Rhizobia: from saprophytes to endosymbionts [Au: It's best if the title starts off with the most important term and if it is as short as possible. Also, I wonder whether 'from saprophytes to endosymbionts' would work better.]

Philip Poole[1], Vinoy Ramachandran[1], Jason Terpolilli[2].
Department of Plant Sciences University of Oxford, South Parks Road, Oxford OX1 3RB, UK[1]. Centre for Rhizobium Studies, School of Veterinary and Life Sciences, Murdoch University, South Street, Perth, Australia[2].

Correspondence to philip.poole@plants.ox.ac.uk

[Au: I've edited the abstract for length and flow. The shorter and clearer the abstract the better.] Abstract | Rhizobia are some of the best studied plant microbiota. These oligotrophic Alpha- or Beta-proteobacteria form symbioses with their legume hosts. Rhizobia must exist in soil and compete with other members of the microbiota before infecting legumes and forming N_2-fixing bacteroids. These dramatic life-style and developmental changes are underpinned by large genomes and even more complex pan-genomes, which encompass the whole population and are subject to rapid genetic exchange. The ability to respond to plant signals and chemoattractants and colonise nutrient-rich roots are crucial for competitive success of these bacteria. The availability of a large body of genomic, physiological, biochemical and ecological studies make rhizobia unique models for investigating community interactions and plant colonisation.

[Au: With ~6300 words your article is substantially over our word limit. I've edited throughout to streamline the article.]

The rhizosphere, which is the region of soil that surrounds plant roots, is under strong selection pressure for particular microorganisms[1]. This selection pressure is even greater in the rhizoplane on the root surface and in the endosphere, the endophytic compartment inside the root cortex **[G]** between plant cells **[Au:OK?]** . The interaction is bidirectional —plants select microorganisms and the microorganisms influence plant health, either being plant-growth promoting or pathogenic. One of the best studied

plant-microorganism interactions is the symbiosis between rhizobia and legumes, such as soybean, chickpea, lentil, pea, common bean, alfalfa and clovers. Rhizobia are nitrogen-fixing bacteria, which elicit the formation of root nodules in which they differentiate into bacteroids **[Au:OK?]**. These symbioses provide a significant proportion of the biosphere's nitrogen[2]. Legumes are some of the world's most important crop and fodder plants[2], although their importance has diminished with the introduction of industrially produced fertilizer, not without consequences for the global nitrogen cycle[3] **[Au:OK?]**.

Rhizobia must survive in soil, colonise roots and gain entry to the plant. They most often enter through root hairs; although infection through epidermal cracks is possible **[Au:OK?]**. When entering root hairs, bacteria grow down a plant-derived infection tube and stay in the extracellular space (Box 1). Concomitantly, nodule development proceeds in the plant cortex and eventually rhizobia enter the cytoplasm of nodule cells. There they differentiate into bacteroids — the N_2-fixing form of rhizobia. Bacteroids, together with bacterial and plant-derived membranes, form the symbiosome. Low oxygen concentrations in the nodule, allow nitrogenase to reduce N_2 to ammonia[4,5]. In this symbiosis, plants provide carbon and energy in the form of dicarboxylic acids and in return the bacteroids secrete ammonia, which the plant uses to synthesize amino acids. Most rhizobia have a narrow host range (Table 1); for example, *Rhizobium leguminosarum* biovar *trifolii* only infecting clover (*Trifollium* species). However, a few rhizobia are promiscuous; for example, *Sinorhizobium* NGR234 infects 112 genera of legumes[6].

Much of the rhizobia research has focused on specific parts of the rhizobial-legume symbioses (reviewed in [5,7-11]), without considering rhizobia as members of complex soil communities. Rhizobia must survive the often harsh conditions in soil and compete with the local microbiota before establishing symbiosis. We lack a clear spatial and temporal map of the various stages in the metabolic and genetic differentiation of rhizobia. Furthermore, most studies use single rhizobial strains under sterile conditions, without tacking competition with the microbiota into account.

The rhizobia are excellent model bacteria for understanding how plants interact with the root microbiota, in particular, due to the different interactions with host and non-host plants. There is a wealth of sequenced rhizobial genomes from defined soils and legume rhizospheres, which show how large pan-genomes **[G]** enable metabolic plasticity in populations. In this Review, we link the genomic data with detailed studies

on rhizobial physiology, biochemistry and genetics to provide a detailed picture of how rhizobia proceed through developmental stages.

[H1] The rhizosphere community

Soil supports a diverse microbial community[12,13], with up to 10^4 bacterial species and 10^9 bacterial cells per g of soil. Abiotic factors, such as soil pH, influence the composition of the soil microbiota, with Acidobacteria becoming dominant under acid conditions, whereas Proteobacteria, Actinobacteria and Firmicutes dominate in neutral or alkaline soils[12,14]. Plants strongly influence the rhizosphere microbiota[12,15-17], which differ substantially from the microbiota in bulk soil. The total microbial diversity in the rhizosphere is usually lower than in soil probably because it is so highly selected by the plant[12,15]. This is particularly evident in the rhizoplane and endosphere. Different *Arabidopsis thaliana* strains select different *Pseudomonas* strains **[Au:OK?]** in their root-associated microbiota **[Au: I've deleted the details about sonication etc to cut lenght. If this is important you could add root-associated microbiota as a glossary term.]**, which suggests that plant genes at least partly control which microorganisms are selected[18]. Similarly, repeated cultivation of a plant in the same soil **[Au: the same soil?]** can result in a suppressive soil, in which build-up of microbiota 'suppresses' the growth of pathogens[19], for example, through the increased abundance of Gammaproteobacteria that produce peptide antibiotics[19]. Legumes, such as pea, *Medicago truncatula*, *Lotus japonicus,* and cereals, such as barley, also have strong effects on the rhizosphere[12,15,20,21] and the strongest selection of the microbiota is usually in the root-associated microbiota. In most soils, rhizobia are common members of the community, regardless of the presence of a legume[22].

[H2] Rhizobial genomic diversity

As saprophytes, rhizobia survive in a complex microbial community by adopting an oligotrophic lifestyle **[G]**. Inside the host legume they differentiate into endosymbiotic bacteroids. This drastic lifestyle change in rhizobia likely underlies the evolution of some of the largest bacterial genomes (~5-10 Mb). Rhizobia encode several genes that are important for endosymbiosis in their chromosome or on plasmids: *nod* genes encode proteins that are involved in lipochitooligosaccharide (LCO) signalling to stimulate nodule formation; *nif* genes encode proteins essential for N_2-fixation, which are shared by most N_2-fixing bacteria; and *fix* genes encode proteins specifically

required by rhizobia for N_2-fixation (Fig 1). Rhizobial plasmids vary in number and size and include core and/or accessory genes. For example, *Sinorhizobium meliloti* 1021 has one chromosome and two large plasmids, pSymA (1.4 Mb) and pSymB (1.7 Mb). pSymA encodes *nod*, *nif* and *fix;* pSymB encodes the exopolysaccharide biosynthesis clusters *exo* and *exp*, which are essential for host infection, *bacA*, which is involved in bacteroid development, *dctABD*, which participates in carbon uptake and tRNA*arg* and *engA*, which are essential for free-living growth **[Au:OK?]** [23]. Thus, pSymA is a symbiotic plasmid, whereas pSymB is considered a chromid[24] — a hybrid replicon with features of both plasmids and chromosomes. The pea-nodulating *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) and common bean-nodulating *Rhizobium etli* CFN42 harbour six plasmids, with most *nod* and *fix* genes on plasmids pRL10 and p42d, respectively[25]. Symbiosis genes in the Betaproteobacteria *Cupriavidus taiwanensis* LMG19424 and *Burkholderia phymatum* STM815 are also encoded on plasmids[26,27]. By contrast, *Mesorhizobium* spp. carry one or two high molecular weight plasmids (>200 kb), with few recognizable symbiotic genes. These rhizobia, as well as *Bradyrhizobium* spp., carry their symbiotic genes on the chromosome[28-31] (Fig 1).

Plasmids, which encode symbiosis genes, can be transferred to other rhizobia, with estimates of transfer rate ranging 10^2–10^9 donor bacteria needed for one successful transfer[32-35]. Plasmid transfer is often controlled by quorum sensing, with a high bacterial cell density inducing transfer[32,35]. In the soil, such high cell densities could be achieved in the rhizosphere or rhizoplane, which suggests that the host influences the spread of symbiosis genes. Indeed, *Rhizobium* spp. and *Sinorhizobium* spp. populations from legume rhizospheres or nodules have much less genetically diverse symbiosis plasmids than chromosomes or accessory plasmids[36-40], probably due to transfer of the symbiosis plasmids in these populations. Furthermore, some rhizobia encode type IV or type III secretion systems, which can increase symbiotic efficiency **[Au:OK?]** [38,41]. The large genomic diversity of rhizobia reflects the multiple soil conditions, competitive challenges and legume hosts that they encounter. No single organism can support the huge metabolic diversity found in a pan-genome but as long as there are efficient transfer methods, favourable traits can rapidly spread through a population.

Although *nod*, *nif* and *fix* genes are chromosomally-encoded in *Mesorhizobium* spp., they are easily transferred, because they are usually located in integrative and conjugative elements (ICEs) **[G]** (Fig 1) **[Au: I've moved the ICE definition to the**

glossary to cut length.]. *Mesorhizobium* spp. ICEs consist of a single contiguous region of ~ 500-kb[42] or can be structurally complex, such as tripartite symbiosis ICEs (or ICE[3]), which exist as three separate regions in the chromosome that recombine into a single element before transfer[43]. *Azorhizobium caulinodans*, one of the few rhizobia that can fix N_2 in its free-living state, also harbours a mobile ICE[44]. This ICE is relatively small (86-kb), encodes *nod* genes (but not *nif* or *fix* genes) and can be transferred to other rhizobial genera[44]. The plant flavonoid naringenin increased transfer of this ICE 50-fold. The effect of legume root exudates on transfer is not known for *Mesorhizobium* spp. ICEs, but transfer rates are 10-100-fold higher in stationary phase cell cultures owing to quorum sensing[45]. *Bradyrhizobium* strains also harbour symbiotic genes in chromosomal ICE-like regions in their genomes[30,46], but it is unclear whether these elements are transmissible.

