
Article:
Fields, B. orcid.org/0000-0003-1251-6867, Moeskjær, S. orcid.org/0000-0003-4987-0130, Deakin, W.J. et al. (5 more authors) (2023) Rhizobium nodule diversity and composition are influenced by clover host selection and local growth conditions. Molecular Ecology. ISSN 0962-1083

https://doi.org/10.1111/mec.17028

Received: 14 February 2023 | Revised: 25 April 2023 | Accepted: 15 May 2023

DOI: 10.1111/mec.17028

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Rhizobium nodule diversity and composition are influenced by clover host selection and local growth conditions

Bryden Fields[1] | Sara Moeskjær[2,3] | William J. Deakin[1] | Emma K. Moffat[1,2] | Niels Roulund[4] | Stig U. Andersen[3] | J. Peter W. Young[1] | Ville-Petri Friman[1,5]

[1]Department of Biology, University of York, York, UK

[2]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

[3]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

[4]DLF, Store Heddinge, Denmark

[5]Department of Microbiology, University of Helsinki, Helsinki, Finland

Correspondence
Ville-Petri Friman, Department of Microbiology, University of Helsinki, Helsinki, Finland.
Email: ville-petri.friman@helsinki.fi

Present address
Bryden Fields, Fera Science Ltd., York Biotech Campus, York, UK

Funding information
Daphne Jackson Trust; Innovationsfonden, Grant/Award Number: 4105-00007A; Royal Society of Biology, Grant/Award Number: CHL\R1\180031 and RSG\R1\180213; UK Research and Innovation, Grant/Award Number: BB/T010606/1

Handling Editor: Jeremy B Yoder

Abstract

While shaping of plant microbiome composition through 'host filtering' is well documented in legume–rhizobium symbioses, it is less clear to what extent different varieties and genotypes of the same plant species differentially influence symbiont community diversity and composition. Here, we compared how clover host varieties and genotypes affect the structure of *Rhizobium* populations in root nodules under conventional field and controlled greenhouse conditions. We first grew four *Trifolium repens* (white clover) F_2 crosses and one variety in a conventional field trial and compared differences in root nodule *Rhizobium leguminosarum* symbiovar *trifolii (Rlt)* genotype diversity using high-throughput amplicon sequencing of chromosomal housekeeping (*rpoB* and *recA*) genes and auxiliary plasmid-borne symbiosis genes (*nodA* and *nodD*). We found that *Rlt* nodule diversities significantly differed between clover crosses, potentially due to host filtering. However, variance in *Rlt* diversity largely overlapped between crosses and was also explained by the spatial distribution of plants in the field, indicative of the role of local environmental conditions for nodule diversity. To test the effect of host filtering, we conducted a controlled greenhouse trial with a diverse *Rlt* inoculum and several host genotypes. We found that different clover varieties and genotypes of the same variety selected for significantly different *Rlt* nodule communities and that the strength of host filtering (deviation from the initial *Rhizobium* inoculant composition) was positively correlated with the efficiency of symbiosis (rate of plant greenness colouration). Together, our results suggest that selection by host genotype and local growth conditions jointly influence white clover *Rlt* nodule diversity and community composition.

KEYWORDS
clover genotype, clover variety, diversity, host filtering, *Rhizobium*, symbiotic specificity

1 | INTRODUCTION

Plant species are well documented to influence their associated microbiota by attracting and repelling specific microorganisms for their productive benefit (Burns et al., 2015; Fitzpatrick et al., 2018; Jones et al., 2019; Quiza et al., 2015; Schmid et al., 2019; Zhalnina et al., 2018). Symbiotic nitrogen-fixing bacteria called rhizobia are one such highly diverse group of soil microbes regularly exposed to host filtering by plants. Plants in the Fabaceae family, commonly known as legumes, can form mutualistic root nodule symbioses with nitrogen-fixing rhizobium species to increase their nitrogen uptake for subsequent growth. To take advantage of this interaction, legume crops are commonly inoculated with rhizobial strains in the field as a sustainable alternative to chemical fertilisation to improve soil nitrogen balance. However, there is variation in the reciprocity of symbiosis; not all symbionts are effective, and this specificity of symbiotic compatibility is well known between different legume species (Bromfield et al., 1995; Laguerre et al., 2003; Wang, Liu, & Zhu, 2018). Legumes are able to discriminate and select for compatible rhizobia through partner choice, whereby the host plant detects incompatible strains and prevents their symbiotic establishment, or by sanctioning, whereby plants restrict resources to ineffective strains once inside the nodule (Kiers & Denison., 2008; Westhoek et al., 2017). Some host-filtering mechanisms used by legumes include symbiotic specificity for nodule occupancy and manipulation of rhizosphere community composition through secretion of specific root exudates (Jones et al., 2019). For example, selection of compatible rhizobia for symbiosis is initially mediated by highly specific interactions between plant flavonoid exudates and corresponding rhizobial symbiosis genes (Clúa et al., 2018; Wang et al., 2012). Therefore, to optimise the efficiency and success of rhizobium inoculation treatments, it is important to select symbiotic bacterial strains that are genetically compatible with the host plant. While it is known that different legume species select for certain rhizobium species, it is less clear to what extent distinct varieties of the same legume species differ in their symbiotic genotype preference.

Studies that have aimed to elucidate the variation in symbiotic specificity between legume cultivars have predominantly used a restricted number of rhizobium strains or a synthetic inoculum community, with potentially limited relevance for field application (Bromfield, 1984; Burghardt et al., 2018; Carelli et al., 2000; Jones & Hardarson, 1979; Rangin et al., 2008; Russell & Jones, 1975; Wadhwa et al., 2011; Xiong et al., 2017; Yang et al., 2017). Under greenhouse conditions, *Vicia faba* (faba bean) and *Pisum sativum* (pea) cultivars have been shown to preferentially select for different rhizobial genotypes (Wadhwa et al., 2011; Xiong et al., 2017). *Medicago sativa* (alfalfa) cultivars displayed the same preferences for *Rhizobium meliloti* strains but only a two-strain inoculant was used (Bromfield, 1984). Conversely, *Trifolium repens* (white clover) varieties were found to display preferences for specific rhizobial genotypes when inoculated with a two-strain *Rhizobium inoculum*, and significant variation in rhizobial ability to nodulate different plant cultivars was identified (Jones & Hardarson, 1979; Russell & Jones, 1975). Despite this, another study showed that white clover cultivars inoculated with a natural soil suspension which likely contained a greater diversity of *Rhizobium* did not display preference for different *Rhizobium leguminosarum* symbiovar *trifolii* (*Rlt*) genotypes under greenhouse conditions (Harrison et al., 1987). However, allozyme variants were used to detect differences in population structure, and these would reflect housekeeping genes rather than those involved in symbiosis. While controlled greenhouse experiments provide vital information for understanding plant microbiome selectivity, 'real' crop systems are important for identifying specific rhizobial inocula that achieve beneficial compatibility with legume host crops in the field (Wadhwa et al., 2011). Moreover, field experiments are required to understand to what extent host filtering is affected by different environmental factors that vary spatially under complex field conditions. For example, interactive effects between the field geography and host genotype could cooperatively manipulate rhizobial diversity as it has been shown, in both soybean and common bean legumes, where soil type predominantly influenced microbiome community and host genotype modified the selectivity (Argaw & Muleta, 2017; Liu et al., 2019). Additionally, rhizobial competitiveness for nodule occupancy in *Trifolium subterraneum* cultivars has been associated with host genotype and bacterial strain compatibility but also with field location (Roughley et al., 1976). However, in other studies, *Rhizobium* genotype preferences have been indistinguishable between *Trifolium* species hosts and soil types when grown across different sites (McGinn et al., 2016). Therefore, more research is needed to better understand the relative effect of white clover genotypes for selecting symbiont genotypes under field and greenhouse conditions.

The compatibility between legume genotypes and *Rhizobium* strains has previously been attributed to differences in both the rhizobial symbiosis plasmid and chromosome (Brewin et al., 1983; Paffetti et al., 1996). A variety of different chromosomal and plasmid-borne gene markers have been used for determining rhizobial population diversity in both nodule and soil samples (Bromfield et al., 1995; Laguerre et al., 2003; McGinn et al., 2016), including *rpoB* (RNA polymerase B subunit) and *recA* (recombinase A) genes, which are robust chromosomal markers for intraspecies diversity (Wang, Cui, et al., 2018; Wang, Liu, & Zhu, 2018; Xiong et al., 2017). Furthermore, *rpoB* and *recA* sequences can be used to assign *R. leguminosarum* isolates to genospecies, which are separated by average nucleotide identity values below 96% (Kumar et al., 2015; Young et al., 2021). Allelic frequencies of several plasmid-borne symbiosis genes have also been used to determine differences in legume–*Rhizobium* partner compatibility, including *nodD* (transcriptional regulator of nodulation *nod* gene activation) (Hassan & Mathesius, 2012; Laguerre et al., 1996; Maj et al., 2010; McGinn et al., 2016; Perret et al., 2000; Redmond et al., 1986; Zézé et al., 2001) and *nodA* (N-acyltransferase essential for successful Nod factor production) genes (Debellé et al., 1996; Igolkina et al., 2019; Lupwayi et al., 2006; Maj et al., 2010; Poinsot et al., 2016; Ritsema et al., 1996; Wang, Liu, & Zhu, 2018). Symbiosis plasmids can transfer between rhizobial strains through horizontal gene transmission on a rapid ecological timescale which

can influence the symbiotic capabilities host specificity of rhizobial populations (Kumar et al., 2015; Wardell et al., 2021). Therefore, by utilising multiple chromosomal and plasmid-borne genetic markers for analysis, a broad insight of intraspecific population diversity driven by both vertical and horizontal genetic transmission can be perceived (Wernegreen & Riley, 1999).

We aimed to test if distinct genetic variants of white clover differ in their host filtering and consequent rhizobial nodule diversity and community composition in field and controlled greenhouse experiments. First, we investigated differences in *Rlt* genotype diversity between root nodule populations from four white clover F_2 crosses and one variety grown under conventional field conditions using high-throughput amplicon sequencing (Fields et al., 2020). We analysed the relative allelic diversity of two *Rlt* chromosomal housekeeping genes, *rpoB* and *recA*, and two plasmid-borne symbiosis genes, *nodA* and *nodD*. Second, we used controlled greenhouse conditions to inoculate different clover varieties and genotypes with a single diverse rhizobium inoculum to determine the variability in white clover host filtering. Our results showed evidence of host filtering under both agricultural and greenhouse conditions. Moreover, the strength of host filtering depended on a given clover variety and varied depending on the spatial location in the field. We observed clear effects of white clover host variety and genotype in a controlled greenhouse experiment, where the strength of host filtering (deviation from the initial rhizobium inoculant composition) was positively correlated with the efficiency of symbiosis (rate of plant greenness colouration). Together, our results suggest that legume selection for symbiotic bacteria is highly specific and determined by genotype-by-genotype-by-environment interactions.

2 | METHODS

2.1 | Field trial plants, nodule sampling and DNA extraction

Genetically distinct F_2 *Trifolium repens* (white clover) variety crosses and one variety (Klondike) were grown in plots in a large conventionally managed field trial at Store Heddinge, Denmark, for a DLF Trifolium research and development plant breeding trial. From this ongoing field trial, we sampled a subset of plots, as indicated in Figure 1, which included four genetically distinct F_2 variety crosses (Cross 1; Cross 2; Cross 3; Cross 4) and one variety (Klondike). No F_2 cross shared a parent variety with another cross. Similarly, the Klondike variety was not a parent in any of the crosses. As a result, all the crosses are genetically distinct from one another, and are here on referred to simply as 'clover crosses'. The varieties used to generate the four crosses have been made confidential in this study due to the intellectual property rights of DLF Trifolium breeder trial developments.

