local soils or genetic drift after long-term ecological isolation and genetic separation. However, in this study this genospecies was found in the nodules of two accessions from *Lathyrus linifoilius,* (JHI1096 and JHI1093) that were collected at James Hutton Institute, Angus, Scotland, a plant host with a typical grassland habitat (Boivin et al., 2020), and one accession collected from nodules on *L. littoralis* in dune slacks in sand and bryophyte carpet, in Whidbey Island, Deception Pass State Park in Washington State, Island County, USA. Unconscious distribution by humans is a strong possibility for the JHI accessions and perhaps also for the rural coastal location of *L. littoralis* on the West coast of Northern America, as Whidbey Island is agricultural, was settled in the late 1800s by Dutch settlers, and has extensive links globally through large naval bases on the island. (Kenicer, 2021).

Nevertheless these bacteria seem to be promiscuous with different Fabeae and enjoy a very broad potential distribution over different habitats.

Some idea of promiscuity might be inferred from the larger clades where same genotypes were recovered from different host species in different Fabeae clades, with closely related rhizobia nodulating phylogenetically distant plants. In particular, the ANI table Figure 1 demonstrates closely similar genomes for three large groups assigned to genospecies K, C and *Rhizobium pisi/fabeae.* These groups cluster as clades in the core 120 genes phylogeny, discussed below. Studies using bacteria from the original isolates to inoculate seedlings of the hosts to assess the ability of the isolates to cross-nodulate with the different host plants would potentially give a clearer picture of bacteria promiscuity.

4.4. Biogeography

4.4.1. Mediterranean

All the Mediterranean accessions here were found to be in the *Rhizobium leguminosarum* clade. Mediterranean species appear to be shown as basal in the core genes phylogeny. (Figure 3 and 4) However, as this is not a fully rooted tree, the topology of the tree should not imply the basal lineages as being the earliest evolved. Although an Eastern Mediterranean origin and centre of diversification has been hypothesised for Fabeae (Kenicer et al., 2005) , the history of diversification and co-evolution of *Rhizobium* along with the legumes is not wholly resolved (Provorov et al., 2020; Sprent, 1994) and the evolutionary origins of the particular bacteria sampled here are not established.

Accessions from nodules of the fairly widespread European species *Vicia hirsuta*, *V. sylvatica* and *Lathyrus sphaericus* were found to be different genospecies in the Rlc and occupy different locations throughout the core genes phylogeny.

Accessions in the clade corresponding to genospecies K (J Peter W Young et al., 2021) consists of accessions from diverse *Vicia* and *Lathyrus* species in this study, mostly from Crete. Many have ruderal habits but there is no discernible pattern to their phylogenetic distribution here, nor does there seem to be a geographical pattern on a finer scale. The type strain for gsK used here is *R. leguminosarum* bv. *phaseolii* FA23 GCF 000419705.1, isolated from *Phaseolus vulgaris* in Gmina Siedliszcze, Poland, there is no way to account for whether this was introduced to the soil by humans, or whether it is simply very common and widespread.

Further groups of accessions from Mediterranean hosts appear in the *Rhizobium leguminosarum* clade phylogeny, also without any apparent pattern. A Spanish accession from *Lathyrus tingitanus* GK287 and a small clade of three accessions from Cretan hosts are unplaced here within the Rlc clade: *Lathyrus oleraceus GK337*, *L. cicera* GK342 and *Vicia bithynica* GK335. These three are from the Katharo plateau in Crete which is geographically isolated and known for traditional farming practices; it is possible that these bacteria may be relict species (Kenicer, 2021) .. However the ANI chart (Figure 1shows that there is a high (>95%) degree of similarity between these and each other, and these all fall within the closely related gsN,O, P, Q, R clade denoted as the F-clade by (J Peter W Young et al., 2021).

4.4.2. Europe

In the large clade of genospecies C, which includes accessions mostly from the UK but also two from Spain and Crete, rhizobia were isolated from host plants across the Fabeae phylogeny, including species from *Vicia* sections Ervilia, Ervum, Cracca, Vicia; and *Lathyrus* sections Lathyrus, Aphaca, Orobus, Pratensis and Pisum (*L.oleraceus.*)

European plants are hypothesised to have dispersed into Europe directly from the centre of origin in the Mediterranean in several separate direct dispersal events.

There is certainly no signal from the ancestry of the plants here. Some plant hosts in this group were grown at research institutions Royal Botanic Garden Edinburgh (RBGE) and fairly nearby James Hutton Institute (JHI) in Scotland. The rhizobial accession from host *Lathyrus polyphyllus* GK12-12(RBGE, native to western USA) is shown as identical with the partner of UK native *Vicia hirsuta*GK367, and that the accession from *Vicia lutea* JHI142 (JHI, native to western Eurasia) is shown as identical with the accession from UK native *Vicia cracca* GK17-1.

4.4.3. North America

Many of the North American host plants here are in *Lathyrus* section Orobus. There are two accessions of *Lathyrus ochroleucus,* collected in Alberta, Canada from the same car park in *Picea* woodland. *Lathyrus ochroleucus* GK 385 presents as gsK and *L. ochroleucus* GK391 is unplaced in Rlc, and in a different clade of the core genes phylogeny. Orobus hosts also included are *Lathyrus nevadensis* gsA and *Lathyrus polyphyllus* gsC. It is hypothesised that these species have migrated from the Mediterranean into northern Eurasia and then over the Beringian land bridge into North America (Kenicer et al., 2005; Schaefer et al., 2012). For a synopsis of this dispersal, see Figure 4(Schaefer et al., 2012)

There are several North American and Canadian accessions here of *Lathyrus japonicus* GK371, 373, 382 and GK416. Also present is an accession isolated from *Lathyrus littoralis*, a similar and possibly closely related species but with a smaller North American and Canadian distribution. This is the only accession from the study set that appears to be gsE, which is *Rhizobium leguminosarum* in the strict sense (J Peter W Young et al., 2021).

4.4.4. South America

South American accessions are clustered together. The accessions from Notolathyrus are all nodulated by a clade of bacteria that could not be identified and may represent new taxa. The South American Notolathyrus group is thought to have migrated to the New World from Eurasia via long distance dispersal. *Lathyrus crassipes* and *L. magellanicus* are hypothesised by to be part of the same dispersal event from the Mediterranean via sea to South America. *Vicia nigricans* is hypothesised to have split form Eurasion *Vicia* lineages via a different long-distance dispersal event from the Mediterranean to South America, although there is a possibility that this dispersal was from Asia via Beringia. (Kenicer et al., 2005; Schaefer et al., 2012)

Here, *Vicia nigricans* GK 460 and *Lathyrus magellanicus* GK 461 were collected from the same location in open forest at Rio Negro, Argentina. My results show that accessions taken from the nodules of these plants are closely related. These form a clade with the accession taken nearby from *Lathyrus magellanicu*s GK 462, and the three Chilean *Lathyrus magellanicus* accessions.

