Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes – a genome comparison

Frances Patricia Jones¹,², Ian M. Clark³, Robert King³, Liz J. Shaw², Martin J. Woodward⁴ & Penny R. Hirsch¹

The slow-growing genus *Bradyrhizobium* is biologically important in soils, with different representatives found to perform a range of biochemical functions including photosynthesis, induction of root nodules and symbiotic nitrogen fixation and denitrification. Consequently, the role of the genus in soil ecology and biogeochemical transformations is of agricultural and environmental significance. Some isolates of *Bradyrhizobium* have been shown to be non-symbiotic and do not possess the ability to form nodules. Here we present the genome and gene annotations of two such free-living *Bradyrhizobium* isolates, named G22 and BF49, from soils with differing long-term management regimes (grassland and bare fallow respectively) in addition to carbon metabolism analysis. These *Bradyrhizobium* isolates are the first to be isolated and sequenced from European soil and are the first free-living *Bradyrhizobium* isolates, lacking both nodulation and nitrogen fixation genes, to have their genomes sequenced and assembled from cultured samples. The G22 and BF49 genomes are distinctly different with respect to size and number of genes; the grassland isolate also contains a plasmid. There are also a number of functional differences between these isolates and other published genomes, suggesting that this ubiquitous genus is extremely heterogeneous and has roles within the community not including symbiotic nitrogen fixation.

The slow-growing bacterial genus *Bradyrhizobium* has been shown to be one of the most abundant groups in soil¹,² including soils sampled from long-term field experiments in the UK (Rothamsted Research, Harpenden)³,⁴. A key characteristic of the order Rhizobiales including the genus *Bradyrhizobium* is the ability to form nitrogen-fixing symbioses with legumes to increase nitrogen availability to plants⁵–⁸. This ability is thought to have evolved through horizontal gene transfer as the genes involved in this process are usually located on symbiosis islands on the chromosomes of bradyrhizobia⁷–⁹ or on symbiotic plasmids in many rhizobia⁷. Some isolates of *Bradyrhizobium* have been shown to be non-symbiotic and do not possess the ability to form nodules. The absence of nodulation ability has been noted in the strain *Bradyrhizobium* sp. S23321 isolated from paddy soil in Japan⁸ although nitrogen fixation (*nif*) genes were present. Recently, *Bradyrhizobium* ecotypes from forest soils have been shown to lack both nodulation and nitrogen fixation genes⁸.

*Bradyrhizobium* is biologically important in soils, with different representatives found to perform a wide range of biochemical functions including photosynthesis, nitrogen fixation during symbioses, denitrification

¹Department of AgroEcology, Rothamsted Research, Harpenden, AL5 2JQ, UK. ²Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AH, UK. ³Department of Computational and Systems Biology, Rothamsted Research, Harpenden, AL5 2JQ, UK. ⁴Department of Food and Nutritional Sciences, University of Reading, Reading, RG6 6AH, UK. Correspondence and requests for materials should be addressed to F.P.J. (email: francis.jones@rothamsted.ac.uk)
Bradyrhizobium

| Taxonomy       | Strain | Host         | Origin    | Genome size (bp) | GC content | Proteins | rRNA operon | tRNA | Gene | Accession number | Reference |
|----------------|--------|--------------|-----------|------------------|------------|----------|-------------|------|------|-----------------|-----------|
| Bradyrhizobium sp. | BF49   | Free-living  | UK        | 7,547,693        | 63.80%     | 7380     | 1           | 48   | 7431 | ERS954599       | This paper |
| Bradyrhizobium sp. | G22    | Free-living  | UK        | 9,022,917        | 63.70%     | 8653     | 1           | 49   | 8706 | ERS956567       | This paper |
| Bradyrhizobium sp. | G22 plasmid | Free-living  | UK        | 364,482          | 60.70%     | 541      | 1           | 2    | 546  | ERS955536       | This paper |
| Bradyrhizobium sp. | G22 unplaced | Free-living  | UK        | 68,438           | 62.10%     | 81       | 0           | 0    | 81   | ERS955657       | This paper |
| Bradyrhizobium sp. | S2321  | Free-living  | Japan     | 7,231,841        | 64.30%     | 6898     | 1           | 45   | 6951 | NC_017082       | 8         |
| B. diazoefficiens USDA 110 | Glycine max | USA        |           | 9,105,828        | 64.10%     | 8317     | 1           | 50   | 8373 | NC_004463       | 10        |
| B. japonicum USDA 6 | Glycine max | Japan      |           | 9,207,384        | 63.70%     | 8829     | 2           | 51   | 8888 | NC_017249       | 59        |
| B. japonicum E109 | Glycine max | Argentina  |           | 9,224,208        | 63.70%     | 8233     | 2           | 54   | 8621 | NZ_CP010313     | 60        |
| B. oligotrophicum S58 | Aeschynomene | Japan   |           | 8,264,165        | 65.10%     | 7228     | 2           | 51   | 7285 | NC_002453       | 11        |
| Bradyrhizobium sp. | ORS278 | Aeschynomene | Africa    | 7,456,587        | 65.50%     | 6752     | 2           | 50   | 6818 | NC_009445       | 6         |
| Bradyrhizobium sp. | BTAi1  | Aeschynomene | N. America | 8,264,687        | 64.90%     | 7394     | 2           | 52   | 7553 | NC_009485       | 6         |
| Bradyrhizobium sp. | BTAi1 plasmid | Aeschynomene | N. America | 228,826          | 60.70%     | 203      | 0           | 0    | 216  | NC_009475       | 6         |

