Title: Different ecological processes drive the assembly of dominant and rare root-associated bacteria in a promiscuous legume

Running title: Root nodule bacterial community assembly

Authorship and affiliations:

Josep Ramoneda¹, Jaco Le Roux², Emmanuel Frossard¹, Beat Frey³, Hannes Andres Gamper⁴

¹Group of Plant Nutrition, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland.

²Department of Biological Sciences, Macquarie University, Sydney, Australia.

³Rhizosphere Processes Group, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland.

⁴Free University of Bolzen-Bolzano, Faculty of Science and Technology, Bolzano, Italy

Corresponding author:

Josep Ramoneda

Email: josep.ramoneda@usys.ethz.ch

Tel: +34 686168574

Abstract

Understanding how plant-associated microbial communities assemble and the roles they play in plant performance are major goals in microbial ecology. For nitrogen-fixing rhizobia, assemblages are mostly determined by filtering by the host as well as abiotic soil conditions. However, for legumes adapted to highly variable environments and nutrient-poor soils, filtering out rhizobial partners may not be an effective strategy to ensure symbiotic benefits. As a consequence, this can lead to net increases in rhizobial diversity driven by stochastic (neutral) assembly processes. Here, we addressed whether symbiotic promiscuity of rooibos (*Aspalathus linearis* Burm. Dahlgren), reflects increases in rhizobial diversity that are independent of the environmental conditions, following a neutral assembly. We performed a common garden experiment to assess whether root system size and location- and habitat-specific rhizobial propagule pools of rooibos affected rhizobial community diversity and composition. We found a dominance of drift in driving taxonomic turnover in the root nodule communities, but operating at different scales in the dominant *Mesorhizobium* symbionts and the rest of bacterial taxa. Specifically, drift triggered differentiation between the core rhizobial symbionts of rooibos due to dispersal limitation on a regional scale, whereas it dominated the assembly of non-dominant rhizobial taxa at the root level. This suggests the existence of a significant neutral component in rhizobial community assembly when selectivity by the host plant is low. We conclude that in this promiscuous legume neutral processes govern bacterial community root nodule community assembly, but that these operate at different scales in dominant and rare rhizobial symbionts of the plant.

Introduction

Exploring ways to manage entire microbial communities with beneficial functions on plant nutrition and health (i.e. functional microbiomes) is a recognized research priority (Lemanceau *et al.*, 2017; Toju *et al.*, 2018). Under field conditions, microbial community assembly from the plant's surrounding environment defines which taxa and potential functions will assist plant health and growth. Therefore, we need a better understanding of how microbial communities assemble in plants, and the roles microbial diversity can play in plant performance, both from ecological and evolutionary standpoints.

Legumes are highly reliant on microbial services and key to ecosystem productivity and agricultural soil fertility (Graham and Vance, 2003). This is because most of them associate with rhizobia from the alpha- and β -Proteobacteria, as well as other microbes that improve growth and alleviate stress impact. Rhizobia fix atmospheric nitrogen (N) in structures called root nodules. Rhizobia thrive as saprobes in soil and the rhizosphere and vastly multiply when in association with their legume hosts, within root structures called root nodules. The plant controls the symbiosis functioning by selectively allocating carbohydrates and other key resources for the endosymbiotic and functionally differentiated rhizobia (Denison and Kiers, 2011).

Surprisingly little is known about rhizobial community assembly in legume roots (Barrett *et al.*, 2015; Sprent *et al.*, 2017; Burghardt *et al.*, 2019). Until most recently, community assembly in rhizobia has received little attention because of presumed low functional diversity, genetically homogeneous rhizobial populations in root nodules, and lack of analytical technology (Burghardt *et al.*, 2019). At the local scale, symbiotic rhizobial community assembly is driven by soil abiotic factors (e.g. pH, humidity, soil texture), the nutritional status of their host plant (Denison and Kiers, 2011; Bever 2015), host plant occurrence and abundance (Denison and Kiers 2011), predators and parasites (Burghardt 2019). Additionally, at the landscape scale differences in population sizes of rhizobia can lead to evolution of geographical variants and ultimately speciation (Bissett *et al.*, 2010; Sprent *et al.*, 2017). Thus, site and host plant adaptation, along with differential propagation of the symbionts, determine the composition of available rhizobial propagules.

As for most symbiotic interactions, the assembly of rhizobial communities in legume roots has both deterministic and stochastic components (Nemergut *et al.*, 2013). These components dominate in so-called trait-based and neutral assembly mechanisms, respectively.

Deterministic assembly is driven by plant-microbe compatibility and adaptation, in which trait differences are decisive. On the other hand, stochastic assembly is reliant on the chance of partner encounters, infection pressure, infection priority or propagule dispersal (Nemergut *et al.*, 2013; Le Roux *et al.*, 2017). Legume-rhizobium partners associate *de novo* at the onset of each new plant generation, but also every time when new root nodules and new roots are formed, and during root growth. These events provide chances for both trait-based and neutral assembly processes to operate. However, their relative importance in legume-rhizobium associations are far from being understood, particularly under agricultural settings.

For legumes thriving in nutrient-poor environments, where extreme N limitation is the norm, it can be advantageous to associate with any, even functionally inferior rhizobia (Kiers *et al.*, 2003). Indeed, legumes in these environments seem to lack several of the mechanisms for symbiont sanctioning (cf. Australian acacias or legumes in the Cape Floristic Region of South Africa: Lemaire *et al.*, 2015; Brink *et al.*, 2017; Vuong *et al.*, 2017; Le Roux *et al.*, 2018). Such legumes associate with higher rhizobial diversity, depending on factors such as drought, plant health, root system size, or plant age (Jiao *et al.*, 2015; Wagner *et al.*, 2016; Dinnage *et al.*, 2019). In the absence of legume control mechanisms, propagule pressure, together with infection priority, may render abundant rhizobium taxa monodominant in root nodules, especially during seedling establishment. Over time, the probability of root-rhizobium encounters may rise as older legume plants associate with increasing numbers of different rhizobia (Dinnage *et al.*, 2019). This may lead to root size-dependent assembly of rhizobial root symbionts, which is apparently neutral, even though infection and maintenance of symbioses at fine (i.e. root) scales may still be trait based (Yu *et al.*, 2018).

To better understand the links between root nodule symbiont diversity and neutral vs non-neutral community assembly processes, we studied rhizobial communities associated with rooibos (*Aspalathus linearis* Burm. Dahlgren), an economically important crop. The species is endemic to the acidic and extremely nutrient-poor soils of the semi-arid South African Fynbos vegetation, and its agricultural production is under threat by climate change and desertification (Malgas *et al.*, 2010; Lötter *et al.*, 2014). Rooibos belongs to the diverse genus *Aspalathus*, which has radiated exclusively in the hyperdiverse Core Cape Subregion (CCR, previously Cape Floristic Region, and now forming part of the Greater Cape Floristic Region) in close association with a wide diversity of rhizobial symbionts (Hassen *et al.*, 2012).

Given the extreme nutrient-poor conditions under which rooibos evolved, we hypothesized that the assembly of the rhizobial root nodule communities of the species would follow

predominantly stochastic rather than deterministic trajectories. Neutral assembly was expected to lead to increased rhizobial diversity when exposed to more diverse inocula and when having larger root systems (i.e. under conditions of increased likelihood of encounter and infection). Using a dual marker approach, we described both symbiotic and non-symbiotic rhizobial communities, and manipulated soil nutrient content and soil inoculum diversity. By using soils from a range of geographical locations, and covering a range of plants with different root sizes, we were able to link assembly processes to regional and root spatial scales.

Materials and Methods

Experimental design and setup

To examine the effects of fertilization and soil mixing on the assembly of root-nodulating rhizobia in rooibos, a cross-factorial pot experiment was setup. It was designed with the factors i) ‘location’ with five levels, represented by five farm origins of the employed soil, ii) ‘soil origin’ with three levels, made up of soil from adjacent cultivated and uncultivated land and a 1:1 (v:v) volumetric mixture of those two, and iii) ‘fertilization’ with two levels, which were either addition of sheep dung or not. The experimental units were pots with one plant individual that were randomly arranged with respect to all 30 resulting experimental treatments (5 locations x 3 soil origins x 2 fertilization treatments) in ten spatial blocks, representing ten replicates, totalling in 300 experimental units represented by individual pots with one single rooibos seedling.