Host plants of *Vicia graminea* GK439 and *Lathyrus nigrivalvis* 438 were growing close together near a small lorry park in seasonally inundated Parana wetland in Argentina, these are part of the same clade and sister taxa here. These host plants and *Lathyrus crassipes* are all from Lowland Eastern Argentina.

4.4.5. Japan

The plant hosts in the study were from *Vicia* section Vicilla and *Lathyrus* section Orobus. The bacterial partners are *Rhizobium pisi* and *R. fabae*, which appear closely related in this study, which is surprising as the accessions were collected from a wide area around the central Pacific coast of Honshu, Japan. It may be that the isolation of the island has some influence on this but it may equally be a sampling artefact as the sample size is small (seven host plants). There is good habitat representation here: *Lathyrus davidii* GK245 being an understory herb in a *Cryptomeria japonica* plantation, with diverse herb community; *Lathyrus quinquenervius* GK255 from periodically flooded riparian grasslands; *Vicia bifolia* GK248 under mixed tree species in the grounds of Nikko botanical garden; *L. japonicus* GK244 from sandy coastal turf. *Rhizobium pisi* was also found nodulating with *Lathyrus japonicus* in a study of the plant and its rhizobia in Japan (Aoki et al., 2010), A larger dataset of sixty-one accessions of *L. japonicus* in their study revealed greater diversity, including several strains of *Rhizobium leguminosarum*, and *Ensifer* and *Mesorhizobium* spp. However this contrasts other studies in Japan and China, where authors found that the plants were mostly nodulating with *Rhizobium anhuiense* (Li et al., 2016)

Lathyrus japonicus is a circumpolar species with long-range sea-drifted distribution (Kenicer and Parsons, 2021). It is a good model for considering the relationships with rhizobia as it grows in a challenging environment that is not cultivated and is naturally dispersed by sea.

For the long-distance dispersed plants, there remain questions. It seems that if the rhizobia are carried along with the seeds and so vertically inherited (Truyens et al., 2015), the plant might have no need to find symbiotic partners immediately on arrival. Likewise if the seed lands on relatively accommodating soil then immediate nodulation might not be necessary.

However a harsh, arid, saline beach soil might be a case whre the symbiont would need to be present and furthermore that compatible nodulation genes would need to be available. Although it seems likely that rhizobia are continually dispersed across continents (Smith David J. et al., 2013) new bacterial arrivals would face selection pressures from soils and climates they were not adapted to, and competition from established microbial communities. For symbionts of the New World plants, this would perhaps be an ancestral problem, however for sea-drifted species such as *Lathyrus japonicus* it may be an issue for each new population. The chief issue here may be the need of the plants to find suitable nodulation genes in the bacteria. Alternatively perhaps the plant itself be the vector here.

4.5. Nodulation genes

Full coverage of nodulation genes A, C and D for all accessions except *L. crassipes* may be taken as some indication that there may be a possible effective symbiotic partnership. (see appendix...X - coverage chart from BLAST), though the presence of these genes is not a guarantee that the process would be effective (Bañuelos-Vazquez et al., 2020)

It is unsurprising that all other accessions were found to be biovar. *viciae*, as Fabeae have been shown to preferentially select these genes (Kumar et al., 2015). It was also expected that they would be closely related. The gene trees (Figure 5)demonstrate the close relationships across the accessions. Clades on very short branches are visible that do not correspond exactly to the core genes phylogenies, suggesting that there may be some genes or alleles that have phenotypic or evolutionary significance for the host plants or environments. Closer examination of the particular differences in these clades might be a subject of further study, perhaps using a more refined ML method such as RAxML.

Outliers in small clades on long branches are the reference genes symbiovar *trifolii* and *phaseolii* and additionally, in the case of *nod*C and *nod*, bv. Trifolii and phaseolii matches for *Lathyrus crassipes*, which would ideally need further confirmation.

Ideally, this study could be extended by fulfilling Koch's postulates by testing the cultured accessions' ability to effectively nodulate with plants and fix nitrogen, the exact nature of their relationships with their host plants is unproven.

5. Conclusion

Results here show that Fabeae are undoubtably quite promiscuous in their choice of rhizobia and that a high degree of host specificity was not evident anywhere in this study for these plants. Host plant phylogeny does not appear to exert much influence on the selection of partners, and biogeography seems to play a greater role in determining which rhizobial species or genospecies will form partnerships with the plants. *Rhizobium leguminosarum* is important in the Northern Hemisphere, but none of the accessions collected from the Southern Hemisphere were found to be within the large and widespread *Rhizobium leguminosarum* complex. However, the sample size for the study is very small. In particular, the South American results are interestingly clustered together, so more sampling is needed to fill in this picture .

All accessions from South America are identified by TYGS as potential new species. The accession *Lathyrus crassipes* GK448 is particular outlier and could be investigated as a new species.

References

Abdelkrim, S., Jebara, S.H., Saadani, O., Abid, G., Taamalli, W., Zemni, H., Mannai, K., Louati, F., Jebara, M., 2020. In situ effects of Lathyrus sativus- PGPR to remediate and restore quality and fertility of Pb and Cd polluted soils. Ecotoxicol. Environ. Saf. 192, 110260. https://doi.org/10.1016/j.ecoenv.2020.110260

Abdelkrim, S., Jebara, S.H., Saadani, O., Chiboub, M., Abid, G., Jebara, M., 2018. Effect of Pb-resistant plant growth-promoting rhizobacteria inoculation on growth and lead uptake by Lathyrus sativus. J. Basic Microbiol. 58, 579–589.

Abdelkrim, S., Jebara, S.H., Saadani, O., Chiboub, M., Abid, G., Mannai, K., Jebara, M., 2019. Heavy metal accumulation in Lathyrus sativus growing in contaminated soils and identification of symbiotic resistant bacteria. Arch. Microbiol. 201, 107–121. https://doi.org/10.1007/s00203-018-1581-4

Adiguzel, A., Ogutcu, H., Baris, O., Karadayi, M., Gulluce, M., Sahin, F., 2010. Isolation and characterization of Rhizobium strains from wild vetch collected from high altitudes in Erzurum-Turkey. Rom Biotechnol Lett 15, 5017–5024.

