A Genomic Encyclopedia of the Root Nodule Bacteria: assessing genetic diversity through a systematic biogeographic survey

Wayne Reeve1*, Julie Ardley1, Rui Tian1, Leila Eshragi1,2, Je Won Yoon1, Pinyaruk Ngamwisetkun1, Rekha Seshadri3, Natalia N Ivanova3 and Nikos C Kyrpides3,4

Abstract

Root nodule bacteria are free-living soil bacteria, belonging to diverse genera within the Alphaproteobacteria and Betaproteobacteria, that have the capacity to form nitrogen-fixing symbioses with legumes. The symbiosis is specific and is governed by signaling molecules produced from both host and bacteria. Sequencing of several model RNB genomes has provided valuable insights into the genetic basis of symbiosis. However, the small number of sequenced RNB genomes available does not currently reflect the phylogenetic diversity of RNB, or the variety of mechanisms that lead to symbiosis in different legume hosts. This prevents a broad understanding of symbiotic interactions and the factors that govern the biogeography of host-microbe symbioses.

Here, we outline a proposal to expand the number of sequenced RNB strains, which aims to capture this phylogenetic and biogeographic diversity. Through the Vavilov centers of diversity (Proposal ID: 231) and GEBA-RNB (Proposal ID: 882) projects we will sequence 107 RNB strains, isolated from diverse legume hosts in various geographic locations around the world. The nominated strains belong to nine of the 16 currently validly described RNB genera. They include 13 type strains, as well as elite inoculant strains of high commercial importance. These projects will strongly support systematic sequence-based studies of RNB and contribute to our understanding of the effects of biogeography on the evolution of different species of RNB, as well as the mechanisms that determine the specificity and effectiveness of nodulation and symbiotic nitrogen fixation by RNB with diverse legume hosts.

Keywords: GEBA-RNB, Root nodule bacteria, Diversity, Symbiosis, Nitrogen fixation

Introduction

The importance of the research

Legumes, with around 20,000 species and over 700 genera, are the third largest flowering plant family and are found on all continents (except Antarctica). They are major components of most of the world’s vegetation types and have important roles in agriculture as both pastures and pulses [1,2]. Most legumes are able to form dinitrogen-fixing symbioses with soil bacteria, collectively known as root nodule bacteria or rhizobia. RNB infection elicits the organogenesis of a unique structure, the nodule, which forms on the root (or less commonly, the stem) of the host plant. The mode of infection and the morphology and structure of the resulting nodule varies within the different legume tribes and has phylogenetic significance [3,4]. Following infection, RNB migrate to the nodule primordium, are endocytosed within the host cell and differentiate into N2-fixing bacteroids.

The availability of utilizable nitrogen is the critical determinant for plant productivity. Legume-RNB symbiotic nitrogen fixation is a vital source of N in both natural and agricultural ecosystems. Based on different estimates, the total annual input of biologically fixed N ranges from 139 to 175 million tons, 35 to 44 million tons of which is attributed to RNB-legume associations growing on arable land, with those in permanent pastures accounting for another 45 million tons of N. N2-fixation by legume pastures and crops provides 65% of the N currently utilized in agricultural production [5,6]. The economic value of legumes on the farm is estimated at $30
billion annually, including $22 billion in the value of legume crops and $8 billion in the value of N₂-fixation. Increasing the efficiency of the legume-RNB symbiosis has been projected to have an annual US benefit of $1,067 million, while transferring SNF technology to cereals and totally eliminating chemical N fertilization of the major crops will have an annual US benefit of $4,484 million [7].