A mixture of 6 g clover cross seed and 20 g of diploid perennial ryegrass varieties, Indiana and Boyn, was sown on each plot on 22 June 2017. All plots were sown in the same field with the same environmental and management conditions. Plots were organised into two blocks, each containing two rows of 18 plots, and plots were sown in dimensions of 8 m by 1.5 m (Figure 1). Within each block, a strip of grassland measuring 3.3 m separated the two rows of 18 plots. A strip of grassland measuring 8 m separated the two blocks of plots. Clover cross plots were sown within blocks in a rectangular Latin plot design (Figure 1); a complete randomised

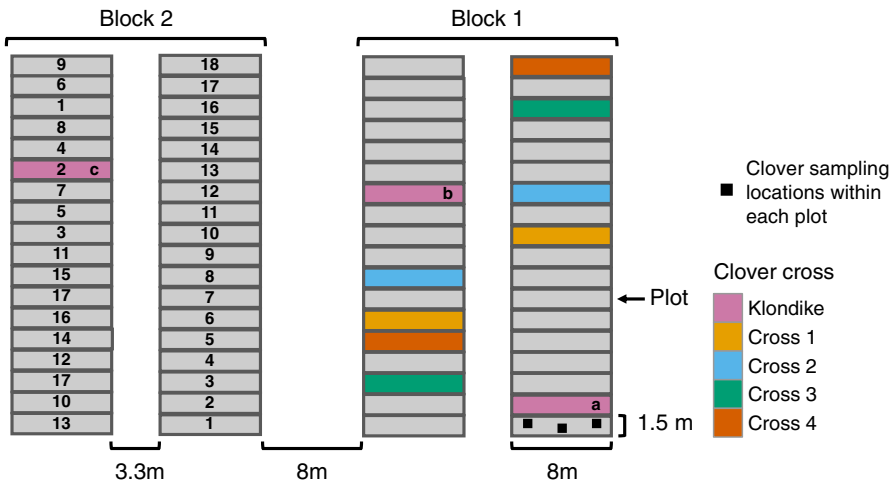


FIGURE 1 Field plot design for sampling. Plots were organised into two blocks, each containing two rows of 18 plots (grey rectangles), sown in dimensions of 8 m by 1.5 m. Blocks were separated by a strip of grassland measuring 8 m across. Additionally, within blocks, the two rows of plots were separated by grassland measuring 3.3 m across. Plots sampled in this study are coloured, respectively, by the white clover cross sown on each plot (see the legend on the right). Two plots were sampled per clover cross from Block 1, except for Klondike where an additional third plot was sampled from Block 2 to enable further analysis of geographic distance (plot c in addition to plots a and b from Block 1). Clovers were sampled from three locations on each plot (Black square dots in the most bottom-right plot). Therefore, six samples were collected for each clover cross from Block 1. A total of 100 nodules were sampled from 3 to 10 clover plants for each sampling point. Clover crosses were sown within blocks in a rectangular Latin plot design, as demonstrated by the numbering system outlined in Block 2, whereby a number represents a clover cross.

block design with a restriction in the randomisation, providing a complete set of all clover crosses in both directions within a block (Figure 1). Within Block 1, clover crosses plots were sown in duplicate (Figure 1). An additional third replicate Klondike plot was sampled from Block 2 to enable further analysis of geographic distance effects (plot c in addition to a and b; Figure 1). No other plant species, or *Rhizobium* inoculations, were added to plots before sowing. In the establishment year, the trial plots received no fertiliser treatment. In the second year, plots received fertiliser treatment four times across the year, which totalled 170 kg N/harvest year across all plots.

Around 3–10 clover plants were sampled from three points in each plot on 23 October 2018 (16 months after sowing). Therefore, for each plot, three independent replicate samples were collected, and six samples were collected in total for each clover cross from Block 1. From Block 2, three independent replicate samples were collected from one Klondike plot only (Figure 1: plot c). For each plot point, clovers were washed, and 100 pink nodules picked and pooled together. Nodules were stored at −20°C. Nodule samples were thawed at room temperature, crushed with a sterile homogeniser stick and mixed with 750 µL Bead Solution from the DNeasy PowerLyzer PowerSoil DNA isolation kit (Qiagen). DNA was then extracted for each within-plot replicate using the DNeasy PowerLyzer PowerSoil DNA isolation kit (Qiagen) following the manufacturer's protocol.

2.2 | Greenhouse trial

Plants were initially grown in the greenhouse as part of a larger greenhouse trial investigating the effects of clover and rhizobium genetic variation through monitoring of plant growth traits (Moeskjær et al., 2022). Mother plants of 20 commercial varieties with known genotypes were initially grown and cuttings were taken. From each variety, either two genotypes (for Aberboost, Aberconcor, Avalon, Barblanca, Borek, Brianna, Chieftain, Coolfin, Cyma, Iona, Klondike, Liblanc, Liflex, Merida, Riesling and Violin varieties) or three genotypes (for Aberpearl, Aran, Rabani and Silvester varieties) were sampled, resulting in a total of 44 white clover genotypes. The genotypes were selected to be morphologically distinct within each variety for the following plant traits: overall plant size, leaf markings, plant density, leaf colour, internode length and stolon density (Moeskjær et al., 2022). The crosses used in the field trial were not used in the greenhouse trial due to the intellectual property rights of DLF Trifolium breeder trial. Three out of eight varieties which were used as parents of the F_2 crosses in the field trial were independently used in the greenhouse experiment, although these varieties were not necessarily the same genotypes as used in the field trial. Therefore, we consider the varieties incomparable due to their genetic distinctness, and for this reason, we have chosen to not form any direct comparisons between varieties used in the greenhouse trial and field trial.

All plants in the greenhouse trial were propagated using four sterilised cuttings from the mother plants. To sterilise the cuttings, cuttings were submerged in 1:4 bleach (Klorin, Colgate-Palmolive Company, USA) for 5 s, washed with tap water and then stored in tap water with Conserve (Dow Agroscience) for up to 10 min until potting. Cuttings were potted in 5 L pots filled with vermiculite (Pull Rhenen B. V.) and grown under nitrogen limited conditions with individual drip irrigation, also supplying nutrient solution. Plants were then grown in Egå, Denmark (56.226° N, 10.259° E) in the growth period 15 August 2018 to 24 October 2018 (68–70 days). Cuttings were kept under white plastic covering at 100% humidity for 2 weeks, acclimatised gradually by cutting holes in the plastic over a week and then transferred to the main greenhouse. Once transferred to the greenhouse, all the plants were then immediately inoculated with a 50-mL inoculum containing a mix of 10 characterised *Rlt* strains (10% each) from the three major genospecies (Cavassim et al., 2020). These strains were chosen with the criteria that all sequences should have a >2 nt difference between them in the amplified region of *recA*, resulting in six genospecies A (gsA), 1 gsB and 3 gsC strains (SM3, SM47, SM88, SM123, SM128A, SM130B, SM132, SM136A, SM140B and SM146A) (Table S1). Strains were grown individually in Tryptone Yeast medium for 48 h at 28°C from glycerol stocks, diluted with tap water (OD_{600} = 0.001) and mixed in equal volumes (10% each).

After the shoots were harvested at the end of the growth period, two within-pot replicates of 50 pink nodules were collected per pot using tweezers. Nodules were washed with water to remove vermiculite and processed for molecular work as described in the section 'Field trial plants, nodule sampling and DNA extraction'.

Plant yield and greenness were measured during the greenhouse experiment, as detailed in Moeskjær et al. (2022) and Tausen et al. (2020) respectively. Briefly, the growth and greenness of plants throughout the greenhouse experiment were monitored using a greenhouse overhead camera imaging system (Moeskjær et al., 2022; Tausen et al., 2020). Firstly, the initial size of each plant was measured by pixel counts from a 512 × 512 pixel mask after the first 10 days of rhizobium mixture inoculation when the symbiotic relationship between clover and rhizobia had not yet been established. The growth per day of each plant was calculated by measuring the dry weight of harvested plants divided by days of growth from inoculation to harvest. Additionally, greenness measurements were used to assess the nitrogen fixation status of the plant during the experiment; clover plants turn yellow under nitrogen starvation, whereas green colour indicates a successful symbiosis with nitrogen-fixing rhizobia (Carter & Knapp, 2001; Sloger, 1969). To provide a quantitative measure of greenness, we calculated the mean and variance of the greenness hue distribution of each plant area and a change in the mean plant colouration over time (Tausen et al., 2020). The rate at which symbiotic nitrogen fixation was established was determined by fitting a linear regression on the greenness measure during the first 20 days after inoculation, when the change in plant greenness colouration

was expected to occur due to establishment of symbiosis (Tausen et al., 2020). The slope of the regression was defined as the rate of change in greenness which reported the rate at which the plant hue changed, indicative of how fast effective nitrogen fixation was established for each plant (Tausen et al., 2020).

2.3 | DNA sample and read processing

Field and greenhouse nodule DNA samples were processed the same way; *Rlt*-specific genes *rpoB*, *recA*, *nodA* and *nodD* were individually PCR amplified and processed for each pooled nodule sample using the MAUI-seq high-throughput amplicon sequencing method, as described in detail previously (Fields et al., 2020). Briefly, in this method, genes are amplified in a nested PCR using primers that contain a region of 12 random bases in the forward inner primer, which generates a unique molecular identifier (UMI) for each initial DNA strand in the first round of PCR amplification. All subsequent daughter DNA strands generated will contain the same UMI as their parent. This consequently means DNA reads with the same UMI can then be grouped and aid an identification of erroneous sequences, such as chimeras and PCR mutations, during sequencing read processing. This was carried out with the aim to better reflect true allelic diversity of samples by filtering out identified errors across samples.

Initial PCRs were carried out individually for each primer set using non-proofreading Platinum Taq DNA polymerase (Thermofisher Scientific Inc.). Equal volumes of the four PCRs produced for each sample were pooled, cleaned (AMPure XP Beads, Beckman Coulter) and indexed for sequencing (Nextera XT DNA Library Preparation Kit v2 set A; Phusion High-Fidelity DNA polymerase, Thermofisher Scientific Inc.) (Fields et al., 2020). All samples were pooled and quality checked by Bioanalyzer 2100 (Agilent), before sequencing using Illumina MiSeq (2×300 bp paired end reads) by the University of York Technology Facility. Full method protocols including PCR reaction mixtures and programmes are detailed in Fields et al. (2020).

Paired-end reads were first merged using the PEAR assembler (Zhang et al., 2014). Reads were then processed using MAUI-seq python scripts to firstly separate reads into the four *Rlt* genes for each sample, and to secondly calculate the abundance of unique UMI reads for each gene in each sample. To calculate total abundances, reads were grouped by UMI, and the most abundant read sequence (primary sequence) was assigned to that UMI, thereby removing PCR errors. Then, the number of UMIs associated with the same primary sequence was counted. Chimeras were detected by comparing the number of times a sequence appears as the most abundant sequence for a UMI compared to appearing as the second most abundant sequence (secondary sequence) in UMI clustering. Additional parameters used to control the stringency of the analysis were all set to default. These include: (1) count a UMI only if the most abundant sequence has two more reads than the second most abundant sequence; (2) reject sequences that occur as secondary sequences at least 0.7 times as often as they appear as primary sequences; (3) discard sequences with an overall relative abundance less than 0.001, when sequences are ordered in rank order. Therefore, sequence counts used in downstream analyses were the number of UMIs associated with each identified sequence (UMI sequence counts), rather than the number of reads for a sequence. MAUI-seq scripts can be found at https://github.com/jpwyoung/MAUI. All raw field trial and greenhouse trial sequencing read data can be found at BioProject PRJNA596932 and BioProject PRJNA934365 respectively.

2.4 | Sequence analysis

For the greenhouse trial, some Rabani and Riesling variety samples did not yield any *recA* or *rpoB* sequence reads, and therefore, it was decided to remove these samples and all samples of their given genotype. This was to enable unbiased genotype comparisons; for example, when looking at differences in diversity between genotypes, there would always be the same number of nodules and pots sampled for each genotype. Therefore, one set of Rabani genotype samples were removed resulting in three Rabani genotypes remaining for analysis. Additionally, all Riesling genotype samples were removed. Therefore, from each variety, either two (Aberboost, Aberconcor, Avalon, Barblanca, Borek, Brianna, Chieftain, Coolfin, Cyma, Iona, Klondike, Liblanc, Liflex, Merida, Rabani, Violin) or three (Aberpearl, Aran, Silvester) genotypes were analysed. Overall, the sequence analysis was carried out on 19 varieties and 41 genotypes. Additionally, of the five *rpoB* sequences identified, the two *rpoB* sequences with the lowest overall relative abundances across samples (seq4 and seq5) did not match to any of the rhizobial inoculant strains, and only matched up to 91% identity to *rpoB* sequences from a collection of 196 *Rlt* genomes (Cavassim et al., 2020). When seq4 and seq5 were further blasted against the NCBI GenBank database, the sequences had 100% identity and 100% coverage to *Agrobacterium*. Seq4 was identified at a very low relative abundance (mean: 0.23%; range: 0%–2.97%) in 53 (64.63%) pots, and seq5 was also found at a very low relative abundance (mean 0.18%; range: 0%–2.48%) in 49 (59.76%) pots. Furthermore, the *rpoB* alleles from all 10 inoculant strains matched one of the three remaining *rpoB* sequences, except for strain SM132 which was not accounted for by any of the five *rpoB* sequences identified by MAUI-seq. Therefore, seq4 and seq5 were likely to be contaminants from the greenhouse which did not nodulate the plants and therefore were removed from the analysis.

The subsequent sequence analysis was carried out the same way for the field and the greenhouse trial data in R (R Core Team, 2020). To enable allele abundance comparison across samples from all genes analysed, UMI sequence counts were converted to relative abundance within each gene for each sample. For the greenhouse trial data, relative abundances for the two within-plot replicates were averaged to produce one set of relative abundances per pot. Sequence presence across clover crosses was displayed with static UpSet plots made using the UpSetR and ComplexUpset packages. For visualisation purposes, to observe relative abundance of allele sequences across samples in a heatmap, relative abundance counts of 0 (occurring when a sequence is not present in a sample

but present in other samples) were converted to one decimal place lower than the lowest relative abundance count and subsequently $\log_{10}$ transformed. Therefore, log transformation produces a negative abundance score, whereby more negative scores denote a lower allele abundance.