Table 1. Summary of the seven published complete Bradyrhizobium genomes and the two novel strains G22 and BF49.

Results and Discussion

General genome description and comparisons. The genome of the grassland isolate G22 is 9,022,917 bp in size while the bare fallow isolate BF49 genome is 7,547,693 bp, constituting a 1.5 Mbp size difference in addition to a 364,482 bp plasmid in G22. The genome size for G22 is similar to nodulating strains B. diazoefficiens USDA 110, B. japonicum USDA 6, B. japonicum E109, B. oligotrophicum S58 and Bradyrhizobium sp. BTAi1, whereas for BF49 it is closer in size to the free-living strain Bradyrhizobium sp. S2321 and the photosynthetic, nodulating Bradyrhizobium sp. LTSP849. Bradyrhizobium sp. LTSP849, Bradyrhizobium sp. LTSP857, Bradyrhizobium sp. LTSP885 and Bradyrhizobium sp. LTSPM299. These genomes were sequenced using shotgun sequencing of the soil community and assembled from near completion. Due to the availability of a diverse array of genome reference sequences, Bradyrhizobium is an appropriate model to study other soil bacteria: understanding the mechanisms of Bradyrhizobium adaptation to independent living in agricultural soils under contrasting management may reveal the genetic potential of this globally important genus.

Here we present the genome and gene annotations and carbon metabolism profiles of two free-living Bradyrhizobium isolates from the Highfield experiment at Rothamsted Research that has three long-term treatment regimes: grassland, arable (wheat) and bare fallow tilled regularly to maintain a plant-free soil. Maintenance of these treatments for 60 years has led to distinct differences in soil properties and the soil microbiome. The first to be isolated and genome sequenced from European soil and the first free-living and non-diazotrophic isolates, without the presence of either nodulation or nitrogen fixation genes, to have their genomes sequenced and assembled from cultured samples. The isolates were interrogated for differences to determine the level of genetic heterogeneity in carbon metabolism between these isolates.

Orthologous gene clusters and core genome phylogeny. G22 and BF49 were compared with the free-living isolate, S2321, and the symbiotic isolate, USDA 110 (Fig. 1). This suggests that there is a core genome of 4562 genes which are present in all four genomes assessed. The 103 genes present only in the USDA 110 genome include those involved in nodulation and uptake hydrogenase. The 171 genes which are only present in
the USDA 110 and S23321 genomes include nitrogen fixation genes. Only a small number of genes are unique to any one isolate (G22: 99, BF49: 90, S23321: 35, USDA 110: 103). OrthoVenn identified a core genome of 3442 homologous gene families across the nine complete genomes available in the database. The core genome SNP analysis (Fig. 2) shows that G22 clusters closest with the *B. japonicum* strains; E109 and USDA 6. BF49 is separate from G22 being more basal. The closest relatives are the free-living strain S23321 and the soybean nodulating type-strain USDA 110. Two of the photosynthetic, *Aeschynomene*-nodulating isolates (BTAi1 and S58) cluster together.

