Discovering the phenotypic variation within *Rhizobium leguminosarum* and determining the best strains for soil inocula

EMMA K. SMITH

B.Sc (Hons)

M.Sc. by Research

University of York

Biology

January 2018

Abstract

Biological nitrogen fixation (BNF) with rhizobia is an environmentally friendly, sustainable alternative to chemical fertilisers. Despite many legumes having the ability to form relationships with a wide range of indigenous soil bacteria, the efficiency of the symbiosis is altered greatly by the combination of partners. Additionally, choosing favourable rhizobia strains for use as soil inocula requires consideration of both the climate and soil conditions. Traditionally, the efficiency of the symbiosis is tested for its versatility under a variety of abiotic and biotic stressors with plant biomass measurements. Consequently, I will measure the ability of free-living rhizobia to survive and grow in a number of conditions associated with the rhizosphere.

Rhizobium leguminosarum bv. *trifolii* (Rlt) forms a symbiotic relationship with *Trifolium repens* (white clover). 192 strains of Rlt were used in the study, isolated in the UK, France and Denmark (both conventional and organic farms) and grouped into genospecies (gsA, gsB, gsC, gsD and gsE) based on their core genome. The strains were cultured onto modified liquid Tryptone-Yeast (TY) media and bacterial density (OD_{600}) was measured at various timepoints post inoculation. When ranking the strains based on their highest bacterial density across experiments, the UK strains were the most consistently high scoring but were significantly impacted by a decrease in nutrient availability. Therefore, based on the experiments conducted, the gsA strains would be the best candidates for soil inoculation across the broadest range of conditions.

Table of Contents

Abstract 2
List of Tables 5
List of Figures 6
Acknowledgements 7
Declaration by author 8
1 Introduction 9
1.1 Symbiosis of rhizobia and leguminous plants 10
1.2 Surviving and competing in the rhizosphere 11
1.2.1 Abiotic stressors 11
1.2.2 Biotic stressors 12
1.2.3 Colonisation ability 12
1.3 Project Outline 12
1.3.1 Background 12
1.3.2 Project Aims and Hypotheses 13
2 Materials and Methods 15
2.1 Microbial strains used in this study 15
2.2 Media 15
2.3 Preparation of working stocks for microbiological measurements 16
2.4 General procedure measuring bacterial density (OD_{600}) 16
2.4.1 Manipulation of nutrient concentration 17
2.4.2 Incubation temperature 17
2.4.3 Media pH 17
2.4.4 Protist predation 17
2.4.5 Biofilm Formation 18
2.5 Metabolism of carbon sources 18
2.6 Statistical Analysis 18
2.6.1 Effect of abiotic and biotic stressors on Rlt growth 19
2.6.2 Biofilm formation 19
2.6.3 Metabolism of carbon sources 19
2.6.4 Ranking the strains 20
3 Results 21
3.1 Effect of abiotic and biotic environment on *Rhizobium* growth and biofilm formation 21
3.2 Nutrient concentration 21
3.2.1 Biofilm Formation 23
3.3 Temperature 25
3.3.1 Biofilm Formation 28
3.4 pH 30
3.4.1 Biofilm Formation 32
3.5 Protists 35
3.6 Principal Component Analysis (PCA) 37
3.7 EcoPlates 39
3.7.1 Average Well Colour Development (AWCD) 40
4 Ranking the strains 41
5 Discussion 43
5.1 Nutrient concentration 43
5.2 Temperature 43
5.3 pH 44
5.4 Biofilm formation 45

5.5 Protist predation 45
5.6 Metabolism of carbon sources 45
5.7 Ranking the best strains 46
5.8 Future research 47
Appendix 48
A.1 Nutrient concentration 48
A.2 Temperature 49
A.3 pH 51
A.4 Protist predation 52
A.5 Biofilm Formation 53
B.1 Principal component analysis 54
C.1 Ranking the strains 57
Abbreviations 61
References 62

List of Tables

Table 1 Where the *Rhizobium leguminosarum* strains were isolated……………………………………… 14

Table 2 Media ingredients………………………………………………………………………………………… 17

Table 3 pH of the soil the Rlt strains were isolated from………………………………………………… 31

Table 4 Statistical summary of the ANCOVA conducted for biofilm formation data……………….. 54

Table 5 Loading for each experiment to the four PCA components……………………………………….. 56

Table 6 Ranking the Rlt strains based on growth…………………………………………………………… 58

List of Figures

Fig. 1	Increasing the concentration of nutrient media increases the rhizobia growth rate……..	23
Fig. 2	gsB/UK strains were inhibited the most at low nutrient concentrations………………………	24
Fig. 3	At nutrient concentrations below 100% TY biofilm formation is unaffected between groups………	25
Fig. 4	Growth is significantly decreased at temperatures below 20°C…………………………………..	26
Fig. 5	The effect of temperature on growth in high and low nutrient conditions…………………	28
Fig. 6	Biofilm formation is not significantly higher for a single group of strains incubated at various temperatures……………………………………………………………………………………………	30
Fig. 7	Soil pH where *Rhizobium leguminosarum* strains were isolated…………………………………	31
Fig. 8	The strains only grew within the range of pH 5 to 8, with the highest growth at pH 6….	32
Fig. 9	The gsB/UK strains has the significantly higher growth at pH 5, 6, and 8………………………	34
Fig. 10	Biofilm formation does not occur at pH 9 or 10 and is the highest at pH 6…………………	35
Fig. 11	Rhizobia growth is statistically decreased with the addition of protist predators, *Tetrahymena pyriformis* and *Chilomonas paramecium*………………………………………………	36
Fig. 12	Overall growth of the gsB/UK strains was least affected by the addition of protist predators………	37
Fig. 13	No separation of the strains was observed from PCA of the growth datasets………………	39
Fig. 14	Metabolism of 31 common carbon sources found in the soil……………………………………...	40
Fig. 15	Average well colour development (AWCD) for genospecies and geographic origins……..	41
Fig. 16	Ranking the strains based on bacterial density (OD600) for each treatment…………………	43
	Appendix	
Fig. 17	Effect of nutrient concentration over time on bacterial density…………………………………	49
Fig. 18	Effect of temperature over time on bacterial density in high nutrient media………………	50
Fig. 19	Effect of temperature over time on bacterial density in low nutrient media……………….	51
Fig. 20	Effect of pH over time on bacterial density………………………………………………………………..	52
Fig. 21	Effect of protist predation over time on bacterial density…………………………………………	53
Fig. 22	Scree plot indicating four components should be extracted………………………………………	55

Acknowledgements

I want to thank my supervisor, Ville-Petri Friman, for giving me the opportunity to do this project at the University of York under his supervision. I want to thank the NCHAIN consortium in general, and specific member who gave me guidance. Additionally, I want to acknowledge the work done previously by NCHAIN. This project would not have been possible without someone securing the funding, setting up a collaboration between institutes and then isolating and sequencing the strains.

Many people helped me get this finished and it wasn't an easy ride. In particular, I want to thank Liam Moffat, Bryden Fields, Claire Brown, Kate Orlandi-Fantini, Angela Smith and John Smith.

Declaration by author

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1 Introduction

An environmentally friendly, sustainable alternative to using chemical fertilisers is the use of Biological Nitrogen Fixation (BNF) [1–3]. BNF represents the introduction of nitrogen fixing microorganisms to the soil. Rhizobia, a group of gram-negative soil bacteria, are an example of a BNF microorganism which form a symbiotic relationship with leguminous plants. The rhizobia convert atmospheric nitrogen (N_2) into ammonia (NH_3) for use by the plant and the plant provides a carbon source and habitat for the bacteria in root nodules.

Beneficial effects of rhizobia to legumes is evident not only for the individual plant [4–8], but also for the health of the soil and surrounding environment [9,10]. Lambers *et al.,* (2009) even suggested that the association between plant and microbiota ultimately contributed to paedogenesis, the formation of soil in evolutionary history [11].

In comparison to inorganic nitrogen fertilisers, BNF by rhizobia is considered to be an inexpensive and environmentally friendly alternative [12–14]. Hassen *et al.,* (2014) estimated the current use of BNF technologies worldwide results in a saving of US $8 billion per year in fertiliser [15]. Thereby, reducing the output of greenhouse gases from the energy expensive Haber-Bosch process and the movement of farming vehicles across the fields [16]. Additionally, it reduces the amount of agricultural run-off leaching into watercourses causing eutrophication of lakes and contributes to marine “dead zones” [2,17].

The facilitation of organic nitrogen uptake by nitrogen fixing bacteria, biofertilisation, has been shown to increase yields [18,19], a highly marketable trait in agriculture. In a sterile soil environment, Biro *et al.,* (2000) found that all treatments with rhizobia increased growth and nutrient uptake of alfalfa [20]. However, Argaw & Minalku (2015) found that only a combination treatment of both rhizobia inoculum and low levels of nitrogen fertiliser would have a positive effect on common bean grain yield due to the inhibitory effect on rhizobia at high nitrogen concentrations [21]. Therefore, BNF may not completely eradicate the need for inorganic fertilisers.

The rhizobia inoculum is generally applied to the legume seeds before sowing but can be applied in larger concentrations directly to the soil. When choosing candidate strains to market as versatile soil inoculums it is important to know, first, the efficiency of nitrogen fixation by a particular rhizobia-legume combination varies enormously. It is estimated that 90% of all inoculants have minimal effects on nodulation and no effect on legume productivity [22,23] and domesticated crops tend to have fewer compatible symbionts [24].

Second, that symbiosis can also be established with indigenous soil bacteria [22,24,25]. Therefore, the strain chosen must be able to outcompete indigenous rhizobia which are adapted to the local environment. It is not often considered the inoculum's tolerance to the environmental stressors or its ability to compete with rhizobia and other microorganisms associated with the soil. This research aims to evaluate the phenotypic diversity within one *Rhizobium* species and propose which strains are best suited to surviving in the largest range of environmental conditions. Improving our understanding and applying precision farming techniques will move us closer to sustainable food security [26–30].

1.1 Symbiosis of rhizobia and leguminous plants

The establishment, maintenance and breakdown of mutualistic interactions is well studied [11,31–36] and the legume-rhizobia relationship is the most studied plant-microbial interaction [32]. Symbiotic, by definition, because of the long-term reciprocal benefit acquired by both organisms in the relationship [31]. In the presence of available fixed nitrogen rhizobia can survive free-living in the soil, however, in the absence of fixed nitrogen they interact with the roots or stems of the leguminous plants symbiotically [37].

In 2017, a list of 176 species spread across 15 genera was published, validated using molecular markers [38], up from 118 described species in 2012 [37]. This shows not only the diversity of rhizobia [39] but also how fast rhizobia research is moving as their agronomic importance is realised. The species *Rhizobium leguminosarum* bv. *trifolii* (Rlt) can be further broken down by means of genetic similarity into five genospecies based on the strain's core genome [40]. The core genome is mostly chromosomal and is shared with all members of that genospecies. The remaining components of the Rlt genome, make up the accessory genome, consisting of DNA located on the mobile genetic elements such as plasmids [41].

The legume-rhizobia symbiosis can be considered to be highly specific; each rhizobia strain interacts with only a specific group of legumes [25]. The relationship can be compared to the specificity of a host-pathogen interaction with the added complexity of the host's immune system not being triggered by friend verses foe [25].

Often symbiosis involves an exchange of goods or services, which results in the acquisition of novel adaptation by at least one partner. Rhizobia fix inorganic nitrogen within root nodules, modified structures on the legume roots. Initiation of nodule development involves molecular recognition between both partners [42]. Flavonoid molecules exuded by plant roots induce expression of

nodulation (Nod) genes by the rhizobia leading to the synthesis of Nod factors [42]. These Nod factors are recognised by the plant, which initiates the production of the root nodule. A root hair curls to trap a small number of bacteria and begins an inverse tip growth forming an infection thread [43]. The infection thread is the site where the bacteria continuously divide whilst the nodule forms around them. Once inside the root nodules the rhizobia undergo a physiological change from their free-living state into "bacteroides" [37]. Only now can the rhizobia begin to convert atmospheric nitrogen into ammonia, freeing the host from its dependence on exogenous sources of nitrogen [44].

Various studies have concentrated on rhizobia genetics to try to identify the genes, which are responsible or essential, to form a symbiosis [45–47]. For example, Nod genes are needed to form nodules with legumes and are important but may not be essential for the symbiosis to take place [42]. Nitrogen fixation (Nif) genes are another important class of genes that are required for successful fixation of nitrogen for the plant. However, it is less clear to what extent variation in other *Rhizobium* genes affect its success to form symbiosis with host plants. One often neglected area is *Rhizobium* survival in the soil and rhizosphere before the actual establishment of symbiosis. For example, many genes and phenotypic traits could be important for resisting abiotic and biotic stress and hence indirectly important for the success of symbiosis.

1.2 Surviving and competing in the rhizosphere

The rhizosphere is defined as the layer of soil influenced by the roots [48]. It is typically richer across all major components of the microbial community including bacteria, fungi, protozoa and archaea, than the surrounding bulk soil as they all benefit from the secretion of metabolites by the plant roots [6,19]. To be able to form beneficial interactions with plant roots, the rhizobia must be able to survive abiotic factors associated with the rhizosphere and bulk soil and compete with other soil microbes to colonise the roots [6].

1.2.1 Abiotic stressors

A number of abiotic factors influence the rhizobia and rhizosphere microbiota: climate (e.g. temperature [49–52] and drought [53]), soil (e.g. nutrient availability [54], salinity [51,55–57], soil type [58] and pH [49,50,59–61]), geography (e.g. latitude [62] and elevation [63]) and farming practices (e.g. fertiliser application [21,64–66], land use and crop management [67,68]). pH is considered the biggest influencer on community composition [59]. However, the phenotypic variation within one species of rhizobia across a number of abiotic stressors has yet to be discovered. Furthermore, studies often consider only the effect of stressors on the symbiosis including few rhizobia isolates.

1.2.2 Biotic stressors

The surrounding rhizosphere community and plant material must also be considered in order to choose the most effective inocula. A number of reviews have summarised our knowledge of the interactions within soil communities but there are still substantial knowledge gaps [49,69–71]. However, there are a number of the challenges in applying biotic stressors experimentally. First, choosing the correct organism to test as the soil is incredibly diverse and in addition to the microorganisms, the community is influenced by the plants [72,73] and mycorrhizal fungi [71]. Second, bacterial communities are phylogenetically clustered [63] which is experimentally difficult to simulate, especially in homogenous liquid media. Finally, inocula must be able to outcompete indigenous rhizobia and survive predation to produce the desired benefits during symbiosis.

1.2.3 Colonisation ability

Bacteria adhere to surfaces in multicellular assemblies known as biofilms [74]. However, to exert their beneficial effects bacteria must first colonise the root surface [6]. Lugtenberg & Kamilova (2009) highlight the importance of understanding root colonisation saying the bacteria must be "rhizosphere competent". Some of the major colonisation traits in tomato appeared to be motility, adhesion to the root and high growth rate in root exudate [6]. Additionally, enhanced root colonising ability is important for the biocontrol of soil-borne diseases by competing for plant exudates and habitat on the root [6,75]. When rhizobia strains were inoculated in competition with the efficient root coloniser *P. fluorescens* strain WCS365, many were outcompeted [76].

1.3 Project Outline

1.3.1 Background

This project will concentrate on *Rhizobium leguminosarum* bv. *Trifolii* (Rlt), which forms a symbiotic relationship with *Trifolium repens* (white clover). The NCHAIN consortium, established at Aarhus University, Denmark, is using this model system to find the best-suited Rlt strains for clover inoculation. They aim to learn more about the Rlt-white clover interaction, and work with industrial partners (DLF Trifolium and Legume Technology Ltd) to develop the most effective clover-grass seed mixtures. Clover-grass leys are used across Europe for cow pasture so this collaboration will have direct agricultural implications.

One of the aims of NCHAIN is to find the best Rlt-white clover genotype match to increase yields as much as possible. It is additionally important to know the stress tolerance of the free-living rhizobia (pre-symbiosis) before implementing it across Europe. Therefore, this project will focus on the ability of the rhizobia to grow in conditions associated with the rhizosphere and the phenotypic variation within Rlt.

192 strains of Rlt were isolated from fields across four locations: UK, France, Denmark conventional farms (DK_C) and Denmark organic farms (DK_O) by NCHAIN (Table 1). Their core and accessory genomes were sequenced by Illumina MySeq outlined in Moeskjær (2017) and then categorised into five distinct genospecies (gsA, gsB, gsC, gsD and gsE) [40,77]. This classification is based on the average nucleotide identity of genes shared amongst all chromatid bearing bacteria, these genes make up the core genome [77].

The collection of strains being analysed are both genetically distinct and geographically disparate, other than the UK strains, which are exclusively gsB (Table 1). Despite this several questions can be asked, “How much of the phenotypic variation between strains is due to their core genome or local adaptation?” and “Are strains from some locations or genospecies more phenotypically robust to survive a range of stressors?”

Table 1 Where the Rhizobium leguminosarum strains were isolated. Broken down into their respective genospecies and geographic origins.

Geographic Origin	Genospecies (gs)					**Total**
	A	B	C	D	E	
United Kingdom (UK)	-	33	-	-	-	**33**
France (F)	1	-	39	-	-	**40**
Denmark conventional farms (DK_C)	5	1	26	4	-	**36**
Denmark organic farms (DK_O)	38	1	39	1	4	**83**
Total	**44**	**35**	**104**	**5**	**4**	**192**

1.3.2 Project Aims and Hypotheses

Through the use of high-throughput environmental manipulation experiments in liquid growth media, I aim to discover the phenotypic diversity within the Rlt species and to find out which strains, from specific locations or genospecies, have the highest growth overall. Then finally, determine which strains should be used as soil inocula based on their ability to withstand a wide range of environmental stressors. This formula could be used to find strains that are specialised to

environmental conditions relevant to specific fields. However, industrially a generalist strain will be more commercially viable.

The majority of the experiments will include individually inoculating the 192 Rlt strains into modified liquid Tryptone-Yeast (TY) media and incubation for 96 hours. At regular intervals during this time, optical density (OD_{600}) will be measured indicating bacterial density of the individual strain. Growth conditions will be manipulated in multiple ways per experiment (treatments), deviating from the optimal laboratory conditions rhizobia are traditionally grown in (100% TY media, 28°C) [78]. At the end of these experiments biofilm formation will be measured as a proxy for colonisation ability.

2 Materials and Methods

2.1 Microbial strains used in this study

All 192 strains of *Rhizobium leguminosarum* bv. *trifolii* (Rlt) were isolated by NCHAIN members at Aarhus University, Denmark, using methods outlined in Moeskjær (2017). They were isolated from the nodules of white clover, *Trifolium repens*, at four main locations referred to as geographic origins: Didbrook, United Kingdom (51.973399, -1.93297); Rennes, France (48.1986, -1.731017); Store Heddinge, Denmark (55.298109, 12.416619); and Jutland, Denmark (various locations). Store Heddinge represents a conventional farm and Jutland represents organic farms in Denmark. The strains each have identification numbers, corresponding to where they were isolated.

All of the strains were sequenced and assembled by NCHAIN members at Aarhus University, Denmark [77]. The strains were then allocated to one of five genospecies (gsA, gsB, gsC, gsD and gsE), based on their core genome [40].

To explore *Rhizobium* strains' ability to resist predation in the soil, protist cultures, *Tetrahymena pyriformis* (*Tetrahymena*) ciliate (CCAP #1630/1W) and *Chilomonas paramecium* (*Chilomonas*) flagellate (CCAP #977/2A), were ordered from the Culture Collection for Algae and Protozoa (CCAP). All selected protist species were originally isolated from aquatic environments. All protist species were cultured based on Friman *et al.,* (2016) [79].

2.2 Media

Single concentration TY media (Table 2) was used to grow the strains for cryopreservation and in all the growth measurement experiments. Single concentrate TY media was diluted for the manipulation nutrient concentration experiment, adding sterile deionised water to make up the various concentrations of TY. The pH of unaltered single concentration TY media (Table 2) is 6.65. To manipulate acidity, Hydrochloric acid (HCl) was added resulting in pH 4, 5 and 6. To manipulate alkalinity, Sodium Hydroxide (NaOH) was added resulting in pH 7, 8, 9 and 10.

Cryopreserve media (Table 2) was used during the preparation of working stocks. PBS buffer (Table 2) was used to replace the TY media used to 'bulk up' the strains before inoculation onto the 96 well EcoPlates (BIOLOG). EcoPlates are modified microplates, which contain three replicates of 31 commonly found carbon sources, plus a control well containing water [80].