To assign a genospecies to each *recA* and *rpoB* allele, BLASTn was used to search for sequences in the genome assemblies of 196 *Rlt* full genome sequenced strains (Cavassim et al., 2020) which are known genospecies strains. Alleles that did not match any of these strains were aligned to the NCBI database using BLASTn (GenBank). BLAST hits were only accepted if they spanned the full length of the query sequence and had 100% sequence identity to known genome assemblies; otherwise, sequences were classed as an 'unassigned genospecies'.

2.5 | *Rhizobium* strain growth phenotype experiments

One hundred and eighty-nine *Rhizobium leguminosarum* symbiovar *trifolii* (*Rlt*) strains were previously isolated from white clover nodules on organic farms and conventional plant-breeding trials across the United Kingdom, France and Denmark (Cavassim et al., 2020). These strains were genetically characterised into genospecies A–E based on core gene phylogeny (Cavassim et al., 2020; Kumar et al., 2015). To test growth differences between genospecies, the strains were inoculated into 200 μL Tryptone Yeast (TY) broth in individual microplate wells using a sterilised metal replicator (around 0.2 μL, Boenik). Plates were incubated at 28°C, and bacterial density (OD_{600}) was measured at 16, 24, 40, 48, 64, 72, 88, 96 h post inoculation using a microplate reader (Tecan infinite 200 plate reader). Each strain was grown in triplicate and the OD_{600} measurements averaged.

Biofilm formation could also be important when rhizobia colonise plant root surfaces before forming nodules. Strains' biofilm formation was measured at the end of the experiment (96 h) by adding 20 μL of 0.1% crystal violet to each plate well, incubating at room temperature for 15 min, then rinsing the wells by submersion in clean water three times, shaking out the water between each submersion. Plates were then dried and 225 μL of absolute ethanol was added to each plate well. The plates were incubated at room temperature for approximately 1 h to dissolve the crystal violet, and then the OD_{600} measurements were taken for each strain. Each strain was grown in triplicate and the OD_{600} measurements averaged. For each strain, the average biofilm OD600 value was then divided by the end time point 96 h average bacterial OD600 value to normalise the data.

2.6 | Statistical analysis

Statistical analyses were carried out the same way for the field trial and the greenhouse trial using R version 4.0.3 (R Core Team, 2020). Allelic similarity between nodule samples was estimated using Bray–Curtis dissimilarity metric, which was calculated for all pairwise sample comparisons with the vegan package using relative abundance count data for all genes. To compare rhizobial diversity between clover crosses and plots (field trial), and between varieties and genotypes (greenhouse trial), non-metric multidimensional scaling (NMDS) was employed on Bray–Curtis dissimilarities using metaMDS in the vegan package. Two dimensions were specified for all NMDS analyses. All NMDS analyses produced a stress score of less than 0.2. Intrinsic sequence variable vectors were fitted with a default of 999 permutations to NMDS coordinates using the env. fit function in the vegan package, to determine which alleles associated significantly with NMDS dimensions. To determine significant differences in rhizobial allele diversity for clover crosses and plots (field trial), and for varieties and genotypes (greenhouse trial), two-way PERMANOVA was undertaken using the adonis vegan function. Additionally, to further identify which clover crosses significantly differed in Bray–Curtis dissimilarity, the adonis. pairwise function with Bonferroni-adjusted *p*-value correction was used from the pairwiseAdonis package. Furthermore, to confirm the PERMANOVA assumption that results were not influenced by dispersion, Bray–Curtis dissimilarity scores were analysed for homogeneity of dispersion (variances) using the betadisper function from the vegan package followed by one-way ANOVA formulated with cross and plot variables (field trial), or varieties and genotypes (greenhouse trial) to test for significance. TukeyHSD post hoc testing from the tidyverse package was additionally implemented to determine which pairwise comparisons displayed significant differences in dispersion. Dispersion tests were found to be not significant unless otherwise indicated.

To identify significant differences in the relative abundance of genospecies across clover samples based on *recA* and *rpoB* sequences, two-way ANOVAs were used. Furthermore, TukeyHSD post hoc testing was used to identify interaction effects between clover samples and relative genospecies abundances.

To determine if similar allelic diversity between samples was associated with geographic distance in the field trial, the Mantel test Pearson's correlation R statistic was calculated between Euclidean geographic distance and allelic (Bray-Curtis) dissimilarity between samples. Geographic distance between sampled plot points was calculated using Euclidean distance based on x–y geographic coordinates. Furthermore, to explain how the variation in *Rlt* allelic abundance was partitioned between clover crosses and the spatial distribution of samples, we used variance partition analysis (vegan package) on the combined relative abundances of all four *Rlt* genes in nodule samples from Block 1. Due to the large proportion of 0's, the relative abundance data were transformed using Hellinger transformation to avoid artificially inflating the similarity between samples due to commonly absent alleles. To determine whether *Rlt* allelic abundances were influenced by nonlinear gradients, we calculated a principal coordinates of neighbour matrices (PCNMs) analysis from the sample spatial distribution data. A subset of significant PCNM axes were chosen for variance partitioning analysis using the ordistep function (vegan package). Individual variance partitioning fractions were tested for significance using permutation test for

redundancy analysis under a reduced model (999 permutations). Additional variance partitioning analyses were also undertaken to examine the allelic abundances of each of the four genes separately.

To determine changes in the rhizobium inoculant community composition during the greenhouse trial, we used Bray–Curtis dissimilarity index to calculate the deviation from the initial genospecies frequency for each plant. This was done using *rpoB* and *recA* alleles separately, using vegdist() from the vegan package, where the degree of deviation (dissimilarity) of the nodule community composition from the initial inoculant community composition indicated the strength of potential host filtering, rhizobia competition or both.

To focus on changes driven by host filtering, we used generalised linear mixed effects models (GLMMs) to test the association between the degree of deviation of the nodule community from the initial inoculant community composition and plant growth traits (growth per day and rate of greenness). In the case of host filtering, positive correlations were expected, where changes in rhizobia nodule community composition would benefit the plants in terms of more efficient symbiosis. To generate GLMMs, we utilised 66 pots with complete data (nodule genospecies' frequencies, growth per day, initial size, hue rate and starting hue values) and some of the plant growth measures were standardised for GLMMs. Firstly, growth per day correlated positively with the initial size of the plant, and hence, the growth per day was corrected with the full effect of initial size using the following equations:

$$y_{gpd} = 1\mu + X_s + e \quad (1)$$

where y_{gpd} reports the observed growth per day values, μ is the intercept, s is the fixed effect of initial size and e is a vector of residuals. X is a design matrix of $n \times 1$ dimension, whereby n is the number of observations with observed initial sizes. In order to fit initial size values as a fixed effect, the lme4 package in R was used (Bates et al., 2015). The values produced by Equation 1 were subsequently used to calculate the growth per day corrected with the full effect of initial size (*gpdCor*):

$$gpdCor = y_{gpd} - Xs - 1\mu \quad (2)$$

where variables and matrices are as reported in Equation 1.

Furthermore, the estimated rate of greenness (hue rate) depended on the starting hue colour of the plant (Tausen et al., 2020). To take this into account, hue rate values were corrected for the full effect of the starting hue in the same way as described for growth per day in Equations 1 and 2.

After these corrections, we produced multiple GLMMs using gsA frequencies, gsB frequencies, gsC frequencies and the deviation from the initial inocula genospecies frequency as response variables (Tables S14–S21). Binomial regression GLMMs were fitted when the response variable was gsA, gsB or gsC frequencies, and was calculated using the glmer() function with a binomial family response from the lme4 package. Beta dispersion regression GLMMs were fitted when the response variable was deviation from the initial inocula genospecies composition, as was calculated using the glmmTMB() function with a beta family response from the glmmTMB package. Growth per day corrected for initial cutting size, and rate of greenness corrected for starting hue colour were included as fixed effects for all models. Clover variety was included as a random effect for all models. The lme4 package utilises general nonlinear optimisation algorithms (optimisers) to calculate the variance estimates of the random effects. The *recA* gsC GLMM failed to converge, and therefore to resolve this, the model was refit using the glmer() function from the lme4 package with the 'nloptwrap' (nonlinear optimisation) optimiser selected as control option flag (Table S21). By using the 'nloptwrap' optimiser instead of the default optimiser a faster alternative solution for model convergence was obtained.

To identify significant differences in competitive growth ability of *Rlt* genospecies, maximum growth rate was determined for 189 *Rlt* strains by calculating the growth rate for each 24 h period in a strains' growth curve and choosing the highest growth rate value across the curve. Kruskal–Wallis test was used to identify if there were significant differences in the maximum growth rate values between genospecies. To determine which genospecies significantly differed in maximum growth rates, Dunn's test for multiple comparisons with Bonferroni *p*-value correction was used. Additionally, to determine significant differences in biofilm formation between genospecies, we first confirmed that there was no colinearity between the 96 h bacterial OD_{600} values and biofilm OD_{600} values for each genospecies. Subsequently, for each strain, the average biofilm OD_{600} value was divided by the 96 h bacterial OD_{600} to normalise the data. To identify significant differences in biofilm formation between genospecies, Kruskal–Wallis test was used on the normalised biofilm values. Dunn's test for multiple comparisons with Bonferroni *p*-value correction was further used to determine genospecies comparisons that significantly differed in biofilm formation. For all strain growth measurements, each strain was grown in triplicate and the OD_{600} measurements averaged before statistical analysis.

R scripts for all statistical analyses and figure generation can be found at https://figshare.com/s/35c51167d97bca241b22.

3 | RESULTS

3.1 | Little variation in *Rhizobium* alleles associating with clover crosses in the field

Using high-throughput amplicon sequencing (Fields et al., 2020), nodule samples from four white clover crosses and one variety (hereon collectively referred to as crosses) grown in conventionally managed plots as part of a larger plant breeding trial (Figure 1) were sequenced for four *Rlt* genes (*rpoB*, *recA*, *nodA* and *nodD*). Per clover cross, six sample replicates were collected across two plots (Figure 1: Block 1) and the number of unique allele sequences identified across all samples was counted for *rpoB*, *recA*, *nodA* and *nodD* genes (Table S2). We identified a greater number of alleles for symbiosis genes than housekeeping genes (Table S2). To determine

whether the variation in the presence of rhizobial alleles was associated with the clover cross host, the total number of unique alleles for each of the five distinct clover crosses across all four genes was counted (Figure 2a). *rpoB* and *recA* allele presence showed little variation across clover crosses (Figure 2b,c). The distribution of *rpoB* alleles was the most homogeneous and all *rpoB* alleles were identified in all clover crosses except for two alleles which were absent from Cross 1 (Figure 2b). Symbiosis genes, *nodA* and *nodD*, displayed more specificity to clover crosses, with some alleles only present in the nodules of a subset of clover crosses (Figure 2d,e). Even so, we identified only one *nodD* allele that was exclusive to a single cross (Figure 2e: Cross 1). Otherwise, all *nodD* and *nodA Rlt* alleles were identified in nodules of at least two clover crosses, and overall, alleles were predominantly found in all clover crosses. Together, these results suggest that clover crosses did not exclusively select for specific *Rlt* alleles based on the four genes tested.

3.2 | *Rhizobium* allele frequencies differ between specific clover crosses in the field

We further investigated the variation in *Rlt* allele frequencies between clover crosses by comparing the relative abundances of *rpoB*, *recA*, *nodA* and *nodD* alleles. The relative abundance of alleles varied between clover crosses for each *Rlt* gene (Figure 3). However, the most abundant alleles were the same across all clover crosses. To study this further, we calculated allelic dissimilarity between all pairwise sample combinations using the Bray–Curtis dissimilarity index based on the combined relative abundances all four *Rlt* genes (Figure 4a), and relative abundances for each gene individually (Figures S1 and S2). When allelic dissimilarity was calculated using all four *Rlt* genes in combination, non-metric multidimensional scaling (NMDS) analysis clustered nodule samples by clover crosses across NMDS coordinate 1 (Figure 4b; all genes PERMANOVA clover cross: $F_{4,29} = 3.7036$, $p < .001$; Table S3). Conversely, samples did not significantly cluster by plot (Table S3). Further post hoc tests identified that Klondike and Cross 2 displayed significant differences in *Rlt* nodule diversity, and that Crosses 1 and 4 were also identified to have significantly different *Rlt* diversities (adjusted $p < .05$). However, the PERMANOVA assumption of homogeneity of dispersion was not upheld between Klondike and Cross 2 (betadisper ANOVA: $F_{4,25} = 3.2518$, $p < .05$; TukeyHSD: adjusted $p < .05$) and this may have slightly influenced the PERMANOVA result. Nevertheless, the overall ordination of nodule samples (Figure 4b) indicates that *Rlt* nodule population diversity was associated with clover cross host when considering all four genes.