**Isolate identification and 16S phylogeny.** The 16S rRNA sequence from G22 shows 100% identity with *Bradyrhizobium* sp. VUPMI37 [Accession number: HG940535] and *Bradyrhizobium* sp. ICMP12674 [Accession number: AY491080] from the NCBI and RDP databases respectively. *Bradyrhizobium* sp. VUPMI37 was originally isolated from *Vigna unguiculata* nodules and *Bradyrhizobium* sp. ICMP12674 was originally isolated from *Ulex europaeus*\(^1\). The BF49 16S rRNA fragment shows 100% identity with *B. canariense* SEMIA928 from the NCBI database [Accession number: FJ390904] and *B. canariense* LG-6 from the RDP database [Accession number: GU306140]. *B. canariense* SEMIA928 was originally isolated from *Lupinus* spp. and *B. canariense* LG-6 was originally isolated from *Lupinus angustifolius* root nodules\(^1\). The 16S phylogeny (Fig. 3) clusters G22 with VUPMI37 and ICMP12674, which were the top hits from the NCBI and RDP databases. Similarly, BF49 clusters with the top hits from the NCBI and RDP databases (SEMPIA928 and LG-6) in addition to *B. lupini* USDA 3051 which was the second hit from the NCBI database [Accession number: NR_134836]. This strain was originally isolated from *Lupinus* and was reclassified from *Rhizobium lupini* to *B. lupini* in 2015\(^2\). Both G22 and BF49 are in the same clade as the free-living strain, S23321, and the soybean-nodulating strains, E109, USDA 6 and USDA 110. Two of the four strains from North American forest soils, *Bradyrhizobium* sp. LTSP849 and LTSPM299 were also in this clade. The photosynthetic strains (ORS278, BTAi1 and S58) are in a separate clade including the remaining two strains from North American forest soils *Bradyrhizobium* sp. LTSP885 and LTSPM299.
Genes involved in nitrogen fixation and nodulation. Nitrogen fixation and nodulation genes including nifDKH, nodD and nodABC, are all absent from both G22 and BF49 (Table 2) and so we suggest that these isolates are not able to either form nodules or fix atmospheric nitrogen. The absence of both nif and nod genes is in contrast to other published, complete Bradyrhizobium genomes and indicates similarity with forest soil bacteria (LTSP849, LTSP857, LTSP885 and LTSPM299). B. diazoefficiens USDA 110 (previously B. japonicum USDA 110) is a known symbiotic strain and contains all nif and nod genes listed. The seven complete Bradyrhizobium genomes which have previously been published either contain both nif and nod genes (the soybean-nodulating strains B. diazoefficiens USDA 110, B. japonicum USDA 6 and B. japonicum E109) or just nif genes (the Aeschynomene-nodulating strains Bradyrhizobium sp. BTA1, Bradyrhizobium sp. S58 and Bradyrhizobium sp. ORS278 and the free-living strain Bradyrhizobium sp. S23321) (Fig. 4). These strains use a nod-independent route for stem and root nodulation of Aeschynomene. The FixLJ two component system is present in both the grassland and bare fallow isolate (Table 2). FixLJ acts in response to low oxygen conditions in soil and in the nodule and controls the expression of genes for both nitrogen fixation and denitrification. This two-component system has also been shown to regulate the response to nitric oxide in Sinorhizobium meliloti being shown to regulate a high proportion of genes induced by the presence of nitric oxide. The presence of FixLJ in G22 and BF49 is consistent with other Bradyrhizobium isolates including all seven completed genomes.

Genes involved in denitrification. Both genomes encode a nitrate reductase, NapA/B, which catalyses the reduction of nitrate to nitrite; the first stage in denitrification. This is common among Bradyrhizobium including all seven complete genomes (Table 2 and Fig. 4) and the four genomes from North American forest soils (Table 2). In addition, all fully-sequenced isolates including both G22 and BF49 contain nirK, encoding a respiratory copper-containing nitrite reductase which is involved in the second stage of denitrification reducing nitrite to nitric oxide. The third stage of denitrification is the conversion of nitric oxide into nitrous oxide, a potent greenhouse gas, and is catalysed by a nitric oxide reductase encoded by norB/C. The presence of a nitric oxide reductase gene has been noted in all previously published and complete Bradyrhizobium genomes in addition to G22 and BF49. The denitrification pathway is not present in the four genomes of strains from North American forest soils.