Table 2 Media ingredients

Ingredient	Single Concentrate TY broth	Cryopreserve Media (Double concentrate TY : 80% Glycerol)	PBS Buffer
Sterile deionised water	1000 ml	600 ml	1000 ml
Tryptone	5.0 g	5.0 g	-
Yeast extract	2.5 g	2.5 g	-
$CaCl_2$	1.47 g	1.47 g	-
Glycerol	-	400 ml	-
KH_2PO_4	-	-	0.24 g
NaCl	-	-	8.0 g
KCl	-	-	0.2 g
Na_2HPO_4	-	-	1.42 g

2.3 Preparation of working stocks for microbiological measurements

The strains were cryopreserved at -80°C in replicate microplates which were used for each of the experiments and discarded after use to avoid the freeze-thaw affecting growth. The replicated microplate stocks contained 192 Rlt strains across two 96-well microplates.

The strains were grown in single concentrate TY media (Table 2), incubated in a rotary shaker (180 rpm, 28°C, 48 hours) and then centrifuged to form a pellet (3000 rpm, 6 min). The supernatant was removed and replaced with an equal amount of cryopreserve media (Table 2). The strains were left for 2 hours at room temperature (approx. 20°C), before being placed in the -80°C freezer.

2.4 General procedure measuring bacterial density (OD_{600})

A basic protocol for measuring the growth of the bacteria was adapted several times to include a number of different conditions. The protocol involved the growth of 192 strains in individual microplate wells containing 200µl of TY media and bacterial density (OD_{600}) was measured at a number of timepoints using a microplate reader (Tecan infinite 200 plate reader). The strains were inoculated into the media using a sterilised metal replicator (around 0.2 microL, Boenik). The plates were placed into grip seal bags with wet tissue paper as a source of moisture and incubated at 28°C. Each strain was grown in triplicate for each treatment and then the OD_{600} measurements averaged. To remove condensation before each measurement, the microplates were placed in a laminar flow with the lids ajar for approximately 10 mins. The last measurements were always at 96 hours post inoculation (hpi). Directly afterwards 20µl of 0.1% crystal violet was added to each well to measure biofilm formation.

2.4.1 Manipulation of nutrient concentration

Nutrients available in the soil are broken down into primary (e.g. Nitrogen, Phosphorous and Potassium), secondary (e.g. Calcium, Magnesium and Sulphur) and micronutrients (e.g. Iron, Zinc etc.). These can be measured using a number of techniques including Near-Infrared Reflectance Spectroscopy and digestion and distillation. However, the nutrient composition of the soil where the strains were isolated was not determined in the initial study by NCHAIN.

Therefore, to approximate the strain's tolerance to low nutrient conditions a frequently used liquid growth media, Tryptone-Yeast, was diluted with distilled water to various concentrations (100, 25, 12.5 and 6.25%). OD_{600} was measured at 16, 24, 40, 48, 64, 72, 88 and 96 hpi. Measurements for replicate one (SM3 to SM107), at 72 hours, were removed from analysis due to an error in data handling.

2.4.2 Incubation temperature

A number of temperatures were chosen as treatments to attempt to imitate rhizosphere temperatures in natural conditions in Denmark, where the inocula are likely to be implemented first. To quantify the effect of temperature on growth, whilst comparing the effect of nutrient concentration, the strains were grown in 100% and 6.25% TY in the following temperatures: 4°C (fridge), 10°C (incubator), 15°C (incubator), 20°C (room temperature) and 28°C (incubator). OD_{600} was measured at 16, 24, 40, 48, 64, 72, 88 and 96 hpi.

2.4.3 Media pH

pH strongly influences the community structure of the rhizosphere [54,58,72,81]. Additionally, knowing the range of pH the rhizobia can grow in has agricultural implications when choosing where the inocula can be used. pH was measured in both acidic and alkaline conditions (pH 4 to pH 10), based on known *Rhizobium* pH tolerances [78]. OD_{600} was measured at 16, 24, 48, 72 and 96 hpi.

2.4.4 Protist predation

To explore how the strains differed in their ability to withstand predation by common soil protists, I measured bacterial density in the presence of *Tetrahymena* and *Chilomonas*. The protist cultures were centrifuged (2.0 G, 5 min, 4°C), the supernatant removed and replaced with an equal amount

of sterile deionised water. 10 µl of this protist suspension was then added to each microplate well before inoculating with the Rlt strains. OD_{600} was measured at 16, 24, 40, 48, 64, 72, 88 and 96 hpi.

2.4.5 Biofilm Formation

Biofilm are communities of microbes attached to surfaces [82], which could be an important ability to have when attempting to colonise the root surface before forming nodules. Biofilm formation was measured at the end of experiments (96 hpi): nutrient concentration, incubation temperature and media pH. This was accomplished by adding 20µl of crystal violet to each well, waiting 15 minutes, then rinsing the wells by submersion in clean water three times, shaking out the water between each submersion. The microplates were then dried and subsequently 225µl of absolute ethanol added. This was left for approximately one hour to dissolve the crystal violet, and then the OD_{600} was measured.

2.5 Metabolism of carbon sources

EcoPlates (BIOLOG) were used to determine the strains' ability to metabolise a range of carbon resources associated with the rhizosphere. Due to cost restrictions, each 96 well EcoPlate was used for three strains, meaning there were no replicates per strain.

The strains were individually grown in 10 ml of 100% TY media for 48 hours (28°C, 180rpm). They were then centrifuged (3000rpm, 28°C, 10min), the supernatant removed and replaced with an equal amount of PBS buffer (Table 2) and then left for 2 hours at room temperature. 120 µl of the cell suspension was added to each of the 31 carbon sources plus an additional well of water. The EcoPlates were placed into grip seal bags, with a piece of wet tissue paper, and incubated at 28°C. Optical densities at 590nm were measured at 24, 48, 72 hpi. Development of a purple colouration (reduction of a tetrazolium dye included in the carbon source) occurred when the strains were able to metabolise the carbon source and begin to respire.

2.6 Statistical Analysis

Graphs were made using SPSS statistics 24 and R- 3.4.3. Data is mean ± standard error, unless stated otherwise. Strains from Denmark conventional farms are often referred to as DK_C in figures, similarly Denmark organic farms are DK_O. When conducting general linear models, Levene's test was occasionally violated but these tests were still conducted, as ANOVA are considered robust [83].

2.6.1 Effect of abiotic and biotic stressors on Rlt growth

For each growth experiment, first, a one-way ANOVA (or non-parametric Kruskal-Wallis test) with post hoc analysis was conducted to determine the optimal conditions for growth for Rlt. Second, two multivariate analysis of variance (MANOVA) were used to analysis the variation between strains from specific genospecies and then geographic origins. MANOVA were used whether the data was parametric or non-parametric, as is known not to be very sensitive to violations of normality.

Additionally, descriptive statistics were calculated using the pH of the soil the strains were isolated from to give an idea of local adaptation (NCHAIN data, unpublished).

Finally, principal component analysis was used to compare the strains across multiple experiments: nutrient concentrations (100, 25, 12.5 and 6.25% TY), temperatures at high (100% TY) and low nutrients (6.25% TY)(4, 10, 15, 20 and 28°C), pH (4, 5, 6, 7, 8, 9 and 10) and protist predation (control, *Tetrahymena* and *Chilomonas*). All OD600 measurements 40 hpi and after were included, as there was no variation between the strains before 40 hpi. The strains will be grouped by their genospecies allocation and geographic origin to determine phenotypic variation through clustering.

2.6.2 Biofilm formation

One-way ANCOVA were conducted per treatment for genospecies and geographic origin adjusted with the bacterial density (OD_{600}) at 96 hpi. The data was adjusted to ensure the results were not just showing the same pattern as the bacterial growth experiments and should now show biofilm formation relative to bacterial density. After adjustment for the bacterial density at 96 hpi, there was a statistically significant different between the biofilm formation of genospecies and geographic origins for all treatments (statistical values summarised in Table 7, Appendix).

2.6.3 Metabolism of carbon sources

First, the raw OD_{590} measurements were standardised for analysis by dividing the value by the control well containing water, to account for any TY media transferred with the strains to the wells. Then the genospecies and geographic origins were compared for each substrate to look for significant differences in carbon source utilisation between the strains.

The average well colour development (AWCD) was calculated using the raw OD_{590} measurements of each EcoPlate well at 72 hpi, in relation to the control [80].

$$AWCD = \frac{\sum(S - C)}{n}$$

Equation 1. S is the OD_{590} value of the substrate well, C is the OD590 value of the control well and n is the number of substrates (i.e. 31)80. Analysing the AWCD of each strain will give an indication of their ability, as a generalist, to utilise a range of carbon sources.

2.6.4 Ranking the strains

The strains were ranked based on their OD_{600} value at 48 hpi per growth experiment, per treatment. The experiments included were nutrient availability (100, 25, 12.5 and 6.25%), temperature in high and low nutrients (15, 20 and 28°C), pH (5, 6, 7 and 8) and protist predation (control, *Tetrahymena* and *Chilomonas*). Each OD_{600} value was scaled as a percentage of the highest OD_{600} measurement for that treatment (Equation 2). These percentages per treatment were then added together and ranked highest to lowest, the highest value strain was then made to be 100% and all other were scaled as a percentage of that value (Equation 3). No statistical analysis was performed on these values as they were created purely for data visualisation purposes.

$$Generalist\ score = \sum\left(\frac{x}{highest\ x} \times 100\right)$$

Equation 2. The generalist score is the sum of all dependent variables changed during the abiotic and biotic experiments described in section 8.4. A high generalist score indicates the strain would be highly suited to a range of environmental conditions. χ = OD_{600} at 48 hpi. “highest χ” = the highest OD_{600} for that treatment.

$$Mean\ fitness\ relative\ to\ best\ strain = \frac{Generalist\ score}{Highest\ Generalist\ score} \times 100$$

Equation 3. Mean fitness relative to best strain is the percentage similarity to the strain with the highest generalist score, as calculated in Equation 2. “Best” is used as to describe the strain which has the highest bacterial density across the widest range of environmental parameters, based only on the experiments conducted in this study.

3 Results

3.1 Effect of abiotic and biotic environment on *Rhizobium* growth and biofilm formation

The objective of these experiments was to find strains which would be able to survive and compete, in the largest range of environmental conditions and thereby be able to be used across Europe as soil inocula for white clover. All experiments were performed in liquid media as it was a high-throughput methodology that could be find a subset of “elite” strains for future *in planta* experimentation. Additionally, a methodology could be developed using these parameters to find new soil inocula for other rhizobia-legume combinations.

A high bacterial density (OD_{600}), also referred to as “growth” indicates the strain is able to survive and replicate within the parameters of the experiment. Increased growth could allow the strain to colonise the rhizosphere and root surface more readily or allow it to outcompete indigenous conspecifics. The strains will be summarised as a whole and grouped based on genospecies allocation, based on their genetics, and the location of their isolation (geographic origin), indicating local adaptation.

Biofilm formation was measured at the end of each 96-hour growth experiment using crystal violet staining. For each experiment, biofilm formation was standardised during analysis by incorporating the bacterial density endpoint (96 hpi) as a covariate. This was to ensure the changes in biofilm formation between groups of strains was due to aggregates of biofilm, not the amount of cells in the well. Biofilm formation is being used here as a proxy for colonisation ability of the strain, as the biofilm allows the enclosed cells to adhere to the root surface [44].

3.2 Nutrient concentration

The rhizosphere and particularly the surrounding bulk soil is nutrient poor in comparison to liquid growth media [6]. The traditional laboratory media used to grow rhizobia, Tryptone-yeast, was dilulted to account for the decrease in soil nutrients of the soil. However, it is not possible to confirm that the nutrient concentration or composition of the media correlates with natural soil conditions. Therefore, what is considered here to be “low” nutrients is relative only to the standard 100% TY media.

A one-way ANOVA with Tukey post hoc analysis at 64 hpi confirmed a statistically significant increase in bacterial density as nutrient concentration increased from 6.25% to 100% TY ($p < 0.001$; 100% = 0.607 ± 0.075; 25% = 0.466 ± 0.057; 12.5% = 0.352 ± 0.038; 6.25% = 0.287 ± 0.028)(Fig. 1). A one-way ANOVA was used as the data was normally distributed for each nutrient concentration ($p > 0.05$) and few outliers were detected by visually assessment of a boxplot.

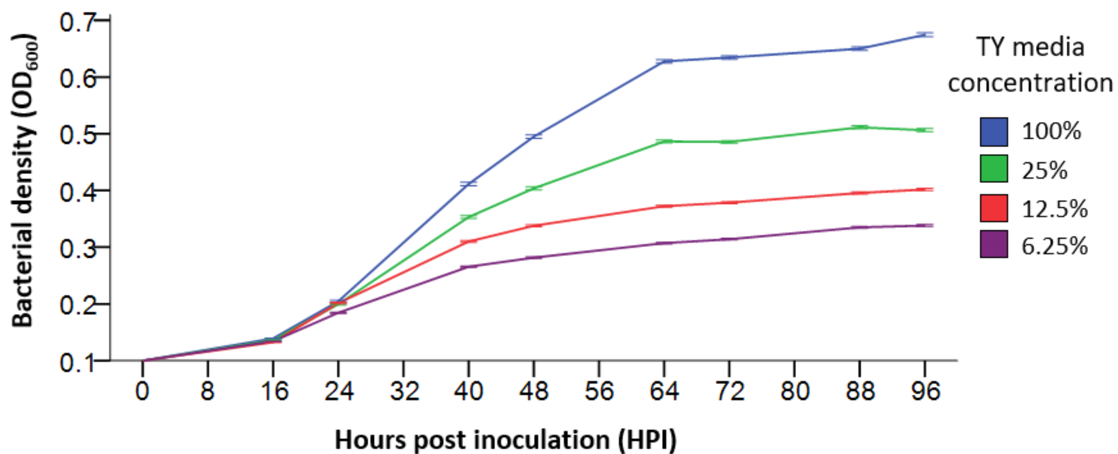

Fig. 1 Increasing the concentration of nutrient media increases the rhizobia growth rate. Mean bacterial density of *Rhizobium leguminosarum* strains (n = 192) when inoculated into TY media of various concentrations, compared over time. One-way ANOVA at 64 hours post inoculation, showed significantly increasing bacterial density as nutrient concentration increased (100% = 0.607 ± 0.075; 25% = 0.466 ± 0.057; 12.5% = 0.352 ± 0.038; 6.25% = 0.287 ± 0.028).

Two multivariate analysis of variance (MANOVA) were used to compare the strain's bacterial density grouped by genospecies or geographic origin, over time for each of the nutrient concentrations (Fig. 19, Appendix). Based on the analysis it is possible to determine which groups of strains are able to utilise the low and high nutrient environments most effeciently, thereby increasing their change of colonising the root surface.

There was a statistically significant interaction between the genospecies over time ($F(112, 5984) = 3.846$, $p < 0.001$; Pillai's Trace = 0.269; partial $\eta 2 = 0.067$) and between the geographic origins over time ($F(84, 6016) = 6.487$, $p < 0.001$; Pillai's Trace = 0.332; partial $\eta 2 = 0.083$). Estimated marginal means generated by the MANOVA analysis are plotted in Fig. 2.

Overall, gsB/UK strains had statistically higher growth than all other genospecies ($p < 0.005$) at high nutrient concentrations (100% = 0.496 ± 0.003; 25% = 0.401 ± 0.003) but significantly lower growth at low nutrient concentrations (6.25% = 0.242 ± 0.002). This indicates specialisation to high nutrient soils. This would not be useful if the inoculum was applied to the bulk soil and the rhizobia were required to live free-living before forming a symbiotic interaction. The gsA and France strains have

generally high growth across all nutrient concentrations indicating generalisation and suitability for use as soil inoculums. gsC had the lowest growth at high and low concentrations, though not statistically significant from gsD and gsE at 100% TY. gsD and gsE had the largest error bars due to the lower sample sizes than the other groups but had statistically higher growth than gsB and gsC at 6.25% TY. DK_C and DK_O alternate as which has the lowest growth, other than at 6.25% TY when DK_C has significantly highest growth (0.262 ± 0.002) from all other locations, showing specialism to low nutrient conditions.

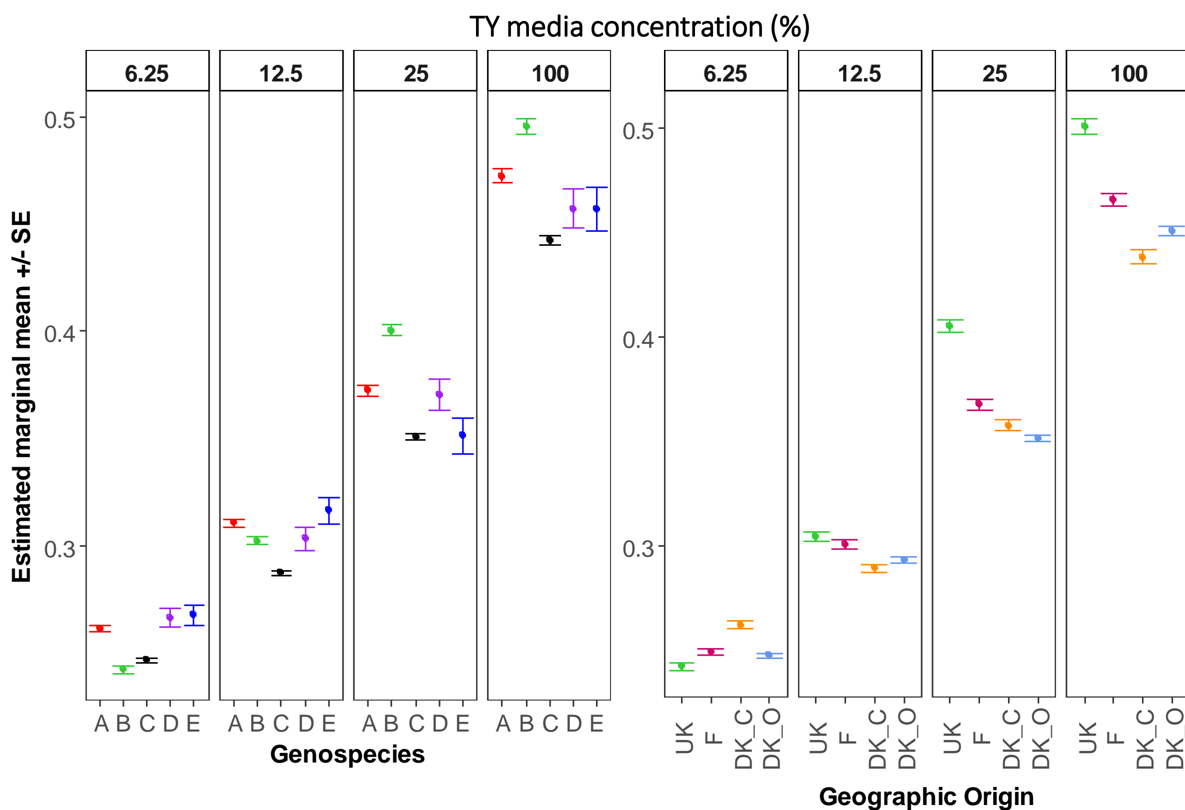

Fig. 2 gsB/UK strains were inhibited the most at low nutrient concentrations. Estimated marginal means of bacterial density (OD_{600}) of *Rhizobium leguminosarum* strains when inoculated into media of high to low nutrient concentrations (100, 25, 12.5 and 6.25% TY media). Values based on two MANOVA, first comparing genospecies, second comparing geographic origin. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.2.1 Biofilm Formation

Analysis of covariance (ANCOVA) was used to compare biofilm formation between genospecies and geographic origins. The covariate incorporated into the analysis was the bacterial density (OD_{600}) at 96 hpi. Therefore, the statistical differences between the groups is not a product of different amount of bacterial growth but the relative amount of biofilm (Fig. 3).

In low nutrient conditions gsB/UK strains has generally high biofilm formation despite their relatively low bacterial density (Fig. 3). Bacterial communities such as biofilms are widely recognized as being important for survival and persistence of bacteria in harsh environments [84]. However, this does not follow as biofilm formation is highest for gsB/UK in the most favourable conditions (100% TY).

For lower nutrient concentrations (6.25, 12.5 and 25% TY), gsA had the highest biofilm formation, followed by gsB and then gsC, gsD and gsE are combined lowest. This could indicate that in less than optimal conditions the strains have genetically adapted to the production of a particular amount of biofilm. However, there is a similar pattern for the geographic origins; in low nutrient concentrations, the UK strains had higher biofilm formation, followed by DK_O and then France and DK_C (Fig. 3).