Incorporating SNF in agricultural systems also reduces energy consumption, compared with systems that rely on chemical N-input. Every ton of manufactured N-fertilizer requires 873 m³ of natural gas and ultimately releases ~2 tons of CO₂ into the air [8]. Furthermore, >50% of US N-fertilizer is imported, which further increases the energy cost of chemical N fertilizer. SNF has the potential to reduce the application of manufactured N-fertilizer by ~160 million tons pa, equating to a reduction of 270 million tons of coal or equivalent fossil fuel consumed in the production process. As well as energy cost savings, this reduces CO₂ greenhouse gas emissions. Legume- and forage-based rotations also reduce CO₂ emission by maintaining high levels of soil organic matter, thus enhancing both soil fertility and carbon storage in soil [9]. There are additional significant environmental costs to the use of N fertilizer: agriculturally based increases in reactive N are substantial and widespread, and lead to losses of biological diversity, compromised air and water quality, and threats to human health [10]. Microbial nitrification and denitrification of soil N are major contributors to emissions of the potent greenhouse gas and air pollutant, nitrous oxide, from agricultural soils [5]. Emission of N₂O is in direct proportion to the amount of fertilizer applied. In addition, fertilizer N not recovered by the crop rapidly enters surface and groundwater pools, leading to drinking water contamination, and eutrophication and hypoxia in aquatic ecosystems [8].

The global increase in population is predicted to double demand for agricultural production by 2050 [11]. To meet this demand without incurring the high and unsustainable costs associated with the increased use of chemical N-fertilizer, the N₂-fixing potential of the legume-RNB symbiosis must be maximized. Achieving this target will require a greater understanding of the molecular mechanisms that govern specificity and effectiveness of N₂ fixation in diverse RNB-legume symbioses.

Here, we outline proposals for two sequencing projects to be undertaken at the DoE Joint Genome Institute that aim to expand the number of sequenced RNB strains in order to capture this phylogenetic and biogeographic diversity. Through the Vavilov centers of diversity (Proposal ID: 231) and GEB-A-RNB (Proposal ID: 882) projects we will sequence 107 RNB strains isolated from diverse legume hosts in various geographic locations in over 30 countries around the world. The sequenced strains belong to nine of the 16 validly described RNB genera and have been isolated from 69 different legume species, representing 39 taxonomically diverse genera, growing in diverse biomes. These proposals will provide unprecedented perspectives on the evolution, ecology and biogeography of legume-RNB symbioses, as no rhizobial sequencing project so far has attempted to relate extensive genomic characterization of RNB strains to comprehensive metadata and thereby identify correlations between the genomes of rhizobial strains, their symbiotic associations with specific legume hosts, and the environmental parameters of their habitats.

**Project design**

**Selection of target organisms**

The proposed RNB genome sequencing projects were designed with two different but complementary objectives in mind. In the “Analysis of the clover, pea/bean and lupin microsymbiont genetic pool by studying isolates from distinct Vavilov centres of diversity” project (Proposal ID: 231), the nominated RNB included clover, pea/Vicia and lupin-nodulating strains, chosen because their hosts are of highly significant commercial importance [12]. The legumes originate from six distinct Vavilov centres of diversity: the Mediterranean basin, high altitude Temperate Europe, North America, South America, highland central Africa and southern Africa [13]. The rhizobial associations in these centers have phenological and geographic specificity for nodulation and nitrogen fixation [14,15]. A detailed analysis of strains representing the six centres of diversity will enable the investigation of the evolution and biodiversity of symbioses from a geographic and phenological viewpoint.

The GEB-A-RNB project falls under the umbrella of the Genomic Encyclopedia of Bacteria and Archaea family projects. The original GEBA project [16] sequenced and analysed the genomes of Bacteria and Archaea species selected to maximize phylogenetic coverage. RNB are polyphyletic, belonging to diverse genera of the Alphaproteobacteria- and Betaproteobacteria; currently, 16 genera and over 100 species have been validly described (ICSP Subcommittee on the taxonomy of Rhizobium and Agrobacterium). Existing RNB sequencing programs have tended to focus on particular organisms or on RNB isolated from
specific hosts. The GEBA-RNB project was therefore designed as a systematic genome sequencing project to capture RNB phylogenetic and symbiotic diversity. RNB strains were selected on the basis of (i) phylogenetic diversity, (ii) legume host diversity, (iii) economic importance and (iv) biogeographic origin. Strains were also required to have comprehensive metadata records and well characterized phenotypes, in particular relating to symbiotic effectiveness. In addition, the phylogenetic divergence of strains from previously sequenced isolates was taken into account.