The soils used for the experiment were collected down to a depth of 30 cm from beneath cultivated and wild rooibos plants not more than 100 m apart from each other, and locations that corresponded with different farms in the Suid Bokkeveld (Western Cape, South Africa). The locations were located 5.5-33 km apart from each other. The experiment was initially setup in a polytunnel and after five months transferred to an outdoor shade house at the Agricultural Research Council Infruitec’s research station in Stellenbosch, South Africa.

Two-liter pots were filled with soil from either cultivated or uncultivated land or a 1:1 volumetric mixture of those. The fertilization treatment consisted in the addition of 7.5 g air-dry dung from a nighttime enclosure (Afrikaans ‘kraal’) of sheep on one of the farms where soils were collected. The sheep dung increased the total N content in the soil of the pots by 30% and the total P content by 50% (see Supplementary Table 1 for the detailed mineral nutrient concentrations of the sheep dung).

Three sulfuric acid-scarified seeds of the sole cultivated rooibos variety *Nortier* were buried about 1 cm below the soil surface in a line perpendicular to that defined by the two patches of sheep dung to avoid direct root-dung contact initially and hence prevent any possible nutrient toxicity. In the last two months post germination, the strongest seedling was kept and any remaining seedling removed to finally have one plant per pot. Other plants growing from the seedbank or from root fragments were periodically removed throughout the duration of the experiment. The experiment ran from the beginning of August 2016 to the middle of April 2017, corresponding to the period between early spring and late autumn under the South African Mediterranean climate.

Harvest and sample preservation for the various analyses

At harvest, individual root systems were well rinsed with pressurized water. Stems, leaves, and roots were separated and the number of root nodules and cluster roots counted. The dry weight of these separate plant fractions was determined after air-drying at 50°C for two days, except for the root nodules which were removed prior to drying. Root nodules were initially air-dried for 1–3 hours before being stored in silica gel. After drying, leaves were removed from stems and branches and weighed separately. The maximum diameter of the taproot was recorded with a caliper. Dried leaves were finely ground for nutrient analyses in wolfram carbide cups on a TissueLyser II swing mill (Qiagen, Hombrechtikon, Switzerland) for 60 s at 28 Hz.

Mineral nutrient and $\delta^{15}\text{N}$ signature analysis

The foliar carbon (C) and nitrogen (N) concentrations and $\delta^{15}\text{N}$ isotopic signature were determined from 4 mg subsamples of milled leaf material on an ion ratio mass spectrometer (Delta V Advantage, ThermoFisher Scientific, Waltham, MA, USA). The other macro- and micronutrients (P, K, Ca, Mg, Mn, Fe) were quantified in 200 mg of ground leaf material, which was digested in 2 ml HNO_3 (69%) and 2 ml deionized H_2O in a microwave liquid digestion system MLS Turbowave (MWS GmbH, Heerbrugg, Switzerland) and analyzed by Inductively Coupled Plasma – Optical Emission Spectrometry (ICP-OES) on a ICPE-9800 apparatus (Shimadzu, Kyoto, Japan).

The soil pH in 0.01M CaCl_2 was measured in triplicate technical replicates. The $\delta^{15}\text{N}$ isotopic signature of triplicate 3 mg subsamples of finely ground sheep dung and macronutrients of finely ground 200 mg subsamples of the soils and sheep dung were

measured after liquid extraction with 4 ml of HNO₃ (69%) and 10 ml deionized H₂O as described above for the plant tissue. All plant nutrient analyses were carried out in the laboratory facilities of the Group of Plant Nutrition at ETH Zurich (Eschikon, Switzerland). The physico-chemical characteristics of the soils were determined by the ISO 17025:2005-certified Bemlab laboratories (Somerset West, South Africa).

DNA extraction, partial gene amplification and sequencing

The root nodules of each individual plant were pooled and milled using a 7 mm diameter glass bead in a 2 ml Eppendorf tube in a TissueLyser II (Qiagen, Hilden Germany). The total DNA was then extracted from 25 mg of nodule powder using the NucleoSpin Plant II Kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol with the sodium dodecyl sulfate-containing lysis buffer PL2 and elution in 100 µl 5mM Tris-EDTA (TE, pH 8.5) elution buffer (PE).

A 817 bp-long fragment of the DNA gyrase B (*gyrB*) gene was amplified using fusion primers with the *gyrB*343F gene-specific primer sequence (Martens *et al.*, 2008) and PacBio's 30mer *universal sequence* as a tag at the 5'-end (5'-GCAGTCGAACATGTAGCTGACTCAGGTCACCTTCGACCAGAAYTCCTAYAAGG-3') and the gene-specific primer *gyrB*1043 (Martens *et al.*, 2008) and PacBio's 30mer *universal sequence* as a tag at the 5'-end (5'-TGGATCACTTGTGCAAGCATCACATCGTAGAGCTTGTCCTTSGTCTGCG-3'). Both primers were amino-blocked at their 5'-end (5' NH₄-C6) to prevent their ligation during library preparation and hence sequencing without a sample-specific barcode. As an easy to align and variable bacterial housekeeping gene, *gyrB* has proven useful as a phylotaxonomic marker to analyze legume root nodule inhabiting bacteria (e.g. Beukes *et al.*, 2016). PCR reactions were performed in triplicates, using a final reaction volume of 12.5 µl. Each reaction contained 4.25 µl PCR-grade water, 6.25 µl Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus Kit), 0.5 µl of each of the 10 µM primers and 1 µl of the template DNA. The amplification reactions were run in a DNA Engine Peltier Thermal Cyclers (Bio-Rad, Hercules, CA, USA), using the following conditions: initial denaturation at 95°C for 5 min followed by 30 cycles of 95°C for 30 s, primer annealing at 58°C for 2 min and primer extension at 72°C for 45 s, followed by a final extension phase at 68°C for 10 min. Reactions were kept at 10°C until storage at -20°C. The three replicated PCR reactions of each sample were pooled, and the amplicons purified with home-made Solid Phase Reversible

Immobilization (SPRI) magnetic beads at a ratio of 0.7 times the reaction volume (Genetic Diversity Center, ETH Zürich, Switzerland).

The amplicons of the different samples were uniquely barcoded with PacBio's 96 Barcoded Universal F/R Primers (Pacific Biosciences, Menlo Park, CA, USA) in a second PCR step for later sample pooling and library construction for Single Molecule Real-Time (SMRT) circular consensus sequencing (ccs). These barcoding primers consist of the complementary 30-mer primer sequence to the universal sequences and contain at their 5'-ends 16-mer unique barcodes. The PCR reaction mix of the barcoding PCR contained 12.25 µl PCR-grade water, 5 µl 5x Q5 reaction buffer (New England Biolabs), 2.5 µl dNTPs (2mM), 2.5 µl of each of the index primers (10 µM), 0.25 µl of Q5 polymerase (New England Biolabs, Ipswich, MA, USA), and 2.5 µl of the bead-purified amplicons diluted to 1 ng µl⁻¹ of the first PCR as template. The following thermal cycling program was used: initial denaturation at 98°C for 30 s, followed by 10 cycles of denaturation at 98°C for 10 s, primer annealing at 71°C for 20 s and primer extension at 72°C for 60 s, with a final further extension at 72°C for 2 min. The barcoded amplicons were again purified using the same SPRI magnetic bead-based protocol. The concentrations after purification were measured with the Qubit assay (Thermo Fisher Scientific Inc., Waltham, MA, USA) on a Spark 10M Multimode Plate Reader (Tecan, Männedorf, Switzerland).

For construction of sequencing libraries, all bead-purified barcoded amplicons were diluted to a concentration of 10 ng µl⁻¹ and multiplexed in pools of 96 samples, whose integrity was checked with an Agilent 2200 Tape Station (Agilent, Santa Clara, CA, USA). Three sequencing libraries were prepared with the SMRTbell™ Template Prep Kit 1.0-SPv3 (Pacific Biosciences) chemistry by the Functional Genomics Center Zurich (University of Zurich, Zurich, Switzerland) and SMRT circular consensus sequenced on individual SMRT cells on a Sequel PacBio sequencer. The average output was 3726 sequence reads per sample, covered by on average 55 and a minimum of 5 polymerase passes, which guarantees a minimum predicted sequencing accuracy of 0.999.