Further analyses of each gene revealed that clustering of clover cross samples was predominantly due to differing allele frequencies of housekeeping genes rather than symbiosis genes (Figure S3; Tables S4–S7). Analysis of rpoB diversity showed that Cross 2 displayed significant differences in *Rlt* nodule diversity compared to Klondike, Cross 1 and Cross 4 (adjusted $p < .05$; Figure S3). For *recA* alleles, the variation observed across NMDS coordinate 1 was biased by Klondike samples (Figure S3), where half of the Klondike samples predominantly contained the overall most abundant *recA* allele and all other samples mainly contained the second overall most abundant allele. While we found no significant differences in *recA* diversity between plots (Table S5), *recA* diversity differed between Crosses 1 and 2, which is congruent with the allelic distinction observed for *rpoB* (Table S5; Figure S3; adjusted $p < .05$). Although we identified significant differences in *nodA* and *nodD* nodule diversity between clover crosses (Tables S6 and S7), significance between pairwise comparisons was lost after post hoc testing *p*-value adjustment. To investigate which allele sequences were driving the sample distribution patterns, intrinsic allele sequence variable vectors were fitted to NMDS coordinates for each of the four *Rlt* genes. The frequencies of the overall most abundant alleles were shown to significantly drive the separation of clover crosses (Figure S3). These results suggest that clover crosses were associated with different frequencies of *Rlt* alleles, and that this variation was dependent on both clover cross and the *Rlt* gene of interest. Specifically, *rpoB* and *recA* alleles showed the greatest allelic distinction, particularly between Cross 1 and Cross 2 nodules.

3.3 | *Rhizobium* genospecies frequencies differ between clover crosses in the field

To study whether the differences in *rpoB* and *recA* frequencies corresponded to clover crosses preferentially selecting for specific *Rlt* genospecies, we assigned *rpoB* and *recA* alleles to genospecies A–E (gsA-E) (Kumar et al., 2015) to calculate the relative genospecies frequencies in nodule samples. We could assign nine of 16 *rpoB* sequences (84.38%–100% of *rpoB* total relative abundance in samples) and five of eight *recA* sequences (94.06%–100% of *recA* total relative abundance in samples) to known genospecies using a 196 *Rlt* genomes data set (Cavassim et al., 2020). The remaining alleles did not largely contribute to the overall relative abundances in samples (0%–15.63% of *rpoB*, and 0%–5.94% of *recA*, total relative abundance in samples), and as they could not be assigned to known genospecies with additional GenBank database comparison, they were classified as 'unassigned genospecies'. The genospecies frequencies of *rpoB* and *recA* alleles were strongly positively correlated (Pearson's Correlation: $R = .951$, $t = 41.193$, $df = 178$, $p < .001$). However, a greater number of *recA* alleles could be associated with gsA, whereas no gsA alleles were identified with *rpoB* (Figure 4c). Noticeably, all clover crosses were significantly dominated by either gsB or gsC in their nodules (Figure 4c; Tables S8 and S9). Similarly, a significant interaction was identified between genospecies abundances and clover cross host for both *recA* and *rpoB* alleles (Tables S8 and S9). In particular, the Klondike variety contained a significantly greater percentage of gsC alleles compared to all other crosses (based on *rpoB* and *recA* alleles: adjusted $p < .05$). Additionally, the significant difference in *rpoB* and *recA* diversity that we identified for Cross 2 compared to Cross 1 and Klondike (Figure S3) was found to be associated

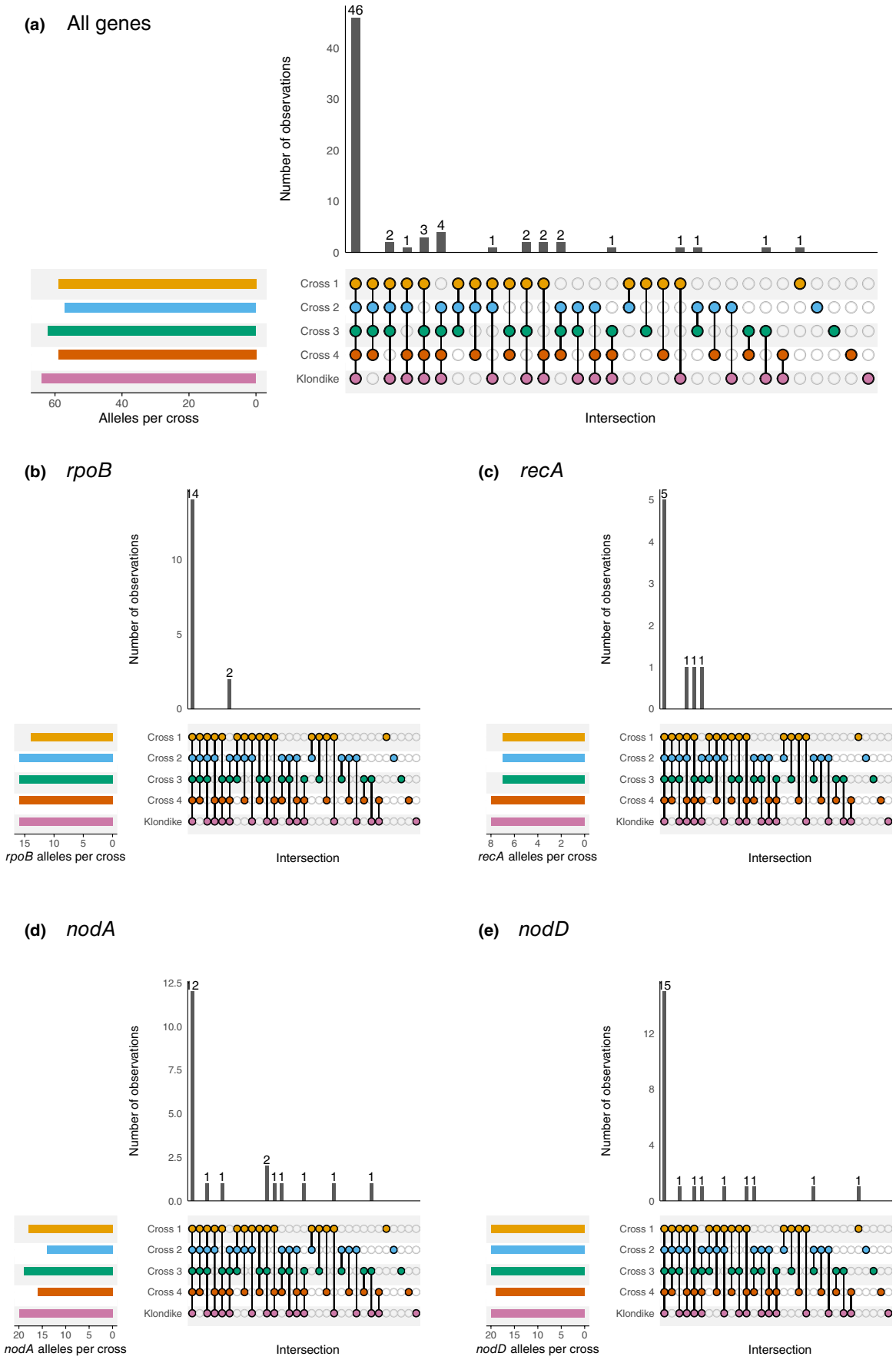


FIGURE 2 Number of unique alleles shared between crosses. Static UpSet plots display the overlap of *Rhizobium leguminosarum* symbiovar *trifolii* alleles identified in each white clover cross for (a) all genes; (b) *rpoB*; (c) *recA*; (d) *nodA* and (e) *nodD* genes. All samples analysed are from Block 1.

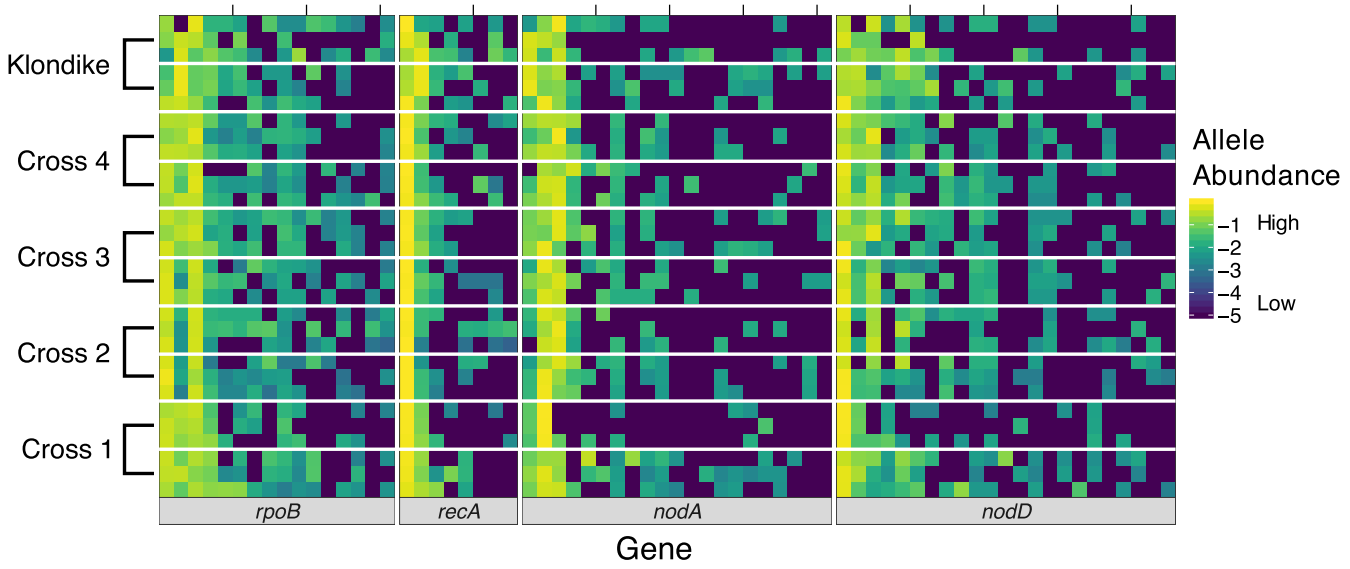


FIGURE 3 Relative abundance of unique alleles identified for four *Rhizobium leguminosarum* symbiovar *trifolii* genes (*rpoB*, *recA*, *nodA* and *nodD*) varied in root nodule populations from five white clover crosses (Cross 1–4, Klondike). Raw UMI sequence counts for each allele were converted to relative abundances (between 0 and 1) and subsequently $\log_{10}$ transformed for visualisation. Log transformation produces a negative abundance score, whereby more negative scores denote a lower allele abundance (yellow = high abundance, blue = low abundance). Clover cross samples were collected across two plots (displayed as separate row sections), and three locations were sampled within each plot (displayed as three rows within each plot section). Only Block 1 samples are shown. Tick marks at the top of the heatmap indicate every five alleles.

with a significantly lower percentage of gsC alleles and greater percentage of gsB alleles in Cross 2 nodules compared to Cross 1 and Klondike (based on *rpoB* alleles: adjusted $p < .05$). In all nodule samples, the frequencies of gsA, gsD and gsE were in low abundance and totalled less than 16% of representative *recA* and *rpoB* alleles. Together, these results show that different clover crosses were associated with distinct rhizobial genospecies, and this was especially clear between Klondike, Cross 1 and Cross 2.

3.4 | Similarity of *Rlt* nodule diversity is associated with both clover cross identity and spatial distribution of plants in the field

To understand how *Rlt* nodule diversity was partitioned between clover crosses and spatial distribution of plants in the field, we carried out variance partition analysis on the combined relative abundances of all four *Rlt* genes. Clover cross and spatial distribution alone explained 4.4% ($p = .075$) and 4.7% ($p < .05$; Table S10) of the total variation in *Rlt* nodule diversity, respectively, while 17.1% of variation was jointly explained by clover cross and spatial distribution of plants (Figure 5a). Furthermore, when variance partitioning analysis was carried out on *rpoB*, *recA*, *nodA* and *nodD* genes individually, clover cross and spatial distribution jointly accounted for the most explained variance (Figure S4; Table S10). However, most of the variation (73.8%) was not explained by the variables included in the models (Figure 5a; Table S10), suggesting that other unidentified abiotic and biotic factors were important for explaining *Rlt* nodule diversity in the field.

To further test the effect of geographic distance for the *Rlt* diversity, we compared the six Klondike nodule samples from Block 1 (Figure 1: Klondike plots a and b) with nodule samples collected from another replicate Klondike plot located within Block 2 which was separated by a larger geographical distance compared to plots within Block 1 (Figure 1: Klondike plot c). We identified that the *Rlt* diversity of the Klondike samples from Block 2 (plot c) was distinctly different to the Klondike samples from Block 1 (plots a and b) (Figure 5b,c). Additionally, allelic dissimilarity calculated from the combined relative abundances of all four genes significantly correlated with geographic distance between nodule samples (Figure 5d; Pearson's correlation $R = .314$, $p < .05$). This indicates that differences in *Rlt* diversity between nodule samples were partially driven by local growth conditions between different blocks, even with the same clover cross host (Figure 5b–d).