Figure 3. Phylogenetic tree for the 16S rRNA gene using Neighbour-Joining clustering method with 1000 bootstraps.
Table 2. Summary of presence or absence of genes of interest in the seven published complete Bradyrhizobium genomes, four isolates from North America forest soils and the two novel strains G22 and BF49 (+ = present, − = absent).  

| Function                  | Gene(s)          | G22 | BF49 | S23321 | ORS278 | BTAa1 | S58 | USDA 110 | USDA 6 | E109 | LTSP849 | LTSP857 | LTSP885 | LTSPM299 |
|---------------------------|------------------|-----|------|--------|--------|-------|-----|----------|--------|------|---------|---------|---------|----------|
| Nitrogen fixation         | nifDKH           | −   | −    | +      | +      | +     | +   | +        | +      | −    | +       | +       | +       | −        |
| Nitrogen fixation         | fixL             | +   | +    | +      | +      | +     | +   | +        | +      | +    | +       | +       | +       | +        |
| Nodulation                | nodD             | −   | −    | −      | +      | +     | +   | +        | −      | −    | −       | −       | −       | −        |
| Nodulation                | nodABC           | −   | −    | −      | +      | +     | +   | +        | −      | −    | −       | −       | −       | −        |
| Denitrification           | Nitrate reductase napA/B | + | + | + | + | + | + | + | + | + | + | + |
| Denitrification           | Nitrite reductase nirK | + | + | + | + | + | + | + | − | − | − | − |
| Denitrification           | Nitric oxide reductase norB/C | + | + | + | + | + | + | + | + | + | + | + |
| Denitrification           | Nitrous oxide reductase nosZ | − | + | − | − | + | + | + | − | − | − | − |
| Uptake hydrogenase        | hop              | −   | −    | −      | +      | +     | +   | +        | +      | −    | −       | −       | −       | −        |
| Photosynthesis            | Bacteriochlorophyll bchCXYZ/FNBHLM | − | − | + | + | + | + | − | − | − | − | − |
| Photosynthesis            | Carotenoids crtEF | − | − | + | + | + | + | − | − | − | − | − |
| Photosynthesis            | Light harvesting complexes pucBAC/pufBA | − | + | + | + | + | + | − | − | − | − | − |
| Photosynthesis            | Reaction centre subunits puhA/pufLM | − | − | + | + | + | + | − | − | − | − | − |
| Carbon fixation           | RuBisCo          | +   | +    | +      | +      | +     | +   | +        | +      | +    | +       | +       | +       | +        |

Carbon metabolism. The principal components analysis (PCA) biplot (Fig. 5A) shows that there is a separation of the time points and USDA 6 from G22 and BF49 across PC1 which accounts for 44.86% of the variation. Across PC2, G22 and BF49 separate and PC2 accounts for 30.52% of the total variation. The first two principal components were visualised as together they accounted for 75.38% of the variation. The third PC accounted for 7.5%. Figure 5B shows the 95 substrates colour coded according to category. More carboxylic acids and amino acid substrates are associated with the separation across PC1; USDA 6 from G22 and BF49. Carbohydrates tend to have a negative direction across PC2 being more closely associated with BF49. One carboxylic acid, malonic acid, was the only substrate which has a positive direction for PC2 and negative for PC1 associating more closely with G22.

Malonic acid can be found in plant tissues, including legumes being first characterised from alfalfa leaves in 1925 and has been found to be in very high concentrations in soybeans. This pathway was found to be closely associated with the symbiotic nitrogen fixation pathway in Rhizobium leguminosarium bv trifolii. Malonic acid is activated before being broken down into acetate and carbon dioxide through decarboxylation. It is often converted to malonyl-CoA by a CoA transferase, which is present in G22 (Phosphoribosyl-dephospho-CoA synthetase).
transferase) and the malonyl-CoA is then decarboxylated by malonate decarboxylase. Malonate decarboxylase has four subunits; alpha, beta, gamma and delta; all of which are present in G22. The grassland isolate also contains two mad genes; madL and madM. These genes have previously been reported to be part of the malonate decarboxylase operon as transport proteins thought to be involved in malonate uptake. The grassland isolate G22 also contains a malonyl CoA acyl carrier protein transacylase and triphosphoribosyl-dephospho-CoA synthetase. All genes involved in malonate decarboxylation are absent in the bare fallow isolate BF49. Malonate transport and utilisation genes are also present in other *Bradyrhizobium* isolates including BTAi1, ORS278, USDA 110, S23321 and all four of the forest strains (LTSP849, LTSP857, LTSP885 and LTSPM299). When grown in malonic acid, only G22 was able to metabolise it whereas BF49 and USDA 6 were not able to utilise this carbon source to the same extent.

The highest loadings and therefore the substrates which make the largest contribution to PC1 were L-pyroglutamic acid, L-leucine and D-galacturonic acid. L-pyroglutamic acid is an amino acid which is also...