A 455 bp-long fragment of the nodulation gene *nodA*, encoding an N-acyltransferase of the rhizobial *nodABC* operon, was amplified with the gene-specific primers *NodA*univF145u (5'-TGGGCSGGNGCNAGRCCBGA-3') and *NodA*Rbrad (5'-TCACARCTCKGGCCCGTTCCG-3') (Moulin *et al.*, 2001). The PCR amplifications were performed in triplicates in final reaction volumes of 12.5 µl. Each reaction contained 3.75 µl

PCR-grade water, 6.25 μ l Multiplex PCR Master Mix (QIAGEN), 0.75 μ l of each of the primers at a concentration of 10 μ M and 1 μ l of the template DNA. The amplification reactions were run in DNA Engine Peltier Thermal Cyclers (Bio-Rad), using the following conditions: initial denaturation at 95°C for 5 min followed by 27 cycles of denaturation at 95°C for 30 s, primer annealing at 69°C for 90 s and primer extension at 72°C for 35 seconds, with a final further extension at 68°C for 10 min. The reaction products were kept at 10°C until storage at -20°C. The three amplification product replicates of each root nodule extract were pooled and purified with home-made SPRI beads at a ratio of 0.8 times the reaction volume. After dilution to 1 ng μ l⁻¹ unique Nextera XT sample indices (Illumina, San Diego, CA, USA) were added by a second PCR in a reaction volume of 25 μ l with 12.5 μ l KAPA HiFi Hot Star Ready Mix (KAPA Biosystems Inc., Wilmington, MA, USA), 5 μ l PCR-grade water, 2.5 μ l of each of the index primers at a concentration of 10 μ M and 2.5 μ l of the purified amplicons as template. The thermal cycling program was the following: initial denaturation at 95°C for 3 min, 8 cycles of denaturation at 95°C, primer annealing at 55°C and primer extension at 72°C for 30 s and a final further extension at 72°C for 5 min. The indexed PCR products were bead-purified again, using the same protocol as for the initial amplicons and quantified, using the Qubit assay (Thermo Fisher Scientific Inc.) on the Spark 10M Multimode Plate Reader (Tecan). The amplicon size and integrity was verified on the Agilent 2200 Tape Station (Agilent, USA) before dilution to 4 ng μ l⁻¹ for equimolar pooling of all 266 samples for library preparation. The amplicons were 2x300 bp paired-end sequenced, using the Illumina MiSeq v2 chemistry and PhiX as internal standard at a concentration of 48.99% on an Illumina MiSeq sequencer at the Genetic Diversity Center of ETH Zürich, Switzerland.

Bioinformatics of sequencing data

For the *gyrB* ccs reads, the SMRT link software (Pacific Biosciences) was used to remove the bell adaptor, barcode and primer sequences, filter the reads for a size range of 700-1000 bp (expected size 817 bp) and demultiplex the sequences of different samples. Read quality assessment and filtering was done in PRINSEQ-lite (Schmieder and Edwards, 2011). All sequences across the entire dataset were demultiplexed in USEARCH, using the function *fastx_uniques* to obtain only the unique sequences for subsequent further quality measures, similarity clustering and phylotaxonomic assignment. The unique sequences were error-corrected and chimera-filtered in UNOISE3, which generated zero-radius operational taxonomic units (ZOTUs). These are expected to represent the true biological sequences and

hence possibly different bacterial strains. To further exclude possible variants due to PCR and sequencing errors, we clustered the ZOTUs at 99% nucleotide sequence identity, using the USEARCH function *cluster_smallmem*. The count table containing the information about the read numbers and thus ZOTU abundance was extracted, using the USEARCH function *otutab*. Taxonomic assignment and a chimera check were done in SINTAX (Edgar, 2016), using all available *gyrB* sequences from NCBI GenBank as the reference database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

For the *nodA* paired-end reads, seqtk (<https://github.com/lh3/seqtk>) was used for 3'-end trimming and FLASH (Magoč and Salzberg, 2011) to join the two ends. The priming sites were removed with USEARCH (v10.0.240) as described in Edgar (2011). PRINSEQ-lite (Schmieder and Edwards, 2011) was used to check the quality of the sequence dataset. The unique sequence reads were obtained with the dereplication command *fastx_uniques* in USEARCH and Illumina sequencing errors were corrected and chimeras removed in UNOISE3. ZOTUs were determined using the same approach as for the *gyrB*. The count table was generated, using the *otutab* command in USEARCH. The phylotaxonomic affiliations of the ZOTUs and a further chimera check were done in SINTAX (Edgar, 2016) in comparison to all *nodA* sequences on NCBI GenBank.

The relative abundances of the ZOTUs were intra- and extrapolated to the median number of reads per sample (1410 reads for *gyrB*, 8720 reads for *nodA*) after excluding four samples with less than 200 reads from the *nodA* dataset and twelve samples with less than 300 reads from the *gyrB* dataset, using the *iNEXT* package (v.2.0.19, Hsieh *et al.*, 2016) in R. *iNEXT* was also used to calculate the α -diversity (richness and Simpson's diversity index). All further raw sequence data processing was done in the *phyloseq* package (v1.16.2, McMurdie *et al.*, 2013) of R. A total of 931,254 *nodA* and 161,155 *gyrB* error-filtered sequence reads were analysed. An indicator species analysis, using the function *multipatt* in the R package *indicspecies* (v1.7.6) was used to infer those ZOTUs whose occurrence and relative abundance was indicative for a certain experimental treatment or plant size class.

Statistical analysis

We ran generalized linear mixed model analyses with location and block as random factors in the R package *lme4* (v1.1-21) to test the effects of Fertilization and Soil origin on the plant growth physiological, symbiotic and alpha diversity parameters. The exact statistical model used was: $y \sim \text{Fertilization} * \text{Soil origin}, \text{random} = \sim 1 / \text{Block} / \text{Location}$. P-values for the results

of the mixed effect models were obtained by running ANOVA on the model fits. Means were compared by post-hoc Tukey's HSD tests at the confidence level of 95%.

β -diversity was analyzed using the relative ZOTU abundances. Analyses on the Hellinger (square-root) transformed abundances did not change the results qualitatively. The effects of the experimental treatments on the composition and structure of the rhizobial communities in the root nodules were visualized by non-metric multidimensional scaling (NMDS) of the Bray-Curtis dissimilarities in the package *vegan* (v2.5-5, Oksanen *et al.*, 2007) of R. The treatment effects were statistically tested by a permutational analysis of variance (PERMANOVA) with the *adonis* function in *vegan*.

Possible causative relationships between soil abiotic parameters and the bacterial communities associated with the root nodules were assessed using correlation analyses. This is because we lacked paired data between soil characteristics and microbial communities and plants (Supplementary Table 1), making statistical testing impossible. Since we had only soil data per Location and Soil origin, the β diversity of the two fertilization treatments and ten experimental replicates was summarized as the distances between the centroids of the Location (5) x Soil origin (3) factor level combinations (15) of a PCoA ordination. These distance values were then correlated to the Euclidean distances between soil parameter values using the *dist_between_centroids* function of the *usedist* package (v0.1.0) of R. Possible effects of differences in the rhizobial community composition and structure (Bray-Curtis dissimilarities) on plant performance (Euclidean distances) were assessed by means of Mantel tests, using the *vegan* function *mantel.test*.

