Aserse et al. Standards in Genomic Sciences (2017) 12:14
DOI 10.1186/s40793-017-0220-z

Standards in Genomic Sciences

EXTENDED GENOME REPORT Open Access

Draft genome sequence of type strain HBR26^T and description of *Rhizobium aethiopicum* sp. nov.

Aregu Amsalu Aserse[1*], Tanja Woyke[2], Nikos C. Kyrpides[2], William B. Whitman[3] and Kristina Lindström[1]

Abstract

Rhizobium aethiopicum sp. nov. is a newly proposed species within the genus *Rhizobium*. This species includes six rhizobial strains; which were isolated from root nodules of the legume plant *Phaseolus vulgaris* growing in soils of Ethiopia. The species fixes nitrogen effectively in symbiosis with the host plant *P. vulgaris,* and is composed of aerobic, Gram-negative staining, rod-shaped bacteria. The genome of type strain HBR26^T of *R. aethiopicum* sp. nov. was one of the rhizobial genomes sequenced as a part of the DOE JGI 2014 Genomic Encyclopedia project designed for soil and plant-associated and newly described type strains. The genome sequence is arranged in 62 scaffolds and consists of 6,557,588 bp length, with a 61% G + C content and 6221 protein-coding and 86 RNAs genes. The genome of HBR26^T contains *repABC* genes (plasmid replication genes) homologous to the genes found in five different *Rhizobium etli* CFN42^T plasmids, suggesting that HBR26^T may have five additional replicons other than the chromosome. In the genome of HBR26^T, the nodulation genes *nodB, nodC, nodS, nodI, nodJ* and *nodD* are located in the same module, and organized in a similar way as *nod* genes found in the genome of other known common bean-nodulating rhizobial species. *nodA* gene is found in a different scaffold, but it is also very similar to *nodA* genes of other bean-nodulating rhizobial strains. Though HBR26^T is distinct on the phylogenetic tree and based on ANI analysis (the highest value 90.2% ANI with CFN42^T) from other bean-nodulating species, these *nod* genes and most nitrogen-fixing genes found in the genome of HBR26^T share high identity with the corresponding genes of known bean-nodulating rhizobial species (96–100% identity). This suggests that symbiotic genes might be shared between bean-nodulating rhizobia through horizontal gene transfer. *R. aethiopicum* sp. nov. was grouped into the genus *Rhizobium* but was distinct from all recognized species of that genus by phylogenetic analyses of combined sequences of the housekeeping genes *recA* and *glnII*. The closest reference type strains for HBR26^T were *R. etli* CFN42^T (94% similarity of the combined *recA* and *glnII* sequences) and *Rhizobium bangladeshense* BLR175^T (93%). Genomic ANI calculation based on protein-coding genes also revealed that the closest reference strains were *R. bangladeshense* BLR175^T and *R. etli* CFN42^T with ANI values 91.8 and 90.2%, respectively. Nevertheless, the ANI values between HBR26^T and BLR175^T or CFN42^T are far lower than the cutoff value of ANI (> = 96%) between strains in the same species, confirming that HBR26^T belongs to a novel species. Thus, on the basis of phylogenetic, comparative genomic analyses and ANI results, we formally propose the creation of *R. aethiopicum* sp. nov. with strain HBR26^T (=HAMBI 3550^T=LMG 29711^T) as the type strain. The genome assembly and annotation data is deposited in the DOE JGI portal and also available at European Nucleotide Archive under accession numbers FMAJ01000001-FMAJ01000062.

Keywords: *Rhizobium aethiopicum*, Ethiopia, Common bean, Symbiotic, Genome, Average Nucleotide Identity

* Correspondence: aregu.aserse@helsinki.fi
[1]Department of Environmental Sciences, University of Helsinki, Viikinkaari 2a, Helsinki, Finland
Full list of author information is available at the end of the article

Introduction

Some bacteria are capable of forming a nitrogen-fixing symbiosis with various herbal and woody legumes. Some other bacterial species involve in nitrogen-fixation as free-living soil organisms [1]. Biological nitrogen fixation by root-nodule forming bacteria in symbiosis with legume plants play significant roles in agricultural systems. The symbiosis provides a nitrogen source for the legumes and consequently improve legume growth and agricultural productivity.