Figure 4. Whole chromosome comparisons between the seven complete *Bradyrhizobium* genomes, BF49 and G22 showing positions of genes involved with nitrogen cycling, nodulation and photosynthesis on the reference genome sequence. Reference sequence: USDA110 (a) and S23321 (b).
known as 5-oxoproline. It is involved in the glutathione pathway and is converted to L-glutamic acid by the enzyme 5-oxoprolinase (EC 3.5.2.9)\(^{40,41}\). This gene is present in USDA 6 but is absent in both G22 and BF49. USDA 6 recorded a larger OD across all time points for this substrate than G22 and BF49. L-leucine is also an amino acid and is involved in numerous pathways including valine, leucine and isoleucine biosynthesis and degradation\(^{40,41}\). USDA 6 was able to utilise this substrate to a much greater extent than G22 and BF49 and the main difference in genes involved in L-leucine metabolism was the presence of leucine transaminase (EC 2.6.1.6) in USDA 6 and absence in G22 or BF49. D-galacturonic acid is a carbohydrate involved in numerous pathways including pentose and glucorionate interconversions, starch and sucrose metabolism and amino sugar and nucleotide sugar metabolism\(^{40,41}\). It is metabolised by USDA 6 more readily than G22 and BF49. This is likely due to the presence of pectinase (EC 3.2.1.15) and urinate dehydrogenase (EC 1.1.1.203) in USDA 6 but not in G22 or BF49. OD curves for the substrates discussed can be found in Supplementary Information S1.

**Plasmid.** Plasmid replication genes repABC and parAB are only found on the G22 plasmid and not on either the G22 or BF49 chromosomes. The trb operon for conjugal transfer consists of 12 genes; traI, trbBCDEJKLF-GHI\(^{42}\). From this operon, 10 out of the 12 genes are present on the G22 chromosome; only trbH and traI are missing. Conjugal transfer genes trbBCDEFGIJL are absent from BF49. Conjugative transfer DNA nicking endonuclease genes traR/traO are present only in the G22 plasmid. The other genes which were unique to the plasmid were two genes involved in purine utilisation (yagS and a putative xanthine transporter) and one for osmotic stress (opgC). The purine utilisation genes, yagS, encodes a periplasmic aromatic aldehyde oxidoreductase. YagS is usually part of an operon yagTSRQ which encodes a molybdenum-containing iron sulphur flavoprotein, where YagS is the FAD-containing subunit. The role of this protein has been suggested to be detoxification of aromatic aldehydes\(^{43,44}\). The flavoprotein produced from yagTSRQ shows homology with xanthine dehydrogenase\(^{43}\). The osmotic stress gene, opgC, is involved in the synthesis of osmoregulated periplasmic glucans (OPGs). The exact role of OPGs is not understood however they have been shown to potentially play a role in the interaction between bacteria and their eukaryotic host\(^{45,46}\). The OpgC protein has been examined in *Rhodobacter sphaeroides* and was shown to encode a succinyltransferase homolog involved in the succinyl modification of OPGs\(^{45,46}\). The other genes contained on the plasmid are also on the chromosome of both G22 and BF49 including genes for DNA ligases, DNA repair, cAMP signalling and RNA processing and modification.

**Conclusions**

The strains described here are the first to be isolated and genome-sequenced from European soil and are unique compared to other completed *Bradyrhizobium* genomes due to the absence of previously characterised genes and gene clusters for symbiosis, nitrogen fixation and photosynthesis. They are also distinct from the North American forest isolates as G22 and BF49 contain genes for denitrification. They represent a major group, likely to play a key role in denitrification. The presence or absence of the terminal denitrification gene, nosZ, may determine whether the end product of denitrification is the potent greenhouse gas, nitrous oxide or the less problematic nitrogen gas. The carbon metabolism analysis shows that G22 and BF49 show different metabolic profiles over time and this are also distinct from a nodulating strain, USDA 6. The genomes and carbon metabolism analysis indicate that the free-living soil *Bradyrhizobium* have the potential to carry out many degradative and transformative functions in soil; the marked differences between two isolates from comparable soils that have undergone different management indicates that they form part of an extremely heterogeneous group.