As a way to measure the relative importance of neutral and trait-based assembly processes, we followed the analytical framework described by Stegen *et al.*, (2012, 2013). The approach is based on the comparison of the phylogenetic distance between closely related taxa across samples to a null distribution. In a first step, we assessed the strength of species turnover using the β -nearest taxon index (β -NTI). This index quantifies how much taxonomic turnover there is between samples, by comparison to a null distribution. This null distribution, which represents the absence of filters in defining taxon presence or absence, is calculated on randomized phylogenetic distances between samples by randomly shuffling taxon identities and abundances across the tips of a phylogeny. β -NTI values below -2 and above +2 indicate that taxa between two communities are more or less phylogenetically related than expected by chance, respectively. These underlie a role for homogenizing and variable selection

respectively (i.e. either the same or different taxa are selected in different communities; Stegen *et al.*, 2015). β -NTI values between -2 and +2 indicate no effects of selection in driving differences between two given communities, leaving only dispersal, dispersal limitation, and drift as possible drivers. To identify these alternative processes, in a second step we calculated the Raup-Crick index, which is the departure from a null distribution of Bray-Curtis dissimilarities between probabilistically generated communities (RC_{Bray}). This distribution is created by generating random communities drawn from a species pool composed by the same species composition, richness and absolute abundance as in the original dataset. The probability of these communities of containing particular taxa is proportional to the abundance and frequency of occurrence of those taxa both in the particular samples being compared and in the entire dataset. This index does not rely on phylogenetic distances because phylogenetic relatedness is considered irrelevant for the probability a species has of being dispersed or subject to ecological drift. RC_{Bray} values below -0.95 or above +0.95 indicate two communities have more or less taxa in common than expected by chance, respectively. These are indicative of homogenizing dispersal and dispersal limitation plus drift, respectively. A RC_{Bray} value between -0.95 and +0.95 indicates two communities have as many taxa in common as expected by chance, and is therefore indicative of ecological drift. The β -NTI and RC_{Bray} were calculated with 999 randomizations (Stegen *et al.*, 2012, 2013). We used the R package *picante* (v1.7) to calculate the Faith's phylogenetic distance and its function *comdistnt* for calculation of the mean nearest taxon index (β -MNTD) needed for calculation of β -NTI.

Results

Rooibos biomass and nutritional response to fertilization and soil origin

The only effect on rooibos growth and nutrition came from fertilization, which doubled rooibos biomass (Figure 1; Table 1). Although plants grown in mixed soils had on average higher biomass, this effect was not significant (Table 1). Moreover, there was no interaction between the fertilization and soil origin treatments. The absence of fertilizer triggered a physiological response in the form of increased cluster root production (Figure 1). In line with this observation, foliar N:P ratios shifted from 26.22 ± 7.15 to 18.57 ± 5.32 when fertilizer was added ($P < 0.001$). Fertilization stimulated root nodule formation by increasing total plant dry matter, including that of the root system. Importantly, the number of root nodules correlated positively with root system dry matter, regardless of the fertilization treatment, while the intensity of nodulation per root dry matter remained the same (Table 1; Supplementary Figure 1). All macro- and micronutrient concentrations in the leaves and roots were higher in the fertilized plants, except the Ca in the shoots and Fe in the roots (Supplementary Table 2).

Alpha diversity of root nodule communities

A total of 93 different *nodA* and 355 *gyrB* ZOTUs were identified in the eight-month old rooibos seedling root nodules. These belonged exclusively to the genus *Mesorhizobium* based on *nodA*, while *gyrB* sequences also corresponded the genera *Rhizobium*, *Bradyrhizobium*, *Agrobacterium* and *Paraburkholderia*, the latter at extremely low abundances, apart from a few non-rhizobial taxa (Figure 2). All genera were present in all treatments, with no evident differences in their distribution with fertilization or soil origin. *Mesorhizobium* were by far the most dominant symbionts in rooibos root nodules, making up the most dominant ZOTU in all but one *gyrB* community. The most dominant ZOTUs associated with individual plants represented 93% and 86% of the average symbiont abundance in *nodA* and *gyrB* respectively. Plants grown in fertilized soils hosted a larger proportion of non-*Mesorhizobium* ZOTUs compared plants grown in non-fertilized soils (Figure 2), while soil origin also influenced their relative abundance. The effect of soil origin on relative abundances was especially strong in non-fertilized soils, with cultivated soils having a higher proportion of non-*Mesorhizobium* ZOTUs than the other soil treatments (Figure 2B).

Relationship between fertilization, soil origin, and rhizobium ZOTU richness and diversity in rooibos root nodules

Fertilization with sheep dung and soil origin had distinct effects on rhizobial richness and Simpson's diversity only on *gyrB* communities (Table 2). Fertilized plants hosted a significantly higher number of different *gyrB* ZOTUs (47.81 ± 19.88 ZOTUs) than unfertilized plants (33.84 ± 16.99 ZOTUs) (Figure 3A; Table 2), which was also influenced, but to a lesser extent, by soil origin. Specifically, cultivated and mixed soils had significantly higher *gyrB* ZOTU richness compared to the uncultivated soils (Tukey's HSD 95% confidence comparisons; Mixed-Cultivated: Estimate = 2.19 ± 2.77 , $P = 0.710$; Mixed-Uncultivated: Estimate = 9.39 ± 2.75 , $P = 0.002$; Cultivated-Uncultivated: Estimate = 7.20 ± 2.81 , $P = 0.030$) (Table 2). Importantly, the effects of soil origin on root nodule community richness were independent from the fertilization treatment.

Plants grown on fertilized soils also had higher root nodule community diversity (D fertilized = 3.34 ± 2.23 , D unfertilized = 2.39 ± 1.32) (Figure 3B; Table 2). Soil origin had a weak but significant effect on *gyrB* ZOTU community diversity as well. In this case, cultivated soils had significantly higher diversity than uncultivated ones, whereas mixed soils did not differentiate from the others (Tukey's HSD 95% confidence comparisons; Mixed-Cultivated: Estimate = -0.536 ± 0.257 , $P = 0.095$; Mixed-Uncultivated: Estimate = 0.554 ± 0.257 , $P = 0.082$; Cultivated-Uncultivated: Estimate = 1.090 ± 0.262 , $P < 0.001$) (Table 2). Like for community richness, the effects of soil origin on community diversity were independent from the fertilization treatment.

Relationship between root biomass and the diversity of the root nodule-associated bacterial communities

The observed increase in *gyrB* ZOTU richness and diversity was positively correlated to root biomass (Figure 4A), as well as to the number of nodules per plant (a correlate of plant biomass) (Figure 4B; Supplementary Figure 1). The mixed effects model identified root biomass category as a significant factor explaining both root nodule community richness ($F_{2,191} = 12.54$, $P < 0.001$) and diversity ($F_{2,193} = 5.05$, $P = 0.007$). Post-hoc Tukey's HSD tests confirmed larger root systems had significantly richer and more diverse root nodule communities than smaller ones (Supplementary Table 3).

Since increased Simpson's diversity implied reduced dominance by *Mesorhizobium*, we correlated the abundance of non-*Mesorhizobium* ZOTUs to root biomass and the number of root nodules per plant. Non-*Mesorhizobium* ZOTUs became more abundant with larger root systems and higher numbers of nodules (Figure 4 C, D). When the relative abundance of non-*Mesorhizobium* strains was directly correlated to richness, the correlation coefficients became even larger (Figure E). This pointed at a prominent contribution of rare ZOTUs to the overall alpha diversity and root nodule community structure of root nodules. Importantly, the strong correlation between non-*Mesorhizobium* relative abundance and the overall root nodule community diversity indicates such rare ZOTUs reduce the dominance of *Mesorhizobium* in larger root systems. We then explored the relative contribution of such non-*Mesorhizobium* ZOTUs to the observed richness increase with larger root systems compared to the dominant *Mesorhizobium*. We found a much stronger correlation between non-*Mesorhizobium* ZOTU richness and the overall root nodule community richness ($r = 0.945$, $P < 0.001$), compared to *Mesorhizobium* richness ($r = 0.381$, $P < 0.001$).

The genera that contributed the most to the higher root nodule *gyrB* community richness in larger root systems were identified using indicator species analysis. In descending order, the top indicator ZOTU genera from plants with large root systems were *Bradyrhizobium* (20 ZOTUs), *Rhizobium* (12 ZOTUs), and *Herbaspirillum* (8 ZOTUs) (all selected based on correlation coefficients $r > 0.8$ and $P < 0.05$). Noteworthy, there were no indicators from medium-sized root systems, while 6 ZOTUs were associated with small root systems. From these 6 ZOTUs, the only 2 ZOTUs that were identified at genus level were both *Bradyrhizobium*, and were only found in a few samples from two particular locations.

Since most of the plants with larger root systems that produced more nodules were from the fertilized treatment, we assessed whether the increase in non-*Mesorhizobium* ZOTU relative abundance could be due to additional taxa introduced directly with the fertilizer (i.e. fertilizer-borne taxa). This confirmed that environmental preference for nutrient-rich soil conditions was an unlikely explanation for the increased richness observed in larger root systems. This is so because, from the 10 ZOTUs that were unique to the fertilized treatment, the one with highest (yet very low) cumulative relative abundance belonged to the genus *Mesorhizobium* (17.56% in cumulative relative abundance), the others being 5.37% or less (Supplementary Table 4). Moreover, these fertilizer-related ZOTUs occurred in very few samples, indicating their low occurrences in *gyrB* communities within fertilized treatments (Supplementary Figure 2). The fact that only 10 ZOTUs were unique to fertilized plants and

their rarity suggests a limited role of environmental preferences in explaining the increased rhizobium diversity in larger root systems.