Ampomah, O., James, E., Iannetta, P., Kenicer, G., Sprent, J., Huss-Danell, K., 2012. Nodulation and ecological significance of indigenous legumes in Scotland and Sweden. Symbiosis 57. https://doi.org/10.1007/s13199-012-0188-9

Ampomah, O.Y., Huss-Danell, K., 2016. Genetic diversity of rhizobia nodulating native Vicia spp. in Sweden. Syst. Appl. Microbiol. 39, 203–210. https://doi.org/10.1016/j.syapm.2016.02.002

Andrews, M., Andrews, M.E., 2017. Specificity in Legume-Rhizobia Symbioses. Int. J. Mol. Sci. 18. https://doi.org/10.3390/ijms18040705

Andronov, E.E., Igolkina, A.A., Kimeklis, A.K., Vorobyov, N.I., Provorov, N.A., 2015. [Characteristics of Natural Selection in Populations of Nodule Bacteria (Rhizobium leguminosarum) Interacting With Different Host Plants]. Genetika 51, 1108–1116.

Aoki, S., Kondo, T., Prévost, D., Nakata, S., Kajita, T., Ito, M., 2010. Genotypic and phenotypic diversity of rhizobia isolated from Lathyrus japonicus indigenous to Japan. Syst. Appl. Microbiol. 33, 383–397. https://doi.org/10.1016/j.syapm.2010.07.001

Asfaw, B., Aserse, A.A., Asefa, F., Yli-Halla, M., Lindström, K., 2020. Genetically diverse lentil- and faba bean-nodulating rhizobia are present in soils across Central and Southern Ethiopia. FEMS Microbiol. Ecol. 96. https://doi.org/10.1093/femsec/fiaa015

Atallah, T., Rizk, H., Cherfane, A., Daher, F.B., El-Alia, R., De Lajuide, P., Hajj, S., 2008. Distribution and Nodulation of Spontaneous Legume Species in Grasslands and Shrublands in Mediterranean Lebanon. Arid Land Res. Manag. 22, 109–122. https://doi.org/10.1080/15324980801957978

Baas-Becking, L.G.M., 1934. Geobiologie; of inleiding tot de milieukunde. WP Van Stockum & Zoon NV.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Bañuelos-Vazquez, L.A., Cazares, D., Rodríguez, S., Cervantes-De la Luz, L., Sánchez-López, R., Castellani, L.G., Tejerizo, G.T., Brom, S., 2020. Transfer of the Symbiotic Plasmid of

Rhizobium etli CFN42 to Endophytic Bacteria Inside Nodules. Front. Microbiol. 11, 1752. https://doi.org/10.3389/fmicb.2020.01752

Baymiev, An. K., Ptitsyn, K.G., Muldashev, A.A., Baymiev, Al. K., 2012. Genetic description of root nodule bacteria of Lathyrus species growing on the territory of the Republic of Bashkortostan. Russ. J. Genet. Appl. Res. 2, 122–126. https://doi.org/10.1134/S2079059712020025

Baymiev, An. Kh., Ivanova, E.S., Ptitsyn, K.G., Chubukova, O.V., Baymiev, Al. Kh., 2011. Phylogenetic analysis of symbiotic genes of nodule bacteria in plants of the genus Lathyrus (L.) (Fabaceae). Mol. Genet. Microbiol. Virol. 26, 154–158. https://doi.org/10.3103/S0891416811040021

Belhadi, D., De Lajudie, P., Ramdani, N., Le Roux, C., Boulila, F., Tisseyre, P., Boulila, A., Benguedouar, A., Kaci, Y., Laguerre, G., 2018. Vicia faba L. in the Bejaia region of Algeria is nodulated by Rhizobium leguminosarum sv. viciae, Rhizobium laguerreae and two new genospecies. Syst. Appl. Microbiol. 41, 122–130.

Boivin, S., Ait Lahmidi, N., Sherlock, D., Bonhomme, M., Dijon, D., Heulin-Gotty, K., Le-Queré, A., Pervent, M., Tauzin, M., Carlsson, G., 2020. Host-specific competitiveness to form nodules in Rhizobium leguminosarum symbiovar viciae. New Phytol. 226, 555–568.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Broughton, W.J., Perret, X., 1999. Genealogy of legume-Rhizobium symbioses. Curr. Opin. Plant Biol. 2, 305–311. https://doi.org/10.1016/S1369-5266(99)80054-5

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Cavassim, M.I.A., Moeskjær, S., Moslemi, C., Fields, B., Bachmann, A., Vilhjálmsson, B.J., Schierup, M.H., W Young, J.P., Andersen, S.U., 2020. Symbiosis genes show a unique pattern of introgression and selection within a Rhizobium leguminosarum species complex. Microb. Genomics 6, e000351. https://doi.org/10.1099/mgen.0.000351

Cevheri, C., 2014. Nodulation study of natural forage legume in semiarid region, Turkey. Pak J Biol Sci 17, 535–539.

Chapman, M.J., Margulis, L., 1998. Morphogenesis by symbiogenesis. Int. Microbiol. Off. J. Span. Soc. Microbiol. 1, 319–326.

Chatterjee, R., 2009. Projecting the future of nitrogen pollution. Environ. Sci. Technol. 43, 1659–1659. https://doi.org/10.1021/es9002179

Chen, W., WANG, E., WANG, S., LI, YINGBO, CHEN, X., LI, YING, 1995. Characteristics of Rhizobium tianshanense sp. nov., a Moderately and Slowly Growing Root Nodule Bacterium Isolated from an Arid Saline Environment in Xinjiang, People's Republic of China. Int. J. Syst. Evol. Microbiol.

Chen, W.-M., Zhu, W.-F., Bontemps, C., Young, J.P.W., Wei, G.H., 2010. Mesorhizobium alhagi sp. nov., isolated from wild Alhagi sparsifolia in north-western China. Int. J. Syst. Evol. Microbiol. 60, 958–962.

Chinnasamy, G., Bal, A.K., 2003. Seasonal changes in carbohydrates of perennial root nodules of beach pea. J. Plant Physiol. 160, 1185–1192. https://doi.org/10.1078/0176-1617-01090

Chinnasamy, G., Davis, P.J., Bal, A.K., 2003. Seasonal changes in oleosomic lipids and fatty acids of perennial root nodules of beach pea. J. Plant Physiol. 160, 355–365.