The map in Figure 1 shows collection sites of strains selected for sequencing. Table 1 lists the strains nominated for sequencing, their country of origin and original host. Extensive metadata is available for all strains and was used to guide strain selection; proposed strains display a wide range of host specificities (from strictly specific to highly promiscuous) and SNF efficiency. The RNB were collected from sites that spanned a broad range of soils and climates (e.g. neutral, acidic or alkaline soil, tropical, arid or temperate climate). These strains differ in their physiological attributes (ability to recycle hydrogen, rhizobitoxine production, salt and acid tolerance, heavy metal resistance, methylotrophy) and some of them display unusual genetic features (unique genotype based on multilocus sequence typing, nodulation phenotype, atypical organization of symbiosis islands or identical symbiosis islands in different genetic backgrounds).

**Organism growth and nucleic acid isolation**

The international consortium, which consists of more than 34 experts in the field from 15 different countries, together with Culture Collections Centers in Australia and Belgium will be growing the 107 different RNB. Quality Control will be performed for all samples before shipping the DNA to the JGI. All samples from members of the consortium that are based in the US, will be sent to Dr Peter van Berkum in Washington DC, and all other samples will be quality controlled at the Centre for Rhizobium Studies, Murdoch University in Australia before shipping to the JGI. Scientists at the Centre for Rhizobium Studies have extensive experience in producing high quality DNA, a skill acquired as a result of a long collaboration with the JGI as is evidenced by collaborative publication [17-22].

**Sequencing approach**

Most RNB strains are characterized by multipartite genomes, the size of which varies between 5-10 Mb, with an average G+C%age of 60-65%. We propose drafting of the

---

**Figure 1** This map shows the location of isolation of root nodule bacteria strains requested for sequencing in this proposal. Color codes are as follows: Alphaproteobacteria; Azorhizobium (yellow), Bradyrhizobium (green), Ensifer (red), Mesorhizobium (light blue), Methylobacterium (purple), Microvirga (orange), Rhizobium (blue) and Betaproteobacteria (circles); Burkholderia (yellow), Cupriavidus (blue).
| Genus              | Species                          | Strain          | Country of isolation | Original host          |
|--------------------|----------------------------------|-----------------|----------------------|------------------------|
| Azorhizobium       | Azorhizobium doebereinerae       | UFLA1-100<sup>T</sup> | Brazil               | Sesbania virgata       |
| Bradyrhizobium     | Bradyrhizobium elkanii           | USDA 76<sup>T</sup>    | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium elkanii           | USDA 94          | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium elkanii           | USDA 3259        | USA                  | Phaseolus lunatus      |
| Bradyrhizobium     | Bradyrhizobium elkanii           | USDA 3254        | USA                  | Phaseolus acutifolius  |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM1741         | Australia            | Rhynchosia minima      |
| Bradyrhizobium     | Bradyrhizobium sp.               | USDA 4           | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium japonicum         | *USDA 6<sup>T</sup> | Japan               | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium japonicum         | USDA 38          | Japan               | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium sp.               | USDA 122         | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium sp.               | USDA 123         | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium sp.               | USDA 124         | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium sp.               | USDA 135         | USA                  | Glycine max            |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM1743          | Australia            | Indigofera sp.         |