[H1] Colonisation of roots and signalling [Au: Main headings have to be <39 characters, including white spaces, to fit into our layout.]

[H3] Root and soil environment. Rhizobia use a wide range of nutrients and detoxify toxins and phytoalexins **[G]** , which are secreted by their host plant[47]. Comparative transcriptomics of *Rhizobium leguminosarum* bv *viciae* 3841 (Rlv3841) during colonisation of pea, alfalfa and sugar beet rhizospheres, revealed that organic and aromatic amino acids fuel bacterial metabolism, as evidenced by the expression of seventy-six ABC and TRAP transporters[48]. Signature-tagged mutagenesis showed that catabolism of arabinose and protocatechuate is essential for fitness in the pea rhizosphere[49]. Transposon insertion sequencing (INSeq) (Box 2), has enabled genome scale identification of essential genes in Rlv3841 during growth in complex media[50], minimal medium[51] and on minimal media supplemented with glucose or succinate at 1% or 21% O_2[52]. Growth on succinate and 1% O_2, mimics the conditions bacteroids encounter in nodules (see below). Transposon insertion sequencing is a powerful technique to identify essential genes during root attachment, rhizosphere colonisation, nodulation and N_2 fixation, and it will transform our understanding of adaptation to these environments.

Genome-scale metabolic modelling of *S. meliloti,* was used to predict adaptations to bulk soil, rhizosphere and nodule environments[53]. Similar reprogramming of metabolic pathways occurred in rhizosphere and bulk soil, except for greater metabolic diversity in the rhizosphere, which enables utilisation of complex

root exudates. The model also indicated that fitness for host-associated niches depends on plasmid genes **[Au:OK?]**. This is consistent with the experimental observation that plasmid pRL8 of Rlv3841 is enriched with genes that are preferentially expressed when the bacterium grows in association with its pea host[49] **[Au:reference?]** . Similarly, pRL8 encodes genes for the utilisation of the pea rhizosphere-specific compound homoserine, and their mutation impaired transfer of pRL8[54]. Recently, a group of solute-specific Lux biosensors was developed based on the transportome profile of *S. meliloti* and *R. leguminosarum* during rhizosphere colonisation[48,55]. These biosensors are non-invasive, do not require added substrate and are environmentally stable[56], which makes them ideal for probing spatial and temporal plant-microorganism interactions.

[H3] Chemoattraction to roots. Until they attach and form a biofilm on roots and root hairs, rhizobia must survive in the soil, using a wide range of carbon and nitrogen sources. Colonisation of roots likely requires bacterial motility and chemotaxis (Fig. 2). Indeed, mutation of *flgE*, which encodes the flagellar hook protein, in *Mesorhizobium tianshanense* compromised attachment to liquorice roots[57]. Furthermore, *R. leguminosarum* 3841 encodes two chemotaxis clusters, *che1* and *che2*[58]. *Che1* is highly conserved across alpha-proteobacteria **[Au: across *Rhizobium* spp.?]** whereas *che2* is less common. In addition, *che1* is needed for competitive nodulation of *R. leguminosarum*, which indicates that chemotaxis is crucial at some stage of colonisation, attachment to root hairs or infection of nodules[58].

[H3] LCO synthesis [Au: Ok? I've changed Nod factor to LCO to make the nomenclature consistent.] . The transcriptional regulator NodD detects plant-derived flavonoids, which induces *nod* expression and the synthesis of LCOs (Fig. 3). **[Au:Definiton of LCOs moved to glossary.]** *NodABC* encode the pathway for the synthesis of the common core chitooligosaccharide, *nodIJ* encode an ABC export system for LCOs and other *nod* genes are responsible for decorating the LCO core. The plant host range can be very narrow, for example *R. leguminosarum b.v. trifolii* produce LCOs that only allow infection of clover **[Au:OK? Or is only one LCO produced?]**. By contrast, *Sinorhizobium* NGR234 produces a range of LCOs and infects 112 genera of legumes[6]. Legumes detect LCOs with a pair of receptors, which initiates the common symbiosis (SYM) signalling pathway shared by arbuscular

mycorrhizae and rhizobia[8] (Box 1). **[Au: I've removed some of the details here to cut length.]** SYM controls flavonoid synthesis[59], which results in a positive feedback loop between LCOs and flavonoids and leads to root hair curling, rhizobial infection and nodule formation (Fig. 2). Indeed, localised bursts of flavonoid synthesis can be detected where rhizobia are attached to roots before the appearance of nodules[60]. Remarkably, intercropping of maize with faba beans substantially enhances root hair curling, nodulation and N_2-fixation by the legume[61]. The likely depends on maize root secretions, which increase flavonoid synthesis by faba bean and consequently rhizobial *nod* gene expression.

[H3] SYM pathway and community composition. Blocking SYM signalling changed the relative abundance **[Au:OK?]** of 14 bacterial orders in both the rhizosphere and root-associated communities in *L. japonicus*[20]. This might be due to biochemical changes associated with nodules, such as the supply of nitrogen and the release of H_2 as a by-product of the nitrogenase reaction. Indeed, H_2 secretion from nodules can change the microbiota[62]. However, microbiota changes occurred even when added nitrate suppressed nodule formation, which suggests that the effect of SYM signalling is independent of nodule formation. The root-associated and nodule communities were selected by parallel rather than sequential pathways **[Au: What do you mean by this? Concurrently?]** , consistent with nodulation relying specifically on root hair infection[63]. Key questions include the role of the SYM pathway in non-legumes and how nodulation specific and mycorrhizal specific pathways interact. *Nin* mutants, which lack nodulation but not mycorrhization, changed the microbiota composition. The effects of mycorrizhal-specific mutants such as *ram*[64,65] or common SYM mutants in cereals, such as rice[66], have yet to be tested. In general, legumes profoundly influence the rhizosphere microbiota with a strong selection for fungi over bacteria[12].

[H3] Plant immunity. Plant immunity regulates selection of the microbiota and rhizobia must modulate the immune response for successful infection of legumes (Fig. 3). Rhizobial LCOs suppress MAMP-triggered immunity in *A. thaliana,* tomato, corn and soybean[67]. Furthermore, rhizobial flagellin does not elicit the immune response caused by most bacterial flagellins[68]. Salicylic acid signalling is important for selection of the microbiota by *A. thaliana*[69]. Signalling between symbiotic interactions and the immune response overlaps. LCOs are detected by a heterocomplex of LysM

receptors[8]. The rice protein OcCERK1 is a LysM receptor[70,71], which, depending on the other LysM receptor it interacts with, detects chitin or initiates mycorrhizal signalling[72,73] **[Au: Ok? I've edited this make it easier to understand. Also, as the plant signalling pathways are not the main focus of the article, less detail is needed.]**. There may be multiple overlapping pairs of LysM proteins recognising LCOs, mycorrhizal (Myc) factor and immune triggers, such as chitin and peptidoglycan. LysM proteins are widely conserved in plants with multiple copies in individual species[74,75], which enables discrimination between symbionts, pathogens and the general microbiota. Perception by legumes of the extracellular polysaccharide layer (EPS) **[Au: EPS is usually used for extracellular polymeric substances; do you mean other polysaccharides here?]** layer of rhizobia is crucial to infection and a bacterial EPS receptor 3 (Epr3) has been identified in *L. japonicus*[76]*.* EPS sensing is required for bacterial passage through the epidermal cell layer and likely involves modulation of the plant immune response. Furthermore, secretion of nodulation effector proteins **[Au:LCOs? Other effectors?]** by type III secretion systems can widen or restrict the host range of rhizobia[77]. More recently it has been shown that the type III secretion system of *B. elkanii* can induce nodulation in the absence of LysM receptors **[Au:OK?]** [78]. **[Au: I've removed some details to cut length]** Furthermore, some photosynthetic bradyrhizobia, which lack *nod* genes, can nodulate *Aeschynomene* plants[79].

[H3] Attachment of rhizobia to roots. Bacterial colonisation of roots and attachment to root hairs (Fig 3), has received less attention than nodulation signalling. As discussed above the processes are interlinked as root association amplifies flavonoid signalling[59,60]. At acidic pH, *R. leguminosarum* uses glucomannan at its cell pole to attach to lectins on pea root hairs[56,80] and the root surface[41]. However, at basic pH, glucomannan is not involved, probably because plant lectins are lost. Attachment to root hairs at alkaline pH remains poorly understood, although it has been proposed that an unidentified calcium-binding protein (rhicadhesin) mediates attachment[81]. Once rhizobia attach to root hairs they aggregate, forming a biofilm, which is encased in cellulose and extracellular polysaccharide and is called a root hair cap[81-83]. Several proteins that are exported through the PrsDE type I secretion system are required for stability of the cap. Most of these secreted proteins contain one or two rhizobia-

adhering or Cadherin-like domains (CHDLs), such as *Rhizobium*-adhering proteins (Raps)[84-87]. However, *prsDE* mutants of Rlv3841 showed no nodulation defect, whereas a double *prsD* and *nodE* mutant showed delayed nodulation[88]. Although this might indicate that nodulation is independent of root colonisation, it is more likely that the extreme competition faced in soil is difficult to replicate in laboratory experiments.

RapA1 is a calcium-binding protein, which is localised at the cell pole[85]. Overexpression of RapA1 in *R. leguminosarum* bv. *trifolii* R200 increased attachment of rhizobia to red clover roots up to five-fold[89]. Similarly, increased attachment occurred in the non-cognate hosts soybean and alfalfa, whereas over-expression of RapA1 from *R. etli* increased attachment to common bean roots. Overexpression did not increase nodulation, but did increase competitiveness for nodule occupancy in clover roots[90]. Similarly, a RapA1 overexpressing strain of *R. leguminosarum* PVu5 formed twice as many nodules as the wild-type strain on kidney beans[91]. In these studies, attachment was not spatially resolved (for example, root hair or epidermal surface), which highlights that root attachment has been measured in different ways, without determining bacterial location. It is unclear whether different pathways are involved in binding to root hairs or the root epidermis, and whether one precedes the other.