Common bean (*Phaseolus vulgaris*) (http://plants.usda.gov/core/profile?symbol=PHVU) is one of the best-known legume plants cultivated worldwide for food. It was originally domesticated in its Mesoamerican gene center, including Mexico, Colombia, Ecuador and northern Peru [2] and in the Andean center in the regions from Southern Peru to northern Argentina [3]. At present, it is widely cultivated in several parts of the tropical, sub-tropical and temperate agricultural systems [4] and used as a vital protein source mainly for low-income Latin Americans and Africans [5]. In Ethiopia, beans are commonly grown as a sole crop or intercropped with cereals, such as sorghum and maize, at altitudes between 1400 and 2000 m above sea level [6]. Bean plants make symbiotic associations promiscuously with several root-nodule forming nitrogen-fixing bacterial species commonly known as rhizobia. Studies thus far show that this legume forms symbiotic associations mainly with rhizobia belong to *Alphaproteobacteria*, such as *Rhizobium phaseoli*, *Rhizobium tropici* [7], *Rhizobium leguminosarum* [8], *Rhizobium etli* [8], *Rhizobium giardinii*, *Rhizobium gallicum* [9], *Rhizobium leucaenae* [10], *Rhizobium lusitanum* [11], *Rhizobium vallis* [12], *Rhizobium ecuadorense* [13]. *Rhizobium mesoamericanum* [14], *Rhizobium freirei* [15], *Rhizobium azibense* [16], *Rhizobium acidisoli* [17], *Ensifer meliloti* [18], *Ensifer fredii* [19], *Ensifer medicae* [20] and *Ensifer americanum* [21]. Rhizobial species belonging to *Betaproteobacteria*, such as *Burkholderia phymatum* [22] was also found capable of forming nodules on common bean plants.

16S rRNA gene sequence similarity and DNA–DNA hybridization techniques have been used as standard methods for describing new bacterial species. However, the 16S rRNA gene sequence divergence between closely related species is low and thus cannot differentiate closely related species found in the same genus [23–25]. The DDH technique was once considered as the gold standard method, and strains classified in the same species should have 70% or greater DDH relatedness among each other [26–29]. However, DDH results vary between different laboratories and this incurs inconsistent classification of the same species [30]. On the other hand, the multilocus sequence analysis method, using the sequences of several housekeeping protein coding genes, have been successfully used for species identification and delineation [24, 25, 31, 32]. The genome-wide ANI method, which was first proposed by Konstantinidis and Tiedje [33] has recently successfully been used for classification of various bacterial species [34, 35]. Depending on the methods used for ANI calculation or the nature of bacterial genome sequences, 95 or 96.5% ANI value [34, 35] corresponds to the classical 70% DNA–DNA relatedness cutoff value for strains of the same species. The advancement of sequencing techniques and its falling price have made genomic data for many bacterial species available for comparison [36]. Consequently, the ANI is becoming the method of choice in current bacterial taxonomic studies.

In our previous study, we isolated a group of rhizobial bacteria from nodules of common bean growing in the soils of Ethiopia. These bacteria formed a unique branch that was distinct from recognized species of the genus *Rhizobium* in phylogenetic trees constructed based on MLSA [24]. In order to compare strains using genome-wide ANI with reference genomes and to describe this group as a new *Rhizobium* species, the representative strain *Rhizobium* sp. HBR26 (hereafter *Rhizobium aethiopicum* sp. nov. HBR26^T) was selected for sequencing. This project was a part of the DOE JGI 2014 Genomic Encyclopedia of Type Strains, Phase III, the genomes of soil and plant-associated and newly described type strains sequencing program [37]. In this study, we present classification and general features of *R. aethiopicum* sp. nov. including the description of the genome sequence and annotation of the type strain HBR26^T.

Organism information

Classification and features

The strain HBR26^T is the type strain of *R. aethiopicum* sp. nov. This strain and other strains in the novel species were isolated from nodules of common bean plants in Ethiopia. Based on multiple housekeeping gene analysis, the closest validly published species was *R. etli* [24]. In this study, a partial 16S rRNA gene tree was constructed by retrieving more and recently published reference sequences from the GenBank database. In the phylogenetic tree, the novel species grouped together and showed high 16S rRNA gene sequence similarity (99%) with strains in the neighbor groups *R. etli* CFN42^T, *Rhizobium vallis* CCBAU65647^T, *Rhizobium phaseoli* CIAT652, *Rhizobium pisi* DSM30132^T, *Rhizobium binae* BlR195^T, and *R. bangladeshense* BLR175^T (Fig. 1). We also analyzed the housekeeping genes *recA* and *glnII* to resolve the relationships between strains in novel species and known species in the *R. leguminosarum* complex group [24]. In the phylogenetic tree reconstructed based