Beta diversity patterns of root nodule communities in rooibos

NMDS ordination plots based on Bray-Curtis *nodA* dissimilarities revealed strong community structure associated with the location from which soils were collected (Figure 5 A-C). Geographical location seemed less important in driving community structure derived from *gyrB* compared to *nodA* (Figure 5 D-F). Mixing of soils from the same location led to an even stronger geographical signal in community dissimilarity (Figure 5C). The PERMANOVA results confirmed the geography-driven β diversity patterns from the ordination plots, also showing that soil origin influenced community composition for both *gyrB* and *nodA* datasets (Table 3). Based on the R^2 values, soil origin was twice as important in structuring *gyrB* than *nodA* communities. Moreover, fertilization had a weak but significant effect on *gyrB* community structure, but no effect on *nodA* communities. Small effects on β diversity were also found for root system size, while there were no effects from the interaction between fertilization and soil origin (Table 3). Despite high stress values, the NMDS plots provide informative insights given the agreement with the PERMANOVA results.

In order to explain the observed β diversity differences, we calculated the distances between centroids of the location x soil clusters based on Bray-Curtis distances for both genes and correlated them to geographical distances between locations and the euclidean distances between soil properties. We found a significant positive correlation ($r = 0.515$, $P < 0.001$) between geographical distances and centroid distances for the *nodA* marker (Figure 6A), but not for the *gyrB* communities ($r = 0.175$, $P = 0.626$; Figure 6B). For the *nodA* communities we only found a significantly positive correlation of distances between centroids and soil Mg concentration ($r=0.227$, $P=0.02$). For *gyrB*, significant correlations were found for water holding capacity ($r=0.303$, $P=0.002$; Figure 6C, D), soil K concentration ($r=0.282$, $P=0.004$), and soil N concentration ($r=0.200$, $P=0.045$).

Finally, following the methods of Stegen et al., (2015), we calculated the relative contribution of neutral (drift and dispersal-based) and deterministic (selection-based) processes in driving species turnover in core symbiotic and bacterial root nodule communities of rooibos. We found a predominance of neutral processes in driving both *gyrB* and *nodA* community turnover (Fig. 7). Root nodule bacterial communities (*gyrB*) were

driven mostly by ecological drift (62.6%), followed by homogeneous selection (25.6%), and the combination of dispersal limitation and drift (6.4%). In contrast, the core symbiotic communities of rooibos (*nodA*) were mostly affected by dispersal limitation and drift (82.7%), followed by drift alone (15.1%) (Fig. 7).

Discussion

In this study we asked whether bacterial communities assemble in neutral or deterministic ways in the root nodules of a promiscuous legume crop adapted to extremely low-nutrient soils. While in rooibos both deterministic and stochastic processes contribute to root nodule community assembly, we found evidence to suggest that it is mostly neutral. Importantly, the core rhizobial symbionts of rooibos (*Mesorhizobium*) were most affected by dispersal limitation in combination with drift, whereas other bacterial members of the root nodule community were more affected by drift alone and selection. Distinct changes in taxonomic diversity with root system biomass, and strong geographical structure in the *Mesorhizobium* group, support such findings.

Contrary to what has been found in most legumes, fertilized rooibos plants produced more nodules on average, and importantly, these harboured richer and more diverse bacterial communities. We suggest that nodule production that is proportional to root biomass production is unsurprising given the low nutrient availability typical of Fynbos soils. This can be seen as an ecological insurance (or “bet-hedging”) strategy (Yachi and Loreau, 1999; Starrfelt and Kokko, 2012), in which initial resource allocation into more nodules may ensure some symbiotic benefits, whether optimal or not, under future conditions of extreme nutrient limitation. Physiological and phenotypic strategies of plants closely related to rooibos with similar adaptations to post-fire recruitment, suggest that such bet-hedging strategies may be common under nutrient-poor conditions (e.g. within the Podalyrieae tribe: *Cyclopia spp.* and *Podalyria spp.*; Maistry *et al.*, 2013; Mndzebele and Dakora, 2017). Such strategy may be related to root nodule bacterial diversity as well (see the balancing nodulation hypothesis, Siler and Friesen, 2017).

Root system biomass strongly correlated with the diversity of bacterial taxa that are not the core symbionts of rooibos. This positive trend happened independently of the soil abiotic conditions, implying it cannot be due to certain strains having environmental preference for particular conditions, but rather a direct consequence of larger root systems or a better health

status of the plant (Dinnage *et al.*, 2019). Mostly rhizobial taxa (*Rhizobium*, *Bradyrhizobium*, *Agrobacterium*), and groups known for their nitrogen-fixation ability (*Herbaspirillum*), colonized nodules from larger root systems. We interpret this as a commensal interaction, whereby taxa supported by neighbouring vegetation may take advantage of the poor ability of rooibos to control nodule infection (Pahua *et al.*, 2018). Under the ecological insurance hypothesis, this *a priori* costly interaction may reward the plant in the form of improved nutrition from drought- or nutrient limitation-adapted taxa on the long run.

Similar to our findings, Dinnage *et al.*, (2019) recently reported increased root nodule rhizobial richness to be linked with root system size in the Australian legume *Acacia acuminata*. This study, based on field data from plants of different age classes, pointed at a role of niche construction during plant growth that would favour the accumulation of rhizobial diversity over time. Under this assumption, root development would make new resources available that could be used by an increasingly wider range of rhizobium strains. While plausible in *Acacia acuminata*, this option is less likely to apply to the present study, as harvested rooibos seedlings were evenly aged and grown in pots. However, Dinnage *et al.*, (2019) also pointed to the role that neutral processes may play in driving the establishment of richer rhizobial communities in larger root systems, which would operate as “islands” (i.e. Theory of Island Biogeography, McArthur and Wilson, 2001). Within such framework, the probability of establishment and maintenance of root nodule bacterial taxa would be proportional to the size of the root system (Dinnage *et al.*, 2019), underlying a neutral community assembly. Support for this island biogeographic model of microbial community assembly is strong from both the α - and β -diversity perspective of the observations made in our study, as discussed below.

The communities of *Mesorhizobium* and rare bacterial taxa in the root nodules were mostly structured by the neutral process of drift, but operating at different spatial scales. The *nodA* communities (true symbionts of rooibos) were mostly structured by geographical origin, which agrees with the predominance of dispersal limitation together with drift in explaining species turnover in these communities. This is further supported by the positive distance-decay in community dissimilarity in the *nodA* community, which is a known indicator of ecological drift (Hanson *et al.*, 2012). This suggests that *nodA* assemblages are most likely shaped by dispersal limitation between distinct rooibos populations. In fact, as a functional marker, the evolution of *nodA* is closely linked to that of its associated host plants (Masson-Boivin *et al.*, 2009; Lemaire *et al.*, 2015). Given the habitat heterogeneity characteristic of the Core Cape Subregion (Cowling *et al.*, 2008), and the known diversity of rooibos ecotypes (Malgas *et al.*,

2010; Hawkins *et al.*, 2011), drift is likely to drive rhizobial community differences between partially isolated rooibos populations (Sprent *et al.*, 2017).

In contrast, the *gyrB* communities, which include a myriad of non-symbiotic taxa, were almost entirely driven by drift alone. When not combined with dispersal limitation, drift underlies the absence of environmental filters (Dumbrell *et al.*, 2010), in this case the absence of filtering by the plant host. As shown before, non-*Mesorhizobium* taxa accumulated proportionally to root system biomass, which agrees with the dominance of drift as there were no environmental filters to root nodule colonization. A proportion of species turnover in these communities was driven by homogenizing selection. Here again, this agrees with the positive relationships observed between *gyrB* community dissimilarity and soil abiotic factors such as WHC. It is known that environmental filtering can favour strains with distinct environmental preferences, for example in response to soil N and P (Vuong *et al.*, 2017), cropping practices (Yan *et al.*, 2014), or surrounding vegetation (Slabbert *et al.*, 2010; Pahua *et al.*, 2018).