3.5 | Clover varieties select for specific rhizobia nodule communities in a controlled greenhouse trial

To reduce the effect of uncontrolled environmental variation typical for field experiments, we quantified the strength of host filtering by sampling potted white clovers from a larger controlled greenhouse trial (Moeskjær et al., 2022; Tausen et al., 2020). Specifically, we inoculated 41 plant genotypes from 19 commercial clover varieties with the same mix of 10 *Rlt* strains (10% each) under controlled nitrogen-limited greenhouse conditions and quantified changes in

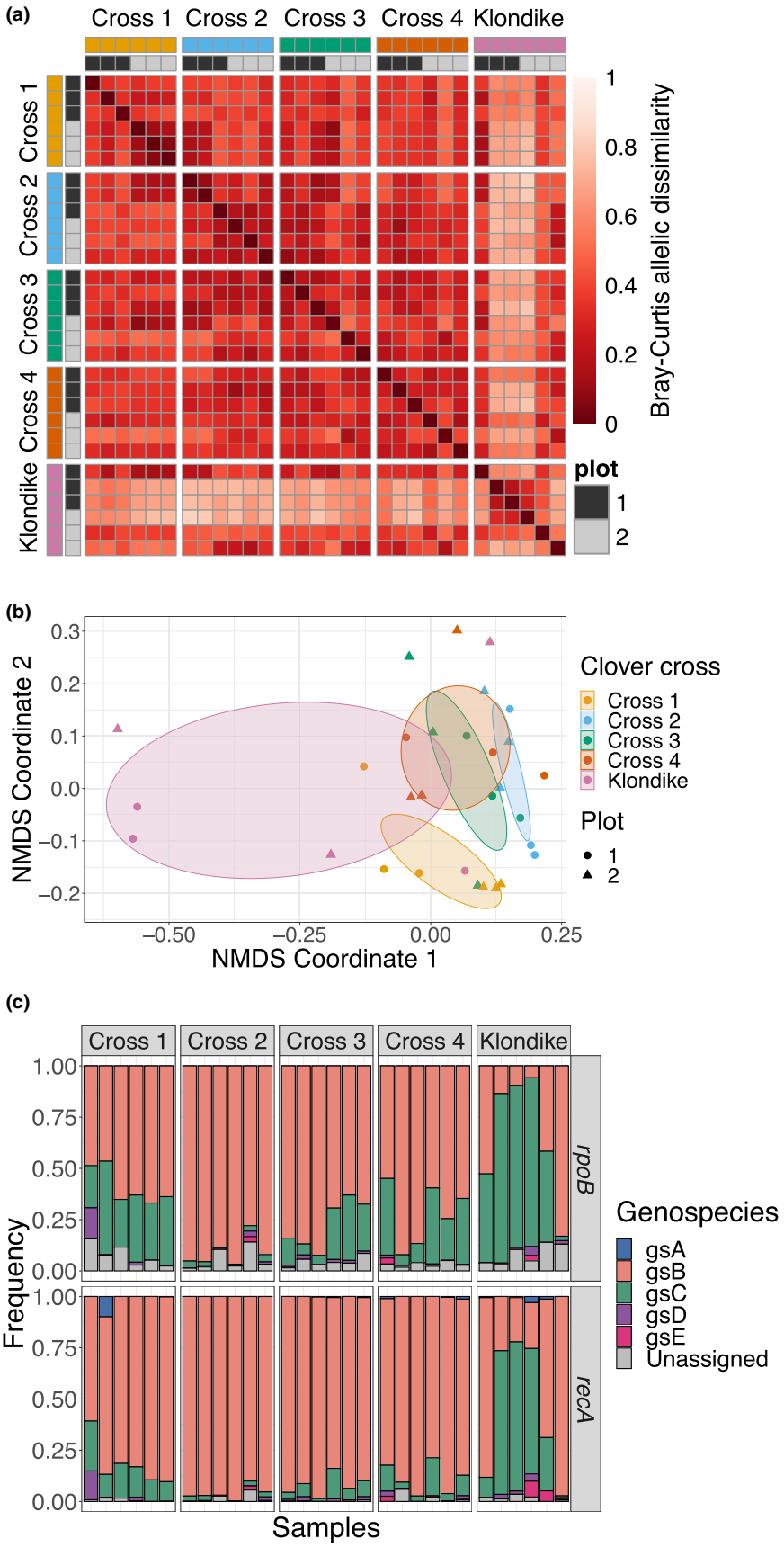


FIGURE 4 The level of observed allelic dissimilarity between clover crosses differed depending on the *Rhizobium leguminosarum* symbiovar *trifolii* (*Rlt*) gene marker. (a) Pairwise allelic dissimilarity of four *Rlt* genes combined (*rpoB*, *recA*, *nodA* and *nodD*) between white clover nodule samples. Bray–Curtis dissimilarity is shown on a scale ranging from low (red) to high (white) allelic dissimilarity. (b) Additionally, non-metric multidimensional scaling (NMDS) analysis of the combined relative abundances of four gene alleles displays the separation of samples based on their allelic dissimilarity. Allelic dissimilarity is displayed across two dimensions, and samples that are closer are more allelically similar. Samples are from Block 1 and are grouped and coloured by their clover cross host, while symbol shapes correspond to different plots. Ellipses show the 95% confidence intervals of the barycenter mean. (c) The relative abundances of *Rhizobium leguminosarum* symbiovar *trifolii* genospecies in the nodules of different clover crosses based on *rpoB* and *recA* alleles. Genospecies A–E allele frequencies are represented by different colours (A = blue, B = orange, C = green, D = purple, E = pink). Allele sequences were assigned to genospecies based on sequence similarity to known genospecies *rpoB* and *recA* sequences (see Section 2). Allele sequences which could not be assigned to a genospecies were labelled as 'Unassigned'. All clover crosses were sampled in replicates of 6 from Block 1.

the *Rlt* allele frequencies using the same molecular methods as in the field trial. The 10 characterised *Rlt* strains included representatives of the three major genospecies and consisted of six gsA, one gsB and three gsC strains which were isolated from white clover nodules sampled from agricultural fields in Denmark, United Kingdom and France (Table S1) (Cavassim et al., 2020). All inoculum strains had

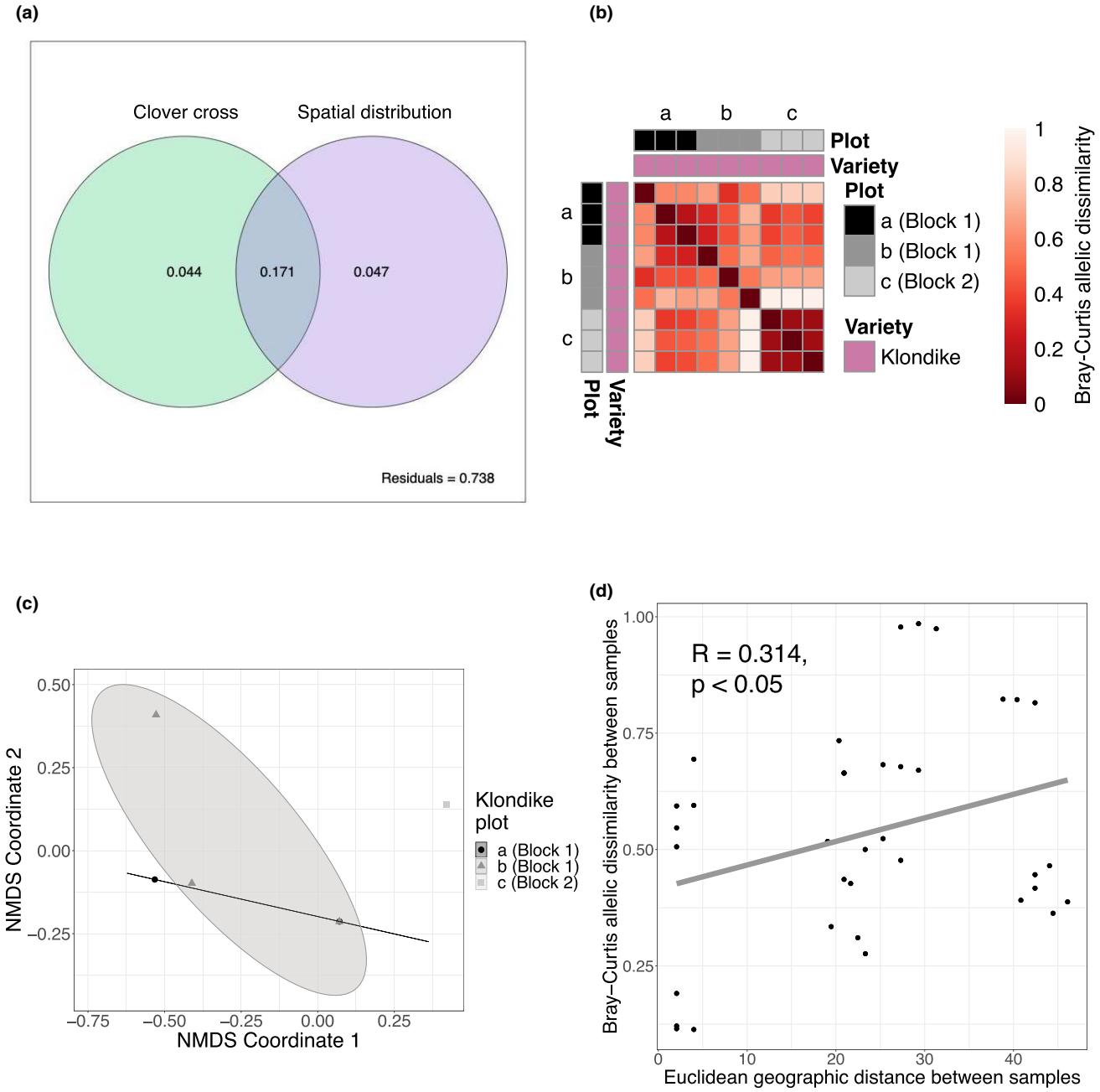


FIGURE 5 Spatial plant distribution was associated with allelic dissimilarity between samples. (a) Variance partitioning on relative allelic abundances of all four *Rhizobium leguminosarum* symbiovar *trifolii* genes together (*rpoB*, *recA*, *nodA*, *nodD*). Variance was partitioned among two predictors: Clover genotype of nodule sample and spatial distribution of samples within the field. (b) Pairwise allelic dissimilarity was calculated between all Klondike samples across three plots from two Blocks. Allelic dissimilarity is shown on a scale ranging from low (red) to high (white) allelic dissimilarity. (c) Non-metric multidimensional scaling (NMDS) of Klondike nodule samples across three plots from two blocks. Allelic dissimilarity for NMDS was calculated using the combined relative abundances of all four genes. Ellipses show the 95% confidence intervals of the barycenter mean. (d) Euclidean geographic distance correlated to allelic dissimilarity (calculated using combined relative abundances of all four genes) between pairwise Klondike sample comparisons across three plots ($R = .314$, $p < .05$).

a *recA* gene distinguishable by >2 nucleotides. From metabarcoding sequencing of greenhouse nodule samples, we identified seven *recA* alleles and three *rpoB* alleles, all of which matched at least one of the original 10 *Rlt* strains inoculated into the pots. From this, we determined that four gsA, one gsB (the most abundant strain across all samples) and two gsC strains were able to form symbiosis with at least nine of 41 genotypes (Figure S5). The presence, and relative abundance, of *recA* alleles differed between clover varieties and between genotypes of the same variety (Figures S5 and S6). Additionally, while all three *rpoB* alleles were present in all 41 genotypes at the end of the greenhouse experiment (Figure S5), the relative abundance of the three *rpoB* alleles varied between different clover varieties and genotypes (Figure S6), indicative of *Rlt* strain selection at the clover genotype level.

To determine the significance of differences in *Rlt* nodule diversity between clover genotypes, NMDS analysis was used (Figure 6a) based on the combined relative frequencies of *rpoB* and *recA* alleles (Figure S7). We observed that some samples were separated by clover variety (Figure 6a; Table S11; PERMANOVA: $F_{18,81}=4.0114$, $R^2=.37587$, $p<.001$), including Aberboost and Aberpearl (NMDS coordinate 1). Overall within varieties, samples were further separated by genotype (Figure 6a; Table S11; PERMANOVA: $F_{22,81}=3.5863$, $R^2=.41071$, $p<.001$), notably with Aberpearl, Merida and Silvester varieties across NMDS coordinate 1. Intrinsic allele sequence variable vectors fitted to NMDS coordinates suggested that the frequencies of almost all alleles significantly drove the separation of clover varieties and genotypes (Figure S8), and overall, clover variety and genotype explained 38% and 41% of the total variance in *Rlt* nodule diversity respectively ($p<.001$; Table S11).