Chirak, E.R., Kimeklis, A.K., Karasev, E.S., Kopat, V.V., Safronova, V.I., Belimov, A.A., Aksenova, T.S., Kabilov, M.R., Provorov, N.A., Andronov, E.E., 2019. Search for ancestral features in genomes of Rhizobium leguminosarum bv. viciae strains isolated from the relict legume Vavilovia formosa. Genes 10, 990.

Chubukova, O.V., Baymiev, Al. Kh., Baymiev, An. Kh., 2011. Polymorphism of lectin genes in Lathyrus plants. Russ. J. Genet. 47, 813–818. https://doi.org/10.1134/S1022795411070040

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.-W., De Meyer, S., Trujillo, M.E., 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol.

ÇİLDİR, H., 2011. Morphology, anatomy and systematics of the genus Lathyrus L.(Leguminosae) in Central Anatolia, Turkey. MIDDLE EAST TECHNICAL UNIVERSITY.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2016. GenBank. Nucleic Acids Res. 44, D67–D72. https://doi.org/10.1093/nar/gkv1276

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633-642. https://doi.org/10.1093/nar/gkt1244

Coulot, P., Rabaute, P., 2016. Monographie des Leguminosae de France, Tome 4: Tribu des Fabeae, des Cicereae et des Genisteae. Société Botanique du Centre Ouest.

de Lajudie, P., Martinez-Romero, E., 2017. International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of Agrobacterium and Rhizobium Minutes of the meeting, 7 September 2014, Tenerife, Spain. Int. J. Syst. Evol. Microbiol.

de Lajudie, P., Mousavi, S.A., Young, J.P.W., 2021. International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Rhizobia and Agrobacteria Minutes of the closed meeting by videoconference, 6 July 2020. Int. J. Syst. Evol. Microbiol.

de Lajudie, P., Young, J.P.W., 2019. International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Rhizobia and Agrobacteria Minutes of the meeting by video conference, 11 July 2018. Int. J. Syst. Evol. Microbiol.

de Lajudie, P.M., Andrews, M., Ardley, J., Eardly, B., Jumas-Bilak, E., Kuzmanović, N., Lassalle, F., Lindström, K., Mhamdi, R., Martínez-Romero, E., Moulin, L., Mousavi, S.A., Nesme, X., Peix, A., Puławska, J., Steenkamp, E., Stępkowski, T., Tian, C.-F., Vinuesa, P., Wei, G., Willems, A., Zilli, J., Young, P., 2019. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. Int. J. Syst. Evol. Microbiol.

De Meyer, S.E., Van Hoorde, K., Vekeman, B., Braeckman, T., Willems, A., 2011. Genetic diversity of rhizobia associated with indigenous legumes in different regions of Flanders (Belgium). Soil Biol. Biochem. 43, 2384–2396. https://doi.org/10.1016/j.soilbio.2011.08.005

De Meyer, S.E., Willems, A., 2012. Multilocus sequence analysis of Bosea species and description of Bosea lupini sp. nov., Bosea lathyri sp. nov. and Bosea robiniae sp. nov., isolated from legumes. Int. J. Syst. Evol. Microbiol.

Dello Jacovo, E., Valentine, T.A., Maluk, M., Toorop, P., Lopez del Egido, L., Frachon, N., Kenicer, G., Park, L., Goff, M., Ferro, V.A., Bonomi, C., James, E.K., Iannetta, P.P.M., 2019. Towards a characterisation of the wild legume bitter vetch (Lathyrus linifolius

L. (Reichard) Bässler): heteromorphic seed germination, root nodule structure and N-fixing rhizobial symbionts. Plant Biol. 21, 523–532. https://doi.org/10.1111/plb.12902

Ding, H., Hynes, M.F., 2009. Plasmid transfer systems in the rhizobia. Can. J. Microbiol. 55, 917–927. https://doi.org/10.1139/W09-056

Drouin, P., Prévost, D., Antoun, H., 2000. Physiological adaptation to low temperatures of strains of Rhizobium leguminosarum bv. viciae associated with Lathyrus spp.1. FEMS Microbiol. Ecol. 32, 111–120. https://doi.org/10.1111/j.1574-6941.2000.tb00705.x

Drouin, P., Prévost, D., Antoun, H., 1996. Classification of bacteria nodulating Lathyrus japonicus and Lathyrus pratensis in northern Quebec as strains of Rhizobium leguminosarum biovar viciae. Int. J. Syst. Bacteriol. 46, 1016–1024. https://doi.org/10.1099/00207713-46-4-1016

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Finlay, B.J., 2002. Global Dispersal of Free-Living Microbial Eukaryote Species. Science 296, 1061. https://doi.org/10.1126/science.1070710

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.

Gritli, T., Ellouze, W., Chihaoui, S.-A., Barhoumi, F., Mhamdi, R., Mnasri, B., 2020. Genotypic and symbiotic diversity of native rhizobia nodulating red pea (Lathyrus cicera L.) in Tunisia. Syst. Appl. Microbiol. 43, 126049. https://doi.org/10.1016/j.syapm.2019.126049

Guefrachi, I., Rejili, M., Mahdhi, M., Mars, M., 2013. Assessing genotypic diversity and symbiotic efficiency of five rhizobial legume interactions under cadmium stress for soil phytoremediation. Int. J. Phytoremediation 15, 938–951. https://doi.org/10.1080/15226514.2012.751350

Guefrachi, I, Rejili, M., Mahdhi, M., Mars, M., 2013. Assessing genotypic diversity and symbiotic efficiency of five rhizobial legume interactions under cadium stress for soil phytoremediation. Int. J. Phytoremediation 15, 938–951.

Güneş, F., 2019. Naturally grow in Turkey some species of Lathyrus L.(Wild pea, Fabaceae) can be used as ornamental plant. Int. J. Environ. Res. Technol. 2, 163–168.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinforma. Oxf. Engl. 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gurusamy, C., Davis, P.J., Bal, A.K., 2000. Seasonal Changes in Perennial Nodules of Beach Pea (Lathyrus maritimus [L.] Bigel.) with Special Reference to Oleosomes. Int. J. Plant Sci. 161, 631–638. https://doi.org/10.1086/314291

Hampe, A., Jump, A.S., 2011. Climate Relicts: Past, Present, Future. Annu. Rev. Ecol. Evol. Syst. 42, 313–333. https://doi.org/10.1146/annurev-ecolsys-102710-145015

Han, T.X., Wang, E.T., Wu, L.J., Chen, W.F., Gu, J.G., Gu, C.T., Tian, C.F., Chen, W.X., 2008. Rhizobium multihospitium sp. nov., isolated from multiple legume species native of Xinjiang, China. Int. J. Syst. Evol. Microbiol.