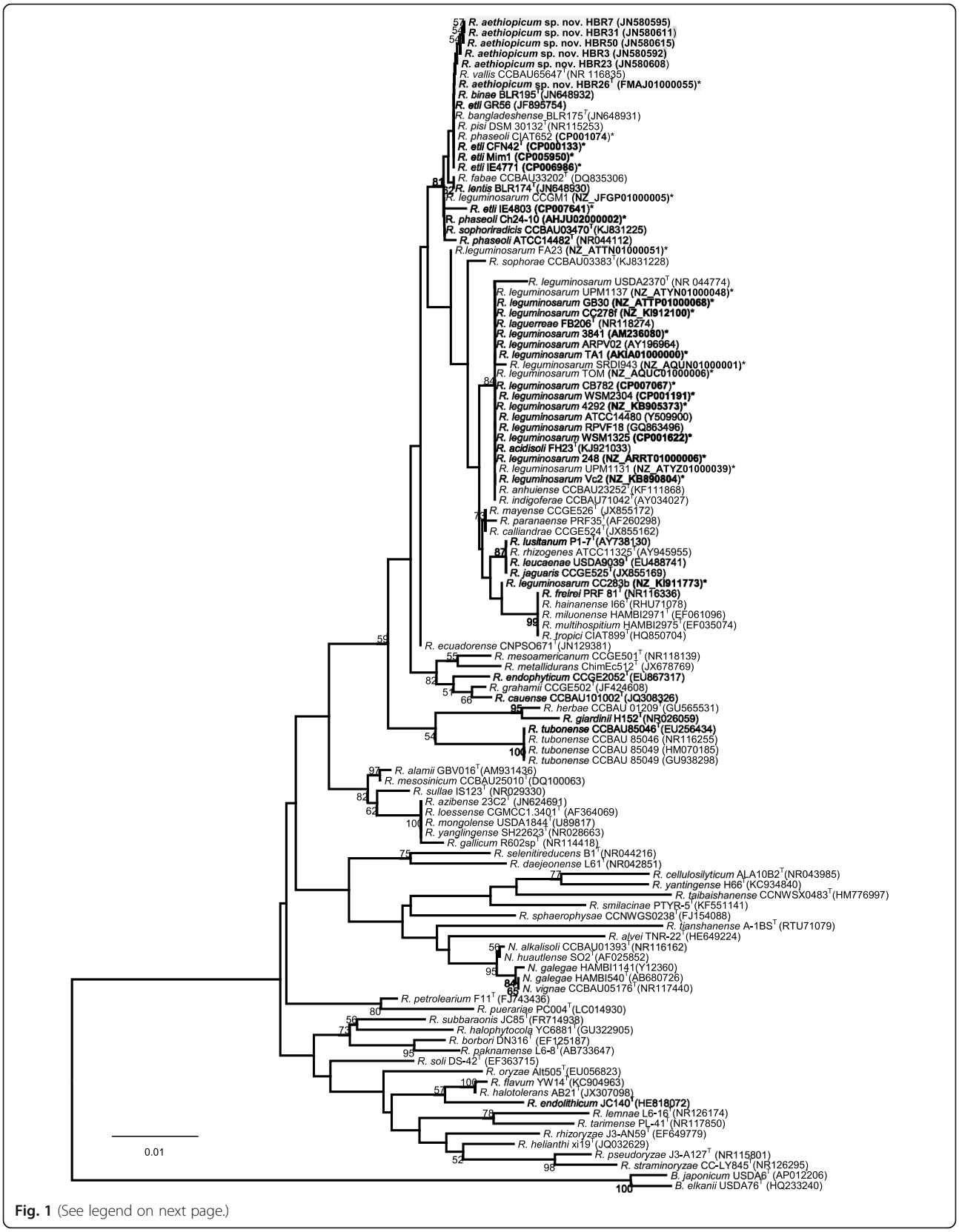

Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 Neighbor-Joining phylogenetic tree reconstructed based partial 16S rRNA gene sequences (801 bp), showing the relationships between *Rhizobium aethiopicum* sp. nov (bold and highlighted) and recognized species of the genus *Rhizobium*. The tree was computed using the Kimura 2-parameter model using MEGA version 7. The rate variation among sites was modeled with a gamma distribution (shape parameter = 4). Bootstrap values (1000 replicates) are shown at the branching points. Reference type strains are indicated with superscript 'T'. Bar, % estimated substitutions. GenBank accession numbers of the sequences are indicated in parentheses next to strains codes. The accession numbers of whole genome sequenced strains are indicated with bold*. Abbreviations: B, *Bradyrhizobium*; R, *Rhizobium*; N, *Neorhizobium*; sp., species

on the concatenated sequences, the novel species formed a clearly distinct group branching from the rhizobial species *R. etli* and *R. bangladeshense* (Fig. 2). This result was in agreement with our previous tree produced from concatenated partial 16S rRNA, *recA*, *rpoB* and *glnII* gene sequences [24]. Strain HBR26^T and other strains in the novel species showed high *recA* and *glnII* gene sequence (892 bp) similarities among each other. The similarities between HBR26^T and the type strains *R. etli* CFN42^T and *R. bangladeshense* BLR175^T ranged from 93 to 94%, CFN42^T being the closest type strain with a sequence similarity of 94%.