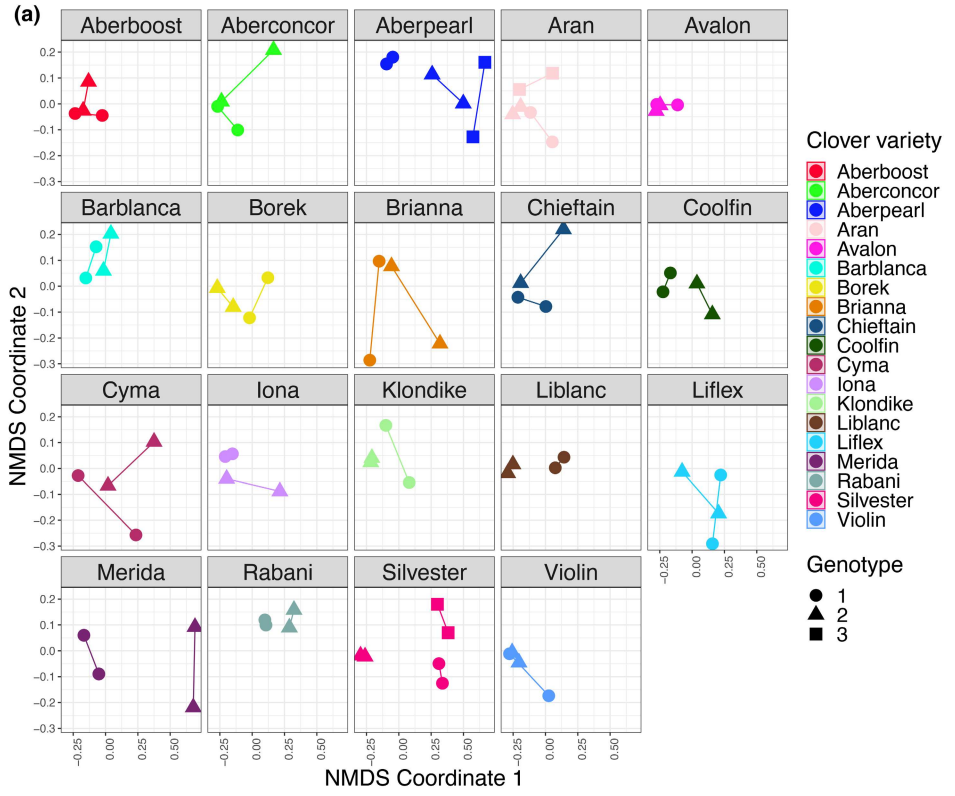
When the relative abundances of *recA* and *rpoB* alleles were grouped by their genospecies classification, gsB was the most abundant despite its lowest relative frequency in the initial inocula, and gsC was the least abundant genospecies across all samples (Figure 6b). The high abundance of gsB observed in the greenhouse experiment is in line with their high abundance observed in the field trial (Figure 4c). As changes in rhizobia genospecies frequencies could be driven by bacterial competition even in the absence of host filtering, we compared the differences in the maximum growth rate and biofilm formation in vitro between 189 *Rlt* gsA-E genospecies strains (Cavassim et al., 2020) present in the larger strain collection, which included the 10 genospecies used in the inoculum. We found that gsB strains had a significantly higher growth rate compared to gsC and gsE (Figure S9; Kruskal–Wallis $X^2=88.831$, $df=4$, $p<.001$), and while gsB strains did not display significantly greater biofilm formation compared to gsC, gsB had higher biofilm formation compared to gsA and gsE (Figure S9; Kruskal–Wallis $X^2=47.513$, $df=4$, $p<.001$). These results suggest that the success of gsB strains could potentially be due to their relatively higher competitive (growth rate) and root colonisation (biofilm) ability.

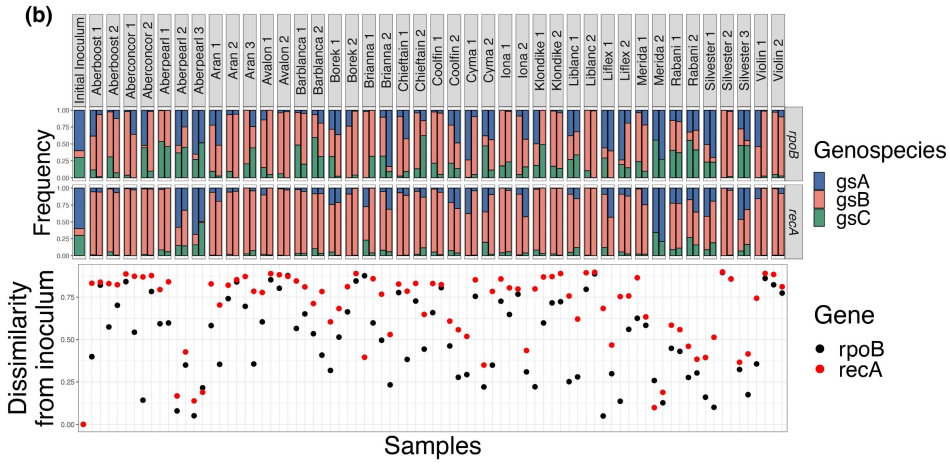
We further compared differences in genospecies community composition between clover varieties and clover genotypes at the end of the greenhouse experiment (Figure 6b, Tables S12 and S13). Surprisingly great variation was found both between and within varieties. For instance, Aberboost nodules contained a greater relative abundance of gsB strains, and lower relative abundance of gsA strains, compared to Aberpearl nodules (Figure 6b; based on *recA* alleles: adjusted $p<.05$). At the genotype level, Merida genotype 1 predominantly selected for gsB alleles whereas Merida genotype 2 mainly selected for gsA alleles (Figure 6b; based on *recA* alleles: adjusted $p<.05$). Moreover, Aberpearl genotypes 2 and 3 selected for a greater abundance of gsA alleles compared to genotype 1 (Figure 6b; based on *recA* alleles: adjusted $p<.05$). In support of our field data, these results suggest that the variation in *Rlt* allele frequencies is significantly associated with both clover variety and clover genotype and not solely explained by relatively higher growth and biofilm formation of gsB strains.

As different clover varieties and genotypes could select for different *Rlt* genospecies, we determined if deviance from the initial inoculum genospecies composition could be positively associated with clover growth traits, including growth per day and rate of change in hue of greenness over time, indicative of increased chlorophyll content and successful symbiotic nitrogen fixation (Figure 6c). In the case of host filtering, positive correlations were expected, where changes in rhizobia community composition would benefit the plants in terms of more efficient symbiosis, regardless if different varieties favour different *Rlt* genospecies. Using generalised linear mixed effects models (Tables S14-S21; Figure 6c), we found a positive relationship between the rate of change in greenness and deviance of the nodule community composition (Tables S14 and S18; Figure 6c). Particularly, the rate of change in greenness was significantly associated with deviations in gsB (*rpoB*) (Table S16) and gsA (*recA*) frequencies (Table S19). However, no significant association between the growth per day and deviance from the initial genospecies composition, or individual genospecies frequencies, was observed (Tables S14-S21). Together, our greenhouse and in vitro data suggest that the efficiency of symbiosis increased when the initial composition of *Rlt* genospecies inoculum changed more during the greenhouse experiment. While the success of gsB strains could also be explained by their potentially higher competitiveness (high growth and biofilm formation), they were not the dominant genospecies with all varieties and genotypes, indicative of active host filtering for compatible symbionts.

4 | DISCUSSION

This study investigated whether distinct clover varieties and genotypes showed differential host filtering for specific rhizobium nodule populations under agricultural and controlled greenhouse conditions. We evaluated *Rlt* nodule populations from five genetically distinct white clover crosses for differences in diversity based on the allele frequencies of two chromosomal and two symbiosis genes (*rpoB*, *recA*, *nodA* and *nodD*) in the field. *Rlt* diversity in samples clustered to some extent by clover cross host, and further analysis identified that different clover crosses were associated with distinct *Rhizobium* genospecies. However, a large proportion of the variance in *Rlt* diversity could not be explained by clover crosses or by sampling distribution in the field, suggesting that the relative importance of symbiont selection by host plants could be lessened under the complex field conditions. To reduce environmental variation, we established a controlled greenhouse trial to test the host filtering between clover varieties and genotypes. We found that *Rlt* nodule diversity significantly differed between both clover varieties and individual genotypes, and this was associated with preferential host filtering of genospecies and relatively higher competitiveness of gsB, which was the most abundant genospecies in both field and greenhouse experiments. Crucially, we observed an increase in the rate of change in clover greenness with increasing deviance of the nodule





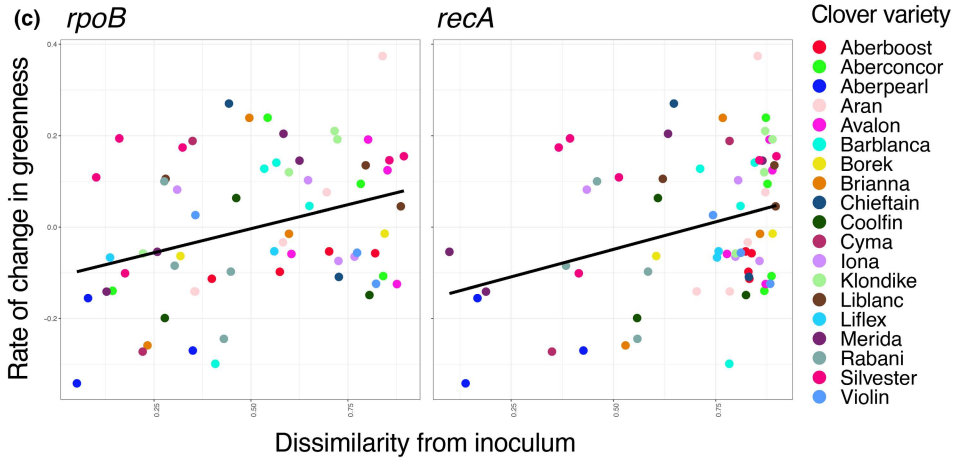


FIGURE 6 Non-metric multidimensional scaling (NMDS) analysis generated from the combined relative abundances of *rpoB* and *recA* alleles displays the separation of greenhouse nodule samples. (a) NMDS analysis displayed across two dimensions, and samples that are closer are more allelically similar. Each point represents one sample, and samples are grouped and shaped by their clover genotype and coloured by their clover variety. While all samples were included in the NMDS analysis, they have been plotted on individual graph panels by their clover variety to clearly display all sample points. (b) The relative abundances of *Rhizobium leguminosarum* symbiovar *trifolii* genospecies in the greenhouse nodule samples of different clover genotypes based on *rpoB* and *recA* alleles. Genospecies A–C allele frequencies are represented by different colours (A = blue, B = orange, C = green). Greenhouse nodule samples are grouped by their clover variety and genotype. Beneath the barchart, the dissimilarity of each sample's nodule community genospecies frequency from the initial inoculum genospecies frequency, calculated as a Bray–Curtis dissimilarity index. Dissimilarity scores determined by *rpoB* and *recA* alleles are indicated as black and red points respectively. (c) The Bray–Curtis dissimilarity from the initial inoculum correlated to the plant's rate of change in greenness over time. Genospecies allele frequencies used to calculate the dissimilarity were determined using *rpoB* and *recA* allele frequencies. The rate of change in greenness is standardised by the starting hue colour of the plant (see Section 2). The black line is fit by linear model. Dots coloured by clover variety.

genospecies community composition from the original inoculum, demonstrating that clover varieties and genotypes selected for compatible symbiont communities that were not always dominated by gsB. However, based on the result of our experiment alone, it is difficult to fully distinguish between the variance in nodule community composition attributed to plant host selection, competition of the rhizobium strains and individual strains' contribution to a plant genotype's performance. Future work could test whether strains inoculated individually and in combination on different host genotypes influence host nodule number, and whether host plant condition potentially affects its choosiness. Together, our results suggest that genotype-by-genotype-by-environment interactions are important for determining successful symbioses with white clover.

In the field trial, *Rlt* diversity and community composition were found to differ between some clover crosses (Figure 4b). However, most of the variation could not be explained by clover host or spatial distribution of field samples indicative of variation in local growth conditions (Figure 5). When we controlled for environmental variation in a controlled greenhouse environment and used a defined *Rlt* community inoculum, clover varieties and distinct clover genotypes within varieties were found to select for different *Rlt* nodule communities (Figure 6). The overall results of this study, where *Rlt* selection varied between clover varieties and between clover genotypes of the same variety, are in line with previous findings (Jones & Hardarson, 1979; Russell & Jones, 1975). Other investigations have similarly found that host genotype significantly influenced rhizobia partner choice and nodule populations in multiple legume species (Bourion et al., 2018; Mytton, 1975; Paffetti et al., 1996; Russell & Jones, 1975; Wadhwa et al., 2011; Xiong et al., 2017). Contrastingly, some studies have not found any associations between cultivar and rhizobium strains (Bromfield, 1984; Buttery et al., 1997; Harrison et al., 1987; McGinn et al., 2016). This suggests that the additional abiotic and biotic complexities of agricultural environments may have lessened any clover host effects in the field, but these could be observable under more controlled conditions akin to our greenhouse experiments where a greater proportion of variation could be explained by clover host than in the field trial. However, as the interaction between legumes and rhizobia is context dependent, we cannot rule out the possibility that plants are more selective when grown under the more favourable greenhouse conditions, and neither can we exclude the possibility that the strains selected by a plant genotype in our greenhouse experiment would continue to be selected by that plant genotype if grown under different environmental conditions.