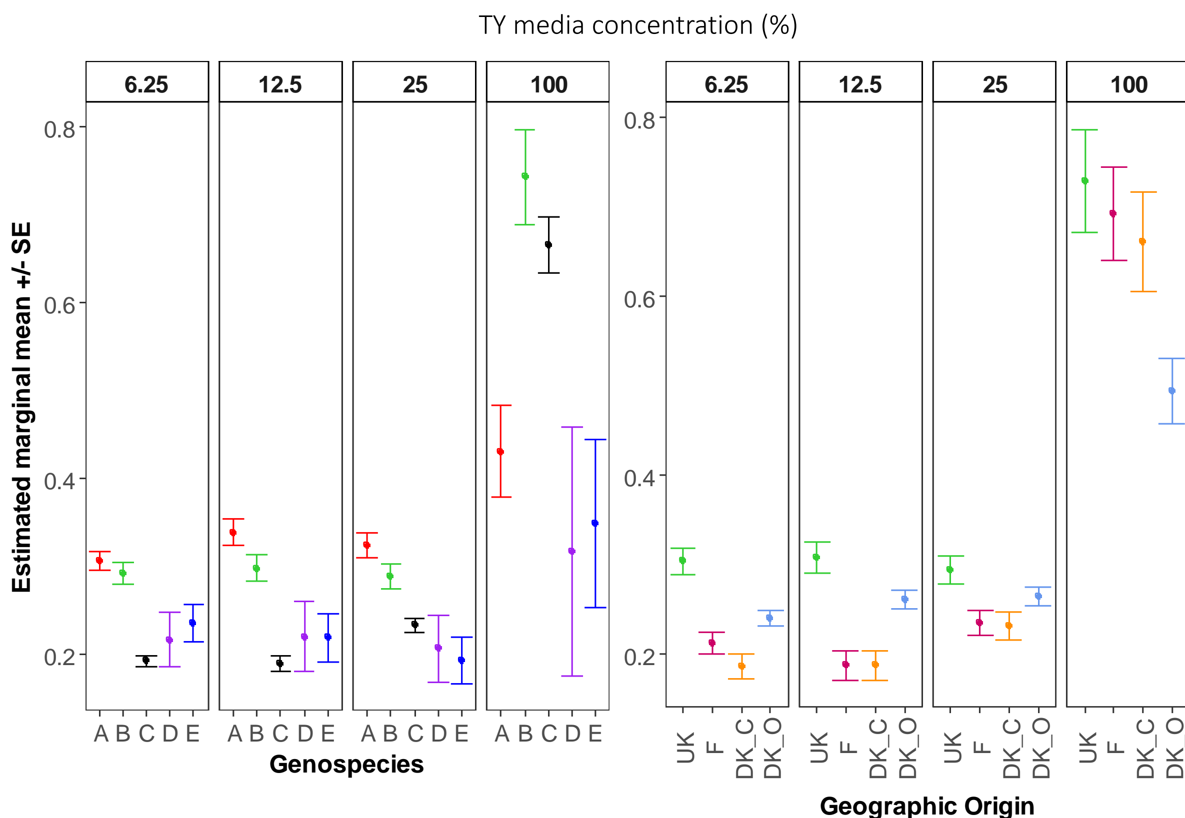

Fig. 3 At nutrient concentrations below 100% TY biofilm formation is unaffected between groups. Biofilm formation of *Rhizobium leguminosarum* strains after 96 hours' growth in various concentrations of TY media from low (6.25% TY) to high (100% TY) nutrients. Estimated marginal means calculated after adjustment using the point of maximum growth (OD_{600}) at 96 hours post inoculation. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.3 Temperature

The strains would be required to tolerate a range of temperatures across Europe and particularly low temperatures when used primarily in Denmark. Therefore, all temperatures chosen as below the laboratory optimum temperatures for rhizobial growth (28°C). First, the growth of the strains was conducted in 100% TY media, as this is a commonly used liquid growth media. Then, low nutrient conditions were used (6.25% TY) to determine if the media concentration rather than the temperature were mostly affecting the strains' bacterial density.

The bacterial density of the strains in high nutrient conditions (100% TY) was analysed with two one-way ANOVA with Tukey post hoc tests (72 and 96 hpi). The data was often normally distributed and it was concluded that an ANOVA would be robust enough to use. At 72 hpi all temperatures were significantly different ($p < 0.05$) except between 4°C and 10°C ($p = 0.930$). At 96 hpi most of the temperatures were significantly different again except between 4°C and 10°C ($p = 0.086$), but 20°C and 28°C were only just significant ($p = 0.048$)(Fig. 4). Over the experiment, the strains showed no growth at 4 and 10°C but generally decreasing temperature from the optimal (28°C) decreased growth. However, a longer experiment could have revealed growth at 10°C and potentially lower. Furthermore, the bacterial density at 20°C increasing above 28°C at 96 hpi indicates a depletion of nutrients that would not be as restrictive in the soil as it is in liquid growth media.

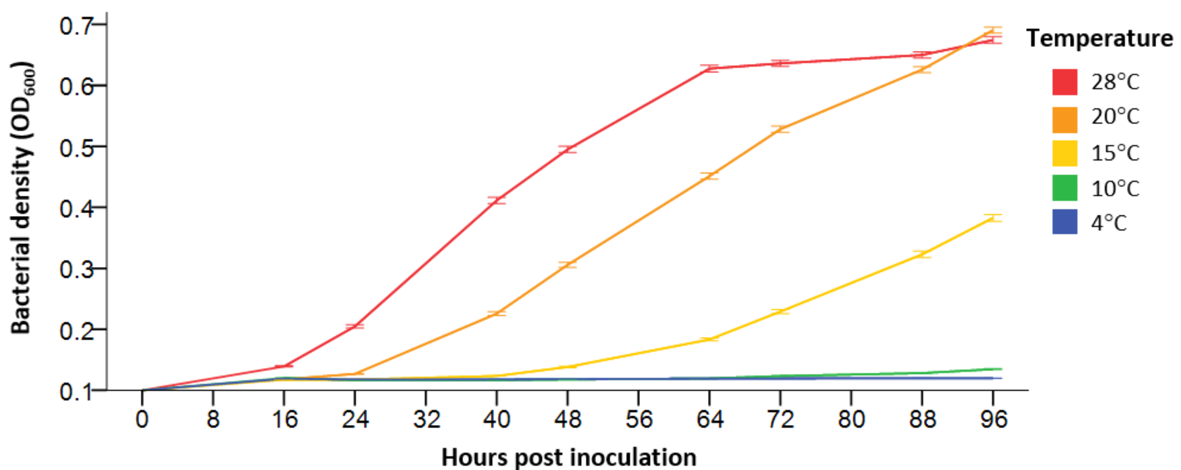

Fig. 4 Growth is significantly decreased at temperatures below 20°C. Mean bacterial density of *Rhizobium leguminosarum* strains (n = 192) when inoculated into 100% TY media and then incubated at various temperatures (4, 10, 15, 20 and 28°C), compared over time. A one-way ANOVA was conducted at 96 hours post inoculation, showed statistically significant difference between all temperatures ($p < 0.05$), except between 4 and 10°C.

Four MANOVA were used to compare the strain's bacterial density over time (Fig. 20 and 21, Appendix) grouped by genospecies or geographic origin and at both low and high nutrient concentrations. At low nutrients, there was a statistically significant interaction between the genospecies over time ($F(140, 7480) = 7.238$, $p < 0.001$; Pillai's Trace = 0.597; partial $\eta 2 = 0.119$) and between the geographic origins over time ($F(105, 7520) = 9.768$, $p < 0.001$; Pillai's Trace = 0.600; partial $\eta 2 = 0.120$). At high nutrients, there was a statistically significant interaction between the genospecies over time ($F(140, 7480) = 8.207$, $p < 0.001$; Pillai's Trace = 0.666; partial $\eta 2 = 0.133$) and between the geographic origins over time ($F(105, 7520) = 7.161$, $p < 0.001$; Pillai's Trace = 0.454; partial $\eta 2 = 0.091$). Estimated marginal means ± standard error are plotted in Fig. 5.

The gsB/UK strains had generally the highest growth across each temperature in the high nutrient conditions (Fig. 3), which correlated with the previous nutrient concentration experiemnt (Fig. 2). Additionally, in low nutrient concentrations gsB is no longer the highest growing genospecies, it is one of the lowest. This could be interpreted as the nutrient availability having a larger effect on growth than the change in temperature (Fig. 3).

The remaining strains from France, DK_C and DK_O were not statistically different at 20°C, high nutrients, and showed similiarities in growth at all temperatures. However, the lowest on average is DK_C which is made up of mainly gsC strains a consistently low performing grouping (Fig. 2; Fig. 5). From the temperature dataset, it can be concluded that gsC is performing consistently poorly whereas gsB/UK strains are specialised to high nutrient conditions but are relatively not affected by a decrease in temperature. At 4°C and 10°C there is more phenotypic variation than in high nutrient concentrations but still difficult to draw biological conclusions from.

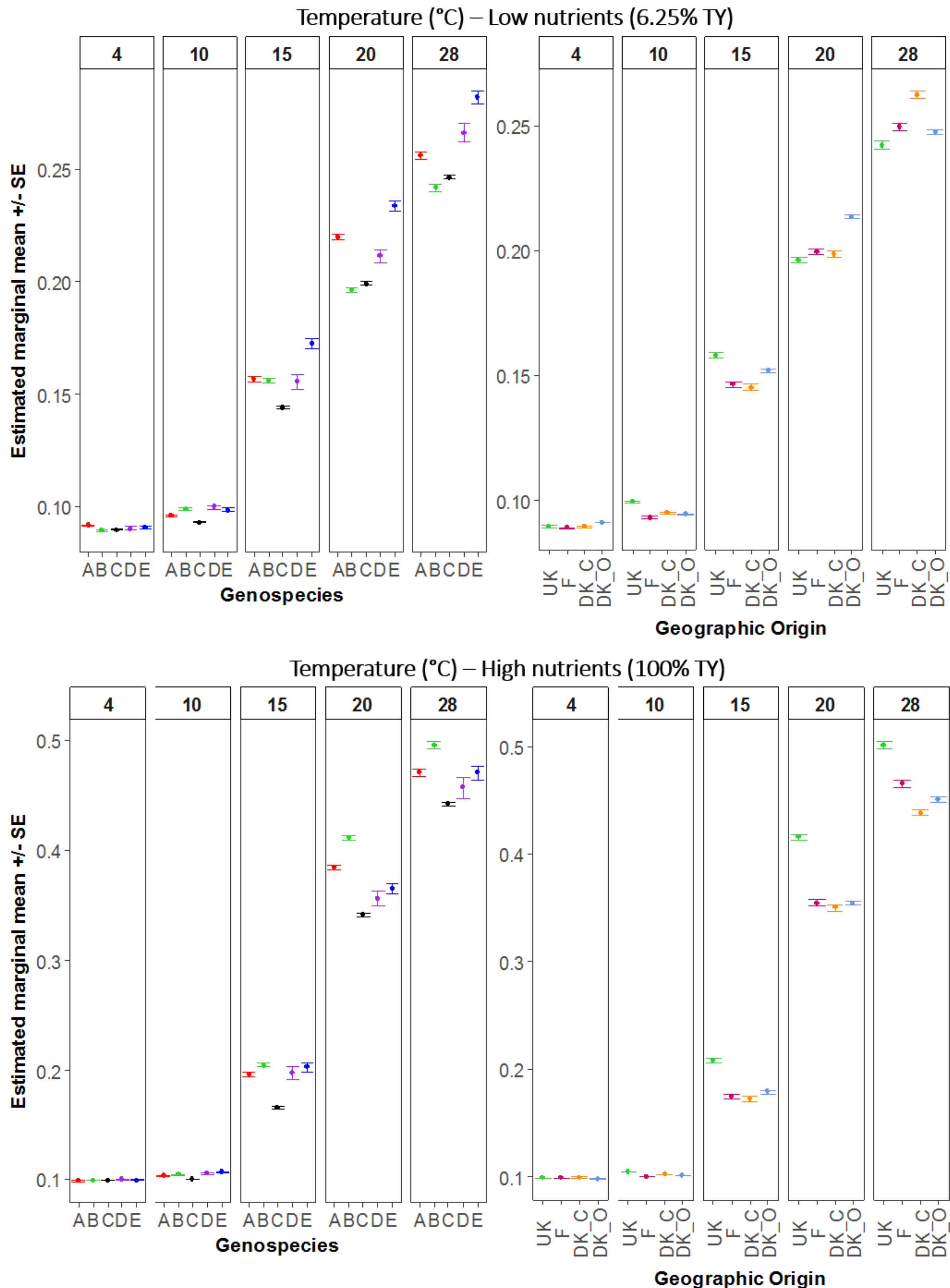

Fig. 5 The effect of temperature on growth in high and low nutrient conditions. Estimated marginal means of bacterial density (OD_{600}) of *Rhizobium leguminosarum* when inoculated into low (above, 6.25% TY) and high (below, 100% TY) nutrient concentration media and incubated at various temperatures (4, 10, 15, 20 and 28°C). Values based on two MANOVA, first comparing genospecies, second comparing geographic origin. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.3.1 Biofilm Formation

In harsh conditions biofilm formation is essential for survive, therefore higher biofilm formation at low temperatures and low nutrient conditions would be a favourable trait for a soil inoculum. Analysis of covariance was conducted on the comparing biofilm formation for each of the genospecies and geographic origin groups of strains (Fig. 5). The covariate was bacterial density at the endpoint of the experiment (96 hpi), therefore the analysis accounts for cell aggregates not purely the number of cells.

There is a huge amount of variation within and across the groups of this dataset but generally in low nutrients biofilm formation is highest for the gsA strains (Fig. 6). However, in high nutrient concentrations biofilm formation by gsA is relatively unaffected by the changes in temperature, indicating that biofilm formation is largely decreased by the lack of nutrients not the temperature decrease (Fig. 6).

At low and high nutrient concentrations the highest biofilm formation was at 20°C rather than at the optimal temperature 28°C (Fig. 6). This could indicate a trade-off between production of the protection extracellular matrix and cell division. However, patterns in this dataset are not clear making biological interpretation difficult and potentially unreliable.

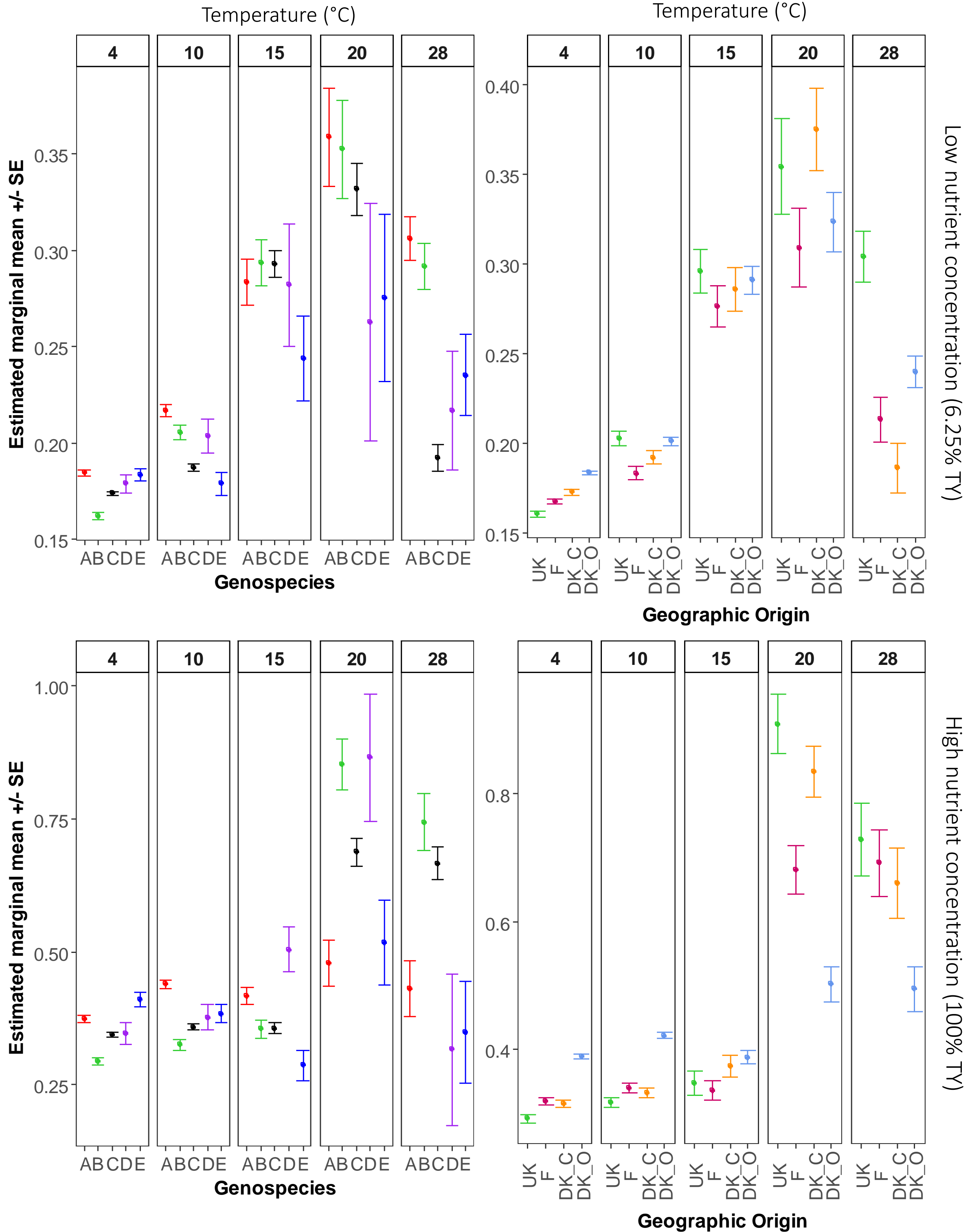

Fig. 6 Biofilm formation is not significantly higher for a single group of strains incubated at various temperatures. Biofilm formation of *Rhizobium leguminosarum* strains after 96 hours incubated at various temperatures whilst inoculated into low (6.25% TY) and high nutrient concentrations (100% TY media), the top and bottom graphs respectively. Estimated marginal means calculated after adjustment using the growth at 96 hours post inoculation. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.4 pH

Rhizobium species are known to tolerate include soil pH of between 4 and 10 [78]. Therefore, the liquid growth media was adjusted using hydrochloric acid (HCl) to increase acidity and sodium hydroxide (NaOH) to increase alkalinity. However, the pH of Danish soil generally ranges from pH 4.0 to 8.3, excluding extreme examples of highly acidic pyrite containing soils (< 4.0) and highly alkaline salt marsh areas (> 8.3) [85]. This correlates with the soil pH where the stains were isolated, which was consistently acidic in Denmark (Table 3; Fig. 8). The strain chosen for use across Denmark and the rest of Europe would be required to tolerate across a range of pH soils and subsequently outcompete locally adapted conspecifics.

Table 3 pH of the soil the Rlt strains were isolated from (data collected by NCHAIN, unpublished).