RapA2 binds directly to acidic EPS and capsular EPS[92], to stabilise biofilms. Deletion of the transcriptional repressor of quorum sensing, *pra*R in Rlv3841 up-regulated *rapA2* and *rapC* expression, which increased root attachment to pea roots five-fold[56]. This effect was abolished in a *praR*, *rapA2* and *rapC* triple mutant, leading to reduced nodule competitiveness **[Au:OK?]** [56]. An important caveat in comparing attachment studies is that bacterial growth, plant growth and attachment measurements may have been performed at different pHs.

Mutation of *pssA*, which regulates EPS biosynthesis, reduces attachment of *R. leguminosarum* to root hairs and cap formation[80]. However, attachment to the root epidermis was still observed when cells were resuspended at acidic or alkaline pH. A *pssA* mutant of *R. leguminosarum* bv. *trifolii* Rt24.2 was highly attenuated in attachment to whole clover roots[93] **[Au:reference?]**. Likewise, deletion of the global regulator *rosR*, which also controls EPS biosynthesis, reduced attachment and biofilm formation on the root surface and root hairs, whereas *rosR* overexpression increased attachment[94,95].

Cellulose fibrils are important for a tight adherence between rhizobial cells during biofilm formation (Fig 3). Mutation of *celA*, which encodes a cellulose synthase, did not alter rhizobial attachment to root hairs, but prevented cap formation under acidic and alkaline conditions[80]. Similarly, deletion of *celC2*, which encodes an endoglucanase, halved the level of attachment to whole roots compared to wild-type. The *celC2* mutant formed irregular aggregates on the root surface and formed a thicker biofilm cap on root hairs[96].

Lipopolysaccharide (LPS) is also important for attachment. *R. leguminosarum* strains with mutations in *lpcA*, *lpcB* or *lpsD*, all of which encode enzymes needed for the synthesis of the O-antigen in the LPS core, formed scattered star-like micro-colonies with poor anchorage to the root surfaces and formation of a thicker root cap[97]. In Rlv3841, mutation in the lipid A component of LPS altered biofilm formation on solid surfaces[98].

Recently, advances have been made in understanding the spatial and temporal regulation of attachment using laser capture microdissection and RNAseq of root nodules and microarray analysis of plant nitrogen fixation mutants **[Au:OK?]** [99,100]. A transformational technique to study the early steps of attachment is a microfluidics system called tracking root interaction system (TRIS), which tracks the colonisation of roots with confocal microscopy[101]. Within 20 min after inoculation, *Bacillus subtilis* was observed to colonize the root elongation zone (REZ) of *A. thaliana*. It is unclear where rhizobia first accumulate and whether binding to the REZ or epidermis precedes attachment to root hairs. **[Au: I've deleted the following sentences about the Lux reporters to cut length and because these are already discussed above.]**

[H1] Bacteroid development

[H3] Plant control and NCR peptides. The differentiation into bacteroids is accompanied by large changes in the transcriptome and proteome: N_2-fixation genes are induced, but most genes required for growth, including those encoding ribosomal proteins and involved in DNA replication and amino acid biosynthesis, are repressed (Fig 4) [102-105]. Legumes of the invert repeat lacking lade (IRLC), such as *M. truncatula*, peas, chickpeas and faba beans, produce up to 700 nodule cysteine rich (NCR) peptides[106,107]. Peptide expression is highly spatially resolved from proximal to distal parts of *M. truncatula* nodules[107-109]. They induce bacterial cell swelling and genome endoreduplication **[G]** and increase permeability of the cell membrane,

which causes terminal differentiation[109]. Peptide NCR247 upregulates the master two component sensor-regulator systems that control exopolysaccharide and cyclic glucan synthesis. It perturbs the cell cycle of *S. meliloti* and inhibits the master cell control regulator CtrA and cell cycle components including FtsZ[110]. Inhibition of CtrA causes cell cycle arrest, swelling and branching in cultured bacteria[111]. NCR247 occurs in three different oxidised regioisomers and a reduced form, which differ in their effects. Blocking secretion of NCR peptides, prevents bacteroid formation in *M. truncatula*[108,112]. Different IRLC legumes cause bacteroids to become either swollen, elongated, spherical, or elongated-branched[113]. The swollen morphotype may be ancestral because it is associated with the fewest changes in the bacteroids and the plants that induce this morphotype produce the fewest NCR peptides.

M. truncatula primarily produces NCR247 in zone II and the II/III interzone (Box 1), where bacterial cell division is arrested and cell elongation occurs[114], although it does have some weak expression in other tissues[107]. It localises to punctate regions in the cytoplasm and interacts with several proteins in *S. meliloti*: the main GroEL chaperonin, CPN60, which is essential for bacteroid development[115], FtsZ, pyruvate dehydrogenase and nitrogenase[114]. In addition, NCR247 can influence many proteins involved in metabolism, N_2 fixation and cell division, maybe in a large N_2-fixing complex. Although NCR035 is produced at the same time and place as NCR247, it targets the division septum of *S. meliloti*[108]. NCR169 and NCR211, which function at a later stage, were the first individual NCR peptides shown to be essential for bacteroid formation[116,117]. They can be found in the symbiosome, and, although NCR211 inhibits growth *in vitro*, both are required for bacteroid persistence. The timing and level of inhibition by particular NCR peptides is probably crucial for bacteroid differentiation.

NCR peptides can also control rhizobial host specificity. *NSF1* and *NSF2* from *M. truncatula* A17 promote cell death and early nodule senescence **[G]** of *S. meliloti* Rm41, resulting in nodules unable to fix N_2 [118,119]. However, numerous other strains of *S. meliloti* can form N_2-fixing nodules on this cultivar and *S. meliloti* Rm41 forms functional nodules on other *M. truncatula* cultivars. In a further twist, ~20% of *S. meliloti* strains produce a metallopeptidase, HrrP, which can degrade NCRs, change the host range and prevent N_2 fixation on specific cultivars[120]. HrrP increases recovery of bacteria ~5-fold from fixing and non-fixing nodules, suggesting HrrP attenuates the antimicrobial effects of NCR peptides. BacA also protects against the antimicrobial effects of NCR peptides, perhaps as a part of an ABC-type exporter[121,122].

Nodule development can occur also without NCR peptides. The non-IRLC legume *Leucaena glauca* does not produce NCR peptides, but forms indeterminate nodules and swollen bacteroids[123]. *S. meliloti* HH103 is resistant to NCR247 and more resistant than *S. meliloti* 1021 to NCR35, irrespective of host[124]. Thus, NCR peptides are a regulatory system imposed by the plant, but are not a fundamental requirement of rhizobia-legume symbioses, which suggests that they have evolved as a control mechanism in some legumes. NCRs may change the efficiency of N_2 fixation. Plants that produce NCR peptides have a lower nodule mass and greater nitrogenase activity than plants that produce no NCR peptides[125]. The mechanisms that underlie this effect are unclear and requires study.

Plants control bacteroid N_2 fixation in bacteroids through homocitrate, which is a ligand of the FeMo cofactor in nitrogenase[126]. Most rhizobia lack *nifV*, which is needed to synthesize homocitrate, and thus cannot fix N_2 *ex-planta*[127]. Another control mechanism is symbiotic auxotrophy — bacteroids depend on the uptake of branched chain amino acids from the plant[128-130]. In summary, many aspects of metabolic control have passed from the bacteria to the plant, supporting the hypothesis that bacteroids are ammoniaplast-like organelles[10].

However, nodule development and bacteroid formation is not entirely plant controlled, as exemplified by the HrrP peptidase[120]. This bacterial enzyme may enable 'cheating' through the formation of ineffective non-fixing nodules and the release of large numbers of bacteria. Furthermore, *S. meliloti* has three cAMP synthases CyaD1, CyaD2 and CyaK, which are activated by a plant metabolite present in nodules[131]. Increased cAMP activates the Crp-like activator Clr, which drives transcription of a gene of unknown function. This cascade suppresses formation of new infection threads by plants. In another example, mature bacteroids of some *S. meliloti* and *R. leguminosarum* strains produce rhizopines[132], which are a potential carbon and nitrogen source for undifferentiated sibling bacteria.

[H3] Regulation of N_2 fixation. Bacteroids are fully differentiated when they commence the reduction of N_2 to ammonia through the nitrogenase enzyme complex (Box 1). The main factor controlling N_2 fixation in bacteroids is O_2 tension, with low O_2 levels in nodules (10-21.5 nM)[133]. In *S. meliloti,* the haem-containing O_2-binding protein FixL initiates a phosphorylation cascade under low O_2 tension, which activates its cognate regulator protein FixJ[134,135] (Fig. 4). FixJ induces FixK, which upregulates

the expression of *fixNOPQ* and associated operons such as *fixGHIS*. *FixNOPQ* encodes the high-affinity terminal oxidase cbb3 complex. In *S. meliloti* the main regulator of *nif* genes, *nifA,* is also partially FixK-regulated with a cognate promoter immediately upstream of *nifA*[136,137]. *NifA* also autoregulates the *fixABCX-nifAB* operon. In *R. leguminosarum* FnrN largely replaces FixL, FixJ and FixK. FnrN probably detects O_2 tension directly and it initiates transcription of *fixNOQP* and *fixGHIS*[138,139]. However, FnrN does not regulate *nifHDK* or *fixABCX-nifAB*, which appear to be strictly regulated by NifA[140]. In *B. japonicum, nifA* and up to 250 other genes are controlled by the two-component redox sensor RegSR[141]. Other plant signals might control *nifA* expression; for example, in *Mesorhizobium loti* R7A FixV is needed to induce *nifA*[142]. The gene that encodes FixV clusters with genes that might metabolise a plant inositol derivative abundant in nodules. If this mechanism can be verified, this would be an exciting example of a nodule-specific compound that controls N_2 fixation.