We further investigated whether the differences in rhizobium nodule diversity between clover hosts under both field and greenhouse conditions could be a result of selection for specific *Rlt* genospecies (Cavassim et al., 2020; Kumar et al., 2015; Young et al., 2021), that vary in their compatibility with distinct clover varieties and genotypes. In the field, genospecies composition differed between clover crosses. Specifically, we found Klondike selected for gsC whereas all other crosses, particularly Cross 2, selected for gsB (Figure 4c). We also found that gsB-dominated samples from Block 1 plots (Figure 4c) and that Block 2 samples were dominated by gsC alleles (Figure 5b,c). The clover preference for gsB is somewhat surprising as gsC has previously been found to be the most prevalent genospecies in clover nodules from Danish soils (Cavassim et al., 2020). Furthermore, nodule samples collected from the same field a year earlier (although from different parts of the same field) were also dominated by gsC alleles (Fields et al., 2020). It is hence possible that composition of rhizobial clover symbionts varies considerably between local microenvironments and years even within the same field. For example, differences in soil type, pH or chemical composition have been associated with diversity and composition of rhizosphere microbial communities in previous studies (Bulgarelli et al., 2012, 2013; Lundberg et al., 2012; Peiffer et al., 2013; Veach et al., 2019; Wang, Cui, et al., 2018; Wang, Liu, & Zhu, 2018), which could determine the initial rhizobial population 'pool' of available genotypes for the plants (Liu et al., 2019; Paffetti et al., 1996; Philippot et al., 2013). Unfortunately, the initial rhizosphere soil populations, and soil physicochemical properties, were not determined in this study. Hence, linking rhizobial genotype distribution with variation in the local growth conditions in the future is required to better understand the biogeographic patterns of symbiont diversity.

In contrast to the field experiment, we found a much stronger signal of host filtering by both clover varieties and genotypes in the controlled greenhouse experiment. For instance, we identified that Aberboost and Aberpearl varieties differed in the relative abundances of gsA and gsB in their nodule communities suggesting differences in host selection (Figure 6b). Furthermore, we also observed

these host-filtering differences between different genotypes of the same variety. For example, we identified that Merida genotype 2 predominantly selected for gsA whereas Merida genotype 1 selected for gsB (Figure 6b). Similar to the field data, gsB was the most abundant genospecies in greenhouse nodule samples. As only one gsB strain was present in the initial inoculum (10% relative abundances), it was able to invade when rare, which could have been due to relatively higher competitive ability compared to other genospecies present in the inoculum. In support of this hypothesis, we found that gsB strains displayed relatively highest maximum growth rates and biofilm formation in TY broth, which could have helped them to establish symbiosis more frequently compared to other genospecies strains (Figure S9). These competitive differences are in line with our previous study where we showed that *R. leguminosarum* genospecies A, C and E display clear differences in competitive fitness and metabolic capacities, despite not differing in their symbiotic performance on white clover (Fields, Moffat, Friman, & Harrison, 2021; Fields, Moffat, Harrison, et al., 2021). However, as gsB strains did not dominate across all varieties and crosses, it is possible that changes in genospecies frequencies were also driven by host preference for compatible symbionts. We identified a positive association between the change in nodule genospecies community composition (deviation from the original inoculum composition) and the rate of change in clover greenness (Figure 6c), indicating that plants which were more selective of rhizobial partners (those with greater deviation from the inoculum) gained more benefit from the symbiosis. In order to directly investigate this, future work could test the benefit gained by each individual strain and compare the benefits gained by the plant. Additionally, this result demonstrates that different plant varieties and genotypes selected for specific symbiont communities based on symbiont compatibility although we cannot exclude that in some cases, the dominance of gsB strains could have been due to their high competitiveness. From an applied perspective, developing inoculants consisting of several *Rlt* genospecies could be an effective strategy to allow plants to select the most compatible symbionts from a diverse pool of rhizobia genotypes.

While sampling distribution also partly explained *Rlt* diversity in the field, most of the variation was left unexplained, potentially due to uncharacterised environmental variables. To address this, it is important in the future to compare variation in the initial rhizosphere *Rlt* population and overall microbiome composition between clover samples (Hibbing et al., 2010) and link this with potential variation in abiotic soil properties (Harrison et al., 1989; Paffetti et al., 1996; Wang, Cui, et al., 2018; Wang, Liu, & Zhu, 2018). Furthermore, clovers were sampled in October in this study and the soil microbial community composition at this time might differ from that in the summer months, after nitrogen fertilisation, or at earlier plant growth stages (Inceoğlu et al., 2010). However, it was previously observed that the same pool of rhizobium genotypes dominated sample populations regardless of the time of year (Duodu et al., 2006), which suggests that the sampling time might have been a lesser problem. Additionally, the genetic similarity between the clover crosses and the extent to which they differed in traits associated with host-filtering mechanisms (Jones et al., 2019) were not determined in this study. It would hence be valuable to investigate whether similarity of nodule symbiont communities correlates with host genotype relatedness in future work. Moreover, the level of observed *Rlt* diversity within and between host crosses was shown to be dependent on the gene markers (Figure 4b; Figure S3). As a result, potential differences in *Rlt* nodule diversity between clovers could have been missed in our investigation because of marker gene choice. We suggest that an area for future work could utilise metagenomics to sequence a larger number of symbiosis-related genes which could provide a higher level of resolution and pinpoint specific genetic determinants of partner choice.

Investigating the extent to which partner choice is advantageous in mutualistic symbioses is critical to aiding our understanding of the evolutionary dynamics and maintenance of symbiosis (Simms & Taylor, 2002). Here, we found that *Rlt* nodule population diversity was observable between clover varieties, and between distinct genotypes within clover varieties, especially in the controlled greenhouse experiment. As several molecular selection processes are involved in *Rlt* x clover genotype compatibility and partner choice, future work should investigate the mechanisms of symbiotic selectivity, such as the importance of extracellular polysaccharides, secretion system specificity and microbe-associated molecular pattern detection by the legume host to consider how these factors may affect the success of both host and symbiont (Perret et al., 2000; Simms & Taylor, 2002; Wang, Liu, & Zhu, 2018). Our previous work has also shown that different *Rlt* genospecies can be highly competitive through resource competition and direct antagonism likely via bacteriocins and quorum-sensing associated mechanisms (Fields, Moffat, Harrison, et al., 2021), which could also partly explain the dominance of gsB strains. Finally, it has previously been shown that legume–rhizobium symbiotic compatibility can evolve rapidly over five plant generations (Batstone et al., 2020). Our findings support this indirectly by showing that the composition of compatible symbiont communities can vary between different plant varieties and genotypes already within one plant generation. The genotype-by-genotype specificity between plant hosts and their symbionts is hence likely to evolve rapidly in both greenhouse and field conditions.

AUTHOR CONTRIBUTIONS

BF, SM and SUA were involved in conceptualisation; BF, SM, SUA, JPWY and VPF were involved in methodology; BF, SM, WJD and EKM were involved in sample collection and performing experiments; BF was involved in data analysis; BF was involved in visualisation; BF and VPF were involved in writing—original draft; all authors were involved in writing—review and editing; SUA, JPWY and VPF were involved in supervision; SUA and VPF were involved in funding acquisition.

ACKNOWLEDGEMENTS

The authors would like to thank DLF for access to clover field trial plots, and the University of York Technology Facility for sequencing. This project was undertaken on the Viking Cluster, which is a high-performance compute facility provided by the University of York.

We are grateful for computational support from the University of York High Performance Computing service, Viking and the Research Computing team.

FUNDING INFORMATION

This work was funded by Innovation Fund Denmark (4105-00007A, SUA). VPF is funded by the Royal Society (RSG\R1\180213 and CHL\R1\180031) and jointly by a grant from UKRI, Defra and the Scottish Government, under the Strategic Priorities Fund Plant Bacterial Diseases programme (BB/T010606/1) at the University of York. WJD is funded by the Daphne Jackson Trust.

CONFLICT OF INTEREST STATEMENT

DLF has developed and marketed the cultivars analysed in this study.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [https://figshare.com/s/35c51167d97bca241b22 and https://doi.org/10.6084/m9.figshare.22069010].

DATA AVAILABILITY STATEMENT

All raw data files, metadata and scripts used to generate analyses can be found at the following FigShare link: https://figshare.com/s/35c51167d97bca241b22 and https://doi.org/10.6084/m9.figshare.22069010. Raw sequence reads are deposited in the SRA (BioProject PRJNA596932 and BioProject PRJNA934365).

BENEFIT-SHARING STATEMENT

Benefits Generated: A research consortium collaboration was established between academic researchers in the United Kingdom and Denmark, and industrial collaborators in Denmark, of which all are included as co-authors. The results of our collaboration have been shared on publicly available repositories for access by the broader scientific community. Our consortium partnership is committed to further developing international academic and industrial research collaborations to deliver solutions for improving global crop and soil health.

ORCID

Bryden Fields https://orcid.org/0000-0003-1251-6867
Sara Moeskjær https://orcid.org/0000-0003-4987-0130

REFERENCES

Argaw, A., & Muleta, D. (2017). Effect of genotypes-Rhizobium-environment interaction on nodulation and productivity of common bean (*Phaseolus vulgaris* L.) in eastern Ethiopia. *Environmental Systems Research*, *6*(14), 1–16. https://doi.org/10.1186/s40068-017-0091-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batstone, R. T., O'Brien, A. M., Harrison, T. L., & Frederickson, M. E. (2020). Experimental evolution makes microbes more cooperative with their local host genotype. *Science*, *370*(6515), 476–478.

Bourion, V., Heulin-Gotty, K., Aubert, V., Tisseyre, P., Chabert-Martinello, M., Pervent, M., Delaitre, C., Vile, D., Siol, M., Duc, G., Brunel, B., Burstin, J., & Lepetit, M. (2018). Co-inoculation of a pea core-collection with diverse rhizobial strains shows competitiveness for nodulation and efficiency of nitrogen fixation are distinct traits in the interaction. *Frontiers in Plant Science*, *8*, 2249. https://doi.org/10.3389/fpls.2017.02249

Brewin, N., Wood, E., & Young, J. P. W. (1983). Contribution of the Symbiotic Plasmid to the Competitiveness of *Rhizobium leguminosarum*. *Journal of General Microbiology*, *12*, 2973–2977.

Bromfield, E. S. P. (1984). Variation in Preference for *Rhizobium meliloti* Within and Between *Medicago sativa* Cultivars Grown in Soil. *Applied and Environmental Microbiology*, *48*(6), 1231–1236.

Bromfield, E. S. P., Barran, L. R., & Wheatcroft, R. (1995). Relative genetic structure of a population of *Rhizobium meliloti* isolated directly from soil and from nodules of alfalfa (*Medicago sativa*) and sweet clover (*Melilotus alba*). *Molecular Ecology*, *4*(2), 183–188. https://doi.org/10.1111/j.1365-294X.1995.tb00207.x

Bulgarelli, D., Rott, M., Schlaeppi, K., Loren, V., van Themaat, E., Ahmadinejad, N., & Assenza, F. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature*, *488*, 91–95. https://doi.org/10.1038/nature11336

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Loren, V., van Themaat, E., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology*, *64*, 807–838. https://doi.org/10.1146/annurev-arplant-050312-120106

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*, *115*(10), 2425–2430. https://doi.org/10.1073/pnas.1714246115

Burns, J. H., Anacker, B. L., Strauss, S. Y., & Burke, D. J. (2015). Soil microbial community variation correlates most strongly with plant species identity, followed by soil chemistry, spatial location and plant genus. *The Open-Access Journal for Plant Sciences*, *5*, plv030.

Buttery, B., Park, S., & van Berkum, P. (1997). Effects of common bean (*Phaseolus vulgaris* L.) cultivar and rhizobium strain on plant growth, seed yield and nitrogen content. *Canadian Journal of Plant Science*, *77*, 347–351.

Carelli, M., Gnocchi, S., Fancelli, S., Mengoni, A., Paffetti, D., Scotti, C., & Bazzicalupo, M. (2000). Genetic diversity and dynamics of *Sinorhizobium meliloti* populations nodulating different alfalfa cultivars in Italian soils. *Applied and Environmental Microbiology*, *66*(11), 4785–4789. https://doi.org/10.1128/aem.66.11.4785-4789.2000

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, *88*, 677–684. https://doi.org/10.2307/2657068

Cavassim, M. I. A., Moeskjaer, S., Moslemi, C., Fields, B., Bachmann, A., Vilhjálmsson, B. J., Schierup, M. H., Young, J. P. W., & Andersen, S. U. (2020). Symbiosis genes show a unique pattern of introgression and selection within a *Rhizobium leguminosarum* species complex. *Microbial Genomics*, *6*(4), e000351. https://doi.org/10.1099/mgen.0.000351

Clúa, J., Roda, C., Zanetti, M. E., & Blanco, F. A. (2018). Compatibility between Legumes and Rhizobia for the Establishment of a Successful Nitrogen-Fixing Symbiosis. *Genes*, *9*(125), 125.