		Mean soil pH ± Standard Error	Soil pH Range	
			Minimum	Maximum
Genospecies	A	6.136 ± 0.059	5.5	6.9
	B	7.491 ± 0.043	6.4	7.6
	C	6.213 ± 0.034	5.3	6.8
	D	6.460 ± 0.051	6.3	6.6
	E	6.000 ± 0.135	5.6	6.2
Geographic Origin	UK	7.552 ± 0.011	7.4	7.6
	France	6.263 ± 0.048	5.9	6.8
	Denmark conventional farms	6.450 ± 0.028	6.2	6.6
	Denmark organic farms	6.057 ± 0.041	5.3	6.9

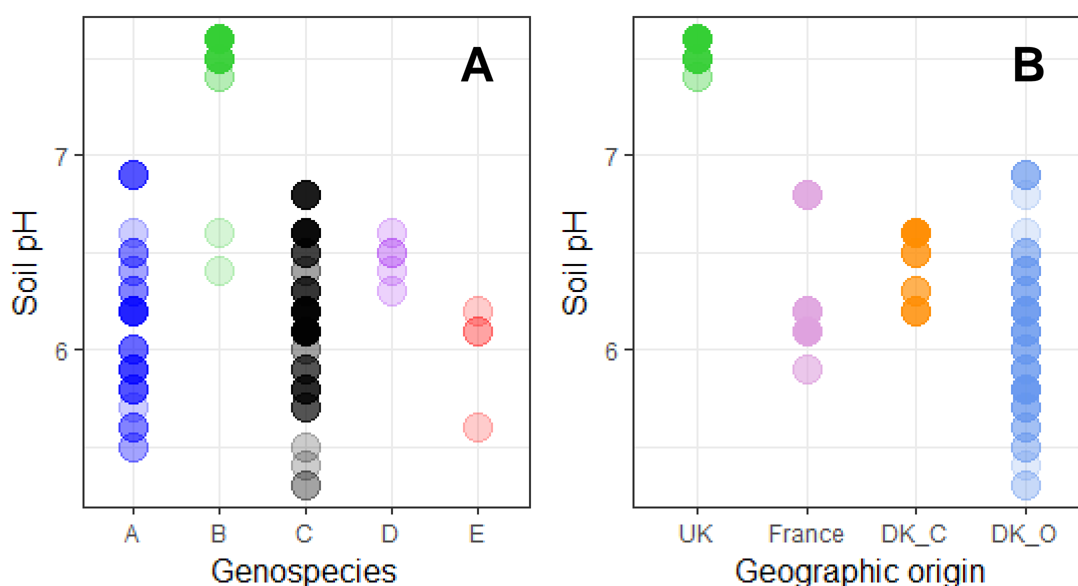

Fig. 7 Soil pH where *Rhizobium leguminosarum* strains (n = 192) were isolated. A; strains grouped based on their genospecies allocation. **B**; strains grouped based on their geographic origin of isolation. DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Points that are not translucent indicate a number of strains coming from the same soil pH stacked on top of one another. (Data collected by NCHAIN, unpublished)

Two Kruskal-Wallis H tests with pairwise comparisons were used to compare both the midpoint (48 hpi) and endpoint (96 hpi) of the experiment. The data was not normally distributed and contained many outliers. At 48 hpi all pH treatments were statistically significant from each other ($p < 0.05$), with pH 6 (0.404 ± 0.004) showing highest bacterial density followed by 8 (0.334 ± 0.000), 7 (0.304 ± 0.005) and 5 (0.216 ± 0.004)(Fig. 8). At 96 hpi, all strains were statistically significant except pH 4 and 10 ($p = 0.562$), and pH 7 and 8 ($p = 0.146$). At the endpoint, pH 6 (0.602 ± 0.004) was found to be optimal for Rlt growth, followed by pH 5 (0.485 ± 0.005), pH 8 (0.445 ± 0.004) and then pH 7 (0.422 ± 0.005). This correlates with Table 3, as the strains had originated in soils with means of around 6.0 to 6.5, except the gsB/UK strains (7.6).

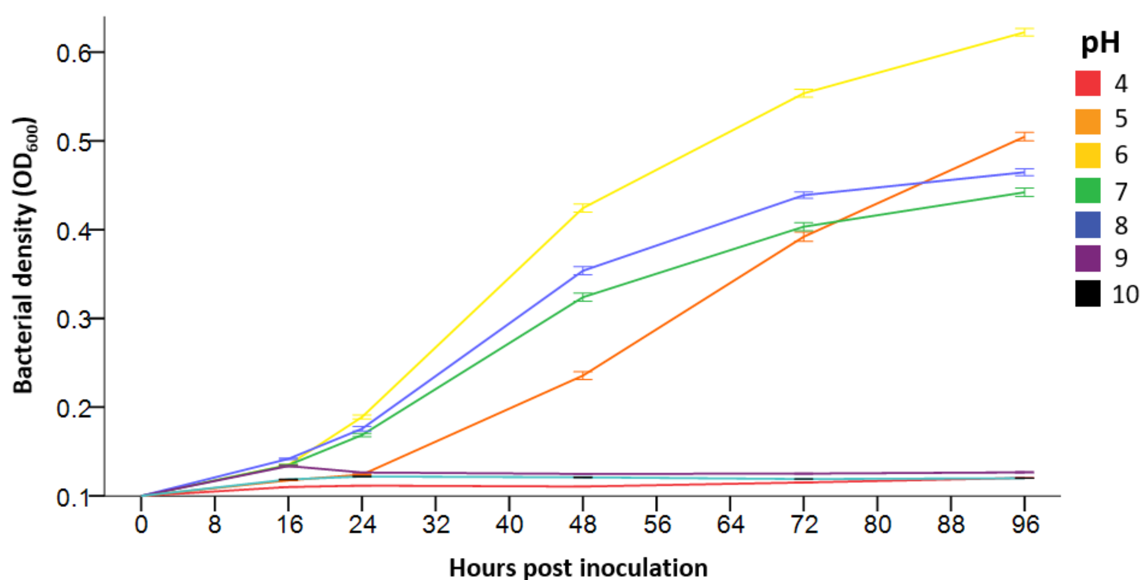

Fig. 8 The strains only grew within the range of pH 5 to 8, with the highest growth at pH 6. Mean bacterial density of *Rhizobium leguminosarum* strains (n = 192) when inoculated into Tryptone-Yeast media of various pH, compared over time. One-way ANOVA at 48 and 96 hours post inoculation, determined statistical significance between pH 5 to 8 ($p < 0.05$).

Two MANOVA were used to compare the strain's bacterial density grouped by genospecies or geographic origin, over time for each of the pH (Fig. 22, Appendix). There was a statistically significant interaction between the genospecies over time ($F(112, 6545) = 7.369$, $p < 0.001$; Pillai's Trace = 0.784; partial $\eta 2 = 0.112$) and between the geographic origins over time ($F(84, 6580) = 12.847$, $p < 0.001$; Pillai's Trace = 0.986; partial $\eta 2 = 0.141$). At pH 4 (0.093 ± 0.000), 9 (0.107 ± 0.000) and 10 (0.100 ± 0.000) there statistically significant differences between genospecies or geographic origins.

Unfortunately, this experiment was only conducted in high nutrient concentration media (100% TY) so it cannot be interpreted how much of the variation between the strains is due to the nutrients or

the pH as is possible with the temperature dataset. Additionally, the growth of strains across all groups is lower at pH 7 than at pH 8 (Fig. 8; Fig. 9) which could indicate an error occurred when using the pH meter.

The UK strains are most distinctly high growing across pH 5 to 8 (Fig. 9), this correlates with previous experiments in high nutrient conditions (Fig. 2; Fig.5). Additionally, the UK stains should be locally adapted to a very narrow soil pH of between 7.4 and 7.6 based on where they were isolated (Table 3; Fig. 7) which indicates that nutrient availability is effecting growth more than the change in pH. This could be equally true for the gsB strains, which are statistically the highest growth at pH 6 (0.404 ± 0.003) and 8 (0.347 ± 0.003), though they have a slightly wider original pH range (6.4 to 7.6).

gsC had either the lowest growth or joint lowest growth across all pH. The remaining locations grew similarly, with no statistically significant difference between France, DK_C and DK_O at pH 5 and 8 and between DK_C and DK_O at pH 6 and 7. The France strains grew better at pH 6 and 7 as this was most similar to the soil pH they were adapted to and there was little variation in pH across the sites that were sampled. However, the DK_O strains came from a diverse range of soil pH therefore the mean is lower at any pH specifically.

3.4.1 Biofilm Formation

ANCOVA analysis was conducted to compare the biofilm formation between groups of strains belonging to each genospecies and geographic origin. There was relatively no biofilm formation in highly alkaline conditions, pH 9 and 10 (Fig. 10). At pH 4, despite there being an insignificant amount of growth (Fig. 8), biofilm formation was equivalent to pH 8. This indicates a range in which the strains can detect stressful pH conditions (pH 4, 5, 7 and 8) and form biofilms but not be destroyed by extreme pH (pH 9 and 10).

gsC has the highest biofilm formation at pH 6, the optimal pH for Rlt growth. This indicates that although gsC frequently has the lowest bacterial density, it produces more biofilm relative to its low growth than the other genospecies. This could suggest variation in resource allocation towards biofilm formation and away from replication.

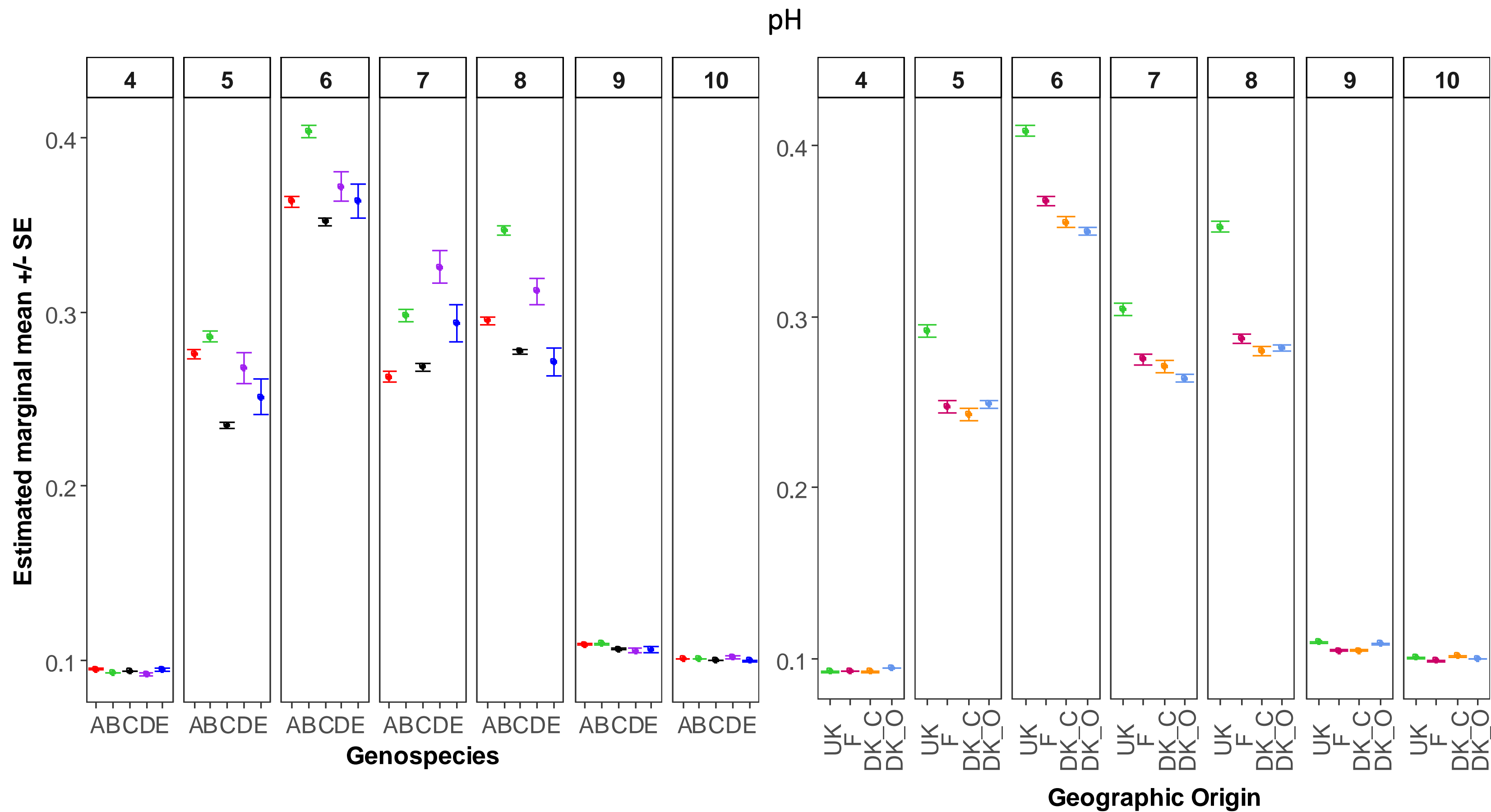

Fig. 9 The gsB/UK strains has the significantly higher growth at pH 5, 6, and 8. Estimated marginal means of bacterial density (OD_{600}) of *Rhizobium leguminosarum* strains grown in media of varying pH (4, 5, 6, 7, 8, 9 and 10). Values based on two MANOVA, first comparing genospecies, second comparing geographic origin. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

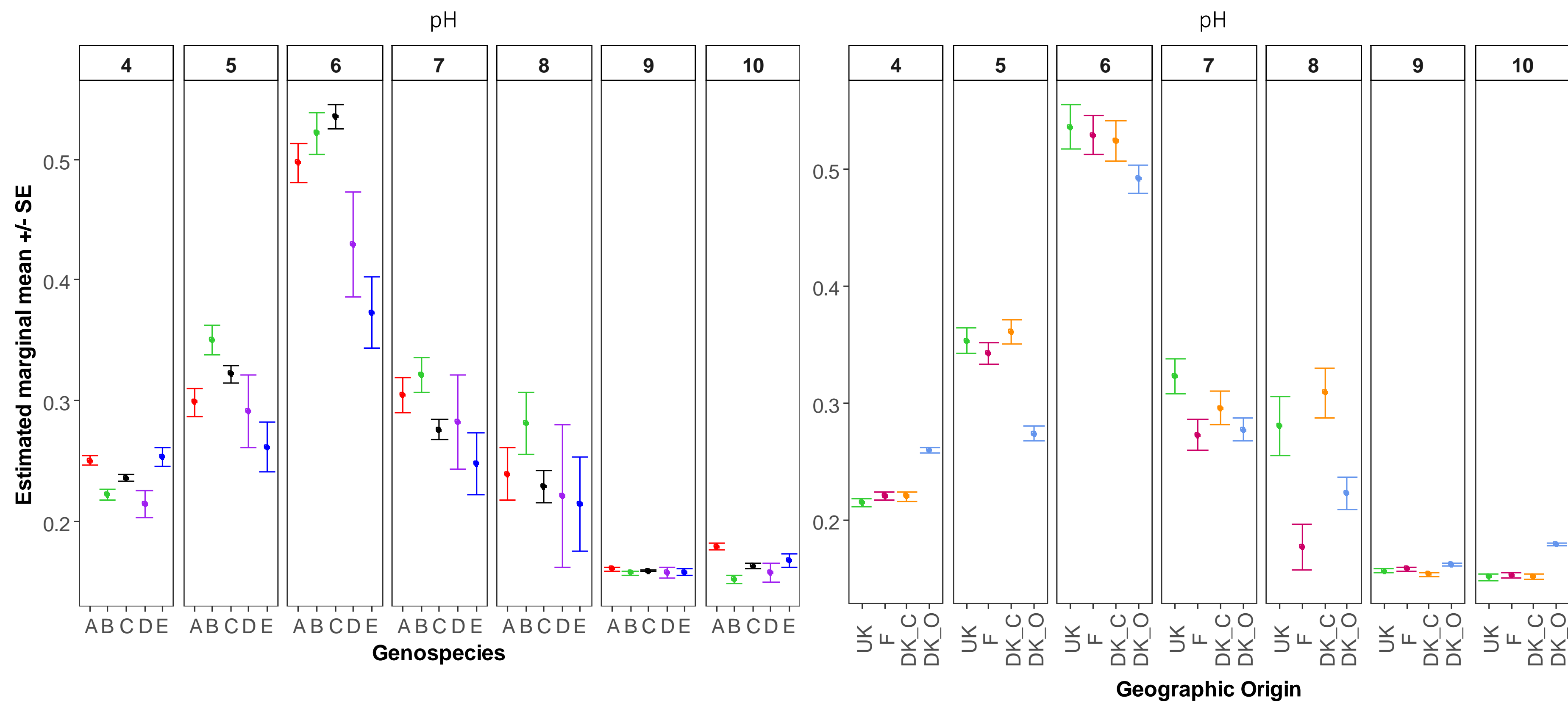

Fig. 10 Biofilm formation does not occur at pH 9 or 10 and is the highest at pH 6. Biofilm formation of *Rhizobium leguminosarum* strains after 96 hours growth in TY media of various pH. Estimated marginal means calculated after adjustment using the growth at 96 hours post inoculation. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.5 Protists

The 192 strains of Rlt were grown in the presence of two predatory protists, *Tetrahymena pyriformis* (*Tetrahymena*) and *Chilomonas paramecium* (*Chilomonas*). A Kruskal-Wallis test with pairwise comparisons was used to compare the bacterial density of the treatments at 64 hpi, as the data was not normally distributed and contained many outliers. All treatments were statistically significant from each other ($p < 0.001$). However, *Tetrahymena* (0.313 ± 0.002) showing the largest bacterial density decrease in comparison to the control (0.574 ± 0.004) with a 45% decrease in growth. *Chilomonas* (0.539 ± 0.005) decreased the growth of the rhizobia by only 6% (Fig. 11).

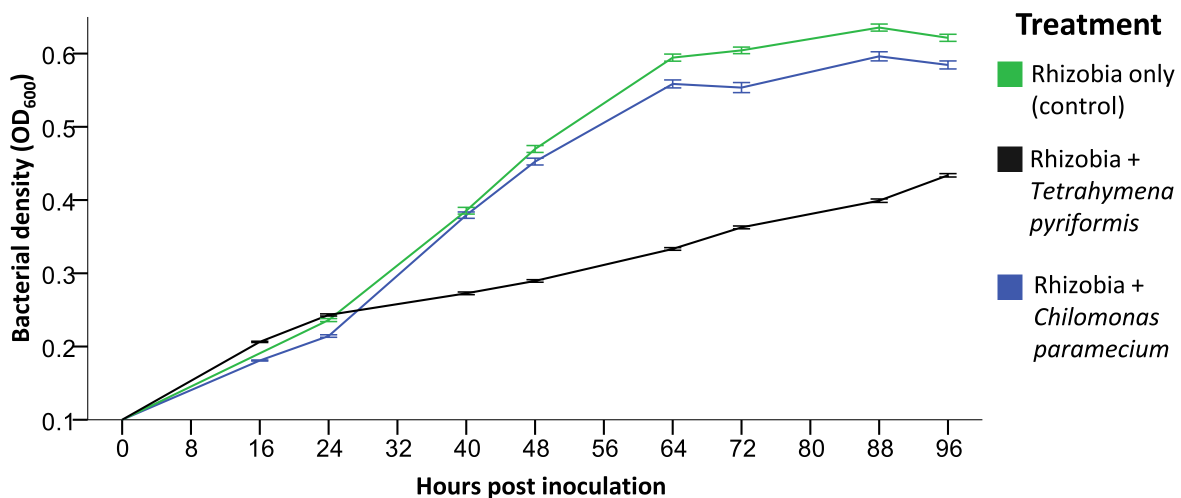

Fig. 11 Rhizobia growth is statistically decreased with the addition of protist predators, *Tetrahymena pyriformis* and *Chilomonas paramecium*. Mean bacterial density of *Rhizobium leguminosarum* strains (n = 192) when inoculated into TY media including no protists (control), *T. pyriformis* or *C. paramecium*. A one-way ANOVA was conducted at 64 hours post inoculation, showing statistically significant variation between treatments ($p < 0.05$). Error bars = mean ± standard error.

Two MANOVA were used to compare the strain's bacterial density grouped by genospecies or geographic origin, over time in the presence of each protist predator (Fig. 22, Appendix). There was a statistically significant interaction between the genospecies over time ($F(54, 3948) = 5.525$, $p < 0.001$; Pillai's Trace = 0.211; partial $\eta2 = 0.070$) and between the geographic origins over time ($F(72, 3927) = 1.923$, $p < 0.001$; Pillai's Trace = 0.102; partial $\eta2 = 0.034$).

gsC and gsE are grew statistically the lowest under *Chilomonas* predation, and gsC grew statistically lowest under *Tetrahymena* predation (Fig. 11). gsC was the only statistically different genospecies with *Tetrahymena*, as the rest all had equally reduced growth. *Chilomonas* predation showed similar variation in growth between groups compared to the control (high nutrients, 28°C). gsD has similar susceptibility and resistance to each predator. The UK strains are statistically less affected by the predators than the other location (Fig. 12). The France strains are most affected by *Tetrahymena*

predation and DK_O most affected by *Chilomonas*, which could indicate an increased susceptibility to these predators (Fig. 12).