We did not find evidence that differences in rhizobium root nodule communities affect the performance of rooibos plants (Figure S3). The maintenance of *Mesorhizobium* dominance throughout the treatments suggests that rooibos derives most of its symbiotic benefits from this genus. Experimental evidence shows that inoculating plants with diverse rhizobial strains decreases symbiotic benefit due to rhizobial strain competition over the short term (Simonsen and Stinchcombe, 2014; Barrett *et al.*, 2015). Despite the action of such biotic interactions within nodules, legumes favour the establishment and replication of the best performing strains (Kiers *et al.*, 2003). When contrasted with the evidence for the “balancing nodulation” hypothesis (Siler and Friesen, 2017), it seems that legumes favour the dominance of very few effective strains to avoid strain competition, while a wide diversity is maintained at very low densities as a potential insurance for the case that alternative symbionts become more beneficial under changing conditions. This complies with the idea that, over the long run, rooibos favours the fitness of *Mesorhizobium* strains within the root nodules, despite a large diversity of other taxa is neutrally maintained at densities low enough to avoid strain competition (Le Roux *et al.*, 2017). Such prediction would be confirmed by studying root nodule assemblages of rooibos in the transition to the nutrient-poor conditions characteristic of its mature life stages, or by performing field studies on distinct mature populations.

Conclusions

Root nodule community assembly in rooibos seedlings is dominated by ecological drift, but it operates at different spatial scales for the core rhizobial symbionts and apparently opportunistic bacterial taxa. Specifically, at the root level, an accumulation of rare bacterial taxa dominated by ecological drift occurs proportionally to the biomass of the root system. Instead, the core rhizobial symbionts of rooibos are structured by dispersal limitation and drift, which operate on a regional scale. The “balancing nodulation” hypothesis (Siler and Friesen, 2017), purports that legumes obtain some long-term benefit from associating with high rhizobium diversity at low densities, such as flexibility to shift the association from one to another symbiont under changing environmental conditions. Our findings suggest under low symbiont selectivity this diversity is maintained by the neutral process of drift. By relating ecological pattern to process, this study shows that dominant symbiotic and rare rhizobial taxa can be structured by different ecological processes. Accounting for these processes can contribute to maintaining root symbiont diversity under productive agricultural conditions, despite the role of root nodule diversity for plant functioning needs to be clarified in the future.

Acknowledgements

This study was funded by the Mercator Research Program of the World Food Systems Center of ETH Zurich. Data produced and analyzed in this paper were generated in collaboration with the Genetic Diversity Center of ETH Zurich (GDC) and the Functional Genomics Center of the University of Zurich (FGCZ). JR acknowledges the assistance from Dr. Jean-Claude Walser, Dr. Andrea Patrignani and Dr. Weihong Qi. JLR acknowledges funding from Macquarie University’s Faculty of Science and Engineering and Department of Biological Sciences. Finally, the authors acknowledge the assistance of Noel Oettlé, Dr. Cecilia Bester, and the community of rooibos farmers in the Suid Bokkeveld for making the sampling in South Africa possible.

References

- Andrews, M., & Andrews, M. E. (2017). Specificity in legume-rhizobia symbioses. *International Journal of Molecular Sciences*, 18(4).
- Barrett, L. G., Bever, J. D., Bissett, A., & Thrall, P. H. (2015). Partner diversity and identity impacts on plant productivity in *Acacia* -rhizobial interactions. *Journal of Ecology*, 103(1), 130–142.
- Bever, J. D. (2015). Preferential allocation, physio-evolutionary feedbacks, and the stability and environmental patterns of mutualism between plants and their root symbionts. *New Phytologist*, 205(4), 1503–1514.
- Bever, J. D., Broadhurst, L. M., & Thrall, P. H. (2013). Microbial phylotype composition and diversity predicts plant productivity and plant-soil feedbacks. *Ecology Letters*, 16(2), 167–174.
- Bissett, A., Richardson, A. E., Baker, G., Wakelin, S., & Thrall, P. H. (2010). Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Molecular Ecology*, 19(19), 4315–4327.
- Brink, C., Postma, A., & Jacobs, K. (2016). Rhizobial diversity and function in rooibos (*Aspalathus linearis*) and honeybush (*Cyclopia spp.*) plants: A review. *South African Journal of Botany*, 110, 80-86.
- Burghardt, L.T. (2019). Evolving together, evolving apart: Measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. *New Phytologist*. <https://doi.org/10.1111/nph.16045>
- Burghardt L. T., Epstein, B., Guhlin, J., Nelson, M. S., Taylor, M. R., Young, N. D., Sadowsky, M. J., & Tiffin, P. (2018). Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proceedings of the National Academy of Sciences USA* 115(10), 2425-2430.

- Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., ... Dangl, J. L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biology*, 15(3), 1–14.
- Cauwenberghe, J. Van, Visch, W., Michiels, J., & Honnay, O. (2016). Selection mosaics differentiate *Rhizobium*-host plant interactions across different nitrogen environments. *Oikos*, 1–7.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–358.
- Cowling, R. M., Procheş, Ş., & Partridge, T. C. (2009). Explaining the uniqueness of the Cape flora: Incorporating geomorphic evolution as a factor for explaining its diversification. *Molecular Phylogenetics and Evolution*, 51(1), 64–74.
- Denison, R. F., & Kiers, E. T. (2011). Life histories of symbiotic rhizobia and mycorrhizal fungi. *Current Biology*, 21(18), 775–785.
- Dinnage, R., Simonsen, A. K., Barrett, L. G., Cardillo, M., Raisbeck-Brown, N., Thrall, P. H., & Prober, S. M. (2019). Larger plants promote a greater diversity of symbiotic nitrogen-fixing soil bacteria associated with an Australian endemic legume. *Journal of Ecology*, 107(2), 977–991.
- Dludlu, M., Chimphango, S., Stirton, C., & Muasya, A. (2017). Differential preference of *Burkholderia* and *Mesorhizobium* to pH and soil types in the Core Cape Subregion, South Africa. *Genes*, 9(1), 2.
- Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C., & Fitter, A. H. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME Journal*, 4(3), 337–345.
- De Cáceres, M. & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90(12): 3566–3574.
- Edgar, R. (2016). SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *BioRxiv*, 074161. <https://doi.org/10.1101/074161>.

- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200.
- Friesen, M. L., & Mathias, A. (2010). Mixed infections may promote diversification of mutualistic symbionts: Why are there ineffective rhizobia? *Journal of Evolutionary Biology*, 23(2), 323–334.
- Gellert, M., Fisher, L. M., & O’Dea, M. H. (1979). DNA gyrase: purification and catalytic properties of a fragment of gyrase B protein. *Proceedings of the National Academy of Sciences*, 76(12), 6289–6293.
- Graham, P. H., & Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant physiology*, 131(3), 872–877.
- Güsewell, S. (2004). N:P ratios in terrestrial plants: Variation and functional significance. *New Phytologist*, 164(2), 243–266.
- Hassen, A. I., Bopape, F. L., Habig, J., & Lamprecht, S. C. (2012). Nodulation of rooibos (*Aspalathus linearis* Burm. f.), an indigenous South African legume, by members of both the α -Proteobacteria and β -Proteobacteria. *Biology and Fertility of Soils*, 48(3), 295–303.
- Haukka, K., Lindström, K., & Young, J. (1998). Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. *Applied and Environmental Microbiology*, 64(2), 419–426.
- Hawkins, H. J., Malgas, R., & Biénabe, E. (2011). Ecotypes of wild rooibos (*Aspalathus linearis* (Burm. F) Dahlg., Fabaceae) are ecologically distinct. *South African Journal of Botany*, 77(2), 360–370.
- Hungria, M., Campo, R. J., & Mendes, I. C. (2003). Benefits of inoculation of the common bean (*Phaseolus vulgaris*) crop with efficient and competitive *Rhizobium tropici* strains. *Biology and Fertility of Soils*, 39(2), 88–93.
- Jiao, Y. S., Liu, Y. H., Yan, H., Wang, E. T., Tian, C. F., Chen, W. X., ... Chen, W. F. (2015). Rhizobial diversity and nodulation characteristics of the extremely promiscuous legume *Sophora flavescens*. *Molecular Plant-Microbe Interactions*, 28(12), 1338–1352.