| Bradyrhizobium     | Bradyrhizobium sp.               | *WSM2254         | Australia            | Acacia dealbata        |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM2793          | South Africa         | Rhynchosia tatta.      |
| Bradyrhizobium     | Bradyrhizobium sp.               | Ai1a-2           | Costa Rica           | Andira inermis         |
| Bradyrhizobium     | Bradyrhizobium sp.               | ARR65            | Australia            | Stylosanthes viscosa   |
| Bradyrhizobium     | Bradyrhizobium sp.               | *CB756           | Zimbabwe             | Macrotyloma africanaum |
| Bradyrhizobium     | Bradyrhizobium sp.               | Cp5.3            | Panama               | Centrosema pubescens   |
| Bradyrhizobium     | Bradyrhizobium sp.               | EC3.3            | Panama               | Erythrina costaricensis|
| Bradyrhizobium     | Bradyrhizobium sp.               | Thb2             | USA                  | Amphicarpaea bracteata |
| Bradyrhizobium     | Bradyrhizobium sp.               | TV2a2            | Panama               | Tachigali versicolor   |
| Bradyrhizobium     | Bradyrhizobium sp.               | *USDA 3384       | Brazil               | Crotalaria paulina     |
| Bradyrhizobium     | Bradyrhizobium sp.               | *WSM471          | Australia            | Ormitopos pinnatus     |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM1253          | Greece               | Ormitopos compressus   |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM1417          | Chile                | Lupinus sp.            |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM2783          | South Africa         | Leobordea carnata      |
| Bradyrhizobium     | Bradyrhizobium sp.               | WSM3983          | Australia            | Kennedia coccinea      |
| Burkholderia       | Burkholderia dillworthii         | WSM3556<sup>T</sup> | South Africa         | Symmatium glabrum      |
| Burkholderia       | Burkholderia mimosarum           | ST3621           | French Guiana        | Lebeckia ambigua       |
| Burkholderia       | Burkholderia mimosarum           | LMG 23256<sup>T</sup> | Taiwan               | Mimosa pudica         |
| Burkholderia       | Burkholderia spiciferae          | WSM5005<sup>T</sup> | South Africa         | Lebeckia ambigua       |
| Burkholderia       | Burkholderia symbiotica          | JPY347           | Brazil               | Mimosa cordistipula    |
| Burkholderia       | Burkholderia tuberum             | WSM4176          | South Africa         | Lebeckia ambigua       |
| Burkholderia       | Burkholderia sp.                 | JPY1251          | Brazil               | Mimosa veloziana       |
| Burkholderia       | Burkholderia sp.                 | JPY366           | Brazil               | Mimosa misera          |
| Burkholderia       | Burkholderia sp.                 | UYP3143          | Uruguay              | Parapiptadenia rigida  |
| Burkholderia       | Burkholderia sp.                 | WSM2230          | Australia            | Kennedia coccinea      |
| Burkholderia       | Burkholderia sp.                 | WSM2232          | Australia            | Gastrolobium capitatum |
| Cupriavidus         | Cupriavidus taiwanensis           | STM6018          | French Guiana        | Mimosa pudica          |
| Cupriavidus         | Cupriavidus taiwanensis           | STM6070          | New Caledonia        | Mimosa pudica          |