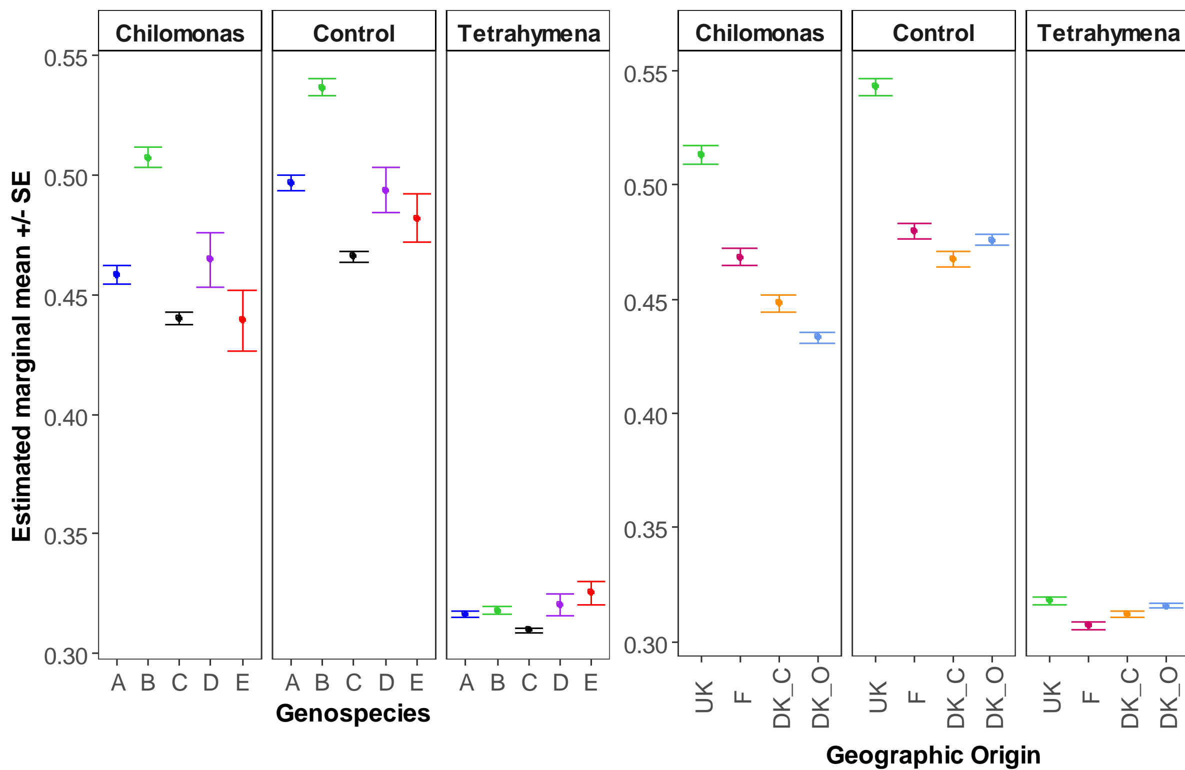

Fig. 12 Overall growth of the gsB/UK strains was least affected by the addition of protist predators. Estimated marginal means of bacterial density (OD_{600}) of *Rhizobium leguminosarum* strains grown in the presence of protist predators (*Chilomonas* and *Tetrahymena*). Values based on two MANOVA, first comparing genospecies, second comparing geographic origin. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = estimated marginal mean ± standard error.

3.6 Principal Component Analysis (PCA)

PCA was conducted to compare the phenotypic variation between strains based on their genospecies allocation or geographic origin. The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.3. The PCA had four components with percentage variance explained greater than 5% (34.1, 10.3, 8.7 and 7.0%). Additionally, these components came before the inflection point on the scree plot further suggesting they should be retained (Fig. 23, Appendix). The total variance explained by these four components was 60.1%. A Varimax orthogonal rotation was employed to aid interpretability. Component loadings are presented in Table 5 (Appendix), after removing the treatments that did not have loading to any of the four components.

Clustering of the strains based on their genospecies or geographic origin would indicate high levels of phenotypic similarity between the strains based on either their core genome or local adaption, respectively. However, when considering all the strains based on genospecies or geographic origin (Fig. 13, panels 1 and 2), there is no distinct clustering based on either group. Further examining the groups individually, gsC includes a representative number of strains from three of the four locations and has no clustering per location. gsE which includes only 4 strains is tightly clustered but this limited amount of strains may not be representative of the genospecies population (Fig. 13, panel 3).

Principal component 1 (PC1) loads highly to treatments that have highest variation amongst the strains during the exponential phase. Thereby it could simply show variation in growth during semi-optimal conditions. PC1 is highly variable for gsA, gsB and gsC and all the geographic origins. PC2 loads highly to high nutrient treatments with high bacterial density, often the endpoints of experiments. Due to gsB's low growth in low nutrient conditions, PC2 could indicate high loading in low nutrient concentrations. This correlated with gsE, high loading to PC2 and high growth in low nutrient concentrations (Fig. 13, panel 3).

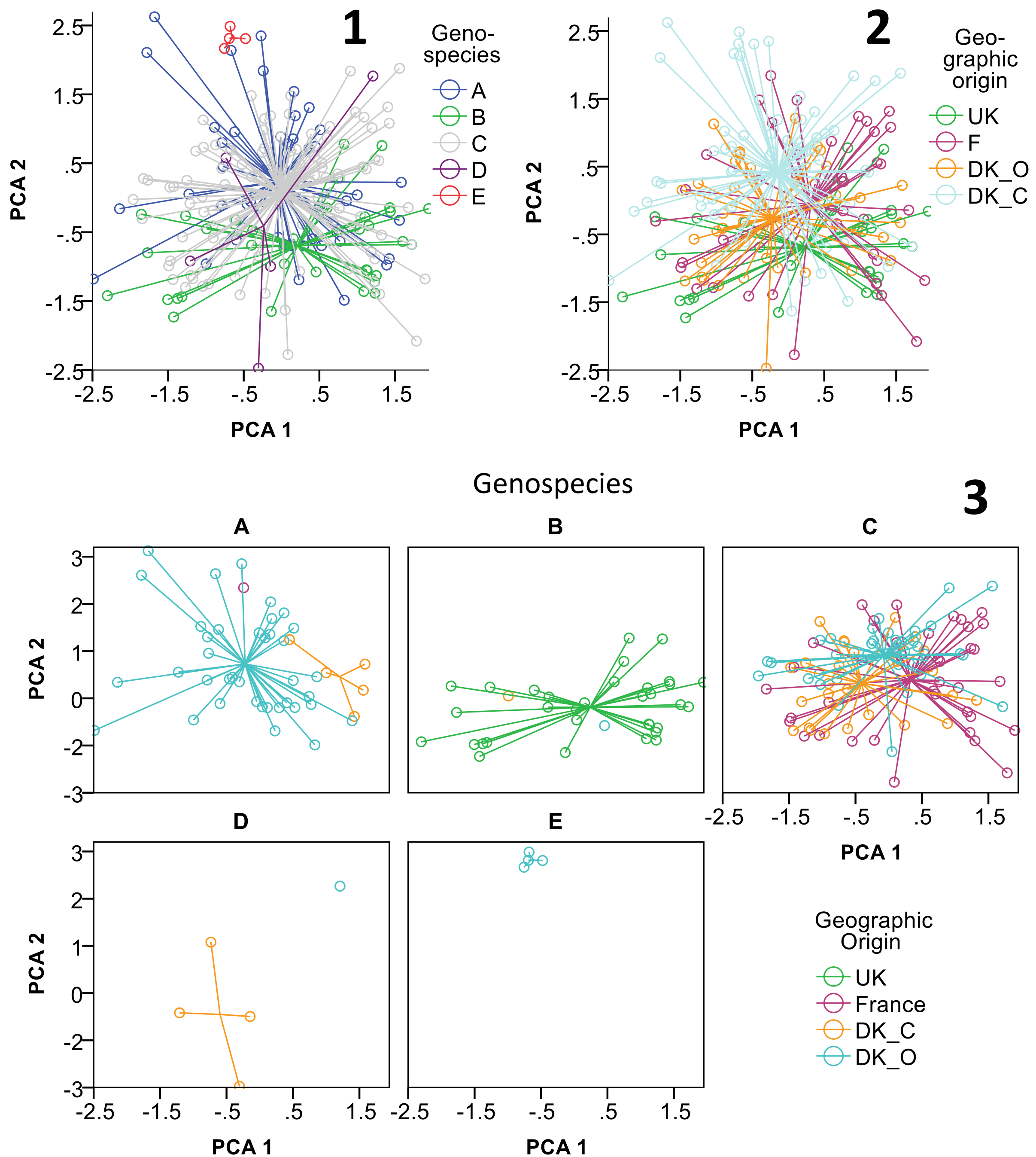

Fig. 13 No separation of the strains was observed from PCA of the growth datasets. Regression plots of loading to extracted components from PCA, each point indicating a *Rhizobium leguminosarum* strain (n = 192), coloured by genospecies and geographic origin. Principal component 1 (PCA 1) and 2 (PCA 2) account for 34.1% and 10.3% variance explained, respectively. Spread of the strains indicates phenotypic variation amongst groupings. **1;** compares the strains belonging to various genospecies. **2;** compares the strains based on the site of isolation (geographic origin). **3;** groups the strains based on their genospecies then into geographic origins. F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms.

3.7 EcoPlates

Each strain of Rlt was inoculated into microplate wells containing one of 31 carbon sources and a final control well containing only water. Based on how much each carbon source is metabolised the well is strained purple through the reduction of a tetrazolium dye. Fig. 14 show the variation in carbon utilisation between each carbon group for all Rlt strains.

29 of the 31 carbon sources were being metabolised more than the control (Fig. 14), indicated by the bar being higher than the 1.0 reference line. The carbon sources that Rlt were unable to metabolise include 2-Hydroy Benzoic acid and α-Ketobutyric acid.

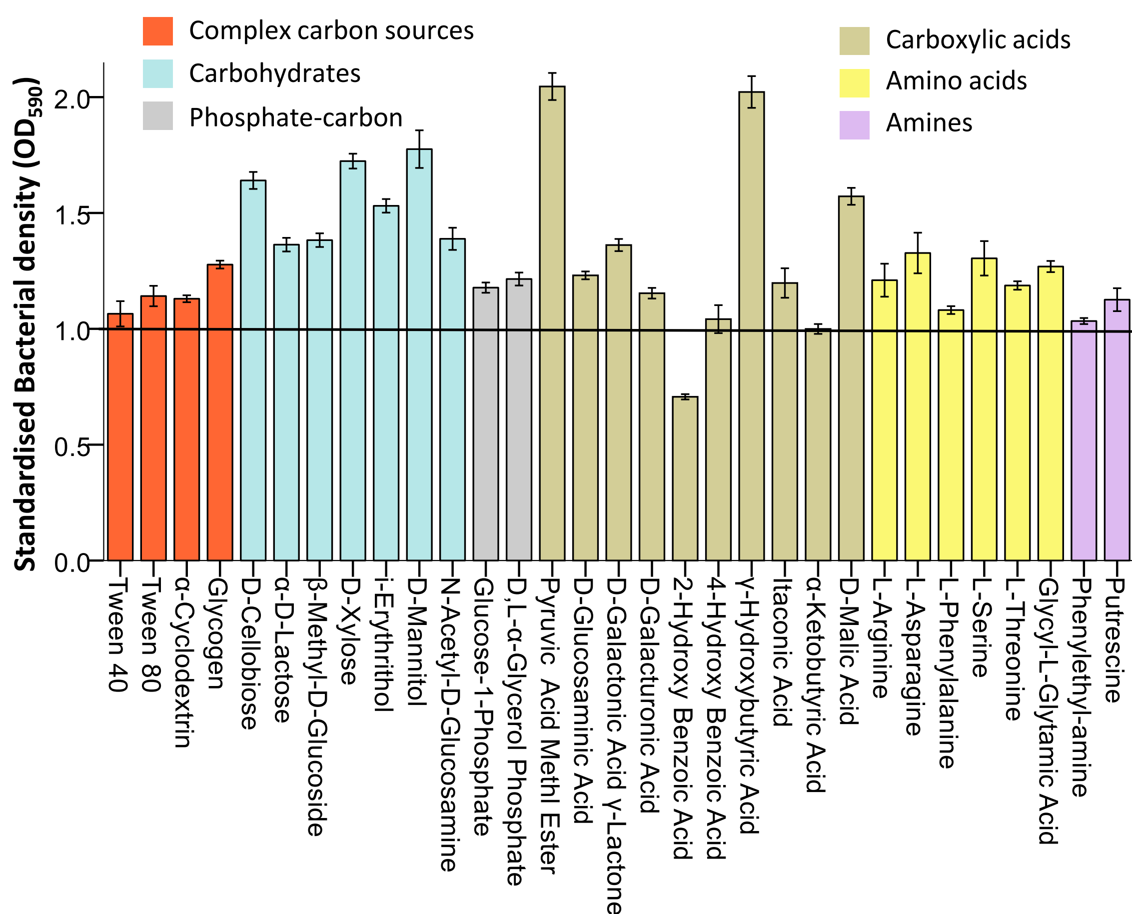

Fig. 14 Metabolism of 31 common carbon sources found in the soil. Utilisation of 31 carbon sources by *Rhizobium leguminosarum* (n = 192), 72 hours post-inoculation. Standardised bacterial density (OD_{590}) represents the ratio of the carbon source OD_{590} to the control well containing water. Any points below 1.0 (indicted with a reference line) had a lower OD_{590} than the control. Error bars = mean ± standard error.

3.7.1 Average Well Colour Development (AWCD)

When considering the strains performance across all carbon sources as AWCD, generalist strains with the ability to utilise a large range of carbon sources are discovered. Kruskal-Wallis tests were used to compare genospecies and geographic origins to find which strains could be inoculated into the largest range of soils types. gsE is the only statistically significant genospecies from the others with a significantly higher AWCD than gsB, gsC and gsD (Fig. 15A). DK_C had a statistically lower AWCD than the other locations (Fig. 15B). A low scoring AWCD could limit the strain's usefulness across a wide range of soils types. Generally, the strains are very similar in which carbon sources they metabolise.

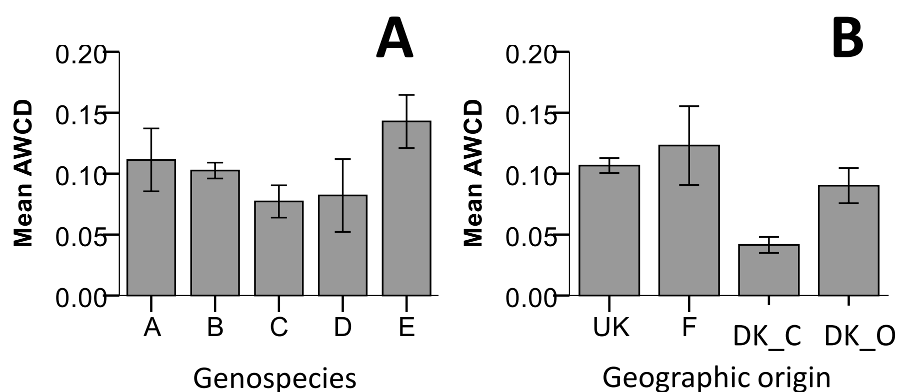

Fig. 15 Average well colour development (AWCD) for genospecies and geographic origins. High AWCD indicates the *Rhizobium leguminosarum* strains can metabolise many of the 31 carbon sources. A; groups strains based on their genospecies. B; groups strains based on their site of isolation (geographic origin). F = France; DK_C = Denmark conventional farms; DK_O = Denmark organic farms. Error bars = mean ± standard error.

4 Ranking the strains

Ranking the strains based on a wide range of treatments will give an indication of their capacity to survive a range of environmental stressors. A marketable soil inoculum must be a generalist, able to form successful symbiotic interactions in soil across Europe.

Several treatments were removed from this analysis because they showed no or little growth through the length of the experiment: 4°C and 10°C (Fig. 4) and pH 4, 9 and 10 (Fig. 8). The "mean fitness relative to best strain" was calculated for each strain using Equations 2 and 3 (complete list of rankings in Table 8, Appendix). The strains were then grouped per genospecies and geographic origin to determine which groups had on average the most generalist strains (Fig. 17 A and B). From strains scoring within 10% of the highest scoring strain were arbitrarily chosen to look at the distribution of "elite" strains across the genospecies and geographic origins (Fig. 17 C).

gsD and gsE had the smallest variation between scores due to their small sample sizes and mainly originating from one geographic origin (Table 1). The highest scoring genospecies were gsA and gsB with the largest amount of "elite" strains between them (Fig. 17 C). gsB showed less variation amongst the individual strain scores with fewer outliers. gsC had the largest variation amongst the genospecies and lowest scores on average; potentially due to originating from three geographic origins (Table 1) each with their own specialisation to a particular environment.

There is a large amount of phenotypic variation between the strains within each geographic origins and genospecies A, B and C (Fig. 13; Fig. 17). Therefore, it cannot be assumed a strain will be suitable for use as a soil inoculum based on where it was isolated and/or its genospecies allocation. Rlt strain "45" is shown in Fig. 17 as the highest scoring strain overall (100%). It is one of the few strains from DK_C that belongs to gsA, therefore if future isolated strains were assessed as having high fitness based on being isolated from DK_C and allocated to gsA, I believe this would be misinformed.

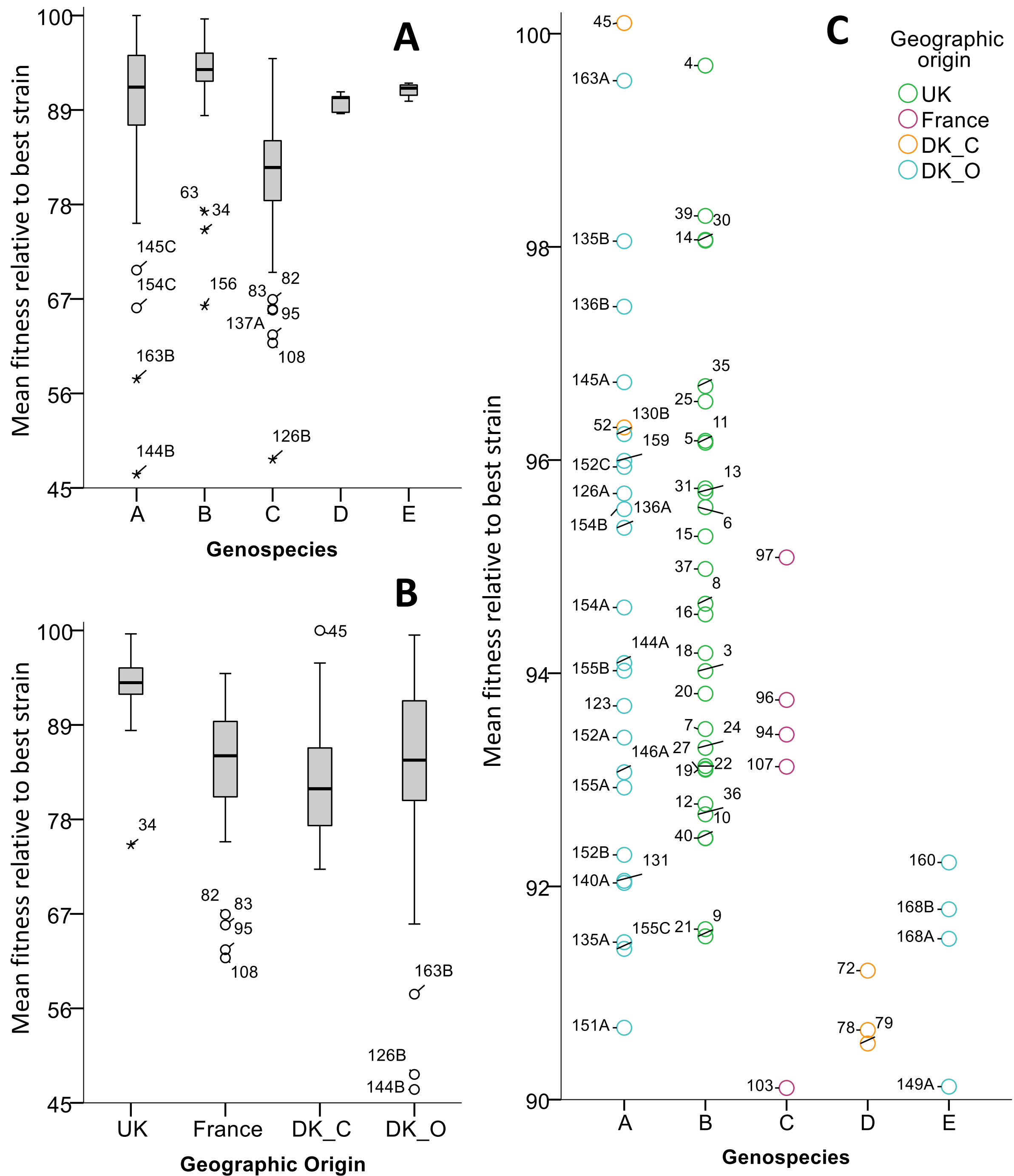

Fig. 16 Ranking the strains based on bacterial density (OD_{600}) for each treatment. The "mean fitness relative to best strain" was calculated using equations 2 and 3. From these calculations it is assumed the highest scoring strains would be best suited to being used as rhizobia soil inoculums. They indicate the strain's ability to grow in a range of environmental conditions. **A;** grouping the scores based on the strain's genospecies allocation. **B;** grouping the scores based on geographic origin of the strains. **C;** from the scores the top 10%, scoring from 90 to 100%, were arbitrarily chosen to look at the distribution of "elite" strains across the genospecies and geographic origins. Numbers attached to points indicate the strain identification number (e.g. 126B). DK_C = Denmark conventional farms; DK_O = Denmark organic farms.