[H3] The benefits of symbiosis — nutrient exchange. Plants provide bacteroids with dicarboxylates, such as L-malate, succinate and fumarate,[143] as energy and carbon sources (reviewed in [10,127,144]), and in turn, bacteroids secrete rather than assimilate ammonium[145,146]. By contrast, free-living diazotrophs assimialte ammonium. Although the activity of enzymes in the tricarboxcylic acid (TCA) cycle increases in bacteroids, bacteroids from beans and soybeans accumulate large amounts of polyhydroxybutyrate (PHB). PHB is made under low O_2 conditions, where acetyl-CoA is metabolised to PHB for storage instead of entering the O_2-limited TCA-cycle. Pea bacteroids also accumulate both PHB and lipid, which also implies O_2-limitation of the TCA-cycle[147]. The increased TCA-cycle enzyme activity might be a response to limitation of key intermediates of the cycle rather than to growth on dicarboxylates. Rlv3841 depends on several metabolic pathways to grow on sugars and dicarboxylates at high and low O_2 tension but PHB synthesis is dispensable[52]. At low O_2 tension growth was slower on succinate compared to glucose. Reoxidation of NAD(P)H produced from dicarboxylate metabolism may be limited at nanomolar concentrations of O_2, which prevents a fully operational TCA-cycle. Furthermore, although NAD(P)H has been considered the electron source for nitrogenase, its redox potential is too positive (-320 mV) to directly reduce ferredoxins (-484 mV) or nitrogenase. Interestingly, FixAB forms an electron transfer flavoprotein complex (ETF), which is essential for N_2 fixation in bacteroids[148,149]. Anaerobic bacterial ETFs

use flavin-dependent electron bifurcation[150,151], in which one electron is donated to a low potential acceptor and one to a high potential acceptor, achieving an overall exergonic reaction **[G]**. FixAB uses the same mechanism[152] although it might receive electrons in a complex with pyruvate dehydrogenase[147]. This was implied by genetic suppression analysis that FixAB and pyruvate dehydrogenase interact[153]. Potential electron acceptors for nitrogenase are ferredoxin and flavodoxin and coenzyme-Q for O_2. This electron bifurcation process also spares acetyl-CoA from entry into the TCA cycle for the synthesis of NADP(H), which makes it available for lipid or PHB synthesis. Given the similarity in structures of the oxoglutarate and pyruvate dehydrogenase complexes, both may interact with FixAB, which requires some acetyl-CoA enter the TCA-cycle but limits production of NAD(P)H. Although genetic suppression analysis suggests a direct interaction between PDH and FixAB[153], it has been shown recently that FixAB of *Azotobacter vinelandii* bifurcates NADH with reduction of flavodoxin and coenzyme-Q[152]. This suggests a tight coupling between production of NADH by PDH and its bifurcation by FixAB.

[H3] Conclusions and outlook

Complex selection and differentiation processes govern the growth of rhizobia in the rhizosphere and their development into N_2-fixing ammoniaplast-like organelles inside legume nodules. There are profound and reciprocal relationships between plants and the rhizosphere microbiota. Research on rhizobia has come a long way from studying a single organism associated with a legume to considering it part of a microbiota. Rhizobia have a rich and varied pan-genome, which is under constant selection in both the soil and by the plant, and thus they are an excellent model to study bacterial genomics.

One of the big challenges for the future is to unify community-focused approaches and the large datasets that they produce with a mechanistic understanding. Genes that have no apparent role when measured in pure cultures may become essential for competition *in vivo*, either with siblings or rival microbial communities in the soil or other niches. Results from techniques such as INSeq (Box 2) promise the birth of a new type of microbiology that integrates community and culture studies.

It is also clear that colonisation of roots by bacteria generally, and specifically the symbiotic interaction between rhizobia and legumes, needs to be understood temporally and spatially. When and in which parts of the roots are solutes and

signalling molecules secreted? Where do bacterial attachment and colonisation begin and how do they proceed? Laser capture microdissection and RNAseq of root nodules or of plant nitrogen fixation mutants in combination with microarray analysis have improved the spatial and temporal resolution of the plant and bacterial genes that are induced[99,100]. These techniques as well as TRIS[101,154] and lux mapping[60] will help identify key genes and the plant-bacteria interactions that determine early stages of colonization. Perhaps when this is combined with the use of synthetic communities, in which competing microorganisms are co-inoculated, we can bridge the gap between high-resolution molecular studies of single organisms and population-wide analysis of community structure.

[Au: Please update the reference and then add the comments highlighting references of particular importance below each.]

1 Philippot, L., Raaijmakers, J. M., Lemanceau, P. & van der Putten, W. H. Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **11**, 789-799, (2013).

2 Herridge, D., Peoples, M. & Boddey, R. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1-18, (2008).

3 Steffen, W. *et al.* Planetary boundaries: Guiding human development on a changing planet. *Science* **347**, (2015).

An attempt to look at the limits of production and consumption and their impact on a sustainable plant. It establishes planetary boundaries of consumption or production that should not be exceeded for key nutrients, CO_2 release and biodiversity loss.

4 Gage, D. J. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* **68**, 280-300, (2004).

5 Gibson, K. E., Kobayashi, H. & Walker, G. C. Molecular determinants of a symbiotic chronic infection. *Annu. Rev. Genetics* **42**, 413-441, (2008).

6 Relic, B. *et al.* Biological activity of rhizobium sp NGR234 nod-factors on *Macroptilium atropurpureum*. *Mol. Plant. Microbe Interact.* **6**, 764-774, (1993).

7 Cao, Y., HalaneWa, M. K., Gassmann, W. & Stacey, G. The role of plant innate immunity in the legume-*Rhizobium* symbiosis. *Annu. Rev. Plant Biol.* **68**, (2017).

8 Oldroyd, G., Murray, J., Poole, P. S. & Downie, J. A. The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genetics* **45**, 119-144., (2011).

9 Oldroyd, G. E. D. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59**, 519-546, (2008).

10 Poole, P. S. & Udvardi, M. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781-805, (2013).

11 Zipfel, C. & Oldroyd, G. E. D. Plant signalling in symbiosis and immunity. *Nature* **543**, 328-336, (2017).
12 Turner, T. R. *et al.* Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *ISME J.* **7**, 2248-2258, (2013).
13 Weinert, N. *et al.* PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS. Microbiol. Ecol.* **75**, 497-506, (2011).
14 Griffiths, R. I. *et al.* The bacterial biogeography of British soils. *Environ. Microbiol.* **13**, 1642-1654, (2011).
15 Tkacz, A., Cheema, J., Chandra, G., Grant, A. & Poole, P. S. Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. *ISME J.* **9**, 2349-2359, (2015).
16 Lundberg, D. S. *et al.* Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**, 86-90, (2012).
17 Bulgarelli, D. *et al.* Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**, 91-95, (2012).
Together with reference 16, this study established the importance of the root associated microbiota as the community most strongly influenced by the plant and has lead to numerous studies on other plants.
18 Haney, C. H., Samuel, B. S., Bush, J. & Ausubel, F. M. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants* **1**, 1-9, (2015).
19 Mendes, R. *et al.* Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**, 1097-1100, (2011).
20 Zgadzaj, R. *et al.* Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proc. Natl Acad. Sci. USA* **113**, E7996-E8005, (2016).
This study showed that common sym pathway has a major influence on the microbiome of the model plant *Lotus japonicus.*
21 Bulgarelli, D. *et al.* Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* **17**, 392-403, (2015).
22 Jones, F. P. *et al.* Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes – a genome comparison. *Sci. Reports* **6**, 25858, (2016).
23 diCenzo, G., Milunovic, B., Cheng, J. & Finan, T. M. The *tRNAarg* gene and *engA* are essential genes on the 1.7-Mb pSymB megaplasmid of *Sinorhizobium meliloti* and were translocated together from the chromosome in an ancestral strain. *J. Bacteriol.* **195**, 202-212, (2013).
24 Harrison, P. W., Lower, R. P., Kim, N. K. & Young, J. P. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* **18**, 141-148, (2010).
25 Young, J. P. *et al.* The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7**, R34, (2006).
26 Amadou, C. *et al.* Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res* **18**, 1472-1483, (2008).
27 Moulin, L. *et al.* Complete genome sequence of *Burkholderia phymatum* STM815^{T}, a broad host range and efficient nitrogen-fixing symbiont of *Mimosa* species. *Stand Genomic Sci* **9**, 763-774, (2014).

28 Brewer, R. J. M., Haskett, T. L., Ramsay, J. P., O'Hara, G. W. & Terpolilli, J. J. Complete genome sequence of *Mesorhizobium ciceri* bv. *biserrulae* WSM1497, an efficient nitrogen-fixing microsymbiont of the forage legume *Biserrula pelecinus*. *Genome Announc.* **5**, (2017).
29 Haskett, T. *et al.* Complete genome sequence of *Mesorhizobium ciceri* strain CC1192, an efficient nitrogen-fixing microsymbiont of *Cicer arietinum*. *Genome Announc.* **4**, (2016).
30 Kaneko, T. *et al.* Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6^{T}. *Genes* **2**, 763-787, (2011).
31 Kaneko, T. *et al.* Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**, 189-197, (2002).
32 Zheng, H. *et al.* The quorum sensing regulator CinR hierarchically regulates two other quorum sensing pathways in ligand-dependent and -independent fashions in *Rhizobium etli*. *J. Bacteriol.* **197**, 1573-1581, (2015).
33 Tun-Garrido, C., Bustos, P., Gonzalez, V. & Brom, S. Conjugative transfer of p42a from *Rhizobium etli* CFN42, which is required for mobilization of the symbiotic plasmid, is regulated by quorum sensing. *J. Bacteriol.* **185**, 1681-1692, (2003).
34 He, X. *et al.* Quorum sensing in *Rhizobium* sp. strain NGR234 regulates conjugal transfer (*tra*) gene expression and influences growth rate. *J. Bacteriol.* **185**, 809-822, (2003).
35 Danino, V. E., Wilkinson, A., Edwards, A. & Downie, J. A. Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum* bv. viciae is regulated by a quorum-sensing relay. *Mol. Microbiol.* **50**, 511-525, (2003).
36 Perez Carrascal, O. M. *et al.* Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. *Environ. Microbiol.* **18**, 2660-2676, (2016).
37 Kumar, N. *et al.* Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol.* **5**, 140133, (2015).
Two (references 36 and 37) studies of sympatric populations of rhizobia revealing greater genetic diversity on rhizobial chromosomes or accessory plasmids than on symbiosis plasmids, suggesting that symbiosis genes are frequently exchanged in the natural environment between different rhizobial genotypes.
38 Sugawara, M. *et al.* Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* **14**, R17, (2013).
39 Epstein, B. *et al.* Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. *PLoS Genet.* **8**, e1002868, (2012).
40 Bailly, X. *et al.* Population genetics of *Sinorhizobium medicae* based on low-coverage sequencing of sympatric isolates. *The ISME Journal* **5**, 1-13, (2011).
41 Nelson, M. S., Chun, C. L. & Sadowsky, M. J. Type IV effector proteins involved in the *Medicago-Sinorhizobium* symbiosis. *Mol. Plant. Microbe Interact.* **30**, 28-34, (2017).
42 Ramsay, J. P., Sullivan, J. T., Stuart, G. S., Lamont, I. L. & Ronson, C. W. Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.* **62**, 723-734, (2006).
43 Haskett, T. L. *et al.* Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc. Natl Acad. Sci. USA* **113**, 12268-12273, (2016).