**Methods**

**Isolation and identification.** *Bradyrhizobium* were isolated from the permanent grassland and bare fallow plots in the Highfield experiment at Rothamsted Research\(^{12}\). Serially diluted soil samples were...
plated onto modified arabinose gluconate (MAG) agar plates incubated at 28 °C. Colonies forming after 7 days were picked, DNA extracted using MicroLYSIS-Plus using manufacturer’s instructions (Microzone, UK) and identified by PCR by the production of a 1360 bp 16S rRNA fragment using custom-designed *Bradyrhizobium* specific primers: Bradj16S70F (5′-GCCGGGTGGCTCCCTCCTGGGTATT-3′) and Bradj16S1430R (5′-GCCGGGCTGCCTCCCTTGCGGGTTA-3′). Each PCR mixture (20 μl) consisted of 1X NH4 reaction buffer, 0.5 mM of each dNTP, 2.5 mM MgCl₂, 0.1 μM of each primer, 1 U Biotaq polymerase (Bioline, UK) and 10 ng of genomic DNA. The PCR conditions were as follows: 95 °C 2 min, 30 cycles of 95 °C 15 sec, 68 °C 15 sec and 72 °C 1 min and extension at 72 °C for 10 min. The PCR products were examined on a 1.5 % agarose gel stained with ethidium bromide (0.2 μg ml⁻¹) at 100 V for 60 min. PCR products were purified using the Wizard SV gel and PCR clean up system (Promega, USA), sequenced as described in Mauchline *et al.* (2014) and identification confirmed using BLAST and searching the NCBI and RDP databases. Two isolates, one from grassland soil and one from bare fallow soil were chosen at random from the generated *Bradyrhizobium* culture collection to have their genomes sequenced *de novo*.

**De novo genome sequencing.** DNA was extracted from isolates grown in MAG broth incubated at 28 °C and shaken at 100 rpm using the GenElute Bacterial Genomic DNA kit (Sigma Aldrich, USA). Extracted DNA was quantified using Qubit Fluorometric Quantification (Life Technologies, USA). Two sequencing platforms were used; Ion Torrent PGM™ (Life Technologies, USA) (located at Rothamsted Research, UK) and Illumina HiSeq 2000 (Illumina Inc, USA) (located at BGI, the Beijing Genomics Institute, China). For Ion Torrent sequencing, two barcoded, unamplified sequencing libraries were constructed using an Ion Plus Fragment Library Kit (Life Technologies, USA) using 1 μg template DNA. Libraries were pooled and sequence template generated using an Ion PGM™ Template OT2 400 Kit. Sequencing was performed using the Ion PGM™ Sequencing 400 Kit and an Ion 318™ Chip. The G22 sample generated 3,652,179 reads and the BF49 sample generated 3,129,252 reads. Illumina sequencing was performed by BGI and 40 μg of genomic DNA was used. A 6 kb mate library with 90 bp read length and 50X coverage was created using the Illumina HiSeq 2000 sequencing platform resulting in 2,997,699 paired reads for G22 and 3,026,005 paired reads for BF49.

**De novo genome assembly and annotation.** The sequence data was assembled using SOAPdenovo2 assembler and gap closer⁴⁹ using a maximum read length of 615 bp and the average insert size of 6 kb. A range of assemblies were produced using kmer values; 61–83, 81, 83, 85, 87, 89, 91. The genomes were manually curated using Geneious (Biomatters Ltd v8.1.5). Gap closing was carried out by PCR. Any remaining contigs from the 61–83 kmer reference assembly over 500 bp were investigated while those under were discarded. The G22 assembly had 19 contigs over 500 bp unplaced within the chromosome, with some containing annotated genes. No contigs over 500 bp remained after the completion of the final BF49 reference and so we can be confident that all genes are present in the assembly. The genomes were uploaded into RAST (Rapid Annotation using Subsystem Technology) for annotation⁵⁰,⁵¹. The default parameters were used which were as follows: Classic RAST annotation scheme, RAST gene caller and Release70 FGam version. Automatically fix errors and backfill gaps were selected. The sequences have been deposited in the ENA database [Study ID: PRJEB10689, Sample ID G22 and unplaced contigs: ERS955657, G22 plasmid: ERS955536, Sample ID BF49: ERS954959]. The raw sequence data was also uploaded to ENA [G22: ERR1110561, BF49: ERR1110562] for Illumina and Ion Torrent [G22: ERR1110625, BF49: ERR1110597].

**Genome comparisons.** Genome annotations were downloaded from RAST and examined manually using KEGG⁵₀,⁵₁. OrthoVenn was used to assess gene orthology using an e-value of 1 e⁻¹⁰⁵². BRIG (BLAST Ring Image Generator) was used to compare the genomes with other *Bradyrhizobium* isolates (NCBI blast version 2.2.31)⁵₃. Large-Scale Genome Alignment Tool (LASTZ) was used to assess genome-wide sequence similarity in *Genus*⁴⁴.