Hely, F., Ofer, I., 1972. Nodulation and frequencies of wild leguminous species in the northern Negev region of Israel. Aust. J. Agric. Res. 23, 267–284.

Herridge, D.F., Peoples, M.B., Boddey, R.M., 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311, 1–18. https://doi.org/10.1007/s11104-008-9668-3

Hördt, A., López, M.G., Meier-Kolthoff, J.P., Schleuning, M., Weinhold, L.-M., Tindall, B.J., Gronow, S., Kyrpides, N.C., Woyke, T., Göker, M., 2020. Analysis of 1,000+ Type-Strain Genomes Substantially Improves Taxonomic Classification of Alphaproteobacteria. Front. Microbiol. 11, 468. https://doi.org/10.3389/fmicb.2020.00468

http://www.microbesng.com [WWW Document], 2021. . MicrobesNG - Unlock Your Microb. Genomes. URL https://microbesng.com/ (accessed 8.17.21).

IPNI, 2021. International Plant Names Index [WWW Document]. URL https://www.ipni.org/ (accessed 8.17.21).

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114. https://doi.org/10.1038/s41467-018-07641-9

James, E.K., Iannetta, P.P., Kenicer, G., Sprent, J.I., Squire, G., 2011. Nodulation of Lathyrus and Vicia spp. in non-agricultural soils in East Scotland. Asp. Appl. Biol. 119–124.

Jensen, E.S., Peoples, M.B., Boddey, R.M., Gresshoff, P.M., Hauggaard-Nielsen, H., J.R. Alves, B., Morrison, M.J., 2012. Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agron. Sustain. Dev. 32, 329–364. https://doi.org/10.1007/s13593-011-0056-7

Jiao, C.-J., Jiang, J.-L., Ke, L.-M., Cheng, W., Li, F.-M., Li, Z.-X., Wang, C.-Y., 2011. Factors affecting β-ODAP content in Lathyrus sativus and their possible physiological mechanisms. Food Chem. Toxicol. 49, 543–549. https://doi.org/10.1016/j.fct.2010.04.050

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. 10, 5029. https://doi.org/10.1038/s41467-019-13036-1

Kan, F.L., Chen, Z.Y., Wang, E.T., Tian, C.F., Sui, X.H., Chen, W.X., 2007. Characterization of symbiotic and endophytic bacteria isolated from root nodules of herbaceous legumes grown in Qinghai–Tibet plateau and in other zones of China. Arch. Microbiol. 188, 103–115.

Kenicer, G., 2008. An Introduction to the Genus Lathyrus L. Curtiss Bot. Mag. 25, 286–295. https://doi.org/10.1111/j.1467-8748.2008.00632.x

Kenicer, G., 2007. Systematics and biogeography of Lathyrus L.(Leguminosae, papilionoideae).

Kenicer, G., Parsons, R., 2021. Lathyrus: The Complete Guide. Royal Horticultural Society, Kew.

Kenicer, G.J., 2021. Personal communication.

Kenicer, G.J., Kajita, T., Pennington, R.T., Murata, J., 2005. Systematics and biogeography of Lathyrus (Leguminosae) based on internal transcribed spacer and cpDNA sequence data. Am. J. Bot. 92, 1199–1209. https://doi.org/10.3732/ajb.92.7.1199

Kimeklis, A.K., Chirak, E.R., Kuznetsova, I.G., Sazanova, A.L., Safronova, V.I., Belimov, A.A., Onishchuk, O.P., Kurchak, O.N., Aksenova, T.S., Pinaev, A.G., Andronov, E.E., Provorov, N.A., 2019. Rhizobia Isolated from the Relict Legume Vavilovia formosa

Represent a Genetically Specific Group within Rhizobium leguminosarum biovar viciae. Genes 10. https://doi.org/10.3390/genes10120991

Küçük, C., Cevheri, C., 2014. Nodulation study of natural forage legume in semiarid region, Turkey. Pak. J. Biol. Sci. PJBS 17, 535–539.

Kumar, N., Lad, G., Giuntini, E., Kaye, M.E., Udomwong, P., Shamsani, N.J., Young, J.P.W., Bailly, X., 2015. Bacterial genospecies that are not ecologically coherent: population genomics of Rhizobium leguminosarum. Open Biol. 5, 140133.

Kumar, S., Gupta, P., Barpete, S., Sarker, A., Amri, A., Mathur, P.N., Baum, M., 2013. 11 - Grass Pea, in: Singh, M., Upadhyaya, H.D., Bisht, I.S. (Eds.), Genetic and Genomic Resources of Grain Legume Improvement. Elsevier, Oxford, pp. 269–292. https://doi.org/10.1016/B978-0-12-397935-3.00011-6

Kupicha, F.K., 1983. Infrageneric structure of Lathyrus. Notes-R. Bot. Gard. Edinb.

Kurchak, O.N., Provorov, N.A., Simarov, B.V., 2011. Comparison of the adaptive potential for Rhizobium leguminosarum bv. viceae nodule bacterial populations isolated in natural ecosystems and agrocenoses. Russ. J. Genet. 47, 425–431. https://doi.org/10.1134/S1022795411040089

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108. https://doi.org/10.1093/nar/gkm160

Laguerre, G., Depret, G., Bourion, V., Duc, G., 2007. Rhizobium leguminosarum bv. viciae genotypes interact with pea plants in developmental responses of nodules, roots and shoots. New Phytol. 176, 680–690. https://doi.org/10.1111/j.1469-8137.2007.02212.x

Laguerre Gisèle, Louvrier Philippe, Allard Marie-Reine, Amarger Noëlle, 2003. Compatibility of Rhizobial Genotypes within Natural Populations of Rhizobium leguminosarum Biovar viciae for Nodulation of Host Legumes. Appl. Environ. Microbiol. 69, 2276–2283. https://doi.org/10.1128/AEM.69.4.2276-2283.2003

Lang, J.M., Darling, A.E., Eisen, J.A., 2013. Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. PloS One 8, e62510.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278.

Lei, X., Wang, E.T., Chen, W.F., Sui, X.H., Chen, W.X., 2008. Diverse bacteria isolated from root nodules of wild Vicia species grown in temperate region of China. Arch. Microbiol. 190, 657–671. https://doi.org/10.1007/s00203-008-0418-y

Li, Yan, Wang, E.T., Liu, Y., Li, X., Yu, B., Ren, C., Liu, W., Li, Yunzhao, Xie, Z., 2016. Rhizobium anhuiense as the predominant microsymbionts of Lathyrus maritimus along the Shandong Peninsula seashore line. Syst. Appl. Microbiol. 39, 384–390. https://doi.org/10.1016/j.syapm.2016.07.001

Lindström, K., Martínez-Romero, M.E., 2005. International Committee on Systematics of Prokaryotes; Subcommittee on the taxonomy of Agrobacterium and Rhizobium. Int. J. Syst. Evol. Microbiol.