In some rhizobia, integrative and conjugative elements (ICE) encoding symbiosis genes exist as three separate chromosomal regions that undergo recombination, assembling into a single circular ICE for conjugative transfer.

44 Ling, J. *et al.* Plant nodulation inducers enhance horizontal gene transfer of *Azorhizobium caulinodans* symbiosis island. *Proc. Natl Acad. Sci. USA* **113**, 13875-13880, (2016).
The first report of symbiosis ICE transfer being enhanced by legume root exudates.

45 Ramsay, J. P. *et al.* Ribosomal frameshifting and dual-target antiactivation restrict quorum-sensing-activated transfer of a mobile genetic element. *Proc. Natl Acad. Sci. USA* **112**, 4104-4109, (2015).

46 Servin-Garciduenas, L. E. *et al.* Complete genome sequence of *Bradyrhizobium* sp. strain CCGE-LA001, isolated from field nodules of the enigmatic wild bean *Phaseolus microcarpus*. *Genome Announc.* **4**, (2016).

47 Albareda, M. *et al.* Factors affecting the attachment of rhizospheric bacteria to bean and soybean roots. *FEMS Microbiol. Lett.* **259**, 67-73, (2006).

48 Ramachandran, V. K., East, A. K., Karunakaran, R., Downie, J. A. & Poole, P. S. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol.* **12**, R106, (2011).
By comparing gene expression in several plant rhizospheres the basis for specificity in interaction with host and non-host plants was established.

49 Garcia-Fraile, P. *et al.* Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar viciae in the pea rhizosphere. *Plant Soil* **390**, 251-264, (2015).

50 Perry, B. J. & Yost, C. K. Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiol.* **14**, 298, (2014).

51 Perry, B. J., Akter, M. S. & Yost, C. K. The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont *Rhizobium leguminosarum*. *Front. Microbiol.* **7**, 1873, (2016).

52 Wheatley, R. M. *et al.* The role of O_2 in the growth of *Rhizobium leguminosarum* bv. viciae 3841 on glucose and succinate. *J. Bacteriol.* **199**, e00572-00516, (2017).

53 diCenzo, G. C. *et al.* Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Comms.* **7**, 12219, (2016).
Using a genome-scale metabolic model, the study showed metabolic reprogramming in rhizobia, when they switch from one niche to another and these adaptation are supported mainly by the secondary replicon.

54 Vanderlinde, E. M., Hynes, M. F. & Yost, C. K. Homoserine catabolism by *Rhizobium leguminosarum* bv. viciae 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environ. Microbiol.* **16**, 205-217, (2014).

55 Mauchline, T. H. *et al.* Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc. Natl Acad. Sci. USA* **103**, 17933-17938, (2006).
A high-througput transcriptional induction study, which aimed to identify the solutes transported by nearly all the transporters of *Sinorhizobium meliloti*. This study shows the metabolic plasticity of soil dwelling rhizobia to adapt and survive wide varitey of niches.

56 Frederix, M. *et al.* Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Mol. Microbiol.* **93**, 464-478, (2014).

57 Zheng, H. *et al.* Flagellar-dependent motility in *Mesorhizobium tianshanense* is involved in the early stage of plant host interaction: study of an *flgE* mutant. *Curr. Microbiol.* **70**, 219-227, (2015).

58 Miller, L. D., Yost, C. K., Hynes, M. F. & Alexandre, G. The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. viciae is essential for competitive nodulation. *Mol. Microbiol.* **63**, 348-362, (2007).

59 Breakspear, A. *et al.* The root rair "Infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for auxin signaling in rhizobial infection. *The Plant Cell* **26**, 4680-4701, (2014).

60 Pini, F. *et al.* Bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion. *Plant Physiol.* **174**, 1289-1306, (2017).

61 Li, B. *et al.* Root exudates drive interspecific facilitation by enhancing nodulation and N_2 fixation. *Proc. Natl Acad. Sci. USA* **113**, 6496-6501, (2016).

62 Maimaiti, J. *et al.* Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environ. Microbiol.* **9**, 435-444, (2007).

63 Madsen, L. H. *et al.* The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nat. Comms.* **1**, (2010).

64 Wang, E. *et al.* A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Curr. Biol.* **22**, 2242-2246, (2012).

65 Gobbato, E. *et al.* A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Curr. Biol.* **22**, 2236-2241, (2012).

66 Gutjahr, C. *et al.* Arbuscular mycorrhiza–specific signaling in rice transcends the common symbiosis signaling pathway. *Plant Cell Online* **20**, 2989-3005, (2008).

67 Liang, Y. *et al.* Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science* **341**, 1384-1387, (2013).

68 Lopez-Gomez, M., Sandal, N., Stougaard, J. & Boller, T. Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. *J. Exp. Bot.* **63**, 393-401, (2012).

69 Lebeis, S. L. *et al.* Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860-864, (2015).

70 Wan, J. *et al.* A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in *Arabidopsis*. *Plant Cell* **20**, 471-481, (2008).

71 Miya, A. *et al.* CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **104**, 19613-19618, (2007).

72 Miyata, K. *et al.* The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol.* **55**, 1864-1872, (2014).

73 Zhang, X. *et al.* The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J.* **81**, 258-267, (2015).

74 Miyata, K. *et al.* Evaluation of the role of the LysM receptor-ilke kinase, OsNFR5/OsRLK2 for AM symbiosis in rice. *Plant Cell Physiol.* **57**, 2283-2290, (2016).

75 Limpens, E., van Zeijl, A. & Geurts, R. Lipochitooligosaccharides modulate plant host immunity to enable endosymbioses. *Annu. Rev. Phytopathol.* **53**, 311-334, (2015).

76 Kawaharada, Y. *et al.* Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**, 308-312, (2015).
77 Schmeisser, C. *et al.* *Rhizobium* sp strain NGR234 possesses a remarkable number of secretion systems. *Appl. Environ. Microbiol.* **75**, 4035-4045, (2009).
78 Okazaki, S., Kaneko, T., Sato, S. & Saeki, K. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proc. Natl Acad. Sci. USA* **110**, 17131-17136, (2013).
79 Okazaki, S. *et al.* *Rhizobium*-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME J.* **10**, 64-74, (2016).
80 Williams, A. *et al.* Glucomannan-mediated attachment of *Rhizobium leguminosarum* to pea root hairs is required for competitive nodule infection. *J. Bacteriol.* **190**, 4706-4715, (2008).
This study demonstarted the polar localization and attachment of rhizobia on the root hair, mediated a glucomannan. They also showed the effect of pH (acid & alkaline) on known attachment factors on rhizobial attachment and colonization of legume roots.
81 Smit, G., Kijne, J. W. & Lugtenberg, B. J. Involvement of both cellulose fibrils and a Ca^{2+}-dependent adhesin in the attachment of *Rhizobium leguminosarum* to pea root hair tips. *J. Bacteriol.* **169**, 4294-4301, (1987).
82 Laus, M. C., van Brussel, A. A. & Kijne, J. W. Role of cellulose fibrils and exopolysaccharides of *Rhizobium leguminosarum* in attachment to and infection of *Vicia sativa* root hairs. *Mol. Plant. Microbe Interact.* **18**, 533-538, (2005).
83 Downie, J. A. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34**, 150-170, (2010).
84 Finnie, C., Zorreguieta, A., Hartley, N. M. & Downie, J. A. Characterization of *Rhizobium leguminosarum* exopolysaccharide glycanases that are secreted via a type I exporter and have a novel heptapeptide repeat motif. *J. Bacteriol.* **180**, 1691-1699, (1998).
85 Ausmees, N., Jacobsson, K. & Lindberg, M. A unipolarly located, cell-surface-associated agglutinin, RapA, belongs to a family of *Rhizobium*-adhering proteins (Rap) in *Rhizobium leguminosarum* bv. trifolii. *Microbiol.* **147**, 549-559, (2001).
86 Romero-Jimenez, L., Rodriguez-Carbonell, D., Gallegos, M. T., Sanjuan, J. & Perez-Mendoza, D. Mini-Tn7 vectors for stable expression of diguanylate cyclase PleD* in Gram-negative bacteria. *BMC Microbiol.* **15**, 190, (2015).
87 Krehenbrink, M. & Downie, J. A. Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. viciae. *BMC Genomics* **9**, 55, (2008).
88 Finnie, C., Hartley, N. M., Findlay, K. C. & Downie, J. A. The *Rhizobium leguminosarum prsDE* genes are required for secretion of several proteins, some of which influence nodulation, symbiotic nitrogen fixation and exopolysaccharide modification. *Mol. Microbiol.* **25**, 135-146, (1997).
89 Mongiardini, E. J. *et al.* The rhizobial adhesion protein RapA1 is involved in adsorption of rhizobia to plant roots but not in nodulation. *FEMS. Microbiol. Ecol.* **65**, 279-288, (2008).
90 Mongiardini, E. J. *et al.* Overproduction of the rhizobial adhesin RapA1 increases competitiveness for nodulation. *Soil Biol. Biochem.* **41**, 2017-2020, (2009).