Reeve et al. Standards in Genomic Sciences 2015, 10:14
http://www.standardsingenomics.com/content/10/1/14
Page 4 of 8
http://www.standardsingenomics.com/content/10/1/14
| Genus             | Species                        | Proposal ID | Country | Host Plant Name          |
|-------------------|--------------------------------|-------------|---------|--------------------------|
| *Cupriavidus*     | sp. AMP6                       | USA         | Mimosa asperata           |
| *Cupriavidus*     | sp. UYP2512                    | Uruguay     | Parapiptadenia rigida     |
| *Ensifer*         | boris                         | Sudan       | Prosopis chilensis        |
| *Ensifer*         | medicae                        | Sardinia    | Medicago arabica          |
| *Ensifer*         | medicae                        | Iraq        | Medicago polymorphia      |
| *Ensifer*         | medicae                        | Greece      | Medicago polymorphia      |
| *Ensifer*         | medicae                        | Sardinia    | Medicago sphaeroacarpus   |
| *Ensifer*         | melliloti                      | Australia   | Melilotus siccus          |
| *Ensifer*         | melliloti                      | Tunisia     | Phaseolus vulgaris        |
| *Ensifer*         | melliloti                      | Kazakhstan  | Medicago lupulina         |
| *Ensifer*         | melliloti                      | Canary Islands (Spain) | Phaseolus vulgaris |
| *Ensifer*         | melliloti                      | Canary Islands (Spain) | Medicago lacinata |
| *Ensifer*         | melliloti                      | Germany     | Medicago sativa           |
| *Ensifer*         | melliloti                      | Australia   | Medicago truncatula       |
| *Ensifer*         | melliloti                      | Greece      | Medicago orbicularis      |
| *Ensifer*         | sp. CB816                      | Brazil      | Leucaena leucocephala     |
| *Ensifer*         | sp. PC2                        | India       | Prosopis cineraria        |
| *Ensifer*         | sp. TW10                       | India       | Tephrosia willichii       |
| *Ensifer*         | sp. USDA 6670 (CC 2017)        | Australia   | Medicago sativa           |
| *Ensifer*         | sp. WSM1721                    | Australia   | Indigofera sp.            |
| *Mesorhizobium*   | ciceri CMG6                    | Tunisia     | Cicer arctinum            |
| *Mesorhizobium*   | ciceri WSM4083                 | Canary Islands (Spain) | Bituminaria bituminosa |
| *Mesorhizobium*   | loti CJ3sym                   | New Zealand | Lotus corniculatus        |
| *Mesorhizobium*   | loti NZP2037                  | New Zealand | Lotus divaricatus          |
| *Mesorhizobium*   | loti R7A                      | New Zealand | Lotus corniculatus        |
| *Mesorhizobium*   | loti R88b                     | New Zealand | Lotus corniculatus        |
| *Mesorhizobium*   | loti USDA 3471T               | New Zealand | Lotus corniculatus        |
| *Mesorhizobium*   | sp. WSM1293                   | Greece      | Lotus sp.                 |
| *Mesorhizobium*   | sp. WSM2561                   | South Africa | Lessertia diffusa         |
| *Mesorhizobium*   | sp. WSM3224                   | South Africa | Ortholobium candidans    |
| *Mesorhizobium*   | sp. WSM3626                   | South Africa | Lessertia diffusa         |
| *Mesorhizobium*   | sp. WSM2598                   | South Africa | Listeria bainesii        |
| *Microvirga*      | lotonanidis WSM557T           | Zambia      | Listeria angolensis       |
| *Microvirga*      | lupini Lut6                    | USA         | Lupinus texensis          |
| *Rhizobium*       | giardinii H152T               | France      | Phaseolus vulgaris        |
| *Rhizobium*       | leguminosarum bv. Phaseoli 4292 | UK          | Phaseolus vulgaris        |
| *Rhizobium*       | leguminosarum bv. Phaseoli FA23 | Poland      | Phaseolus vulgaris        |
| *Rhizobium*       | leguminosarum bv. Trifolii *C8782 | Kenya      | Trifolium semipilum          |
| *Rhizobium*       | leguminosarum bv. Trifolii CC278b | USA        | Trifolium narum          |
| *Rhizobium*       | leguminosarum bv. Trifolii *CC283b | Russia     | Trifolium ambiguum        |
| *Rhizobium*       | leguminosarum bv. Trifolii SRI0565 | Australia | Trifolium subtenanenum   |
| *Rhizobium*       | leguminosarum bv. Trifolii SRI0943 | Australia | Trifolium subtenanenum   |
| *Rhizobium*       | leguminosarum bv. Trifolii *TA1 | Australia | Trifolium subtenanenum   |
| *Rhizobium*       | leguminosarum bv. Trifolii WSM597 | Uruguay  | Trifolium palidum        |
107 RNB genomes using Illumina, PacBio or Roche sequencing platforms. All genomes will be completed to at least the stage of high quality draft. As most RNB strains carry their symbiotic genes on plasmids or within mobile islands that can be integrated in different sites on the chromosome, accurate scaffolding information is important for separation of chromosomal and plasmid-borne genes of interest.