Mahdavi, B., 2007. Nodulation and Root Traits in Four Grasspea (Lathyrus sativus) Ecotypes under Root-Zone Temperatures. Pak. J. Biol. Sci. 10, 1243–1249.

Mantelin, S., Saux, M.F.-L., Zakhia, F., Béna, G., Bonneau, S., Jeder, H., de Lajudie, P., Cleyet-Marel, J.-C., 2006. Emended description of the genus Phyllobacterium and description of four novel species associated with plant roots: Phyllobacterium bourgognense sp. nov., Phyllobacterium ifriqiyense sp. nov., Phyllobacterium leguminum sp. nov. and Phyllobacterium brassicacearum sp. nov. Int. J. Syst. Evol. Microbiol. 56, 827–839. https://doi.org/10.1099/ijs.0.63911-0

Martínez-Hidalgo, P., Hirsch, A.M., 2017. The Nodule Microbiome: N2-Fixing Rhizobia Do Not Live Alone. Phytobiomes J. 1, 70–82. https://doi.org/10.1094/PBIOMES-12-16-0019-RVW

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Øvreås, L., Reysenbach, A.-L., Smith, V.H., Staley, J.T., 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. 4, 102–112. https://doi.org/10.1038/nrmicro1341

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60. https://doi.org/10.1186/1471-2105-14-60

Meier-Kolthoff, J.P., Göker, M., 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 10, 2182. https://doi.org/10.1038/s41467-019-10210-3

Mikić, A., Smýkal, P., Kenicer, G., Vishnyakova, M., Sarukhanyan, N., Akopian, J.A., Vanyan, A., Gabrielyan, I., Smýkalová, I., Sherbakova, E., Zorić, L., Atlagić, J., Zeremski-Škorić, T., Cupina, B., Krstić, D., Jajić, I., Antanasović, S., Dorđević, V., Mihailović, V., Ivanov, A., Ochatt, S., Toker, C., Zlatković, B., Ambrose, M., 2014. Beauty will save the world, but will the world save beauty? The case of the highly endangered Vavilovia formosa (Stev.) Fed. Planta 240, 1139–1146. https://doi.org/10.1007/s00425-014-2136-9

Mohammed, M.A., Chernet, M.T., Tuji, F.A., 2020. Phenotypic, stress tolerance, and plant growth promoting characteristics of rhizobial isolates of grass pea. Int. Microbiol. Off. J. Span. Soc. Microbiol. 23, 607–618. https://doi.org/10.1007/s10123-020-00131-3

Moschetti, G., Peluso, A., Protopapa, A., Anastasio, M., Pepe, O., Defez, R., 2005. Use of nodulation pattern, stress tolerance, nodC gene amplification, RAPD-PCR and RFLP–16S rDNA analysis to discriminate genotypes of Rhizobium leguminosarum biovar viciae. Syst. Appl. Microbiol. 28, 619–631. https://doi.org/10.1016/j.syapm.2005.03.009

Muresu, R., Polone, E., Sulas, L., Baldan, B., Tondello, A., Delogu, G., Cappuccinelli, P., Alberghini, S., Benhizia, Y., Benhizia, H., 2008. Coexistence of predominantly nonculturable rhizobia with diverse, endophytic bacterial taxa within nodules of wild legumes. FEMS Microbiol. Ecol. 63, 383–400.

Mutch, L.A., Young, J.P.W., 2004. Diversity and specificity of Rhizobium leguminosarum biovar viciae on wild and cultivated legumes. Mol. Ecol. 13, 2435–2444.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132. https://doi.org/10.1186/s13059-016-0997-x

Pandey, R., 2000. Descriptors for Lathyrus spp. Bioversity International.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004. https://doi.org/10.1038/nbt.4229

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542.

Peoples, M.B., Herridge, D.F., Ladha, J.K., 1995. Biological nitrogen fixation: An efficient source of nitrogen for sustainable agricultural production?, in: Ladha, J.K., Peoples, M.B. (Eds.), Management of Biological Nitrogen Fixation for the Development of More Productive and Sustainable Agricultural Systems: Extended Versions of Papers Presented at the Symposium on Biological Nitrogen Fixation for Sustainable Agriculture at the 15th Congress of Soil Science, Acapulco, Mexico, 1994. Springer Netherlands, Dordrecht, pp. 3–28. https://doi.org/10.1007/978-94-011-0055-7_1

POWO, 2019. Plants of the World Online | Kew Science [WWW Document]. Plants World Online. URL http://www.plantsoftheworldonline.org/cite-us (accessed 8.4.21).

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One 5, e9490.

Provorov, N.A., Andronov, E.E., Kimeklis, A.K., Chirak, E.R., Karasev, E.S., Aksenova, T.S., Kopat, V.V., 2020. Evolutionary Geography of Root Nodule Bacteria: Speciation Directed by the Host Plants. Microbiology 89, 1–12. https://doi.org/10.1134/S0026261720010129

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez-Bahena, M.H., García-Fraile, P., Peix, A., Valverde, A., Rivas, R., Igual, J.M., Mateos, P.F., Martínez-Molina, E., Velázquez, E., 2008a. Revision of the taxonomic status of the species Rhizobium leguminosarum (Frank 1879) Frank 1889AL, Rhizobium phaseoli Dangeard 1926AL and Rhizobium trifolii Dangeard 1926AL. R. trifolii is a later synonym of R. leguminosarum. Reclassification of the strain R. leguminosarum DSM 30132 (= NCIMB 11478) as Rhizobium pisi sp. nov. Int. J. Syst. Evol. Microbiol. 58, 2484–2490.

Ramírez-Bahena, M.H., García-Fraile, P., Peix, A., Valverde, A., Rivas, R., Igual, J.M., Mateos, P.F., Martínez-Molina, E., Velázquez, E., 2008b. Revision of the taxonomic status of the species Rhizobium leguminosarum (Frank 1879) Frank 1889AL, Rhizobium phaseoli Dangeard 1926AL and Rhizobium trifolii Dangeard 1926AL. R. trifolii is a later synonym of R. leguminosarum. Reclassification of the strain R. leguminosarum DSM 30132 (= NCIMB 11478) as Rhizobium pisi sp. nov. Int. J. Syst. Evol. Microbiol. 58, 2484–2490.