91 Nigmatullina, L. R., Lavina, A. M., Vershinina, Z. R. & Baimiev, A. Role of bacterial adhesin RAPA1 in formation of efficient symbiosis of *Rhizobium leguminosarum* with bean plants. *Mikrobiologiia* **84**, 705-711, (2015).

92 Abdian, P. L., Caramelo, J. J., Ausmees, N. & Zorreguieta, A. RapA2 is a calcium-binding lectin composed of two highly conserved cadherin-like domains that specifically recognize *Rhizobium leguminosarum* acidic exopolysaccharides. *J. Biol. Chem.* **288**, 2893-2904, (2013).

93 Janczarek, M. & Rachwal, K. Mutation in the *pssA* gene Involved in exopolysaccharide synthesis leads to several physiological and symbiotic defects in *Rhizobium leguminosarum* bv. *trifolii*. *Int. J. Mol. Sci.* **14**, 23711-23735, (2013).

94 Janczarek, M. & Skorupska, A. The *Rhizobium leguminosarum* bv. trifolii RosR: transcriptional regulator involved in exopolysaccharide production. *Mol. Plant. Microbe Interact.* **20**, 867-881, (2007).

95 Janczarek, M., Kutkowska, J., Piersiak, T. & Skorupska, A. *Rhizobium leguminosarum* bv. trifolii *rosR* is required for interaction with clover, biofilm formation and adaptation to the environment. *BMC Microbiol.* **10**, 284, (2010).

96 Robledo, M. *et al.* Role of *Rhizobium* endoglucanase CelC2 in cellulose biosynthesis and biofilm formation on plant roots and abiotic surfaces. *Microb. Cell Fact.* **11**, 125, (2012).

97 Russo, D. M. *et al.* Lipopolysaccharide O-chain core region required for cellular cohesion and compaction of in vitro and root biofilms developed by *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* **81**, 1013-1023, (2015).

98 Vanderlinde, E. M. *et al. Rhizobium leguminosarum* biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility. *Microbiol.* **155**, 3055-3069, (2009).

99 Lang, C. & Long, S. R. Transcriptomic analysis of *Sinorhizobium meliloti* and *Medicago truncatula* symbiosis using nitrogen fixation-deficient nodules. *Mol. Plant. Microbe Interact.* **28**, 856-868, (2015).

100 Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant Journal* **77**, 817-837, (2014).

101 Massalha, H., Korenblum, E., Malitsky, S., Shapiro, O. & Aharoni, A. Live imaging of root-bacteria interactions in a microfluidics set-up. *Proc. Natl Acad. Sci. USA* **114**, 4549-4554, (2017).

102 Pessi, G. *et al.* Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol. Plant Micro. Interact.* **20**, 1353-1363, (2007).

103 Barnett, M. J., Tolman, C. J., Fisher, R. F. & Long, S. R. A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proc. Natl Acad. Sci. USA* **101**, 16636-16641, (2004).

104 Karunakaran, R. *et al.* Transcriptomic analysis of *Rhizobium leguminosarum* b.v. *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**, 4002-4014, (2009).

105 Becker, A. *et al.* Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol. Plant Micro. Interact.* **17**, 292-303, (2004).

106 Mergaert, P. *et al.* A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* **132**, 161-173, (2003).
107 Guefrachi, I. *et al.* Extreme specificity of NCR gene expression in *Medicago truncatula*. *BMC Genomics* **15**, 712, (2014).
108 Van de Velde, W. *et al.* Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**, 1122-1126, (2010).
109 Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl Acad. Sci. USA* **103**, 5230-5235, (2006). **The breakthrough paper that established that members of the IRLC legumes produce up to hundreds of nodule cysteine rich peptides that take control of the bacterial cell cycle and development.**
110 Penterman, J. *et al.* Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc. Natl Acad. Sci. USA* **111**, 3561-3566, (2014).
111 Pini, F. *et al.* Cell cycle control by the master regulator CtrA in *Sinorhizobium meliloti*. *PLoS Genet.* **11**, e1005232, (2015).
112 Wang, D. *et al.* A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. *Science* **327**, 1126-1129, (2010).
113 Montiel, J. *et al.* Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl Acad. Sci. USA* **114**, 5041-5046, (2017).
114 Farkas, A. *et al. Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc. Natl Acad. Sci. USA* **111**, 5183-5188, (2014).
115 Bittner, A. N., Foltz, A. & Oke, V. Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *J. Bacteriol.* **189**, 1884-1889, (2007).
116 Kim, M. *et al.* An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc. Natl Acad. Sci. USA* **112**, 15238-15243, (2015).
117 Horváth, B. *et al.* Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. *Proc. Natl Acad. Sci. USA* **112**, 15232-15237, (2015).
118 Wang, Q. *et al.* Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc. Natl Acad. Sci. USA* **114**, 6854-6859, (2017).
119 Yang, S. *et al.* Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc. Natl Acad. Sci. USA* **114**, 6848-6853, (2017).
120 Price, P. A. *et al.* Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proc. Natl Acad. Sci. USA* **112**, 15244-15249, (2015).
121 Haag, A. F. *et al.* Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* **9**, e1001169, (2011).
122 Karunakaran, R. *et al.* BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. *J. Bacteriol.* **192**, 2920-2928, (2010).
123 Ishihara, H. *et al.* Characteristics of bacteroids in indeterminate nodules of the leguminous Tree *Leucaena glauca*. *Microbes. Environ.* **26**, 156-159, (2011).

124 Crespo-Rivas, J. C. *et al.* *Sinorhizobium fredii* HH103 bacteroids are not terminally differentiated and show altered O-antigen in nodules of the Inverted Repeat-Lacking Clade legume *Glycyrrhiza uralensis*. *Environ. Microbiol.* **18**, 2392-2404, (2016).

125 Oono, R. & Denison, R. F. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* **154**, 1541-1548, (2010).

126 Hakoyama, T. *et al.* Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* **462**, 514-518, (2009).

Most rhizobia cannot make a function nitrogenase because they lack NifV and the ability to synthesize homocitrate. Instead they rely on the plant to make homocitrate via the product of Fen1.

127 Terpolilli, J. J., Hood, G. A. & Poole, P. S. What determines the efficiency of N_2-fixing *Rhizobium*-legume symbioses? *Advances in Microbial Physiology* **Volume 60**, 325-389, (2012).

128 Prell, J. *et al.* Role of symbiotic auxotrophy in the *Rhizobium*-Legume symbioses. *Plos One* **5**, (2010).

129 Prell, J. *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl Acad. Sci. USA* **106**, 12477-12482, (2009).

Bacteroids shut down many nutrient synthesis pathways, such as branceh chain amino acid synthesis and become symbiotic auxotrophs dependent on the plant.

130 Lodwig, E. M. *et al.* Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* **422**, 722-726, (2003).

131 Tian, C. F., Garnerone, A.-M., Mathieu-Demazière, C., Masson-Boivin, C. & Batut, J. Plant-activated bacterial receptor adenylate cyclases modulate epidermal infection in the *Sinorhizobium meliloti–Medicago* symbiosis. *Proc. Natl Acad. Sci. USA* **109**, 6751-6756, (2012).

132 Murphy, P. J. *et al.* Synthesis of an opine-like compound, a rhizopine, in alfalfa nodules is symbiotically regulated. *Proc. Natl Acad. Sci. USA* **85**, 9133-9137, (1988).

133 King, B. J. *et al.* Regulation of O_2 concentration in soybean nodules observed by *in situ* spectroscopic measurement of leghemoglobin oxygenation. *Plant Physiol.* **87**, 296-299, (1988).

134 Cherfils, J., Gibrat, J. F., Levin, J., Batut, J. & Kahn, D. Model-building of Fnr and FixK DNA-binding domains suggests a basis for specific DNA recognition. *J. Mol. Recognit.* **2**, 114-121, (1989).

135 de Philip, P., Batut, J. & Boistard, P. *Rhizobium meliloti* Fix L is an oxygen sensor and regulates *R. meliloti nifA* and *fixK* genes differently in *Escherichia coli*. *J. Bacteriol.* **172**, 4255-4262, (1990).

136 David, M. *et al.* Cascade regulation of *nif* gene-expression in *Rhizobium meliloti*. *Cell* **54**, 671-683, (1988).

137 Ditta, G., Virts, E., Palomares, A. & Kim, C. H. The *nifA* gene of *Rhizobium meliloti* is oxygen regulated. *J. Bacteriol.* **169**, 3217-3223, (1987).

138 Boesten, B. & Priefer, U. B. The C-terminal receiver domain of the *Rhizobium leguminosarum* bv. viciae FixL protein is required for free-living microaerobic induction of the *fnrN* promoter. *Microbiol.* **150**, 3703-3713, (2004).

139 Patschkowski, T., Schluter, A. & Priefer, U. B. *Rhizobium leguminosarum* bv *viciae* contains a 2nd *fnr/fixK*-like gene and an unusual FixL homolog. *Mol. Microbiol.* **21**, 267-280, (1996).