Annotation and comparative analysis
The microbial genome annotation pipeline at the JGI has been scaled to handle hundreds of microbial genomes per month [23-25].

Publication of analyzed genomes
As many genomes as possible that are of publication quality will be published in Standards in Genomic Sciences [26,27].

The scientific questions we expect to answer
The genome sequences of RNB generated in this project will be used to identify the core genomes of different RNB species, as well as dispensable parts of species pan-genomes and their distribution between strains from different locales and/or plant hosts. Symbiotically relevant sets of genes such as those participating in adhesion, biosynthesis of nodulation factors, SNF, energy metabolism and exopolysaccharide biosynthesis will be characterized in detail. This will include the genes’ evolutionary histories and genome dynamics, such as localization on plasmids or within genomic islands and relation to mobile genetic elements. Statistical analyses will be performed in order to identify genes and gene sets that correlate with host specificity, nodulation and SNF efficiency and with various environmental metadata such as edaphic and climatic constraints. Within RNB strains of the same species, but from different environmental sites and/or legume hosts, genes that are under selective pressure will be identified and characterized by analysis of synonymous and non-synonymous substitution rates.

These analyses will be informed by the comprehensive metadata that are available for each strain, including data on the strains’ collection site, host specificity, nodulation and SNF efficiency. Considerable efforts have been devoted to sourcing strains from different geographical locations in order to improve legume productivity across a range of environments, and the project takes advantage of the particularly well characterized RNB that have been sourced from several culture collections around the globe. Biogeographic considerations are particularly relevant to the RNB as their survival and persistence as soil

Table 1 Root nodule bacteria strains used in the Vavilov centers of diversity (Proposal ID: 231) and GEBA-RNB (Proposal ID: 882) genome sequencing projects (Continued)

| Rhizobium                     | Rhizobium leguminosarum bv. Trifolii | WSM1689 | Greece | Trifolium uniflorum |
|-------------------------------|--------------------------------------|---------|--------|---------------------|
| Rhizobium                     | Rhizobium leguminosarum bv. Trifolii | WSM2012 | Ethiopia | Trifolium rupepellum |
| Rhizobium                     | Rhizobium leguminosarum bv. Trifolii | WSM2297 | South Africa | Trifolium africanaum |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | 248     | UK     | Vicia faba          |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | 128C53  | UK     | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | GB30    | Poland | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | Ps8     | UK     | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | TOM     | Turkey | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | UPM1131 | Italy  | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | UPM1137 | Italy  | Pisum sativum       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | Vc2     | UK     | Vicia cracca        |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | VF39    | Germany| Vicia faba          |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | Vh3     | UK     | Vicia hirsuta       |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | WSM1455 | Greece | Vicia faba          |
| Rhizobium                     | Rhizobium leguminosarum bv. Viciae   | WSM1481 | Greece | Vicia faba          |
| Rhizobium                     | Rhizobium leucaenae                 | *USDA 9039 \(^\text{1}\) | Brazil | Phaseolus vulgaris  |
| Rhizobium                     | Rhizobium mesoamericanum            | STM6155 | New Caledonia | Mimosa pudica |
| Rhizobium                     | Rhizobium mongolense                | USDA 1844 \(^\text{1}\) | China  | Medicago ruthenica  |
| Rhizobium                     | Rhizobium sullae                    | *WSM1592 | Italy  | Hedysarum coronarium|
| Rhizobium                     | Rhizobium tibeticum                 | OR 191  | USA    | Medicago sativa     |

| *type strain; \(^\text{1}\)commercial inoculant. |
The economic and ecological importance of symbiotic nitrogen fixation is significant. The legume-RNB symbiosis is one of the best-studied agricultural symbioses and microbe-eukaryote interactions. This project will support systematic sequence-based studies of the RNB and contribute to our understanding of the biogeographic effects on the evolution of different rhizobial species, as well as the mechanisms determining the specificity and efficiency of nodulation and N₂-fixation by RNB.

**The relevance of the project to problems of societal importance**

The symbiotic nitrogen fixation by RNB is a significant asset for world agricultural productivity, farming economy and environmental sustainability. Large-scale agricultural use of highly effective N₂-fixing legumes will be critical for sustainable food production for livestock and humans. Increased incorporation of SNF into agricultural systems reduces the requirement for inputs of economically and environmentally costly nitrogenous fertilizer. Currently, ~1–2% of the world’s annual energy supply is used in the Haber-Bosch process to manufacture chemical N, at a cost of $US 6.8 billion pa. In addition, SNF significantly reduces greenhouse gas emissions compared to intensive agriculture practice, which requires large inputs of chemical N. SNF also benefits the environment by helping to reduce dry-land salinity, increase soil fertility, promote carbon sequestration and prevent eutrophication of waterways. Recent publications have also emphasized the importance of providing renewable sources of biofuels [29,30], and a detailed understanding of endosymbionts and SNF will aid this quest. *Pongamia pinnata*, for example, is a leguminous tree that is important for the biofuel industry and is nodulated by a *Bradyrhizobium* strain [31] that has been included for sequencing in this proposal.

Apart from their economic importance, RNB also represent a uniquely tractable biological system that can offer insights into the shared genetic mechanisms between fungal and bacterial root endosymbioses [32] and between intracellular pathogens and endocytosed RNB microsymbionts. The latter have been shown to share similar host-adapted strategies in their infection processes and adaptation to growth within the cytoplasm of a eukaryotic host [33,34]. An understanding of these mechanisms will facilitate the quest to extend N₂-fixation to cereals, a goal which is being vigorously pursued and which has been described as essential for future sustainable food production [35].