5 Discussion

The primary aim of this study was to determine which *Rhizobium leguminosarum* strains would be best suited for use as soil inoculums and whether the strains from particular genospecies or locations have distinct phenotypes. In addition, the environmental ranges the strains can grow in was established.

The optimal growth conditions used in laboratory experiments of rhizobium species are not similar to natural soil conditions. Generally, the media is nutrient rich, the strains are inoculated at 28°C in slightly acidic media (TY media is 6.65 pH) creating a 100-fold increase in bacteria concentration compared to the rhizosphere [6]. Therefore, it is important to establish how this research relates to free-living conditions in the soil to be able to use this methodology to assess the suitability of strains for use as soil inocula. For the purposes of this discussion, natural conditions will be Denmark, as the soil inocula produced using these strains will be implemented in Denmark first.

5.1 Nutrient concentration

The composition of soil nutrients available to various plant species and genotypes differs widely [54], suggesting a limited value of soil chemical analysis attempting to determine plant-available nutrients for each geographic origin. Therefore it is assumed that the “low” nutrient concentration (6.25% and 12.% TY media) are more accurate to natural conditions, though differing in chemical composition [54].

In high nutrient concentrations (100 and 25% TY), the gsB/UK strains had the highest bacterial density, whereas at low nutrient concentrations (12.5 and 6.25% TY) they had the lowest growth (Fig. 2), indicating specialism to high nutrient conditions. This could be due to them being isolated in a nutrient rich soil. gsA is observed to be a generalist across all nutrient concentration treatments and have the highest bacterial density at 6.25% TY (Fig. 2), potentially making these strains more versatile across a range of soil conditions.

5.2 Temperature

The average temperature in Denmark ranges from 1.2°C in February to 17.4°C in July [86]. Whereas, the optimal conditions for growth were either 20°C or 28 °C depending on the length of the experiment (Fig. 4). The strains did not show any growth at 4 or 10°C (Fig. 4); this is surprising based on the average temperatures in Denmark. However, the strains subjected to 10°C could have just been experiencing a prolonged lag phase, as the bacterial density is slightly increasing between 88

and 96 hpi (Fig. 4). Therefore, if the experiment were repeated it should be continued for longer than 96 hours.

Nuccio *et al.,* 2016 found the rhizosphere community to be most influenced by regional climate such as changes in soil moisture and temperature [49]. Whereas the background communities were most affected by soil characteristics (e.g. pH). This highlights the importance of measuring the growth of the strains across a range of temperatures but also brings into question the ability of these strains to growth in temperatures below 15°C.

Additionally, in this study the high nutrient conditions appear to influence the variation between the strains more than the temperature (Fig. 5). The same strains are favoured at 15, 20 and 28°C for both the genospecies and geographic origins. gsB/UK have the highest growth followed by gsA, gsD/gsE and then gsC, respectively, highly similar to the nutrient concentration experiment (Fig. 2). In conclusion it is difficult to attribute changes in growth between genospecies/geographic origins to a change in temperature.

In addition to designing the best soils inocula for the present, the future and future climate should be considered. A shift in the function and composition of microbial communities could be altered by soil warming by increasing the available carbon for microbial respiration [52].

5.3 pH

The strains could not tolerate or remained dormant in media of pH 4, 9 and 10 (Fig. 9). This correlates with Adhikari *et al.,* (2012) who found no growth of their strains at pH 4.0 despite corresponding soil pH ranging from 3.6 to 6.4 [50].

Changes in rhizosphere community diversity and richness can largely be explained by changes in pH [59,60]. This can be attributed to the narrow pH ranges for optimal growth of bacteria [60]. However, these results show the same strains are favoured across a range of pH (Fig. 10). As with the temperature experiment this could be due to the high nutrient conditions disproportionally favouring of certain strains (Fig. 10). This experiment was only conducted in high nutrient conditions (100% TY) so it is impossible to say whether the gsB/UK strains are best suited for soil inoculum because of their high growth across a range of pH (5 to 8).

5.4 Biofilm formation

Rhizosphere biofilms are determined by species richness, diversity and relative population densities and have a significant effect on soil ecology [87]. Additionally, biofilm is rich in microbiota; therefore, biofilm formation is a good indication of colonisation potential. Simons *et al.,* (1996) pioneer of a high throughput alternative to soil colonisation assays, concluded that fast colonisation is a key advantage to being outcompeted by faster growers [76].

However, despite including the covariant “bacterial density at 96 hpi” for each set of analysis, it seems that biofilm formation was significantly higher for all treatments with high growth (Fig. 3; Fig. 6; Fig. 10). Scepticism of the results comes from seeing no detectable patterns in the temperature (Fig. 6) and pH (Fig. 10) experiments. Therefore, in subsequent biofilm formation assays a more robust protocol would be attempted, removing the supernatant before adding crystal violet to decrease background variation [82].

5.5 Protist predation

During the protist predation experiment, the rhizobia were heavy predated by *Tetrahymena* with no variation between genospecies or geographic origins (Fig. 12). This indicates none of the groups has any specialised protection mechanisms or were favoured by the predators. This is also true for the *Chilomonas* treatment but with a reduced intensity of predation. The bacterial density followed the same pattern as the control but with a slight decrease caused by the strains being predated equally by the *Chilomonas* (Fig. 12). Therefore, no genospecies/geographic origin specific conclusions can be extracted.

Alternatively, Habte & Alexander (1978) found the number of rhizobia cells to stabilise and co-existence to occur with *T. pyriformis* [88]. This indicates that over a longer experiment or at a lower concentration of protist inoculum, the differences between the genospecies/geographic origins could have become apparent.

5.6 Metabolism of carbon sources

A combination of soil properties, plant characteristics, and the interaction of roots with microorganisms control the availability of nutrients in the rhizosphere [89]. Sugiyama *et al.* (2014) found a 1.5 to 3-fold increase in AWCD for the rhizosphere soil in comparison to the bulk soil [90]. Additionally, most of the carbon sources were highly metabolised by the rhizosphere soils with the exception of 2-hydroxybenzoic acid, as can be seen in Figure 14. However, one of the irregularities

with the data is that the average OD_{590} 2-hydroxybenzoic acid is below the reference line indicating that the strains must have metabolised in the water well. This could be due to growth media transfer into the EcoPlate wells when the strains are introduced.

The carbon source group Rlt can metabolise most readily are the carbohydrates and the carboxylic acids but there is very little phenotypic variation between AWCD for the genospecies or geographic origins (Fig. 15). Therefore, no individual group of strains can be considered more adequate at metabolising a larger range of metabolites.

5.7 Ranking the best strains

High ranking strains would indicate resilience against a range of abiotic and biotic stressors and hints that these strains could efficiently colonise the rhizosphere. Another option is to combine strains which are suited to a wider range of environments together. Kyei-Boahen et. al. (2005) found his made little difference on yield because all strains area equally effective, which would not be true in this case [91].

Overall, the highest preforming strains were from the UK strains, consisting of only gsB. The UK strains have the highest average score when the strains were ranked across the growth experiments (Fig. 16). This indicates that based on the stressors tested, the UK strains would be the best candidates for use as part of a soil inoculum. The UK strains, had the highest growth in the highest nutrient concentration (100% TY) at incubation temperatures of 15, 20 and 28°C (Fig. 5) and pH 5, 6, 7 and 8 (Fig. 9). Biofilm formation was highest for the UK strains at all nutrient concentrations and among the highest across all temperature and pH treatments (Fig. 21; Fig. 22; Fig. 23). These results indicate a strong argument to investigate the use of these UK strains for soil inocula. However, these experiments were all conducted in high nutrient conditions potentially favouring the gsB/UK strains.

At low nutrient concentrations (6.25% TY) the gsB/UK strains were the worst preforming group of strains. This highlights a need for further research into whether the gsB/UK strains could be used as soil inoculums applied directly to the low nutrient conditions of the bulk soil. Potentially, these strains could still be inoculated onto seed before sowing, increasing their chances of forming root nodules before the plants is in contact with indigenous soil bacteria. Experimentation in low nutrient soil with soil inoculation and pre-inoculated seeds would be required to confirm the success of the gsB/UK strains.

5.8 Future research

The success of particular groups of strains within these experiments should first be correlated with their symbiotic efficiency *in planta*. Other members of NCHAIN are inoculated white clover to find out which strains result in the highest plant biomass (unpublished). Once these studies are combined, a high throughput methodology for determining the success of strains as soil inoculums can be established. Secondly, further analysis should be conducted measuring competition between the strains and other microorganisms from the rhizosphere. Finally, the genomes can be assessed to dissect genes responsible for particular phenotypes or are present in only "elite" strains.

Appendix

A.1 Nutrient concentration

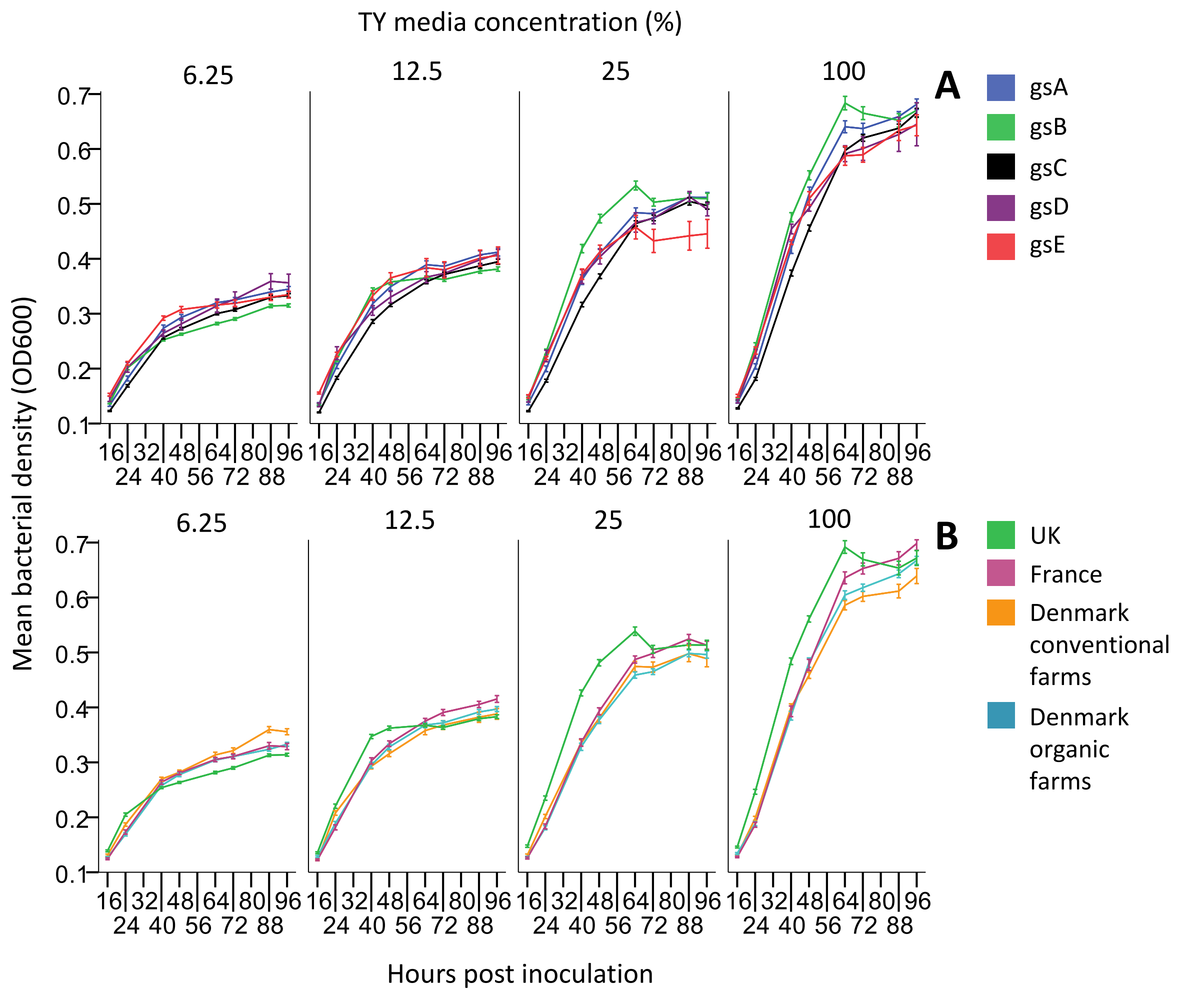

Fig. 17 Effect of nutrient concentration over time on bacterial density. *Rhizobium leguminosarum* (n = 192) strains, inoculated into high to low nutrient concentrations (100, 25, 12.5 and 6.25% Tryptone-Yeast (TY) media). **A;** grouping the strains based on their genospecies allocation. **B;** grouping the strain based on their geographic origin of isolation. Error bars = mean ± standard error.

A.2 Temperature

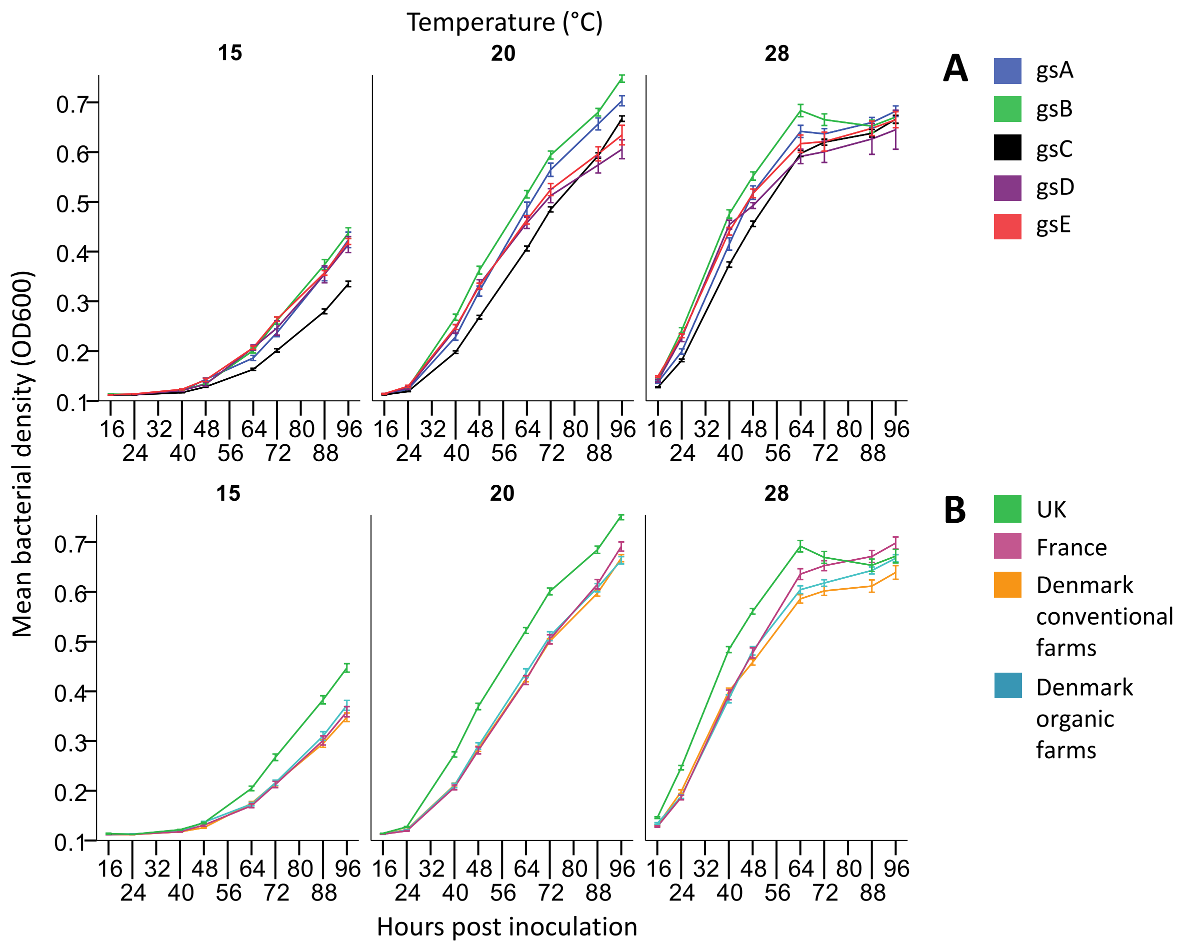

Fig. 18 Effect of temperature over time on bacterial density in high nutrient media. *Rhizobium leguminosarum* (n = 192) strains inoculated into high nutrient concentration media (100% TY) and incubated at various temperatures. 28°C is considered optimal growth conditions for rhizobia. 20°C, 15°C, 10°C and 4°C were used to mimic temperatures associated with natural conditions. 10°C and 4°C are not shown, as there was no increase in bacterial density throughout the experiment (Fig. 4). **A;** grouping the strains based on their genospecies allocation. **B;** grouping the strains based on their geographic origin. Error bars = mean ± standard error.

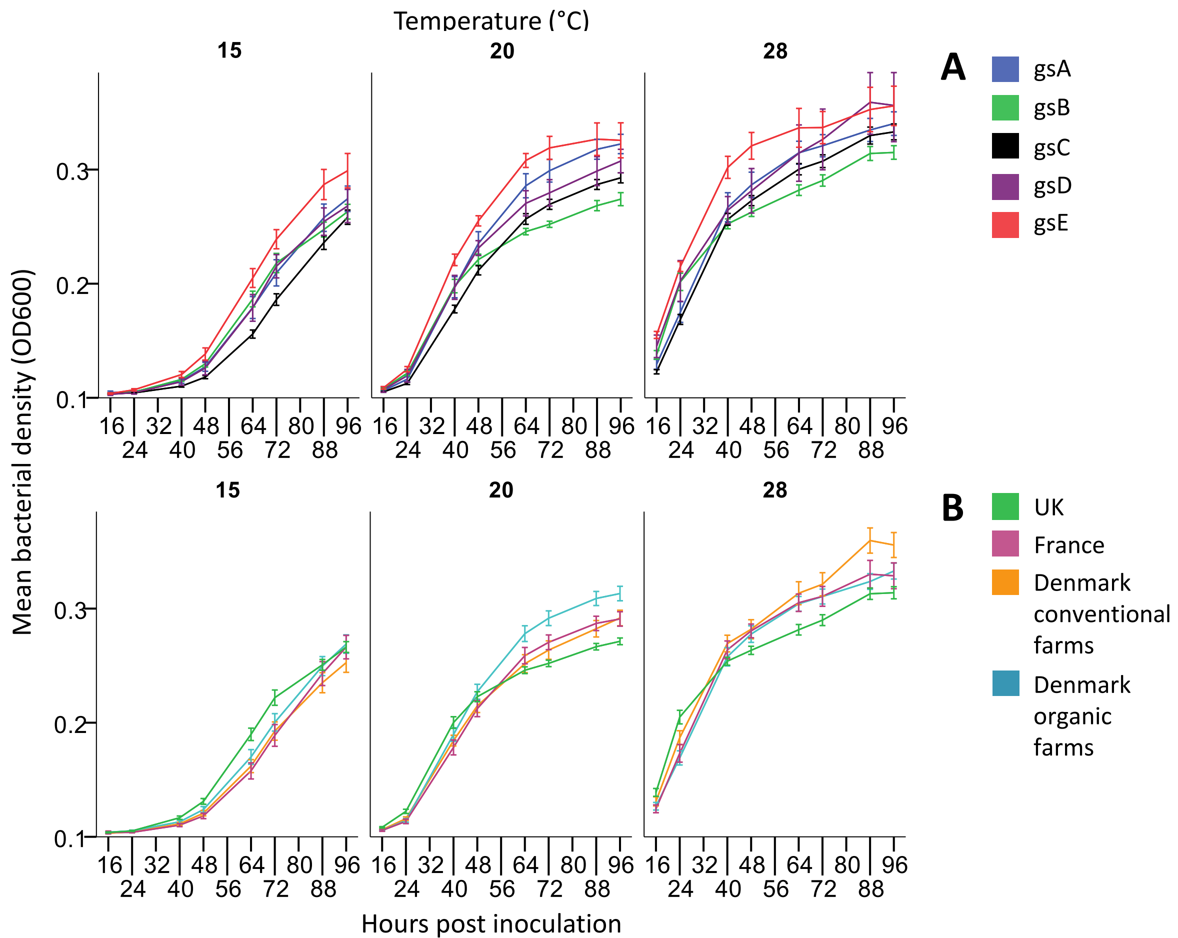

Fig. 19 Effect of temperature over time on bacterial density in low nutrient media. *Rhizobium leguminosarum* (n = 192) inoculated into low nutrient concentration media (6.25% TY) and incubated at various temperatures. 28°C is considered optimal growth conditions for rhizobia. 20°C, 15°C, 10°C and 4°C were used to mimic temperatures associated with natural conditions. 10°C and 4°C are not shown, as there was no increase in bacterial density throughout the experiment (Fig. 4). **A;** grouping the strains based on their genospecies allocation. **B;** grouping the strains based on their geographic origin. Error bars = mean ± standard error.