140 Martinez, M., Palacios, J. M., Imperial, J. & Ruiz-Argueso, T. Symbiotic autoregulation of *nifA* expression in *Rhizobium leguminosarum* bv. viciae. *J. Bacteriol.* **186**, 6586-6594, (2004).
141 Lindemann, A. *et al.* New target genes controlled by the *Bradyrhizobium japonicum* two-component regulatory system RegSR. *J. Bacteriol.* **189**, 8928-8943, (2007).
142 Sullivan, J. T., Brown, S. D. & Ronson, C. W. The NifA-RpoN regulon of *Mesorhizobium loti* strain R7A and its symbiotic activation by a novel LacI/GalR-family regulator. *Plos One* **8**, e53762, (2013).
143 Mitsch, M. J., diCenzo, G. C., Cowie, A. & Finan, T. M. Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* nor *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* **doi 10.1128/aem.01561-17**, (2017).
144 Geddes, B. A. & Oresnik, I. J. Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*. *Can. J. Microbiol.* **60**, 491-507, (2014).
145 Mulley, G. *et al.* Mutation of GOGAT prevents pea bacteroid formation and N_2 fixation by globally downregulating transport of organic nitrogen sources. *Mol. Microbiol.* **80**, 149-167, (2011).
146 Patriarca, E. J., Tate, R. & Iaccarino, M. Key role of bacterial NH_4^+ metabolism in *Rhizobium*-plant symbiosis. *Microbiol. Mol. Biol. Rev.* **66**, 203-222, (2002).
147 Terpolilli, J. J. *et al.* Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. *J. Bacteriol.* **198**, 2864-2875, (2016).
148 Gubler, M., Zurcher, T. & Hennecke, H. The *Bradyrhizobium japonicum fixBCX* operon: identification of *fixX* and of a 5' mRNA region affecting the level of the *fixBCX* transcript. *Mol. Microbiol.* **3**, 141-148, (1989).
149 Earl, C. D., Ronson, C. W. & Ausubel, F. M. Genetic and structural analysis of the *Rhizobium meliloti fixA, fixB, fixC*, and *fixX* genes. *J. Bacteriol.* **169**, 1127-1136, (1987).
150 Buckel, W. & Thauer, R. K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. *Biochim. Biophys. Acta* **1827**, 94-113, (2013).
151 Weghoff, M. C., Bertsch, J. & Müller, V. A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environmental Microbiology* **17**, 670-677, (2014).
152 Ledbetter, R. N. *et al.* The electron bifurcating FixABCX protein complex from *Azotobacter vinelandii*: Generation of low-potential reducing equivalents for nitrogenase catalysis. *Biochem.* **56**, 4177-4190, (2017).
153 Scott, J. D. & Ludwig, R. A. *Azorhizobium caulinodans* electron-transferring flavoprotein N electrochemically couples pyruvate dehydrogenase complex activity to N_2 fixation. *Microbiol.* **150**, 117-126, (2004).
154 Poole, P. Shining a light on the dark world of plant root–microbe interactions. *Proc. Natl Acad. Sci. USA* **114**, 4281-4283, (2017).
155 Robledo, M. *et al. Rhizobium* cellulase CelC2 is essential for primary symbiotic infection of legume host roots. *Proc. Natl Acad. Sci. USA* **105**, 7064-7069, (2008).
156 Ivanov, S. *et al. Rhizobium*-legume symbiosis shares an exocytotic pathway required for arbuscule formation. *Proc. Natl Acad. Sci. USA* **109**, 8316-8321, (2012).
157 Turner, T., James, E. & Poole, P. The plant microbiome. *Genome Biol.* **14**, 209, (2013).
158 Bai, Y. *et al.* Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**, 364-369, (2015).

159 Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306-00315, (2015).
160 Andrews, M. & Andrews, M. E. Specificity in legume-rhizobia symbioses. *Int. J. Mol. Sci.* **18**, (2017).
161 Sprent, J. I., Ardley, J. & James, E. K. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* **215**, 40-56, (2017).

Acknowledgments
We would like to thank Timothy Haskett and Beatriz Jorrín for help with figures and Alison East and Andrzej Tkacz for comments on the manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council [grant numbers BB/K001868/1, BB/K001868/2, BB/J007749/1, BB/J007749/2, BB/K006134/1, BB/L011484/1, BB/N003608/1, BB/N013387/1]

Author contributions

P.P., V.R. and J.T. substantially contributed to the discussion of content, wrote the article and reviewed and edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Box 1: Legume symbiosis and root nodulation

Formation of legume nodules is initiated by release of plant flavonoids, which induce *nod* genes in rhizobia that synthesise lipochitooligosaccharides (LCOs, also known as Nod factors) **[Au:OK?]** . Rhizobia attach to root hairs and are entrapped by root hair curling, which results in the formation of an infection pocket. The bacteria secrete a cellulase to bore a hole in the root hair through which they enter into a plant-derived infection thread, which separates the rhizobia from plant cells[155]. Plant detect LCOs with a heteroduplex of LysM receptors, for example, Lyk3-NFP in *Medicago truncatula* and NFR1-NFR5 in *Lotus japonicus*[8]. Mycorrhizal signalling and plant immunity probably use similar LysM receptors and may share some receptors with the Nod pathway. The LysM receptors transmit the signal through the common SYM pathway. This leads to calcium oscillations in the nucleus, which are decoded by a calcium

calmodulin kinase (CCaMK). Downstream of this the signals diverge between mycorrhizal and nodulation pathways[8]. For successful nodulation, plants also must initiate a nodule meristem, which contains dividing cells. As the nodule develops the infection threads branch and carry rhizobia into the developing nodule. Eventually, bacteria are released and engulfed by plant cells in a process that requires vSnares[156]. At this stage, bacteria differentiate into N_2-fixing bacteroids, which are surrounded by symbiosome membranes. The symbiosome is the engine of N_2-fixation; the bacteroids are supplied with dicarboxylates as carbon sources and they secrete ammonium to the plant[10].

Nodules can be either determinate, as in beans, soybeans and *L. japonicus*, or indeterminate, as in alfalfa, pea and clover. In determinate nodules the meristem dies, with all cells at the same development stage and nodules grow larger by cell expansion, usually with several bacteroids enclosed by a single symbiosome membrane. In indeterminate nodules, the meristem persists, which leads to development zones in nodules containing a single bacteroid (see figure, reviewed in [8-10]). The distal zone I contains the nodule meristem, in which new plant cells are produced; zone II contains infection threads full of bacteria; in the interzone between zone II and III bacteria are released from infection threads and engulfed by plant cells; zone III contains the mature N_2-fixing bacteroids, and in zone IV bacteroids senesce.

Box 2. New approaches to study plant-bacteria interactions

High throughput sequencing has revolutionised the analysis of root microbiota[157]. **[Au: I've removed the following general sentences about sequencing. Most readers will be familiar with the process.]** To examine the plant microbiota, it is crucial to separate the microbial community into specific fractions: the bulk soil community; the rhizosphere community, which encompasses microorganisms in the soil that tightly adhere to roots upon harvest (typically 1-3 mm around the root); the rhizoplane community, which are microorganisms that require sonication for removal from the root; and the root-associated community, which is only released after maceration of roots. However, there is no universally accepted definition or protocol for these microbiota fractions, partly because different plants differ so much in their resistance to these treatments. More recently there is a drive to isolate

microorganisms from the root environment and, as most of them are culturable[158], they can be used to generate synthetic communities for experimentation.

Perhaps the principal limitation of sequencing is that spatial information is usually absent from studies on community composition. The recent development of microfluidic chambers to track movement of fluorescently labelled bacteria towards roots and Lux imaging of biosensor bacteria to detect both colonisation and the presence of particular metabolites are powerful new tools to overcome this limitation[60,101]. Finally, the gap between characterising community composition and function can be narrowed by insertion sequencing (InSeq, also known as transposon insertion sequencing (Tn-seq)). For this, large libraries of strains that have been mutated with a transposon, such as mariner or Tn5, are introduced into an environment of choice, recovered and digested with a type IIS restriction enzyme whose recognition sequence is at the transposon end. IIS enzymes cut away from the recognition sequence, yielding genomic tags which are sequenced to yield the transposons location. Mutants that reduce growth in the environment will be lost and therefore recovery of their associated sequence tags is reduced. Mariner and Tn5 transposons can also be bar coded , which enables testing of multiple strains and species at the same time[159].

Figure 1. Rhizobial genome organisation. Rhizobial genomes are generally large (≈5-10 Mb), often with complex architecture consisting of a chromosome plus numerous additional plasmids and/or chromids of varying sizes. Rhizobial symbiosis genes (*nod, nif* and *fix*) may be chromosomal or plasmid encoded. **a** | Plasmids, which carry, symbiosis genes, can be very large (for example, pSymA in *Sinorhizobium meliloti* 1021 is ~1.4 Mb). Together with the chromosome, multiple additional chromids and/or accessory plasmids, they form the genome of alphaproteabacterial and betaproteobacterial rhizobia, for example, in the *Rhizobium* and *Sinorhizobium (*also known as *Ensifer)* genera and in the *Cupriavidus* and *Burkholderia* genera, respectively*. **[Au:OK?]*** **b** | Symbiosis genes in some alphaproteobacterial genera, including *Mesorhizobium*, *Azorhizobium* and *Bradyrhizobium,* are chromosomally encoded. For *Mesorhizobium* and *Azorhizobium*, chromosomal symbiosis gene are encoded on mobile integrative and conjugative elements (ICEs). ICEs can excise from the chromosome, forming a plasmid-like replicon, which can transfer through conjugation to a recipient cell. There, it integrates into the host genome at conserved

attachment sites, which are often located within amino-acyl tRNA genes. Symbiosis ICEs may consist of a single contiguous region of DNA or, as in the case of the tripartite ICEs (ICE^3), may be composed of three separate chromosomally integrated regions that recombine into a single element prior to transfer. For *Bradyrhizobium* spp., it is currently unclear whether their chromosomal ICE-like symbiosis regions are transmissible.

Figure 2. Rhizobial attachment and colonization of legume roots. a | Rhizobia are attracted towards legume roots by chemoattractants in root exudates. **b |** Once rhizobia are in close proximity to root hairs, flavonoids from the root hair induce bacterial *nod* genes. This leads to the production of lipochitooligosaccharides (LCOs; also known as Nod factors). Binding of LCOs to LysM receptors on root hairs initiates early signalling events. **c |** Plant inducers in root exudates trigger the transfer of integrative and conjugative elements or plasmids to compatible rhizobia (green). **d |** It is unclear whether rhizobia first attach to the root elongation zone (REZ) (Root Elongation Zone) or root epidermis and then to root hairs or directly to root hairs. At acidic pH, rhizobial polar glucomannan (red), binds to plant lectin (green) on the root hair tip. **e |** Rhizobia on the root surface and REZ or on the root hair form a biofilm, with *Rhizobium*-adhering proteins (Raps), extracellular polymeric substances (EPS) and cellulose fibrils. The biofilm structure on the root hair is called a 'root hair cap'. There probably is probably strong induction of flavonoid and LCO signalling when rhizobia are attached to root hairs, which leads to root hair curling and entrapment of the rhizobia.

[Au: I've moved Fig 2B to Box 1 as this is where the zones are explained]

Figure 3. Molecular mechanisms of plant-rhizobia signalling.