Rathi, D., Chakraborty, S., Chakraborty, N., 2021. Grasspea, a critical recruit among neglected and underutilized legumes, for tapping genomic resources. Curr. Plant Biol. 26, 100200. https://doi.org/10.1016/j.cpb.2021.100200

Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. 106, 19126. https://doi.org/10.1073/pnas.0906412106

Rivas, R., García-Fraile, P., Velázquez, E., 2009. Taxonomy of bacteria nodulating legumes. Microbiol. Insights 2, MBI-S3137.

Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., Fricko, O., Gusti, M., Harris, N., Hasegawa, T., Hausfather, Z., Havlík, P., House, J., Nabuurs, G.-J., Popp, A., Sánchez, M.J.S., Sanderman, J., Smith, P., Stehfest, E., Lawrence, D., 2019. Contribution of the land sector to a 1.5 °C world. Nat. Clim. Change 9, 817–828. https://doi.org/10.1038/s41558-019-0591-9

Rogel, M.A., Ormeño-Orrillo, E., Romero, E.M., 2011. Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst. Appl. Microbiol. 34, 96–104. https://doi.org/10.1016/j.syapm.2010.11.015

Ruiz-Díez, B., Fajardo, S., Felipe, M. del R. de, Fernández-Pascual, M., 2012. Characterization of rhizobia from legumes of agronomic interest grown in semi-arid areas of Central Spain relates genetic differences to soil properties. J. Basic Microbiol. 52, 66–78. https://doi.org/10.1002/jobm.201100058

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945.

Safronova, V., Belimov, A., Andronov, E., Popova, J., Tikhomirova, N., Orlova, O., Verkhozina, A., Chimitov, D., Tikhonovich, I., 2017. Method for obtaining root nodules of the Baikal relict legumes in laboratory pot experiments. Int. J. Environ. Stud. 74, 700–705. https://doi.org/10.1080/00207233.2017.1283948

Safronova, V.I., Kimeklis, A.K., Chizhevskaya, E.P., Belimov, A.A., Andronov, E.E., Pinaev, A.G., Pukhaev, A.R., Popov, K.P., Tikhonovich, I.A., 2014. Genetic diversity of rhizobia isolated from nodules of the relic species Vavilovia formosa (Stev.) Fed. Antonie Van Leeuwenhoek 105, 389–399. https://doi.org/10.1007/s10482-013-0089-9

Safronova, V.I., Kuznetsova, I.G., Sazanova, A.L., Belimov, A.A., Andronov, E.E., Chirak, E.R., Osledkin, Y.S., Onishchuk, O.P., Kurchak, O.N., Shaposhnikov, A.I., Willems, A., Tikhonovich, I.A., 2017. Microvirga ossetica sp. nov., a species of rhizobia isolated from root nodules of the legume species Vicia alpestris Steven. Int. J. Syst. Evol. Microbiol.

Safronova, V.I., Kuznetsova, I.G., Sazanova, A.L., Kimeklis, A.K., Belimov, A.A., Andronov, E.E., Pinaev, A.G., Chizhevskaya, E.P., Pukhaev, A.R., Popov, K.P., Willems, A., Tikhonovich, I.A., 2015. Bosea vaviloviae sp. nov., a new species of slow-growing rhizobia isolated from nodules of the relict species Vavilovia formosa (Stev.) Fed. Antonie Van Leeuwenhoek 107, 911–920. https://doi.org/10.1007/s10482-015-0383-9

Sánchez-Cañizares, C., Jorrín, B., Durán, D., Nadendla, S., Albareda, M., Rubio-Sanz, L., Lanza, M., González-Guerrero, M., Prieto, R.I., Brito, B., 2018. Genomic diversity in the endosymbiotic bacterium Rhizobium leguminosarum. Genes 9, 60.

Schaefer, H., Hechenleitner, P., Santos-Guerra, A., Menezes de Sequeira, 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 Evol. Biol. 12, 250–250. https://doi.org/10.1186/1471-2148-12-250

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Shade, A., Jacques, M.-A., Barret, M., 2017. Ecological patterns of seed microbiome diversity, transmission, and assembly. Curr. Opin. Microbiol. 37, 15–22. https://doi.org/10.1016/j.mib.2017.03.010

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75

Smith David J., Timonen Hilkka J., Jaffe Daniel A., Griffin Dale W., Birmele Michele N., Perry Kevin D., Ward Peter D., Roberts Michael S., 2013. Intercontinental Dispersal of Bacteria and Archaea by Transpacific Winds. Appl. Environ. Microbiol. 79, 1134–1139. https://doi.org/10.1128/AEM.03029-12

Smýkal, P., 2011. Phylogeny, phylogeography and genetic diversity of the Pisum genus. Plant Genet. Resour. Charact. Util. v. 9, 4–0. https://doi.org/10.1017/S147926211000033X

Smýkal, P., Chaloupská, M., Bariotakis, M., Marečková, L., Sinjushin, A., Gabrielyan, I., Akopian, J., Toker, C., Kenicer, G., Kitner, M., Pirintsos, S., 2017. Spatial patterns and intraspecific diversity of the glacial relict legume species Vavilovia formosa (Stev.) Fed. in Eurasia. Plant Syst. Evol. 303, 267–282. https://doi.org/10.1007/s00606-016-1368-5

Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W., Berger, J., Greene, S.L., Nelson, M.N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R.K., Roorkiwal, M., Pandey, M.K., Hu, J., Li, Y.H., Wang, L.X., Guo, Y., Qiu, L.J., Redden, R.J., Varshney, R.K., 2015. Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. Crit. Rev. Plant Sci. 34, 43–104. https://doi.org/10.1080/07352689.2014.897904

Soussana, J.-F., Loiseau, P., Vuichard, N., Ceschia, E., Balesdent, J., Chevallier, T., Arrouays, D., 2004. Carbon cycling and sequestration opportunities in temperate grasslands. Soil Use Manag. 20, 219–230. https://doi.org/10.1111/j.1475-2743.2004.tb00362.x

Sprent, J.I., 1994. Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? Plant Soil 161, 1–10. https://doi.org/10.1007/BF02183080

Sprent, J.I., Ardley, J., James, E.K., 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytol. 215, 40–56. https://doi.org/10.1111/nph.14474

Sui, X., Han, L., Wang, E., Jiang, F., Liu, Y., Chen, W., 2009. Novel associations between rhizobial populations and legume species within the genera Lathyrus and Oxytropis grown in the temperate region of China. Sci. China C Life Sci. 52, 182–192. https://doi.org/10.1007/s11427-008-0132-7