**16S rRNA phylogeny analysis.** The 16S rRNA gene sequences for G22 and BF49 were compared with other *Bradyrhizobium* sequences in the NCBI database. All *Bradyrhizobium* 16S sequences from the NCBI RefSeq database and the four *Bradyrhizobium* isolates from North American forest soils were downloaded. These sequences were aligned using MUSCLE using 8 iterations⁵₅. The aligned region was extracted (1220 bp) and a phylogenetic tree was created using the neighbour joining clustering method with 1000 bootstraps. A 75 % support threshold was used for drawing the phylogenetic tree. Accession numbers for the sequences used in this analysis can be found in Supplementary Information S2.

**Core genome phylogenetic analysis.** OrthoVenn was used to identify genes which were present in all 9 genomes and were considered the "core genome". The inflation value was set to 1.5 and an e-value of 1 e⁻¹⁰. These genes were uploaded into the call SNPs and infer phylogeny (CSI) tool⁶ hosted by the Center for Genomic Epidemiology⁷ using the default options using *B. diazoefficiens* USDA 110 as the reference sequence. Default options were as follows: minimum of 10x depth at SNP positions, 10% relative minimum depth at SNP positions, a minimum of 10 bp distance between SNPs, minimum SNP quality score of 30, minimum read mapping quality score of 25 and minimum z-score of 1.96. Altered FastTree was selected. The core genome phylogenetic tree (Newick file) was visualised in Geneious.

**Biolog carbon assays.** Isolates G22, BF49 and *B. japonicum* USDA 6 were grown in MAG broth, incubated at 28 °C and shaken at 100 rpm. Cell density was estimated from 1 ml of culture stained with 0.05% methylene blue using a haemocytometer. The cultures were diluted to a cell density of 10⁶ μl⁻¹. The diluted cultures were centrifuged at 14000 × g for 1 minute, the supernatant was removed and the cells re-suspended in sterile deionised water.
water. Each well of a Biolog GN2 MicroPlate™ was inoculated with 140 μl (10⁶ cells/140 μl) of bacterial culture.

Three replicate plates per isolate were used. The optical density (OD) was read using a VersaScan SkanIt plate reader (Thermo Fisher Scientific Inc.) at 590 nm and 25 °C every 24 hours for a total of 98 hours (0, 24, 48, 72 and 98 hours). The plates were incubated at 25 °C and shaken at 100 rpm. Full list of substrates can be seen in Supplementary Information S3.

Statistical analyses. The Biolog data was analysed using principal components analysis (PCA) using the inbuilt function, prcomp, in R (version 3.2.2). For visualisation of the PCA, biplots were drawn using PC1 and PC2 and the loadings matrix was extracted (see Supplementary Information S3). This identified specific substrates which were associated with the isolates and also substrates making the greatest contribution to the principal component. These pathways were then examined in the genomes. The substrates were grouped into carbohydrates, carboxylic acids, amines and amides, amino acids, polyesters and miscellaneous according to categories identified in previously published work[23].