Many of the plant-rhizobia interactions take place on root hairs, where bacteria will form a root cap, which is crosslinked by extracellular polymeric substances (EPS), *Rhizobium*-adhering proteins (Raps) and cellulose. Lipochitooligosaccharides (LCOs; also known as Nod factors) supress immunity in non-legumes, therefore, we

speculate that it binds to a pattern recognition receptor (PRR) to suppress the response to a microorganism-associated molecular pattern (MAMP). Likewise, we assume that the binding of EPS to EPS receptor 3 (Epr3) supresses immunity. Plant flavonoids induce the production of LCOs, which in turn bind LysM receptor heterocomplex, such as NFR-NFP5 and Lyk3-NRF1 **[Au:OK? We usually write complexes as abc-xyz.]**. This activates the leucine rich repeat protein SymRK and in turn the SYM signalling pathway. These signalling events result in the induction of nodule formation, including production of nodule cysteine rich (NCR) peptides, nodulins and flavonoids. NCR peptides are only made in a few legumes, including those in the invert repeat lacking clade (IRLC). Bacteroid development is coupled to nodule formation with low O_2 being the main signal. Some bacteria inject effector proteins (NOPs) through type III secretion systems (T3SS) and T4SS. NOPs can increase the host range, presumably by suppressing plant immunity, although they may also bind to a plant R protein to stimulate immunity and restrict host range.

Figure 4. Nutrient exchange and regulation of bacteroid development. O_2 is the main signal that regulates bacteroid development, through two signalling circuits. The first involves FixLJK in *Sinorhizobium meliloti* and FnrN in *Rhizobium leguminosarum*, which induce the expression of *fixNOQP*, and the second is NifA, which induces *nifHDK,* encoding nitrogenase, and autoinduces *fixABCX-nifAB*. These circuits seem to be completely separate in *R. leguminosarum* but partly overlap in *S. meliloti* with weak induction of *fixABCX-nifAB* by FixK. Other regulatory circuits exist such as RegSR in *Bradyrhizobium* spp. (not shown). FixABCX is an electron bifurcating complex **[Au:OK?], probably donating electrons to ferredoxin or a flavodoxin and CoQ**. Pyruvate dehydrogenase (PDH), the tricarboxylic acid (TCA)-cycle and 2-oxoglutarate dehydrogenase might form a complex with FixABCX to achieve this or possibly NAD(P)H provide electrons directly. FixNOQP is a high affinity electron transport complex needed in O_2-limited cells. Low O_2 limits the TCA-cycle causing acetyl-CoA to be used to produce lipids and polyhydroxybutyrate (PHB). The peptidase HppR is speculated to change host specificity by degrading nodule cysteine rich (NCR) peptides. The HppR-NCR peptide interactions in the figure are hypothetical **[Au:OK?]** and whether the shown peptides are important targets or in the correct compartment (for example, cytoplasm, symbiosome space or division septum) is unclear. The major chaperonin in bacteroids is CPN60 (pale red).

It can interact with multiple proteins, including PDH, NifHDK and FixAB. Aap: amino acid permease; Asn: asparagine; Bra: branched amino adic permease; DCT: dicarboxylate transporter; Gln: glutamine; Mal: malate; Pyr: pyruvate. **[Au: Definitions ok?]**

Table 1: Rhizobia and their legume hosts

[Au: Tables need to have three columns, otherwise we can't lay them out. Thus, I've added the alpha/beta classification to the table.]

Rhizobia[#]	Plant host species	Bacterial class
Azorhizobium caulinodans	*Sesbania rostrata*	Alphaproteobacteria
Azorhizobium doebereinerae	*Sesbania virgata*	Alphaproteobacteria
Bradyrhizobium japonicum	*Glycine max*	Alphaproteobacteria
Bradyrhizobium spp.,	*Cajanus Cajun, Glycine max, Aeschynomene Americana, Aeschynomene indica, Aeschynomene rudis, Arachis duranensis, Arachis hypogaea*	Alphaproteobacteria
Burkholderia spp.	*Mimosa pudica*	Betaproteobacteria
Cupriavidus sp. Amp6	*Mimosa asperata*	Betaproteobacteria
Cupriavidus taiwanensis	*Mimosa pudica*	Betaproteobacteria
Ensifer meliloti	*Medicago sativa, Medicago truncatula*	Alphaproteobacteria
Mesorhizobium haukuii	*Sesbania sesban*	Alphaproteobacteria
Mesorhizobium loti	*Lotus japonicus, Lotus corniculatus*	Alphaproteobacteria
Rhizobium etli	*Phaseolus vulgaris*	Alphaproteobacteria
Rhizobium gallicum	*Phaseolus vulgaris*	Alphaproteobacteria
Rhizobium leguminosarum bv phaseoli	*Phaseolus vulgaris*	Alphaproteobacteria
Rhizobium leguminosarum bv trifolii	*Trifollium* spp.	Alphaproteobacteria
Rhizobium leguminosarum bv viciae	*Pisum sativum, Lens culinaris, Vicia cracca, Vicia hirsuta, Vicia faba*	Alphaproteobacteria
Rhizobium sp. strain IRBG74	*Sesbania cannabina, Sesbania bispinosa, Sesbania cannabina, Sesbania exasperata, Sesbania formosa, Sesbania grandiflora, Sesbania madagascariensis, Sesbania macrantha, Sesbania pachycarpa*	Alphaproteobacteria

Rhizobium tropici	*Phaseolus vulgaris*	Alphaproteobacteria
Sinorhizobium fredii NGR234	112 genera	Alphaproteobacteria
Sinorhizobium fredii USDA257	79 genera	Alphaproteobacteria
Sinorhizobium medicae	*Medicago sativa, Medicago truncatula*	Alphaproteobacteria

#This table lists representative rhizobial species and strains; a comprehensive list of most of the known rhizobia and their plant hosts have recently been compiled[160,161].

Glossary terms

[Au: I've deleted terms that are now explained in the main text and added a few new ones.]

Root cortex: The outermost layer of the plant root, which lays between the epidermal cells on the outside and vascular cells on the inside. **[Au:OK?]**

Pan-genomes: The complete set of genes present in the members of a certain group, for example, the sum of all genes found in bacterial strains belonging to a species. **[Au:OK?]**

Oligotrophic lifestyle: The usage of a broad range of carbon sources in a nutritionally-limited environment. **[Au: Changed to match wording in main text.]**

Lipochitooligosaccharide (LCO): Microbial signalling molecule with a 1,4-linked acetylglucosamine backbone, which induce nodule formation. Species-dependent side decorations determine plant specificity. **[Au:OK?]**

Integrative and conjugative elements (ICEs): ICEs are mobile genetic elements, which can excise from the host chromosome to form a plasmid-like entity, capable of catalysing its own transfer through conjugation. In recipient cells, ICEs integrate site-specifically into the chromosome, usually at conserved sites within an aminoacyl-tRNA gene.

Saprophytes: Organisms that live on dead and decaying organic matter.

Nodule senescence: Old nodules cease N_2-fixation and viable rhizobia are released back into the soil. In terminally differentiated rhizobia (for example, bacteroids from IRLC legumes) only undifferentiated bacteria from infection threads will be viable.

Phytoalexins: Antimicrobial compounds produced by plants to protect them from pathogens.

Endoreduplication: Repeated cycles of DNA replication without cell division, which lead to extensive amplification of the entire genome.

Exergonic reaction: A chemical reaction that releases free energy. **[Au:OK?]**

Key points

Root secretion and plant immunity are key factors in controlling the assembly of root-associated microbiotas of which rhizobia are key members.

Rhizobia exist in soil and compete with the general microbiota before infecting legumes, typically through root hairs, and forming N_2-fixing bacteroids.

Rhizobia have complex pan-genomes. Some strains also have large plasmids or symbiosis islands, which are crucial for fitness, nodulation and N_2 fixation.

Rhizobia have specific host plants, which makes them excellent models for studying the mechanisms, timing and location of root colonisation in host and non-host plants.

Some legumes, such as members of the invert repeat lacking clade (IRLC), produce up to several hundred antimicrobial peptides to control bacteroid cell division and development.

Bacteroids receive carbon as dicarboxylates from legumes and in exchange they fix N_2 in a low O_2-environment and secrete ammonia to the plant. Bacteroids must balance electron flow to nitrogenase, lipids, polyhydroxybutyrate and O_2, and coordinate this process with reductant production by the TCA-cycle.

Author biographies

Philip Poole is Professor of Plant Microbiology in the Department of Plant Sciences at Oxford. His principal interests are in the regulation of N_2-fixation in rhizobia and how this is controlled in legume nodules. In addition, he studies how bacteria colonise roots and what factors determine colonisation competitiveness. This has led to a broader interest in how plants influence the microbiota of the rhizosphere and roots.

Jason Terpolilli is Lecturer in Biochemistry and Systems Biology in the Centre for Rhizobium Studies at Murdoch University. Prior to this appointment, he was a postdoctoral research fellow in the laboratory of Philip Poole. His key research interests are in how bacteroid metabolism is regulated and how the lateral transfer of rhizobial symbiosis genes impacts the efficiency of N_2-fixing symbioses.

Vinoy Ramachandran, Senior Researcher, Philip Poole's group in the Department of Plant Sciences at Oxford. His research interests are in bacterial attachment and

colonisation to legume roots. He employs modern omics- technologies such as RNA-Seq, INSeq to understand and identify determinants, which contribute to root attachment and colonisation.

Subject categories

Biological sciences / Microbiology / Bacteria / Symbiosis

[URI /631/326/41/547]

Biological sciences / Microbiology / Environmental microbiology / Soil microbiology

[URI /631/326/171/1818]

Biological sciences / Microbiology / Bacteria / Bacterial physiology

[URI /631/326/41/1969]

Biological sciences / Plant sciences / Plant symbiosis / Rhizobial symbiosis

[URI /631/449/2676/2678]

ToC blurb

Rhizobia can exist as both free-living soil microbiota or as plant-associated endosymbionts, which form nitrogen-fixing root nodules. In this Review, Poole, Ramachandran and Terpolilli explore the drastic life-style shift that underlies this transition and the associated plant-bacteria interactions. **[Au:OK? For your information, the blurb can be a maximum of 40 words.]**