A.3 pH

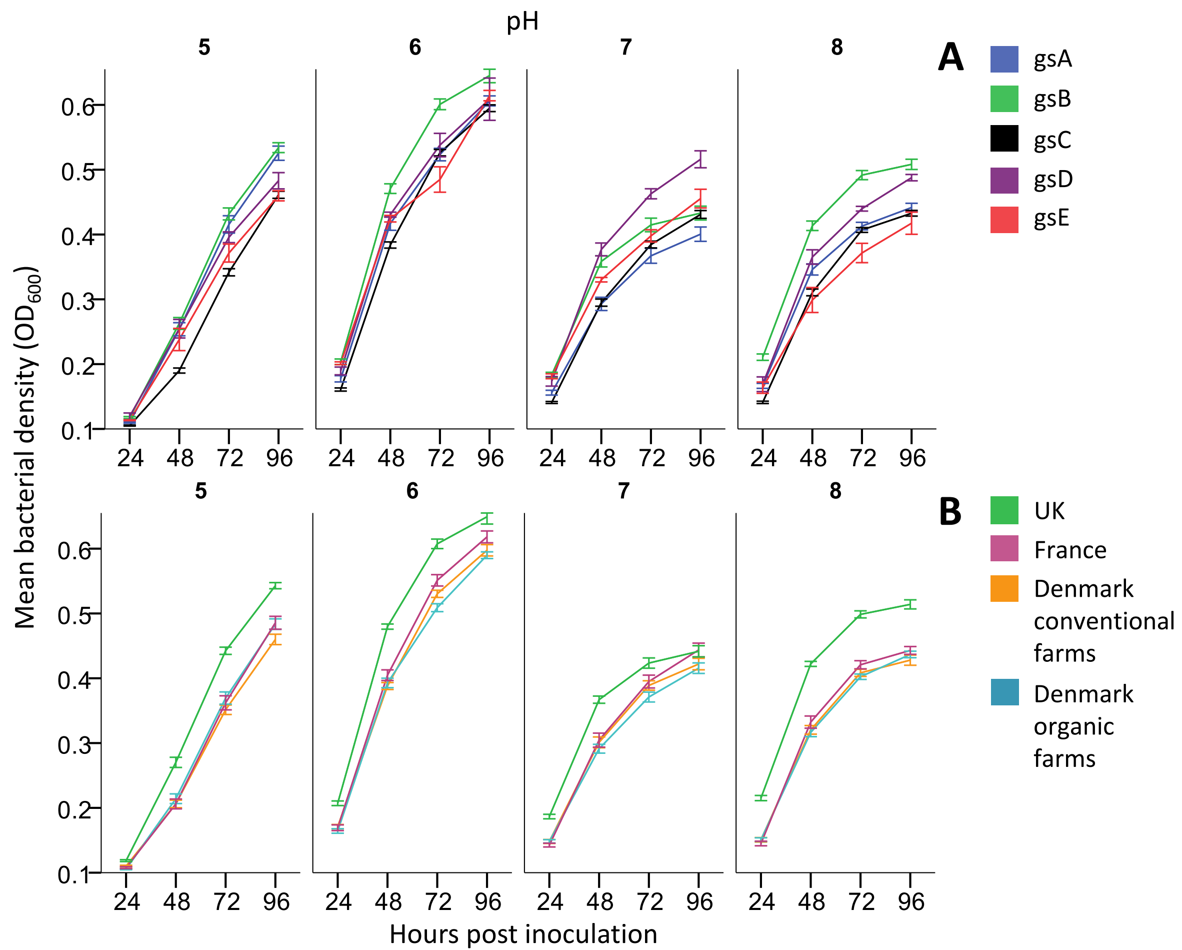

Fig. 20 Effect of pH over time on bacterial density. *Rhizobium leguminosarum* (n = 192) strains inoculated into TY media of various pH. The pH of the media before manipulation was 6.65 with either HCl or NaOH. pH of the media was either 4, 5, 6, 7, 8, 9 or 10. pH 4, 9 and 10 are not shown as there was no increase in bacterial density throughout the experiment (Fig. 9). **A;** grouping the strains based on their genospecies allocation. **B;** grouping the strain based on their geographic origin of isolation. Error bars = mean ± standard error.

A.4 Protist predation

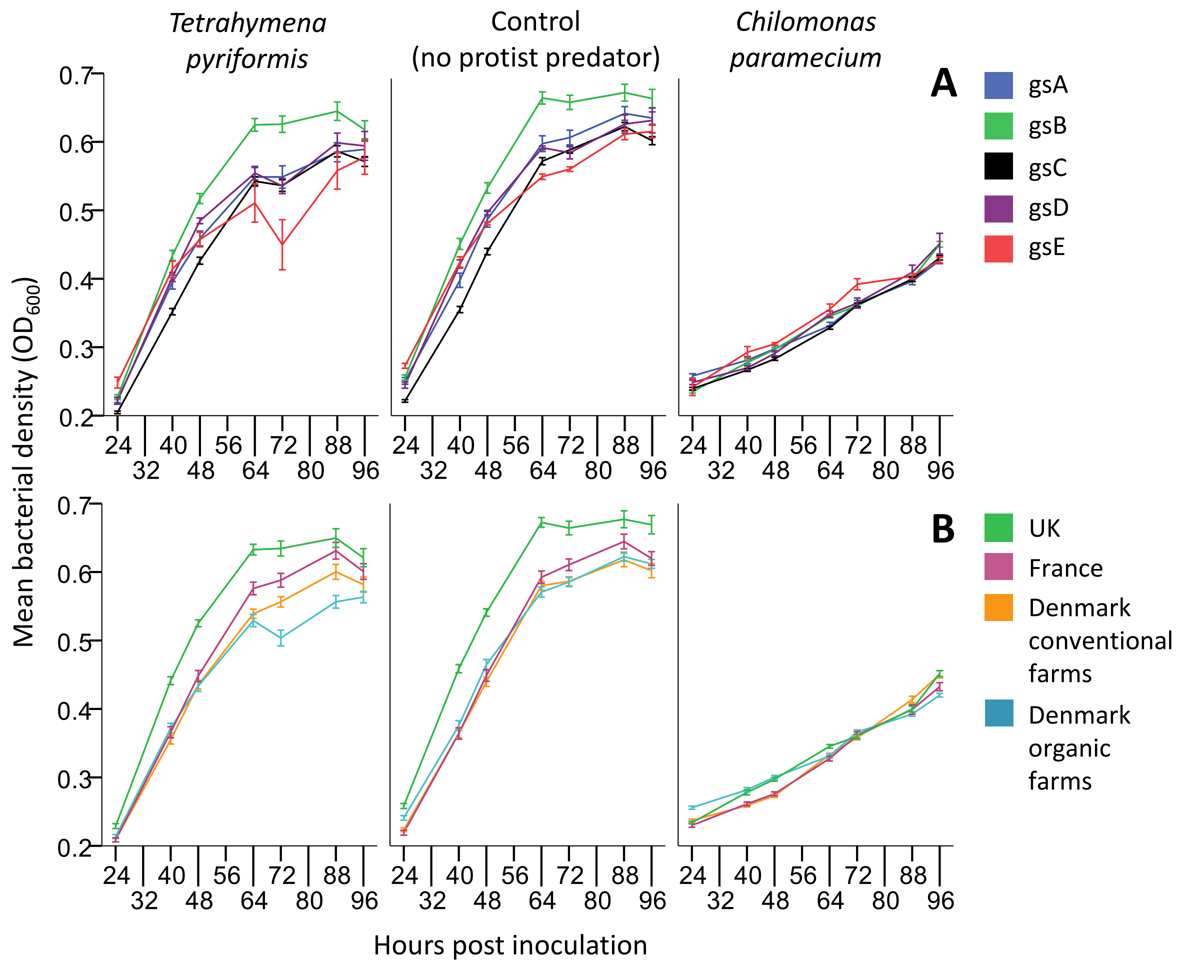

Fig. 21 Effect of protist predation over time on bacterial density. *Rhizobium leguminosarum* strains (n = 192), inoculated into TY media including no protist predator (control), *Tetrahymena pyriformis* or *Chilomonas paramecium*. **A;** grouping the strains based on their genospecies allocation. **B;** grouping the strain based on their geographic origin of isolation. Error bars = mean ± standard error.

A.5 Biofilm Formation

Table 4 Statistical summary of the ANCOVA conducted for biofilm formation data

Experiment	Treatment		Df		F	Sig	Partial η2
TY	100	Genospecies	4	186	8.275	0.000	0.151
		Geographic Origin	3	187	6.031	0.001	0.088
	25	Geographic Origin	3	187	3.684	0.013	0.056
		Genospecies	4	186	10.786	0.000	0.188
	12.5	Geographic Origin	3	187	13.348	0.000	0.176
		Genospecies	4	186	23.884	0.000	0.339
	6.25	Geographic Origin	3	187	12.715	0.000	0.169
		Genospecies	4	186	25.811	0.000	0.357
Temperature (low nutrients)	4	Geographic Origin	3	187	57.317	0.000	0.479
		Genospecies	4	186	20.273	0.000	0.304
	10	Geographic Origin	3	187	7.067	0.000	0.102
		Genospecies	4	186	17.742	0.000	0.276
	15	Geographic Origin	3	187	0.557	0.644	0.009
		Genospecies	4	186	1.172	0.325	0.025
	20	Geographic Origin	3	187	1.773	0.154	0.028
		Genospecies	4	186	1.270	0.283	0.027
	28	Geographic Origin	3	187	12.715	0.000	0.169
		Genospecies	4	186	25.811	0.000	0.357
Temperature (high nutrients)	4	Geographic Origin	3	187	83.972	0.000	0.574
		Genospecies	4	186	20.603	0.000	0.307
	10	Geographic Origin	3	187	64.623	0.000	0.509
		Genospecies	4	186	23.286	0.000	0.334
	15	Geographic Origin	3	187	2.989	0.032	0.046
		Genospecies	4	186	7.436	0.000	0.138
	20	Geographic Origin	3	187	26.209	0.000	0.296
		Genospecies	4	186	10.831	0.000	0.189
	28	Geographic Origin	3	187	6.031	0.001	0.088
		Genospecies	4	186	8.275	0.000	0.151
pH	4	Geographic Origin	3	187	39.469	0.000	0.388
		Genospecies	4	186	6.165	0.000	0.117
	5	Geographic Origin	3	187	28.175	0.000	0.311
		Genospecies	4	186	4.943	0.001	0.096
	6	Geographic Origin	3	187	1.966	0.121	0.031
		Genospecies	4	186	8.064	0.000	0.148
	7	Geographic Origin	3	187	2.829	0.040	0.043
		Genospecies	4	186	2.775	0.028	0.056
	8	Geographic Origin	3	187	8.776	0.000	0.123
		Genospecies	4	186	0.913	0.458	0.019
	9	Geographic Origin	3	187	7.776	0.000	0.111
		Genospecies	4	186	0.650	0.628	0.014
	10	Geographic Origin	3	187	59.174	0.000	0.487
		Genospecies	4	186	10.932	0.000	0.190

B.1 Principal component analysis

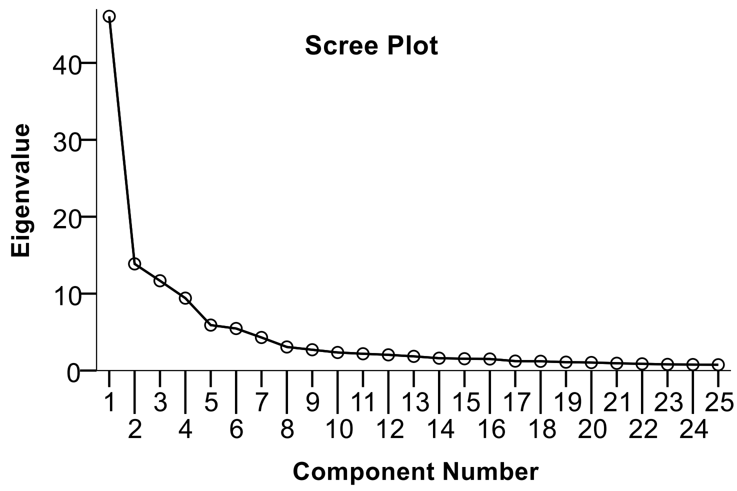

Fig. 22 Scree plot indicating four components should be extracted.

Table 5 Loading for each experiment to the four PCA components. Four components with percentage variance explained greater than 5% were extracted, representing 34.1%, 10.3%, 8.7% and 7.0%, respectively.

Experiment (TY%_hpi_treatment)	Component			
	1	2	3	4
6%_96hours_4C_Temperature	0.301			
6%_48hours_10C_Temperature	0.516			
6%_64hours_10C_Temperature	0.696			
6%_88hours_10C_Temperature	0.815			
6%_96hours_10C_Temperature	0.832			
6%_40hours_15C_Temperature	0.783			
6%_48hours_15C_Temperature	0.836			
6%_64hours_15C_Temperature	0.892			
6%_72hours_15C_Temperature	0.892			
6%_88hours_15C_Temperature	0.75			
6%_96hours_15C_Temperature	0.672			
6%_40hours_20C_Temperature	0.869			
6%_48hours_20C_Temperature	0.729			
6%_64hours_20C_Temperature	0.341		0.397	
6%_72hours_20C_Temperature			0.398	
6%_88hours_20C_Temperature			0.319	
100%_40hours_4C_Temperature				0.819
100%_48hours_4C_Temperature				0.849
100%_64hours_4C_Temperature				0.918
100%_72hours_4C_Temperature				0.916
100%_88hours_4C_Temperature				0.878
100%_96hours_4C_Temperature				0.875
100%_40hours_10C_Temperature	0.354			0.352
100%_48hours_10C_Temperature	0.518			0.327
100%_64hours_10C_Temperature	0.638			
100%_72hours_10C_Temperature	0.432			
100%_88hours_10C_Temperature	0.773			
100%_96hours_10C_Temperature	0.801			
100%_40hours_15C_Temperature	0.805			
100%_48hours_15C_Temperature	0.846			
100%_64hours_15C_Temperature	0.907			
100%_72hours_15C_Temperature	0.945			
100%_88hours_15C_Temperature	0.942			
100%_96hours_15C_Temperature	0.931			
100%_40hours_20C_Temperature	0.937			
100%_48hours_20C_Temperature	0.945			
100%_64hours_20C_Temperature	0.881			
100%_72hours_20C_Temperature	0.837			
100%_88hours_20C_Temperature	0.705	0.385		
100%_96hours_20C_Temperature	0.593	0.462		
6%_40hours_TY_concentration	0.507		0.335	
6%_48hours_TY_concentration	0.362		0.372	
12%_40hours_TY_concentration	0.853			
12%_48hours_TY_concentration	0.722			
12%_64hours_TY_concentration	0.328			
25%_40hours_TY_concentration	0.912			
25%_48hours_TY_concentration	0.851			
25%_64hours_TY_concentration	0.508			

25%_72hours_TY_concentration	0.326	0.36		
25%_88hours_TY_concentration		0.326		
25%_96hours_TY_concentration		0.34		
100%_40hours_TY_concentration	0.92			
100%_48hours_TY_concentration	0.889			
100%_64hours_TY_concentration	0.647	0.589		
100%_72hours_TY_concentration	0.405	0.765		
100%_88hours_TY_concentration		0.85		
100%_96hours_TY_concentration		0.874		
6%_48hours_protist_control			0.459	
6%_40hours_protist_Chilomonas			0.682	
6%_48hours_protist_Chilomonas			0.806	
6%_64hours_protist_Chilomonas			0.844	
6%_72hours_protist_Chilomonas			0.906	
6%_88hours_protist_Chilomonas			0.885	
6%_96hours_protist_Chilomonas			0.879	
6%_64hours_protist_Tetrahymena				0.362
6%_96hours_protist_Tetrahymena	0.356			
100%_40hours_protist_control	0.904			
100%_48hours_protist_control	0.859			
100%_64hours_protist_control	0.652	0.559		
100%_72hours_protist_control	0.486	0.733		
100%_88hours_protist_control		0.874		
100%_96hours_protist_control		0.841		
100%_40hours_protist_Chilomonas	0.912			
100%_48hours_protist_Chilomonas	0.861	0.303		
100%_64hours_protist_Chilomonas	0.552	0.548		
100%_72hours_protist_Chilomonas	0.425	0.596		
100%_88hours_protist_Chilomonas		0.755		
100%_96hours_protist_Chilomonas		0.801		
100%_40hours_protist_Tetrahymena	0.386			
100%_72hours_pH4			0.36	
100%_48hours_pH5	0.872			
100%_72hours_pH5	0.826			
100%_96hours_pH5	0.706	0.381		
100%_48hours_pH6	0.87			
100%_72hours_pH6	0.48	0.506		
100%_96hours_pH6	0.338	0.734		
100%_48hours_acid_control	0.709			
100%_72hours_acid_control	0.431	0.692		
100%_96hours_acid_control		0.757		
100%_48hours_alkali_control	0.784			
100%_72hours_alkali_control	0.54	0.509		
100%_96hours_alkali_control	0.362	0.638		
100%_48hours_pH7	0.743			
100%_72hours_pH7	0.517	0.47		
100%_96hours_pH7	0.314	0.589		
100%_48hours_pH8	0.815	0.316		
100%_72hours_pH8	0.546	0.53		
100%_96hours_pH8	0.356	0.585		
100%_48hours_pH9	0.315			

C.1 Ranking the strains

Table 6 Ranking the Rlt strains based on growth. A percentage of the highest growing strain for every treatment, adding those percentages together to get the "Total" and taking the highest "Total" and turning each value into a percentage of that (Percentage column).