References
1. Uroz, S., Buee, M., Murat, C., Frey-Klett, P. & Martin, F. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ. Microbiol. Rep.* **2**, 281–288 (2010).
2. Vaninsbergh, D. et al. Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME J.* **9**, 2435–2441 (2015).
3. Delmont, T. O. et al. Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* **6**, 1677–1687 (2012).
4. Zhalina, K. et al. *Ca. Nitrosothiobacillus* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. *Front. Microbiol.* **4**, 104–104 (2013).
5. Van Rijn, P. & Vanderleyden, J. The *Rhzobium*-plant symbiosis. *Microbiol. Rev.* **59**, 124–142 (1995).
6. Giraud, E. et al. Legumes symbioses: Absence of *nod* genes in photosynthetic bradyrhizobia. *Science.* **316**, 1307–1312 (2007).
7. Rivas, R., Martens, M., De Lajudie, P. & Willems, A. Multilocus sequence analysis of the genus *Bradyrhizobium*. *Syst. Appl. Microbiol.* **32**, 101–110 (2009).
8. Okubo, T. et al. Complete genome sequence of *Bradyrhizobium* sp S23321: Insights into symbiosis evolution in soil oligotrophs. *Microbes Environ.* **27**, 306–315 (2012).
9. Sachs, J. I., Ehinger, M. O. & Simms, E. L. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. *J. Evol. Biol.* **23**, 1075–1089 (2010).
10. Kaneke, T. et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**, 189–197 (2002).
11. Okubo, T. et al. Genome analysis suggests that the soil oligotrophic bacterium *Agronomas oligotropha* (*Bradyrhizobium oligotrophicum*) is a nitrogen-fixing symbiont of *Aesculynome indica*. *Appl. Environ. Microbiol.* **79**, 2542–2551 (2013).
12. Hirsh, P. R. et al. Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. *Syst. Biol.* **61**, 2021–2024 (2012).
13. Weir, B. S., Turner, S. J., Silvester, W. B., Park, D. C. & Young, J. A. Unexpectedly diverse *Mesorhizobium* strains and *Rhzobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species. *Appl. Environ. Microbiol.* **70**, 5980–5987 (2004).
14. Bejarano, A., Ramirez-Bahena, M.-H., Velazquez, E. & Peix, A. *Vigna unguiculata* is nodulated in Spain by endosymbionts of *Genistae* legumes and by a new symbiobacter (vignae) of the genus *Bradyrhizobium*. *Syst. Appl. Microbiol.* **37**, 533–540 (2014).
15. Menna, P., Barcellos, F. G. & Hungria, M. Phylogeny and taxonomy of a diverse collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the 16S RNA gene, ITS region and *glnII*, *recA*, atpD and *dnaK* genes. *Int. J. Syst. Evol. Microbiol.* **59**, 2934–2950 (2009).
16. Menna, P. & Hungria, M. Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer. *Int. J. Syst. Evol. Microbiol.* **61**, 3052–3067 (2011).
17. Peix, A. et al. Revision of the taxonomic status of the species *Rhzobium lupini* and reclassification as *Bradyrhizobium lupini* comb. nov. *Int. J. Syst. Evol. Microbiol.* **65**, 1213–1219 (2015).
18. Ferrieres, L., Francez-Charlot, A., Gouzy, J., Rouille, S. & Kahn, D. Fix-regulated genes evolved through promoter duplication in *Sinorhizobium meliloti*. *Microbiology* **150**, 2335–2345 (2004).
19. Bobik, C., Meilhoc, E. & Batut, J. Fix: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *Environ. Microbiol.* **188**, 4890–4902 (2006).
20. Meilhoc, E., Cam, Y., Skapski, A. & Brund, C. The response to nitric oxide of the nitrogen-fixing symbiont *Sinorhizobium meliloti*. *Mol. Plant. Microbe Interact.* **23**, 748–759 (2010).
21. Delgado, J. J., Bonnard, N., Tresierra-Ayala, A., Bedmar, E. J. & Muller, P. The *Bradyrhizobium japonicum* napEDABC genes encoding the periplasmic nitrate reductase are essential for nitrate respiration. *Microbiology* **149**, 3395–3403 (2003).
22. Bedmar, E. J., Robles, E. & Delgado, M. J. The complete denitrification pathway of the symbiotic, nitrogen-fixing bacterium *Bradyrhizobium japonicum* sp. *J. Bacteriol.* **189**, 1307–1312 (2007).
23. Fernandez, L. A., Beatriz Perotti, E., Antonio Sagardoy, M. & Anahi Gomez, M. Denitrification activity of *PC2* and the loadings matrix was extracted (see Supplementary Information S3). This identified specific substrates which were associated with the isolates and also substrates making the greatest contribution to the principal component. These pathways were then examined in the genomes. The substrates were grouped into carbohydrates, carboxylic acids, amines and amides, amino acids, polyesters and miscellaneous according to categories identified in previously published work[23].
Bradyrhizobium et al.: Jones, F. P.

Novel European free-living, non-diazotrophic.../Sci. Rep. 6:25858; doi: 10.1038/srep25858 (2016).

How to cite this article

Competing financial interests: The authors declare no competing financial interests.

Acknowledgements

We would like to thank Dr Steve Hanley, Rothamsted Research, for carrying out the Ion Torrent sequencing and Dr Kirsty Hassell, Rothamsted Research, for statistical advice. This study was supported by a University of Reading, Lawes Agricultural Trust and BBSRC DTP studentship and BBSRC (BBS/E/C/00005196). Rothamsted Research receives strategic funding from the BBSRC.

Author Contributions

F.P.J., I.M.C., E.J.S., M.J.W. and P.R.H. were involved in the project conception, initiation and planning. F.P.J. isolated the strains and prepared the samples for sequencing. F.P.J. and R.K. performed the genome assembly and analysis. F.P.J. wrote the main manuscript text. All authors edited and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/