**Conclusion**

The legume-RNB symbiosis is one of the best-studied associations between microbes and eukaryotes, due to the economic and ecological importance of symbiotic nitrogen fixation. Targeting RNB for sequencing on the basis of firstly, phylogenetic diversity and secondly, isolation from taxonomically distinct host legumes growing in diverse biomes offers significant benefits. Previous RNB sequencing projects have tended to focus on a narrow range of model organisms. By setting a goal of maximizing the phylogenetic diversity of sequenced RNB strains, these projects, in keeping with the other members of the GEB family of projects, aid the development of a phylogenetically balanced genomic representation of the microbial tree of life and allow for the large-scale discovery of novel rhizobial genes and functions. The chosen RNB strains are available to the global research community and are stored in culture collections that are dedicated to long-term storage and distribution. A wealth of experimental data and metadata is available for each strain, which will inform analyses to identify genes and gene sets that correlate with rhizobial adaptation to diverse biomes, to the nodule environments found in taxonomically distinct legume hosts and to the effectiveness of nitrogen fixation within these nodules. Moreover, the legume-RNB symbiosis is an excellent model system to study plant-bacterial associations, including symbiotic signaling, cell differentiation and the mechanisms of endocytosis. The sequenced RNB genomes will not only provide a greater understanding of legume-RNB associations, but can be used to gain insights into the evolution of N₂-fixing symbioses and microbe-eukaryote interactions.

**Abbreviations**

RNB: Root nodule bacteria; N₂: Dinitrogen; N: nitrogen; SNF: Symbiotic nitrogen fixation; N₂O: Nitrous oxide; GEB: Genomic Encyclopedia for Bacteria and Archaea.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

WR and JA supplied background information for this project, TR supplied DNA to the JGI, WR and JA drafted the paper, JWY and PN supplied figures and all other authors were involved in sequencing the genomes and/or editing the final paper. All authors read and approved the final manuscript.

**Acknowledgements**

This work was performed under the auspices of the US Department of Energy Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231.

**Author details**

1Centre for Rhizobium Studies, Murdoch University, Murdoch, Western Australia, Australia. 2Centre for Phytophthora Science and Management, Murdoch University, Murdoch, Western Australia, Australia. 3DOE Joint Genome Institute, Walnut Creek, CA, USA. 4Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 26 November 2014 Accepted: 5 December 2014 Published: 9 February 2015

**References**

1. Lewis G, Schrire B, Mackinder B, Lock M. Legumes of the World. Richmond, Surrey: Royal Botanic Gardens, Kew; 2005.
2. Graham PH, Vance CP. Legumes: importance and constraints to greater use. Plant Physiol. 2003;131:872–7.

3. Sprent JI, James EK. Legume evolution: where do nodules and mycorrhizas fit in? Plant Physiol. 2007;144:57–81.

4. Sprent JI, Ardley JK, James EK. From North to South: a latitudinal look at legume nodulation processes. S Afr J Bot. 2013;89:31–41.

5. Socolow RH. Nitrogen management and the future of food: Lessons from the management of energy and carbon. Proc Natl Acad Sci U S A. 1999;96:6001–8.

6. Galloway JN, Schlesinger WH, Levy H, Michaels A, Schnoor JL. Nitrogen: an essential resource. Annu Rev Env Resour. 2010;34:97–125.

7. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to meet food demands. Proc Natl Acad Sci U S A. 2008;105:15744–8.

8. Hill J, Nelson E, Dolak S, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A. 2006;103:11206–10.

9. Deakin WJ, Broughton WJ. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat Rev Micro. 2009;7:312–20.

10. Toft C, Anderson SG. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11:465–75.

11. Charpentier M, Oldroyd G. How close are we to nitrogen-fixing cereals? Curr Opin Plant Biol. 2010;13:556–64.

12. Revere WG, Nandasena K, Yates R, Tiwari R, O’Hara G, Ninawi M, et al. Complete genome sequence of Rhizobium leguminosarum bv trifolii strain WSM2304, an effective microsymbiont of the South American clover Trifolium polymorphum. Stand Genom Sci. 2010;2:455–64.