Truyens, S., Weyens, N., Cuypers, A., Vangronsveld, J., 2015. Bacterial seed endophytes: genera, vertical transmission and interaction with plants. Environ. Microbiol. Rep. 7, 40–50. https://doi.org/10.1111/1758-2229.12181

Villadas, P.J., Lasa, A.V., Martínez-Hidalgo, P., Flores-Félix, J.D., Martínez-Molina, E., Toro, N., Velázquez, E., Fernández-López, M., 2017. Analysis of rhizobial endosymbionts of Vicia, Lathyrus and Trifolium species used to maintain mountain firewalls in Sierra Nevada National Park (South Spain). Syst. Appl. Microbiol. 40, 92–101. https://doi.org/10.1016/j.syapm.2016.11.008

Vitousek, P.M., Cassman, K., Cleveland, C., Crews, T., Field, C.B., Grimm, N.B., Howarth, R.W., Marino, R., Martinelli, L., Rastetter, E.B., Sprent, J.I., 2002. Towards an ecological understanding of biological nitrogen fixation, in: Boyer, E.W., Howarth, R.W. (Eds.), The Nitrogen Cycle at Regional to Global Scales. Springer Netherlands, Dordrecht, pp. 1–45. https://doi.org/10.1007/978-94-017-3405-9_1

Wayne, L., Brenner, D., Colwell, R., Grimont, P., Kandler, O., Krichevsky, M., Moore, L., Moore, W., Murray, Rge., Stackebrandt, E., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Evol. Microbiol. 37, 463–464.

Xiong, J.-L., Bai, X., Batool, A., Kong, H.-Y., Tan, R.-Y., Wang, Y.-F., Li, Z.-X., Xiong, Y.-C., 2014. Ecological function and application of toxin beta-ODAP in grass pea (Lathyrus sativus). Ying Yong Sheng Tai Xue Bao J. Appl. Ecol. 25, 1197–1205.

Young, John M, J.M., Kuykendall, L.D., Martínez-Romero, E., Kerr, A., Sawada, H., 2001. A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Allorhizobium undicola de Lajudie et al. 1998 as new combinations: Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola and R. vitis. Int. J. Syst. Evol. Microbiol.

Young, J.P.W., 2021. Personal communication.

Young, J.P.W., Crossman, L.C., Johnston, A.W., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R., Todd, J.D., Poole, P.S., Mauchline, T.H., East, A.K., Quail, M.A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S., Parkhill, J., 2006. The genome of Rhizobium leguminosarum has recognizable core and accessory components. Genome Biol. 7, R34. https://doi.org/10.1186/gb-2006-7-4-r34

Young, J. Peter W., Moeskjær, S., Afonin, A., Rahi, P., Maluk, M., James, E.K., Cavassim, M.I.A., Rashid, M.H., Aserse, A.A., Perry, B.J., Wang, E.T., Velázquez, E., Andronov, E.E., Tampakaki, A., Flores Félix, J.D., Rivas González, R., Youseif, S.H., Lepetit, M., Boivin, S., Jorrin, B., Kenicer, G.J., Peix, Á., Hynes, M.F., Ramírez-Bahena, M.H., Gulati, A., Tian, C.-F., 2021. Defining the Rhizobium leguminosarum Species Complex. Genes 12. https://doi.org/10.3390/genes12010111

Young, J Peter W, Moeskjær, S., Afonin, A., Rahi, P., Maluk, M., James, E.K., Cavassim, M.I.A., Rashid, M.H.-O., Aserse, A.A., Perry, B.J., Wang, E.T., Velázquez, E., Andronov, E.E., Tampakaki, A., Flores Félix, J.D., Rivas González, R., Youseif, S.H., Lepetit, M., Boivin, S., Jorrin, B., Kenicer, G.J., Peix, Á., Hynes, M.F., Ramírez-Bahena, M.H., Gulati, A., Tian, C.-F., 2021. Defining the Rhizobium leguminosarum Species Complex. Genes 12, 111. https://doi.org/10.3390/genes12010111

Zahran, H.H., 2001. Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. Co-Oper. Eur. Dev. Ctries. Res. Biol. Nitrogen Fixat.-Needa Means 91, 143–153. https://doi.org/10.1016/S0168-1656(01)00342-X

Zahran, H.H., 1999. Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiol. Mol. Biol. Rev. MMBR 63, 968–989, table of contents. https://doi.org/10.1128/MMBR.63.4.968-989.1999

Zakhia, F., Jeder, H., Domergue, O., Willems, A., Cleyet-Marel, J.-C., Gillis, M., Dreyfus, B., Lajudie, P. de, 2004. Characterisation of Wild Legume Nodulating Bacteria (LNB) in the Infra-arid Zone of Tunisia. Syst. Appl. Microbiol. 27, 380–395. https://doi.org/10.1078/0723-2020-00273

Zézé, A., Mutch, L.A., Young, J.P.W., 2001. Direct amplification of nodD from community DNA reveals the genetic diversity of Rhizobium leguminosarum in soil. Environ. Microbiol. 3, 363–370. https://doi.org/10.1046/j.1462-2920.2001.00202.x

Zhang, X.X., Kosier, B., Priefer, U.B., 2001. Genetic diversity of indigenous Rhizobium leguminosarum bv. viciae isolates nodulating two different host plants during soil restoration with alfalfa. Mol. Ecol. 10, 2297–2305. https://doi.org/10.1046/j.0962-1083.2001.01364.x

Zhang, Y.J., Zheng, W.T., Everall, I., Young, J.P.W., Zhang, X.X., Tian, C.F., Sui, X.H., Wang, E.T., Chen, W.X., 2015. Rhizobium anhuiense sp. nov., isolated from effective nodules of Vicia faba and Pisum sativum. Int. J. Syst. Evol. Microbiol.

Zohary, D., Hopf, M., 1973. Domestication of Pulses in the Old World. Science 182, 887. https://doi.org/10.1126/science.182.4115.887

Appendix

Uploaded as separate files to Propagate:

1. Figure 1. ANI table, taxonomic annotations

2.Figure 2. ANI table, geographical annotations

3. Figure 3. ML phylogeny, 120 core genes, taxomnomic

5. Taxon/accession names, accession numbers, collection data: Accessions_references. xlsx

6. Spreadsheet of hits for biovars NodACD_symbiovars.xlsx

7. Quast_report_averages.xlsx

8. J. Peter Young's phython scripts, adapted.