Debellé, F., Plazanet, C., Roche, P., Pujol, C., Savagnac, A., Rosenberg, C., Promé, J. C., & Dénarié, J. (1996). The NodA proteins of *Rhizobium meliloti* and *Rhizobium tropici* specify the N-acylation of Nod factors by different fatty acids. *Molecular Microbiology*, *22*(2), 303–314. https://doi.org/10.1046/j.1365-2958.1996.00069.x

Duodu, S., Nsiah, E. K., Bhuvaneswari, T. V., & Svenning, M. M. (2006). Genetic diversity of a natural population of *Rhizobium leguminosarum* biovar trifolii analysed from field nodules and by a plant infection technique. *Soil Biology and Biochemistry*, *38*(5), 1162–1165. https://doi.org/10.1016/j.soilbio.2005.07.015

Fields, B., Moeskjær, S., Friman, V.-P., Andersen, S. U., & Young, J. P. W. (2020). MAUI-seq: Metabarcoding using amplicons with unique molecular identifiers to improve error correction. *Molecular Ecology Resources*., *21*(3), 703–720. https://doi.org/10.1111/1755-0998.13294

Fields, B., Moffat, E. K., Friman, V.-P., & Harrison, E. (2021). The impact of intra-specific diversity in the Rhizobia-legume symbiosis. *Microbiology*, *167*(4), 001051. https://doi.org/10.1099/mic.0.001051

Fields, B., Moffat, E. K., Harrison, E., Andersen, S. U., Young, J. P. W., & Friman, V.-P. (2021). Genetic variation is associated with differences in facilitative and competitive interactions in the *Rhizobium leguminosarum* species complex. *Environmental Microbiology*, *24*(8), 3463–3485. https://doi.org/10.1111/1462-2920.15720

Fitzpatrick, C. R., Copeland, J., Wang, P. W., Guttman, D. S., Kotanen, P. M., & Johnson, M. T. J. (2018). Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences*, *115*(6), E1157–E1165. https://doi.org/10.1073/pnas.1717617115

Harrison, S. P., Jones, D. G., & Young, J. P. W. (1989). Rhizobium population genetics: Genetic variation within and between populations from diverse locations. *Microbiology*, *135*(5), 1061–1069.

Harrison, S. P., Young, J. P. W., & Jones, D. G. (1987). Rhizobium population genetics: Effect of clover variety and inoculum dilution on the genetic diversity sampled from natural populations. *Plant and Soil*, *103*, 147–150.

Hassan, S., & Mathesius, U. (2012). The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany*, *63*(9), 3429–3444. https://doi.org/10.1093/jxb/err430

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*, 15–25.

Igolkina, A. A., Bazykin, G. A., Chizhevskaya, E. P., Provorov, N. A., & Andronov, E. E. (2019). Matching population diversity of rhizobial nodA and legume NFR5 genes in plant–microbe symbiosis. *Ecology and Evolution*, *9*(18), 10377–10386.

Inceoğlu, Ö., Salles, J. F., van Overbeek, L., & van Elsas, J. D. (2010). Effects of plant genotype and growth stage on the betaproteobacterial communities associated with different potato cultivars in two fields. *Applied and Environmental Microbiology*, *76*(11), 3675–3684. https://doi.org/10.1128/AEM.00040-10

Jones, D. G., & Hardarson, G. (1979). Variation within and between white clover varieties in their preference for strains of Rhizobium trifolii. *Annals of Applied Biology*, *92*(2), 221–228. https://doi.org/10.1111/j.1744-7348.1979.tb03867.x

Jones, P., Garcia, B. J., Furches, A., Tuskan, G. A., & Jacobson, D. (2019). Plant host-associated mechanisms for microbial selection. *Frontiers in Plant Science*, *10*, 862. https://doi.org/10.3389/fpls.2019.00862

Kiers, T. E., & Denison, R. F. (2008). Sanctions, Cooperation, and the Stability of Plant-Rhizosphere Mutualisms. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 215–236. https://doi.org/10.1146/annurev.ecolsys.39.110707.173423

Kumar, N. L., 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 Biology*, *5*(1), 140133. https://doi.org/10.1098/rsob.140133

Laguerre, G., Louvrier, P., Allard, M.-R., & Amarger, N. (2003). Compatibility of rhizobial genotypes within natural populations of *Rhizobium leguminosarum* biovar viciae for nodulation of host legumes. *Applied and Environmental Microbiology*, *69*(4), 2276–2283.

Laguerre, G., Mavingui, P., Allard, M.-R., Charnay, M.-P., Louvrier, P., Mazurier, S.-I., Rigottier-Gois, L., & Amarger, N. (1996). Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: Application to *Rhizobium leguminosarum* and Its Different Biovars. *Applied and Environmental Microbiology*, *62*(6), 2029–2036.

Liu, F., Hewezi, T., Lebeis, S. L., Pantalone, V., Grewal, P. S., & Staton, M. E. (2019). Soil indigenous microbiome and plant genotypes cooperatively modify soybean rhizosphere microbiome assembly. *BMC Microbiology*, *19*(1), 201.

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., Glavina del Rio, T., Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Green Tringe, S., & Dangl, J. L. (2012). Defining the core Arabidopsis thaliana root microbiome. *Nature*, *488*, 86–90. https://doi.org/10.1038/nature11237

Lupwayi, N. Z., Clayton, G. W., & Rice, W. A. (2006). Rhizobial inoculants for legume crops. *Journal of Crop Improvement*, *15*(2), 289–321. https://doi.org/10.1300/J411v15n02_09

Maj, D., Wielbo, J., Marek-Kozaczuk, M., & Skorupska, A. (2010). Response to flavonoids as a factor influencing competitiveness and symbiotic activity of *Rhizobium leguminosarum*. *Microbiological Research*, *165*(1), 50–60. https://doi.org/10.1016/j.micres.2008.06.002

McGinn, K. J., van der Putten, W. H., Duncan, R. P., Shelby, N., Weser, C., & Hulme, P. E. (2016). Trifolium species associate with a similar richness of soil-borne mutualists in their introduced and native ranges. *Journal of Biogeography*, *43*(5), 944–954.

Moeskjær, S., Kiel Skovbjerg, C., Tausen, M., Wind, R., Roulund, N., Janss, L., & Andersen, S. U. (2022). Major effect loci for plant size before onset of nitrogen fixation allow accurate prediction of yield in white clover. *Theoretical and Applied Genetics*, *135*, 125–143. https://doi.org/10.1007/s00122-021-03955-3

Mytton, L. R. (1975). Plant genotype x rhizobium strain interactions in white clover. *Annals of Applied Biology*, *80*, 103–107.

Paffetti, D., Scotti, C., Gnocchi, S., Fancelli, S., & Bazzicalupo, M. (1996). Genetic diversity of an Italian *Rhizobium meliloti* population from different *Medicago sativa* varieties. *Applied and environmental microbiology*, *62*(7), 2279–2285.

Peiffer, J. A., Spor, A., Koren, O., & Ley, R. E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences*, *110*, 6548–6553. https://doi.org/10.1073/pnas.1302837110

Perret, X., Staehelin, C., & Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews*, *64*(1), 180–201.

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*, 789–799.

Poinsot, V., Crook, M. B., Erdn, S., Maillet, F., Bascaules, A., & Jean-Michel, A. (2016). New insights into Nod factor biosynthesis: Analyses of chitooligomers and lipo-chitooligomers of *Rhizobium* sp. IRBG74 mutants. *Carbohydrate Research*, *434*, 83–93. https://doi.org/10.1016/j.carres.2016.08.001

Quiza, L., St-Arnaud, M., & Yergeau, E. (2015). Harnessing phytomicrobiome signaling for rhizosphere microbiome engineering. *Frontiers in Plant Science*, *6*, 507. https://doi.org/10.3389/fpls.2015.00507

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rangin, C., Brunel, B., Cleyet-Marel, J.-C., Perrineau, M.-M., & Béna, G. (2008). Effects of Medicago truncatula genetic diversity, rhizobial competition, and strain effectiveness on the diversity of a natural Sinorhizobium species community. *Applied and Environmental Microbiology*, *74*(18), 5653–5661. https://doi.org/10.1128/AEM.01107-08

Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., & Rolfe, B. G. (1986). Flavones induce expression of nodulation genes in Rhizobium. *Nature*, *323*(6089), 632–635.

Ritsema, T., Wijfjes, A. H., Lugtenberg, B. J., & Spaink, H. P. (1996). *Rhizobium* nodulation protein NodA is a host-specific determinant of the transfer of fatty acids in Nod factor biosynthesis. *Molecular and General Genetics*, *251*, 44–51.

Roughley, R. J., Blowes, W. M., & Hurridge, D. F. (1976). Nodulation of *Trifolium subterraneum* by introduced rhizobia in competition with naturalized strains. *Soil Biology and Biochemistry*, *8*(5), 403–407. https://doi.org/10.1016/0038-0717(76)90041-9

Russell, P. E., & Jones, D. G. (1975). Variation in the selection of Rhizobium trifolii by varieties of red and white clover. *Soil Biology and Biochemistry*, *7*(1), 15–18. https://doi.org/10.1016/0038-0717(75)90024-3

Schmid, M. W., Hahl, T., van Moorsel, S. J., Wagg, C., De Deyn, G. B., & Schmid, B. (2019). Feedbacks of plant identity and diversity on the diversity and community composition of rhizosphere microbiomes from a long-term biodiversity experiment. *Molecular Ecology*, *28*(4), 863–878. https://doi.org/10.1111/mec.14987

Simms, E. L., & Taylor, D. L. (2002). Partner choice in Nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative Biology*, *42*(2), 369–380. https://doi.org/10.1093/icb/42.2.369

Sloger, C. (1969). Symbiotic effectiveness and N 2 fixation in nodulated soybean. *Plant Physiology*, *44*, 1666–1668. https://doi.org/10.1104/pp.44.12.1666

Tausen, M., Clausen, M., Moeskjær, S., Shihavuddin, A. S. M., Dahl, A. B., Janss, L., & Andersen, S. U. (2020). Greenotyper: Image-based plant phenotyping using distributed computing and deep learning. *Frontiers in Plant Science*, *11*, 1181. https://doi.org/10.3389/fpls.2020.01181

Veach, A. M., Morris, R., Yip, D. Z., Yang, Z. K., Engle, N. L., Cregger, M. A., Tschaplinski, T. J., & Schadt, C. W. (2019). Rhizosphere microbiomes diverge among Populus trichocarpa plant-host genotypes and chemotypes, but it depends on soil origin. *Microbiome*, *7*(1), 76. https://doi.org/10.1186/s40168-019-0668-8

Wadhwa, K., Dudeja, S. S., & Yadav, R. K. (2011). Molecular diversity of native rhizobia trapped by five field pea genotypes in Indian soils. *Journal of Basic Microbiology*, *51*(1), 89–97. https://doi.org/10.1002/jobm.201000065

Wang, D., Yang, S., Tang, F., & Zhu, H. (2012). Symbiosis specificity in the legume: rhizobial mutualism. *Cellular Microbiology*, *14*(3), 334–342. https://doi.org/10.1111/j.1462-5822.2011.01736.x

Wang, Q., Liu, J., & Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Frontiers in Plant Science*, *9*, 313. https://doi.org/10.3389/fpls.2018.00313

Wang, X. L., Cui, W. J., Feng, X. Y., Zhong, Z. M., Li, Y., Chen, W. X., Chen, W. F., Shao, X. M., & Fu Tian, C. F. (2018). Rhizobia inhabiting nodules and rhizosphere soils of alfalfa: A strong selection of facultative microsymbionts. *Soil Biology and Biochemistry*, *116*, 340–350. https://doi.org/10.1016/j.soilbio.2017.10.033

Wardell, G. E., Hynes, M. F., Young, J. P. W., & Harrison, E. (2021). Why are rhizobial symbiosis genes mobile? *Philosophical Transactions of the Royal Society London B*, *377*(1842), 20200471. https://doi.org/10.1098/rstb.2020.0471

Wernegreen, J. J., & Riley, M. A. (1999). Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: A case for the genetic coherence of rhizobial lineages. *Molecular Biology and Evolution*, *16*(1), 98–113. https://doi.org/10.1093/oxfordjournals.molbev.a026041

Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P. S., & Turnbull, L. A. (2017). Policing the legume-Rhizobium symbiosis: A critical test of partner choice. *Scientific Reports*, *7*, 1419.

Xiong, H. Y., Zhang, X. X., Guo, H. J., Ji, Y. Y., Li, Y., Wang, X. L., Zhao, W., Mo, F. Y., Chen, J. C., Yang, T., Zong, Y., Chen, W. X., & Tian, C. F. (2017). The epidemicity of facultative microsymbionts in faba bean rhizosphere soils. *Soil Biology and Biochemistry*, *115*, 243–252.

Yang, C., Bueckert, R., Schoenau, J., Diederichsen, A., Zakeri, H., & Warkentin, T. (2017). Symbiosis of selected *Rhizobium leguminosarum* bv. viciae strains with diverse pea genotypes: Effects on biological nitrogen fixation. *Canadian journal of microbiology*, *63*, 909–919. https://doi.org/10.1139/cjm-2017-0281

Young, J. P. 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., Félix, J. D. F., González, R. R., Youseif, S. H., Lepetit, M., Boivin, S., … Tian, C.-F. (2021). Defining the *Rhizobium leguminosarum* Species Complex. *Genes*, *12*(1), 111. https://doi.org/10.3390/genes12010111

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. *Environmental Microbiology*, *3*(6), 363–370. https://doi.org/10.1046/j.1462-2920.2001.00202.x

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., Cho, H., Karaoz, U., Loqué, D., Bowen, B. P., Firestone, M. K., Northen, T. R., & Brodie, E. L. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology*, *3*, 470–480. https://doi.org/10.1038/s41564-018-0129-3

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fields, B., Moeskjær, S., Deakin, W. J., Moffat, E. K., Roulund, N., Andersen, S. U., Young, J. P. W., & Friman, V.-P. (2023). *Rhizobium* nodule diversity and composition are influenced by clover host selection and local growth conditions. *Molecular Ecology*, *00*, XX–19. https://doi.org/10.1111/mec.17028