Ranking (Best to Worst)	Strain (SM)	Genospecies	Geographic origin	Percentage
1	45	A	DK_C	100
2	4	B	UK	99.6
3	163A	A	DK_O	99.46
4	39	B	UK	98.19
5	14	B	UK	97.97
6	30	B	UK	97.96
7	135B	A	DK_O	97.95
8	136B	A	DK_O	97.34
9	145A	A	DK_O	96.63
10	35	B	UK	96.59
11	25	B	UK	96.45
12	52	A	DK_C	96.21
13	130B	A	DK_O	96.14
14	5	B	UK	96.08
15	11	B	UK	96.06
16	159	A	DK_O	95.89
17	152C	A	DK_O	95.84
18	31	B	UK	95.63
19	13	B	UK	95.6
20	126A	A	DK_O	95.59
21	6	B	UK	95.46
22	154B	A	DK_O	95.44
23	136A	A	DK_O	95.26
24	15	B	UK	95.18
25	97	C	France	94.99
26	37	B	UK	94.88
27	8	B	UK	94.55
28	154A	A	DK_O	94.52
29	16	B	UK	94.45
30	18	B	UK	94.09
31	144A	A	DK_O	94
32	155B	A	DK_O	93.92
33	3	B	UK	93.92
34	20	B	UK	93.71
35	96	C	France	93.65
36	123	A	DK_O	93.59
37	7	B	UK	93.38
38	94	C	France	93.32
39	152A	A	DK_O	93.3
40	24	B	UK	93.2
41	22	B	UK	93.03
42	107	C	France	93.02

43	27	B	UK	93
44	19	B	UK	93
45	146A	A	DK_O	92.97
46	155A	A	DK_O	92.83
47	12	B	UK	92.67
48	36	B	UK	92.58
49	40	B	UK	92.35
50	10	B	UK	92.35
51	152B	A	DK_O	92.2
52	160	E	DK_O	92.13
53	131	A	DK_O	91.95
54	140A	A	DK_O	91.93
55	168B	E	DK_O	91.69
56	21	B	UK	91.5
57	9	B	UK	91.43
58	168A	E	DK_O	91.41
59	135A	A	DK_O	91.38
60	155C	A	DK_O	91.31
61	72	D	DK_C	91.11
62	151A	A	DK_O	90.57
63	78	D	DK_C	90.55
64	79	D	DK_C	90.43
65	149A	E	DK_O	90.02
66	103	C	France	90.01
67	98	C	France	89.98
68	115	C	France	89.76
69	99B	C	France	89.7
70	120	A	France	89.62
71	85	C	France	89.45
72	84	C	France	89.37
73	151B	A	DK_O	89.18
74	32	B	UK	89.03
75	73	C	DK_C	88.98
76	17	B	UK	88.74
77	164B	D	DK_O	88.73
78	128B	A	DK_O	88.62
79	51	D	DK_C	88.58
80	91	C	France	88.56
81	109	C	France	88.5
82	38	B	UK	88.35
83	145B	A	DK_O	88.33
84	43	C	DK_C	88.14
85	146B	A	DK_O	87.87
86	141B	A	DK_O	87.87
87	133A	C	DK_O	87.87
88	90	C	France	87.79
89	137B	A	DK_O	87.66
90	42	C	DK_C	87.64
91	128A	A	DK_O	87.61
92	92	C	France	87.44
93	110	C	France	87.12

94	138A	A	DK_O	86.87
95	129	C	DK_O	86.85
96	147A	C	DK_O	86.47
97	130A	A	DK_O	86.39
98	113	C	France	86.32
99	104	C	France	86.08
100	81	C	France	86.05
101	111	C	France	85.95
102	166A	C	DK_O	85.84
103	55	C	DK_C	85
104	138B	A	DK_O	84.89
105	117	C	France	84.88
106	147B	C	DK_O	84.83
107	86	C	France	84.81
108	89	C	France	84.78
109	134A	C	DK_O	84.3
110	168C	C	DK_O	84.3
111	164A	C	DK_O	84.1
112	116	C	France	84
113	60	A	DK_C	83.81
114	170C	C	DK_O	83.58
115	106	C	France	83.57
116	100	C	France	83.47
117	114	C	France	83.44
118	153D	C	DK_O	83.41
119	118	C	France	83.36
120	58	C	DK_C	83.24
121	122A	C	DK_O	83.15
122	165A	C	DK_O	83.06
123	76	C	DK_C	82.82
124	125	C	DK_O	82.79
125	59	C	DK_C	82.66
126	57	C	DK_C	82.65
127	166B	C	DK_O	82.58
128	53	C	DK_C	82.41
129	68	C	DK_C	82.4
130	101	C	France	82.37
131	88	C	France	82.23
132	169	C	DK_O	82.1
133	153A	C	DK_O	82.1
134	148B	C	DK_O	81.93
135	56	C	DK_C	81.75
136	164C	C	DK_O	81.67
137	150	C	DK_O	81.48
138	67	C	DK_C	81.38
139	41	C	DK_C	81.07
140	149C	C	DK_O	81.05
141	170A	C	DK_O	80.95
142	69	C	DK_C	80.72
143	167	C	DK_O	80.53
144	70	C	DK_C	80.5

145	149B	C	DK_O	80.49
146	151C	C	DK_O	80.25
147	134B	C	DK_O	80.17
148	74	C	DK_C	79.81
149	157B	C	DK_O	79.62
150	44	C	DK_C	79.56
151	105	C	France	79.03
152	47	A	DK_C	78.94
153	148A	C	DK_O	78.77
154	87	C	France	78.72
155	121A	C	DK_O	78.71
156	132	C	DK_O	78.51
157	157A	C	DK_O	78.5
158	112	C	France	78.4
159	158	C	DK_O	78.28
160	54	C	DK_C	78.17
161	93	C	France	77.85
162	77	C	DK_C	77.29
163	61	C	DK_C	77.27
164	63	B	DK_C	77.15
165	49	A	DK_C	77.05
166	119	C	France	77.03
167	71	C	DK_C	76.36
168	127	C	DK_O	76.05
169	140B	A	DK_O	75.8
170	66	C	DK_C	75.49
171	102	C	France	75.4
172	80	C	DK_C	75.32
173	46	C	DK_C	75.19
174	34	B	UK	75.01
175	50	C	DK_C	74.84
176	143	C	DK_O	73.29
177	153C	C	DK_O	72.71
178	141A	C	DK_O	72.43
179	161	C	DK_O	72.25
180	48	C	DK_C	72.2
181	145C	A	DK_O	70.36
182	165B	C	DK_O	70.09
183	82	C	France	66.95
184	156	B	DK_O	66.16
185	154C	A	DK_O	65.96
186	137A	C	DK_O	65.82
187	83	C	France	65.7
188	95	C	France	62.83
189	108	C	France	61.87
190	163B	A	DK_O	57.65
191	126B	C	DK_O	48.31
192	144B	A	DK_O	46.56

Abbreviations

ANCOVA	analysis of covariance
AWCD	average well colour development
BNF	biological nitrogen fertiliser
DK_O	Denmark organic farms
DK_C	Denmark conventional farms
F	France
hpi	hours post inoculation
MANOVA	multivariate analysis of variance
OD_{600}	optical density at 600 nm wavelength
PCA	principal component analysis
PGPR	plant-growth promoting rhizobacteria
Rlt	*Rhizobium leguminosarum* bv. *Trifolii*
TY	Tryptone-Yeast (media)
UK	United Kingdom

References

1. Bohlool, B. B., Ladha, J. K., Garrity, D. P. & George, T. Biological nitrogen fixation for sustainable agriculture: A perspective. *Plant Soil* **141**, 1–11 (1992).
2. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–9 (2008).
3. Cameron, K. C., Di, H. J. & Moir, J. L. Nitrogen losses from the soil/plant system: a review. *Ann. Appl. Biol.* **162**, 145–173 (2013).
4. Yang, J., Kloepper, J. W. & Ryu, C.-M. Rhizosphere bacteria help plants tolerate abiotic stress. *Trends Plant Sci.* **14**, 1–4 (2009).
5. Van Loon, L. C. Plant responses to plant growth-promoting rhizobacteria. *Eur. J. Plant Pathol.* **119**, 243–254 (2007).
6. Lugtenberg, B. & Kamilova, F. Plant-Growth-Promoting Rhizobacteria. *Annu. Rev. Microbiol.* **63**, 541–556 (2009).
7. Gopalakrishnan, S. *et al.* Plant growth promoting rhizobia: challenges and opportunities. *Biotech* **5**, 355–377 (2015).
8. Hayat, R., Ali, S., Amara, U., Khalid, R. & Ahmed, I. Soil beneficial bacteria and their role in plant growth promotion: A review. *Ann. Microbiol.* **60**, 579–598 (2010).
9. Sturz, A. V., Christie, B. R. & Nowak, J. Bacterial Endophytes: Potential Role in Developing Sustainable Systems of Crop Production. *CRC. Crit. Rev. Plant Sci.* **19**, 1–30 (2000).
10. Jensen, E. S. *et al.* Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development* **32**, (2012).
11. Lambers, H., Mougel, C., Jaillard, B. & Hinsinger, P. Plant-microbe-soil interactions in the rhizosphere: An evolutionary perspective. *Plant Soil* **321**, 83–115 (2009).
12. Herridge, D. F., Peoples, M. B. & Boddey, R. M. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1–18 (2008).
13. Galloway, J. N. *et al.* Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70**, 153–226 (2004).
14. Galloway, J. N. *et al.* Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions. *Science* **320**, 889–892 (2008).
15. Hassen, A. I., Bopape, F. L. & Trytsman, M. Nodulation Study and Characterization of Rhizobial Microsymbionts of Forage and Pasture Legumes in South Africa. *World J. Agric. Res.* **2**, 93–100 (2014).

16. Byrnes, B. H. Environmental effects of N fertilizer use - An overview. Fertil. Res. **26**, 209–215 (1990).

17. Cameron, K. C., Di, H. J. & Moir, J. L. Nitrogen losses from the soil/plant system: a review. *Ann. Appl. Biol.* **162**, 145–173 (2013).

18. Vessey, J. K. Plant growth promoting rhizobacteria as biofertilizers. *Plant Soil* **255**, 571–586 (2003).

19. Marschner, H. Marschner's Mineral Nutrition of Higher Plants. (*Academic Press INC*, 2011).

20. Biró, B. *et al.* Interrelations between *Azospirillum* and *Rhizobium* nitrogen-fixers and arbuscular mycorrhizal fungi in the rhizosphere of alfalfa in sterile, AMF-free or normal soil conditions. *Appl. Soil Ecol.* **15**, 159–168 (2000).

21. Argaw, A. & Minalku, A. Inherent Soil Fertility as Affected by *Rhizobium* Inoculation and Inorganic N Application on Common Bean Common bean is considered as poor-fixer of atmospheric N compared with other. *Ethiop. J. Agric. Sci.* **26**, 27–48 (2015).

22. Schumpp, O. & Deakin, W. J. How inefficient rhizobia prolong their existence within nodules. *Trends Plant Sci.* **15**, 189–195 (2010).

23. Brockwell, J. & Bottomley, P. J. Recent advances in inoculant technology and prospects for the future. *Soil Biol. Biochem.* **27**, 683–697 (1995).

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

25. Wang, D., Yang, S., Tang, F. & Zhu, H. Symbiosis specificity in the legume - rhizobial mutualism. *Cell. Microbiol*. **14**, 334–342 (2012).

26. Flood, J. The importance of plant health to food security. *Food Secur.* **2**, 215–231 (2010).

27. Godfray, H. C. J. *et al.* Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).

28. Cassman, K. G. Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5952–5959 (1999).

29. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* **108**, 20260–20264 (2011).

30. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).

31. Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14**, 49–53 (1999).

32. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume–*rhizobium* mutualism. *Nature* **425**, 78–81 (2003).

33. Denison, R. F. & Kiers, E. T. Lifestyle alternatives for rhizobia: Mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiol. Lett.* **237**, 187–193 (2004).

34. Kiers, E. T., Palmer, T. M., Ives, A. R., Bruno, J. F. & Bronstein, J. L. Mutualisms in a changing world: An evolutionary perspective. *Ecol. Lett.* **13**, 1459–1474 (2010).

35. West, S. A., Kiers, T. E., Pen, I. & Denison, R. F. Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15**, 830–837 (2002).

36. Klinger, C. R., Lau, J. A. & Heath, K. D. Ecological genomics of mutualism decline in nitrogen-fixing bacteria. *Proc. R. Soc. B* **283**, 2015–2563 (2016).

37. Pongsilip, N. in Phenotypic and Genotypic Diversity of Rhizobia 3–48 (*Bentham Science Publishers,* 2012). doi:10.2174/97816080546191120101

38. Tak, A., Gehlot, P., Pathak, R. & Singh, S. K. in *Rhizobium* Biology and Biotechnology 215–245 (*Springer,* 2017). doi:10.1007/978-3-319-64982-5_11

39. Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. Establishing nitrogen-fixing symbiosis with legumes: how many *rhizobium* recipes? *Trends Microbiol.* **17**, 458–466 (2009).

40. Kumar, N. *et al.* Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol.* **5**, 140133 (2015).

41. Young, J. P. W. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7**, R34.1-20 (2006).

42. Giraud, E. *et al.* Legumes Symbioses: Absence of Nod Genes in Photosynthetic *Bradyrhizobia*. *Science* **316**, 1307–1312 (2007).

43. Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).

44. Demason, D. A., Hirsch, A. M., Fujishige, N. A., Kapadia, N. N. & Hirsch, A. M. A feeling for the micro-organism: structure on a small scale. Biofilms on plant roots. *Bot. J. Linn. Soc.* **150**, 79–88 (2006).

45. Mavingui, P. *et al.* Genes essential for nod factor production and nodulation are located on a symbiotic amplicon (AMPRtrCFN299pc60) in *Rhizobium tropici*. *J. Bacteriol.* **180**, 2866–74 (1998).

46. Long, S. R. Genes and signals in the *rhizobium*-legume symbiosis. *Plant Physiol.* **125**, 69–72 (2001).

47. Freiberg, C. *et al.* Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394–401 (1997).

48. Hartmann, A., Rothballer, M. & Schmid, M. Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. *Plant Soil* **312**, 7–14 (2008).

49. Nuccio, E. E. *et al.* Climate and edaphic controllers influence rhizosphere community assembly for a wild annual grass. *Ecology* **97**, 1307–1318 (2016).

50. Adhikari, D. *et al.* Genetic diversity of soybean-nodulating rhizobia in Nepal in relation to climate and soil properties. *Plant Soil* **357**, 131–145 (2012).

51. Vriezen, J. A. C., De Bruijn, F. J. & Nüsslein, K. Desiccation responses and survival of *Sinorhizobium meliloti* USDA 1021 in relation to growth phase, temperature, chloride and sulfate availability. *Lett. Appl. Microbiol.* **42**, 172–178 (2006).

52. Zogg, G. P. *et al.* Compositional and Functional Shifts in Microbial Communities Due to Soil Warming. *Soil Sci. Soc. Am. J.* **61**, 475 (1997).

53. Kunert, K. J. *et al.* Drought Stress Responses in Soybean Roots and Nodules. *Front. Plant Sci.* **7**, 1–7 (2016).

54. Rengel, Z. & Marschner, P. Nutrient availability and management in the rhizosphere: Exploiting genotypic differences. *New Phytol.* **168**, 305–312 (2005).

55. Bouhmouch, I., Souad-Mouhsine, B., Brhada, F. & Aurag, J. Influence of host cultivars and Rhizobium species on the growth and symbiotic performance of *Phaseolus vulgaris* under salt stress. *J. Plant Physiol.* **162**, 1103–1113 (2005).

56. Aydi, S., Drevon, J. J. & Abdelly, C. Effect of salinity on root-nodule conductance to the oxygen diffusion in the *Medicago truncatula-Sinorhizobium meliloti* symbiosis. *Plant Physiol. Biochem.* **42**, 833–840 (2004).

57. Hatimi, A. Effect of salinity on the association between root symbionts and *Acacia cyanophylla* Lind.: growth and nutrition. *Plant Soil* **216**, 93–101 (1999).

58. Marschner, P., Crowley, D. & Rengel, Z. Rhizosphere interactions between microorganisms and plants govern iron and phosphorus acquisition along the root axis - model and research methods. *Soil Biol. Biochem*. **43**, 883–894 (2011).

59. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 626–631 (2006).

60. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).

61. Fernández-Calviño, D. & Bååth, E. Growth response of the bacterial community to pH in soils differing in pH. *FEMS Microbiol. Ecol.* **73,** 149–156 (2010).

62. Shiro, S. *et al.* Genetic diversity and geographical distribution of indigenous soybean-nodulating *bradyrhizobia* in the United States. *Appl. Environ. Microbiol.* **79**, 3610–3618 (2013).

63. Bryant, J. A. *et al.* Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proc. Natl. Acad. Sci.* **105**, 11505–11511 (2008).

64. Weese, D. J., Heath, K. D., Dentinger, B. T. M. & Lau, J. A. Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* **69,** 631–642 (2015).

65. Heath, K. D., Stock, A. J. & Stinchcombe, J. R. Mutualism variation in the nodulation response to nitrate. *J. Evol. Biol.* **23**, 2494–2500 (2010).

66. Caballero-mellado, J. & Martinez-romero, E. Soil Fertilization Limits the Genetic Diversity of *Rhizobium* in Bean Nodules. *Symbiosis* **26**, 111–121 (1999).

67. Yan, J. *et al.* Abundance and diversity of soybean-nodulating rhizobia in black soil are impacted by land use and crop management. *Appl. Environ. Microbiol.* **80**, 5394–5402 (2014).

68. Depret, G. *et al.* Long-term effects of crop management on *Rhizobium leguminosarum* biovar *viciae* populations. *FEMS Microbiol. Ecol.* **51**, 87–97 (2004).

69. Van Der Heijden, M. G. A. *et al.* Symbiotic bacteria as a determinant of plant community structure and plant productivity in dune grassland. *FEMS Microbiol. Ecol.* **56**, 178–187 (2006).

70. McInnes, A., Thies, J., Abbott, L. & Howieson, J. Structure and diversity among rhizobial strains, populations and communities–a review. *Soil Biol. Biochem.* **36**, 1295–1308 (2004).

71. Bonfante, P. & Anca, I.-A. Plants, Mycorrhizal Fungi, and Bacteria: A Network of Interactions. *Annu. Rev. Microbiol*. **63**, 363–383 (2009).

72. Marschner, P., Crowley, D. & Yang, C. H. Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. *Plant Soil* **261**, 199–208 (2004).

73. Grayston, S. J., Wang, S., Campbell, C. D. & Edwards, A. C. Selective influence of plant species on microbial diversity in the rhizosphere. *Soil Biol. Biochem.* **30**, 369–378 (1998).

74. Ramey, B. E., Koutsoudis, M., Bodman, S. B. von & Fuqua, C. Biofilm formation in plant–microbe associations. *Curr. Opin. Microbiol.* **7**, 602–609 (2004).

75. Kamilova, F., Validov, S., Azarova, T., Mulders, I. & Lugtenberg, B. Enrichment for enhanced competitive plant root tip colonizers selects for a new class of biocontrol bacteria. *Environ. Microbiol.* **7**, 1809–1817 (2005).

76. Simons, M. *et al.* Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. *Mol. Plant. Microbe. Interact.* **9**, 600–7 (1996).

77. Moeskjær, S. Quantitative genetics of white clover nitrogen fixation efficiency Understanding the interactions of *Rhizobium leguminosarum* and *Trifolium repens* through quantitative genetics. (*Aarhus University*, 2017).

78. Howieson, J. G. & Dilworth, M. J. Working with rhizobia. (2016).

79. Friman, V.-P., Dupont, A., Bass, D., Murrell, D. J. & Bell, T. Relative importance of evolutionary dynamics depends on the composition of microbial predator–prey community. *ISME J*. **10**, 1352–1362 (2016).

80. Garland, J. L. & Mills, A. L. Classification and characterisation of heterotrophic microbial communities on the basis of pattern of community-level sole-carbon-source utilization. *Appl. Environ. Microbiol.* **57**, 2351–2359 (1991).

81. Marschner, P., Yang, C. H., Lieberei, R. & Crowley, D. E. Soil and plant specific effects on bacterial community composition in the rhizosphere. *Soil Biol. Biochem.* **33**, 1437–1445 (2001).

82. O'Toole, G. A. Microtiter dish biofilm formation assay. *J. Vis. Exp.* (2011). doi:10.3791/2437

83. Bradley, J. V. Nonrobustness in Z, t, and F tests at large sample sizes. *Bull. Psychon. Soc.* **16**, 333–336 (1980).

84. Johnson, L. R. Microcolony and biofilm formation as a survival strategy for bacteria. *J. Theor. Biol.* **251**, 24–34 (2008).

85. Reuter, H. I., Rodriguez Lado, L., Hengl, T. & Montanarella, L. Continental-Scale digital soil mapping using European Soil Profile Data: Soil pH. *Hamburg. Beiträge zur Phys. Geogr. und Landschaftsökologie* **19**, (2008).

86. Laursen, E. V., Thomsen, R. S. & Cappelen, J. Observed Air Temperature, Humidity, Pressure, Cloud Cover and Weather in Denmark with Climatological Standard Normals, 1961-90. *Danish Meteorol. Inst.* (1999).

87. Lappin-Scott, H. M. & Costerton, J. W. in Microbial Biofilms 207–220 (*Cambridge University Press*, 1995).

88. Habte, M. & Alexander, M. Protozoan Density and the Coexistence of Protozoan Predators and Bacterial Prey. *Ecology* **59**, 140–146 (1978).

89. Jones, D. L., Hodge, A. & Kuzyakov, Y. Plant and mycorrhizal regulation of rhizodeposition. *New Phytol.* **163**, 459–480 (2004).

90. Sugiyama, A., Ueda, Y., Zushi, T., Takase, H. & Yazaki, K. Changes in the Bacterial Community of Soybean Rhizospheres during Growth in the Field. *PLoS One* **9**, e100709 (2014).

91. Kyei-Boahen, S., Nleya, T., Hynes, R. & Walley, F. L. Single and Multistrain Rhizobial Inocula for Pinto and Black Bean Cultivars. *J. Plant Nutr.* **28**, 1679–1692 (2005).