- Keller, K. R., & Lau, J. A. (2018). When mutualisms matter: Rhizobia effects on plant communities depend on host plant population and soil nitrogen availability. *Journal of Ecology*, 106(3), 1046–1056.
- Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume–rhizobium mutualism. *Nature*, 425(9), 78–81.
- Kiers, E. T., Hutton, M. G., & Denison, R. F. (2007). Human selection and the relaxation of legume defences against ineffective rhizobia. *Proceedings of the Royal Society B: Biological Sciences*, 274(1629), 3119–3126.
- Lambers, H., Bishop, J. G., Hopper, S. D., Laliberté, E., & Zuniga-Feest, A. (2012). Phosphorus-mobilization ecosystem engineering: the roles of cluster roots and carboxylate exudation in young P-limited ecosystems. *Annals of Botany*, 110(2), 329–348.
- Lamont, B. B., & Pérez-Fernández, M. (2016). Total growth and root-cluster production by legumes and proteas depends on rhizobacterial strain, host species and nitrogen level. *Annals of Botany*, 118(4), 725–732.
- Le Roux, J. J., Keet, J. H., Mutiti, B., & Ellis, A. G. (2017). Cultivation may not dramatically alter rhizobial community diversity or structure associated with rooibos tea (*Aspalathus linearis* Burm.f.) in South Africa. *South African Journal of Botany*, 110, 87–96.
- Le Roux, J. J., Ellis, A. G., van Zyl, L. M., Hosking, N. D., Keet, J. H., & Yannelli, F. A. (2018). Importance of soil legacy effects and successful mutualistic interactions during Australian acacia invasions in nutrient-poor environments. *Journal of Ecology*, 106(5), 2071–2081.
- Lemaire, B., Dlodlo, O., Chimphango, S., Stirton, C., Schrire, B., Boatwright, J. S., ... Muasya, A. M. (2015). Symbiotic diversity, specificity and distribution of rhizobia in native legumes of the Core Cape Subregion (South Africa). *FEMS Microbiology Ecology*, 91(2), 1–17.
- Lemanceau, P., Blouin, M., Muller, D., & Moënne-Loccoz, Y. (2017). Let the Core Microbiota Be Functional. *Trends in Plant Science*, 22(7), 583–595.

- Lötter, D., & le Maitre, D. (2014). Modelling the distribution of *Aspalathus linearis* (Rooibos tea): Implications of climate change for livelihoods dependent on both cultivation and harvesting from the wild. *Ecology and Evolution*, 4(8), 1209-1221.
- Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963.
- Maistry, P. M., Muasya, A. M., Valentine, A. J., & Chimphango, S. B. (2015). Increasing nitrogen supply stimulates phosphorus acquisition mechanisms in the fynbos species *Aspalathus linearis*. *Functional Plant Biology*, 42(1), 52-62.
- Maistry, P. M., Cramer, M. D., & Chimphango, S. B. M. (2013). N and P colimitation of N₂-fixing and N-supplied fynbos legumes from the Cape Floristic Region. *Plant and Soil*, 373(1–2), 217–228.
- Malgas, R. R., Potts, A. J., Oetl, N. M., Koelle, B., Todd, S. W., Verboom, G. A., & Hoffman, M. T. (2010). Distribution, quantitative morphological variation and preliminary molecular analysis of different growth forms of wild rooibos (*Aspalathus linearis*) in the northern Cederberg and on the Bokkeveld Plateau. *South African Journal of Botany*, 76(1), 72–81.
- Martens M., Dawyndt P., Coopman R., Gillis M., De Vos P. & Willems A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *International Journal of Systematic and Evolutionary Biology*, 58, 200–214.
- Masson-Boivin, C., Giraud, E., Perret, X., & Batut, J. (2009). Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends in Microbiology*, 17(10), 458–466.
- McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4).
- Mndzebele, B. M., & Dakora, F. D. (2017). Plant growth and N₂ fixation in *Cyclopia longifolia* (Vogel L.) supplied with mineral nutrients in pot and field experiments. *South African Journal of Botany*, 110, 97–102.

- Moulin, L., Munive, A., Dreyfus, B. and Boivin-Masson, C. (2001). Nodulation of legumes by members of the 13-subclass of Proteobacteria. *Nature* 411, 948-950.
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., ... Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, 77(3), 342–356.
- Neumann, G., & Martinoia, E. (2002). Cluster roots - An underground adaptation for survival in extreme environments. *Trends in Plant Science*, 7(4), 162–167.
- Oksanen, A. J., Kindt, R., Legendre, P., Hara, B. O., Simpson, G. L., Stevens, M. H. H., ... Solymos, P. M. (2008). The vegan package. *The Vegan Package*, 190.
- Pahua, V. J., Stokes, P. J. N., Hollowell, A. C., Regus, J. U., Gano-Cohen, K. A., Wendlandt, C. E., ... Sachs, J. L. (2018). Fitness variation among host species and the paradox of ineffective rhizobia. *Journal of Evolutionary Biology*, 31(4), 599–610.
- Philippot, L., Raaijmakers, J. M., Lemanceau, P., & van der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews. Microbiology*, 11(11), 789–799.
- Ramoneda, J., Le Roux, J., Frossard, E., Bester, C., Oetlé, N., Frey, B., & Gamper, H. (2019). Insights from invasion ecology: Can consideration of eco-evolutionary experience promote benefits from root mutualisms in plant production? *AoB Plants*, DOI: 10.1093/aobpla/plz060/5572512.
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863–864.
- Shane, M. W., & Lambers, H. (2005). Manganese accumulation in leaves of *Hakea prostrata* (Proteaceae) and the significance of cluster roots for micronutrient uptake as dependent on phosphorus supply. *Physiologia Plantarum*, 124(4), 441–450.
- Siler, E., & Friesen, M. L. (2017). Widespread Negative Frequency-Dependent Selection Maintains Diversity in the Legume-Rhizobia Symbiosis Balancing nodulation may

- explain the paradox of rhizobium diversity. *BioRxiv*, 1–15.
<https://doi.org/10.1101/153866>.
- Simberloff, D. (2009). The Role of Propagule Pressure in Biological Invasions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 81–102.
- Simonsen, A. K., & Stinchcombe, J. R. (2014). Standing genetic variation in host preference for mutualist microbial symbionts. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797).
- Slabbert, E., Kongor, R. Y., Esler, K. J., & Jacobs, K. (2010). Microbial diversity and community structure in Fynbos soil. *Molecular Ecology*, 19(5), 1031–1041.
- Sprent, J. I., Ardley, J., & James, E. K. (2017). Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytologist*, 215(1), 40–56.
- Starrfelt, J., & Kokko, H. (2012). Bet-hedging-a triple trade-off between means, variances and correlations. *Biological Reviews*, 87(3), 742–755.
- Team R Core. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. <https://www.r-project.org/>.
- Thrall, P. H., Laine, A. L., Broadhurst, L. M., Bagnall, D. J., & Brockwell, J. (2011). Symbiotic effectiveness of rhizobial mutualists varies in interactions with native Australian legume genera. *PloS one*, 6(8), e23545.
- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., ... & Yoshida, K. (2018). Core microbiomes for sustainable agroecosystems. *Nature Plants*, 4(5), 247.
- Van Klinken, R. D., Lukitsch, B., & Cook, C. (2008). Interaction between seed dormancy-release mechanism, environment and seed bank strategy for a widely distributed perennial legume, *Parkinsonia aculeata* (Caesalpinaceae). *Annals of Botany*, 102(2), 255–264.
- Vannette, R. L., & Fukami, T. (2014). Historical contingency in species interactions: Towards niche-based predictions. *Ecology Letters*, 17(1), 115–124.

- Vuong, H. B., Thrall, P. H., & Barrett, L. G. (2017). Host species and environmental variation can influence rhizobial community composition. *Journal of Ecology*, 105(2), 540–548.
- Wagner, M. R., Lundberg, D. S., Del Rio, T. G., Tringe, S. G., Dangl, J. L., & Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications*, 7, 1–15.
- Wang, Q., Liu, J., & Zhu, H. (2018). Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions. *Frontiers in Plant Science*, 9(3), 1–8.
- Wendlandt, C.E., Regus, J.U., Gano-Cohen, K.A., Hollowell, A.C., Quides, K.W., Lyu, J.Y., Adinata, E.S., Sachs, J.L. (2019). Host investment into symbiosis varies among genotypes of the legume *Acmispon strigosus*, but host sanctions are uniform. *New Phytologist*, 221(1), 446-458.
- West, S. A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: Why do rhizobia fix nitrogen? *Proceedings of the Royal Society B: Biological Sciences*, 269(1492), 685–694.
- Yan, J., Han, X. Z., Ji, Z. J., Li, Y., Wang, E. T., Xie, Z. H., & Chen, W. F. (2014). Abundance and diversity of soybean-nodulating rhizobia in black soil are impacted by land use and crop management. *Applied and Environmental Microbiology*, 80(17), 5394–5402.
- Young, J. P. W., & Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. *Trends in Ecology and Evolution*, 4(11), 341–349.
- Yu, P., Wang, C., Baldauf, J.A., Tai, H., Gutjahr, C., Hochholdinger, F., Li, C. (2018). Root type and soil phosphate determine the taxonomic landscape of colonizing fungi and the transcriptome of field-grown maize roots. *New Phytologist*, 217(3), 1240-1253.