13. Deakin WJ, Broughton WJ. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat Rev Micro. 2009;7:312–20.

14. Toft C, Anderson SG. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11:465–75.

15. Charpentier M, Oldroyd G. How close are we to nitrogen-fixing cereals? Curr Opin Plant Biol. 2010;13:556–64.

16. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature. 2009;462:1056–60.

17. Revere W, Chain P, O’Hara G, Ardley J, Nandasena K, Brau L, et al. Complete genome sequence of the Medicago microsymbiont Ensifer (Sinorhizobium) medicae strain WSM419. Stand Genom Sci. 2010;2:77–86.

18. Revere W, O’Hara G, Chain P, Ardley J, Brau L, Nandasena K, et al. Complete genome sequence of Rhizobium leguminosarum bv. trifoli strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. Stand Genom Sci. 2010;2:347–56.

19. Revere W, O’Hara G, Chain P, Ardley J, Brau L, Nandasena K, et al. Complete genome sequence of Rhizobium leguminosarum bv. trifoli strain WSM2304, an effective microsymbiont of the South American clover Trifolium polymorphum. Stand Genom Sci. 2010;2:455–64.

20. Revere WG, Nandasena K, Yates R, Tiwari R, O’Hara G, Ninawi M, et al. Complete genome sequence of Mesorhizobium opportunism type strain WSM2073T. Stand Genom Sci. 2013;3:294–303.

21. Revere W, Nandasena K, Yates R, Tiwari R, O’Hara G, Ninawi M, et al. Complete genome sequence of Mesorhizobium atrualicum type strain (WSM2073T). Stand Genom Sci. 2013;3:910–9.

22. Nandasena K, Yates R, Tiwari R, O’Hara G, Howieson J, Ninawi J, et al. Complete genome sequence of Mesorhizobium ciceri bv. bisnuleae type strain (WSM1271T). Stand Genom Sci. 2013;3:462–72.

23. Mavromatis K, Ivanova NN, Chen IM, Sierotzki H, Markowitz VM, Kyrpides NC. The DOE-IMG standard operating procedure for the annotations of microbial genomes. Stand Genom Sci. 2009;1:63–7.

24. Parniske M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Curr Opin Plant Biol. 2010;13:556–64.

25. Garrity GM. The state of standards in genomic sciences. Stand Genom Sci. 2011;5:262–8.

26. Poole PS, Hynes MF, Johnston AWB, Tiwari RP, Revere WG, Dawnie JA. Physiology of root nodule bacteria. In: Dillworth MJ, James EK, Sprent JI, Newton WE, editors. Nitrogen-Fixing Legume Symbioses, vol. 7. Dordrecht, The Netherlands: Springer; 2008. p. 241–92.

27. Farrow AE, Plevin RJ, Turner BT, Jones AD, O’Hare M, Kammen DM. Ethanol can contribute to energy and environmental goals. Science. 2006;311:506–8.

28. Oldroyd G, Hara GW, Carr SJ. Changing roles for legumes in biogeochemical cycles. Curr Opin Plant Biol. 2000;3:107–14.

29. Sprent JI, James EK. Legume evolution: where do nodules and mycorrhizas fit in? Plant Physiol. 2007;144:57–81.

30. Sprent JI, Ardley JK, James EK. From North to South: a latitudinal look at legume nodulation processes. S Afr J Bot. 2013;89:31–41.

31. Socolow RH. Nitrogen management and the future of food: Lessons from the management of energy and carbon. Proc Natl Acad Sci U S A. 1999;96:6001–8.

32. Galloway JN, Schlesinger WH, Levy H, Michaels A, Schnoor JL. Nitrogen: an essential resource. Annu Rev Env Resour. 2010;34:97–125.

33. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. Plos One. 2013;8:e66428.

34. Hill J, Nelson E, Dolak S, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A. 2006;103:11206–10.

35. Toft C, Anderson SG. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11:465–75.

36. Charpentier M, Oldroyd G. How close are we to nitrogen-fixing cereals? Curr Opin Plant Biol. 2010;13:556–64.