Display items

Figure 1 Total dry plant biomass of rooibos grown in pots for eight months as affected by fertilization with sheep dung and by soil from either plantations, wild populations, or a mixture of these. The soils were collected from five locations/farms with adjacent cultivated and uncultivated rooibos populations. Box plots show the median and first and third quartile and the whiskers 1.5 standard deviations. Black dots indicate data from plants with cluster roots and grey dots data from plants without cluster roots. The statistical summary reports on differences in plant biomass between the fertilization treatments (n=271).

Figure 2 Relative abundances of different rhizobial genera from pools of root nodules of eight-month-old rooibos grown in soils from five plantations (cultivated) and adjacent wild (uncultivated) populations and a 1:1 (v:v) mixture of soil pairs (mixed). Soils were either fertilized with sheep dung (A) or not (B). The analysis relied on *gyrB* phylotaxonomic marker sequences and their comparison to a reference database of all publicly available and taxonomically assigned sequences.

Figure 3 Effects of fertilization and soil origin on richness and diversity of bacterial operational taxonomic units (ZOTUs) in pools of nodules of rooibos grown in cultivated and wild soils and the mixture of these two soils. Soils were taken from five locations/farms, and were either fertilized with sheep dung or not. The ZOTUs were determined by similarity clustering of an 817 bp fragment of the phylotaxonomic marker gene *gyrB* at 99% sequence identity. The result of the mixed model testing the effects of fertilization on richness and diversity is reported on the graphs.

Figure 5 Non-metric multidimensional scaling (NMDS) of Bray-Curtis dissimilarities between the root nodule community structures from pools of root nodules of single rooibos plants grown in soil from five locations/farms for eight months. Half of the plants were fertilized with sheep dung. Root nodule community structures were characterized using the nodulation gene, *nodA*, and the DNA gyrase subunit B phylotaxonomic marker, *gyrB*. Ordination plot of bacterial communities analysed with *nodA* from cultivated (pane A, Stress = 0.24) and uncultivated soils (pane B, Stress = 0.23). C) Ordination plot of bacterial communities of plants grown in a 1:1 (v:v) mixture of soil from plantations and adjacent uncultivated populations of rooibos analysed with *nodA* (Stress = 0.22). Ordination plot of root nodule community structures analysed with *gyrB* from cultivated (pane D, Stress = 0.16) and uncultivated soils (E, Stress = 0.16). F) Ordination plot of bacterial communities of plants grown in a 1:1 (v:v) mixture of soil from plantations and adjacent uncultivated populations of rooibos analysed with *gyrB* (Stress = 0.14), 95% confidence ellipses are shown when differences between clusters explained more than 10% of the variation.

Figure 6 Relationship between the differences among the rhizobial community structures of root nodule pools of rooibos and geographical distances between locations (A, B), and water holding capacities (WHC) of different soils (C, D). The rhizobial community structures were characterized using the nodulation gene *nodA* (A, C) and the DNA gyrase subunit B phylotaxonomic gene, *gyrB* (B, D).

Figure 7 Relative contribution of neutral (drift and dispersal-related) and deterministic (selection-related) ecological processes explaining differences in root nodule community structure. The core rhizobial community was characterized using the nodulation gene *nodA* and the bacterial root nodule community was characterized using the *gyrB* gene. Proportions are based on β -NTI and RC_{Bray} values obtained by comparison to null models as in Stegen et al., (2015).

Response variable	Experimental factors					
	Fertilization (df = 1)		Soil origin (df = 2)		Fertilization:Soil origin (df = 2)	
	F-value	P-value	F-value	P-value	F-value	P-value
Total plant dry biomass (g)	72.87	<0.001	2.31	0.102	0.82	0.440
Root dry biomass (g)	60.43	<0.001	2.01	0.137	0.76	0.468
Leaf N content (mg)	109.24	<0.001	1.71	0.183	0.51	0.600
Leaf P content (mg)	181.82	<0.001	3.26	0.040	0.94	0.392
Leaf N:P ratio	105.88	<0.001	11.24	<0.001	0.22	0.300
□ ¹⁵ N (‰)	211.76	<0.001	0.34	0.714	1.18	0.308
# Cluster roots	31.72	<0.001	0.30	0.740	0.57	0.566
# Nodules	59.98	<0.001	3.46	0.033	2.09	0.126
Nodule density (# nodules root dry biomass ⁻¹)	0.20	0.653	13.14	<0.001	2.64	0.073

Table 1 Results of mixed model analyses on the effects of fertilization, soil origin and their interaction on plant biomass, nitrogen (N) and phosphorus (P) nutritional parameters as well as cluster root and root nodule formation of rooibos. The experimental factor ‘fertilization’ involved sheep dung addition to soil, and ‘soil origin’ comprises use of soils from five pairs of plantations and adjacent uncultivated populations of rooibos and a 1:1 (v:v) mixture of those. The five locations (farms) were treated as a random and nested factor within fertilization and soil origin. There were 10 pot replicates per experimental treatment.

<i>Factor</i>	<i>Richness</i>				<i>Simpson's diversity</i>			
	<i>gyrB</i>		<i>nodA</i>		<i>gyrB</i>		<i>nodA</i>	
	<i>F-value</i>	<i>P-value</i>	<i>F-value</i>	<i>P-value</i>	<i>F-value</i>	<i>P-value</i>	<i>F-value</i>	<i>P-value</i>
Fertilization	34.94	<0.001	0.75	0.387	21.27	<0.001	3.10	0.080
Soil origin	5.45	0.005	0.49	0.613	8.44	<0.001	2.32	0.101
Fertilization :Soil origin	1.08	0.343	0.45	0.638	1.29	0.277	1.74	0.178

Table 2 Results of mixed model analyses on the effects of fertilization, soil origin and their interaction on the richness and diversity of rhizobia in root nodules of rooibos. The root nodules were analysed using *gyrB* and *nodA* DNA sequences, which was used to calculate zero-sums operational taxonomic units (ZOTU) based on 99% sequence identity. Richness represents the total number of ZOTUs detected. The experimental factor ‘fertilization’ involved sheep dung addition to soil, and ‘soil origin’ comprises the use of soils from five pairs of plantations and adjacent uncultivated populations of rooibos and a 1:1 (v:v) mixture of those. The five locations (farms) were treated as a random and nested factor within fertilization and soil origin. There were 10 pot replicates per experimental treatment.

<i>Factor</i>	<i>nodA</i>				<i>gyrB</i>		
	<i>df</i>	<i>F-value</i>	<i>R</i> ²	<i>P-value</i>	<i>F-value</i>	<i>R</i> ²	<i>P-value</i>
Location	4	30.28	0.246	<0.001	17.56	0.174	<0.001
Soil origin	2	10.29	0.042	<0.001	14.29	0.071	<0.001
Fertilization	2	1.27	0.003	0.204	3.36	0.008	0.010
Root system size	2	2.01	0.008	0.003	2.64	0.013	0.008
Location:Soil origin	8	11.88	0.193	<0.001	10.19	0.202	<0.001
Fertilization:Soil origin	2	1.28	0.004	0.171	1.58	0.008	0.106

Table 3 Results of a permutational analysis of variance (PERMANOVA) on the effects of location, soil origin, fertilization and root system size and their interactions on the rhizobial community structures in root nodules of rooibos). DNA sequence data of the *gyrB* and *nodA* genes were used to determine operational taxonomic units at 99% sequence identity whose relative abundances were used to calculate Bray-Curtis dissimilarities. The analysis was run with 9999 permutations.