Minimum Information about the Genome Sequence is provided in Table 1 and the Additional file 1: Table S1. *R. aethiopicum* sp. nov. HBR26^T is fast-growing, forming moist, raised and smooth colonies 3–5 mm in diameter within 3–4 days on YEM agar plates at 28 °C. It is able to grow in the 15 °C to 30 °C temperature range, but its optimal growth was at 28 °C. The organism is able to grow at NaCl concentrations of 0–0.5% and at pH values in the range 5–10. Growth at pH4, at 4 °C and at 37 °C, and in 1-5% NaCl was recorded negative (Additional file 1: Table S1). This bacterial species is Gram-negative and rod shaped with a size of 1.0-2.4 μM in length (Fig. 3). HBR26^T and other strains in the novel species were able to respire many carbon sources when assessed by Biolog GN2 plates following the manufacturer's instructions [38]. In brief, colonies grown on YEM agar were transferred to and incubated for 48–96 h at 28 °C on freshly prepared R2A media consisting of yeast extract 0.5 g, proteose peptone 0.5 g, casamino acids 0.5 g, glucose 0.5 g, soluble starch 0.5 g, sodium pyruvate 0.3 g, K_2HPO_4 0.3 g, $MgSO_4.7H_2O$ 0.05 g, and noble agar 15 g per liter of distilled H_2O at pH7.2. Then colonies were suspended in 0.5% (w/v) saline (turbidity level of 52% transmittance), and 150 μl of the saline suspension was transferred to each of 96 wells of the Biolog GN2 Microplate. The plates were incubated at 28 °C, and results were checked after 4, 24, and 48 h. Positive results were recorded when the wells turned purple. All tested *R. aethiopicum* sp. nov. strains could respire 40 of the substrates in common, but 21 carbon sources were not respired by any of the tested strains. While the test strains did not show much diversity among themselves in substrate utilization pattern, they were distinctly different from carbon source respiration pattern of the closest reference *R. etli* CFN42^T; the test strains responded positively for seven carbon sources that were not used by *R. etli* CFN42^T. Substrates D-galactonic acid, lactone, sebacic acid and D- and L-α-glycerol phosphate were used exclusively by HBR26^T. Quinic acid and glycyl-L-aspartic acid were used solely by *R. aethiopicum* sp. nov. HBR31. The details of carbon source assimilation results are presented in Additional file 2: Table S2.

Symbiotaxonomy

HBR26^T including other strains in the *R. aethiopicum* sp. nov. are nodule forming and nitrogen-fixing on common bean host plants. The strains were originally isolated from root nodules of common bean plants growing in soils of Ethiopia [24]. In this study, the nodulation and nitrogen fixation capability was tested on legumes plants common bean, faba bean (*Vicia faba*) (http://plants.usda.gov/core/profile?symbol=VIFA), field pea (*Pisum sativum*) (http://plants.usda.gov/core/profile?symbol=PISA6) and lentil (*Lens culinaris*) (http://plants.usda.gov/core/profile?symbol=LECU2) on a sand, vermiculite and gravel mixture plant medium (5:3:3 ratio, respectively) in a growth chamber as previously described [24]. The test revealed that the strains were able to form effective nitrogen-fixing nodules in symbioses with common bean host plants. Nevertheless, the strains were not able to form symbiotic associations with faba bean, field pea and lentil. The nodulation and symbiotic characteristics results are summarized in Additional file 1: Table S1.

Genome sequencing information

Genome project history

In our previous study [24], the organism showed a unique phylogenetic position which most likely represented a new species. Thus, it was chosen for genome sequencing in order to describe a new species by comparing its genome sequence with the genome sequences of other close *Rhizobium* species. This project was a part of the DOE JGI 2014 Genomic Encyclopedia of Type Strains, Phase III the genomes of soil and plant-associated and newly described type strains sequencing program. The genome project is deposited at the DOE JGI genome portal [39] and also available at European Nucleotide Archive [40] under accession numbers FMAJ01000001-FMAJ01000062. Sequencing, assembling, and annotation were done by the

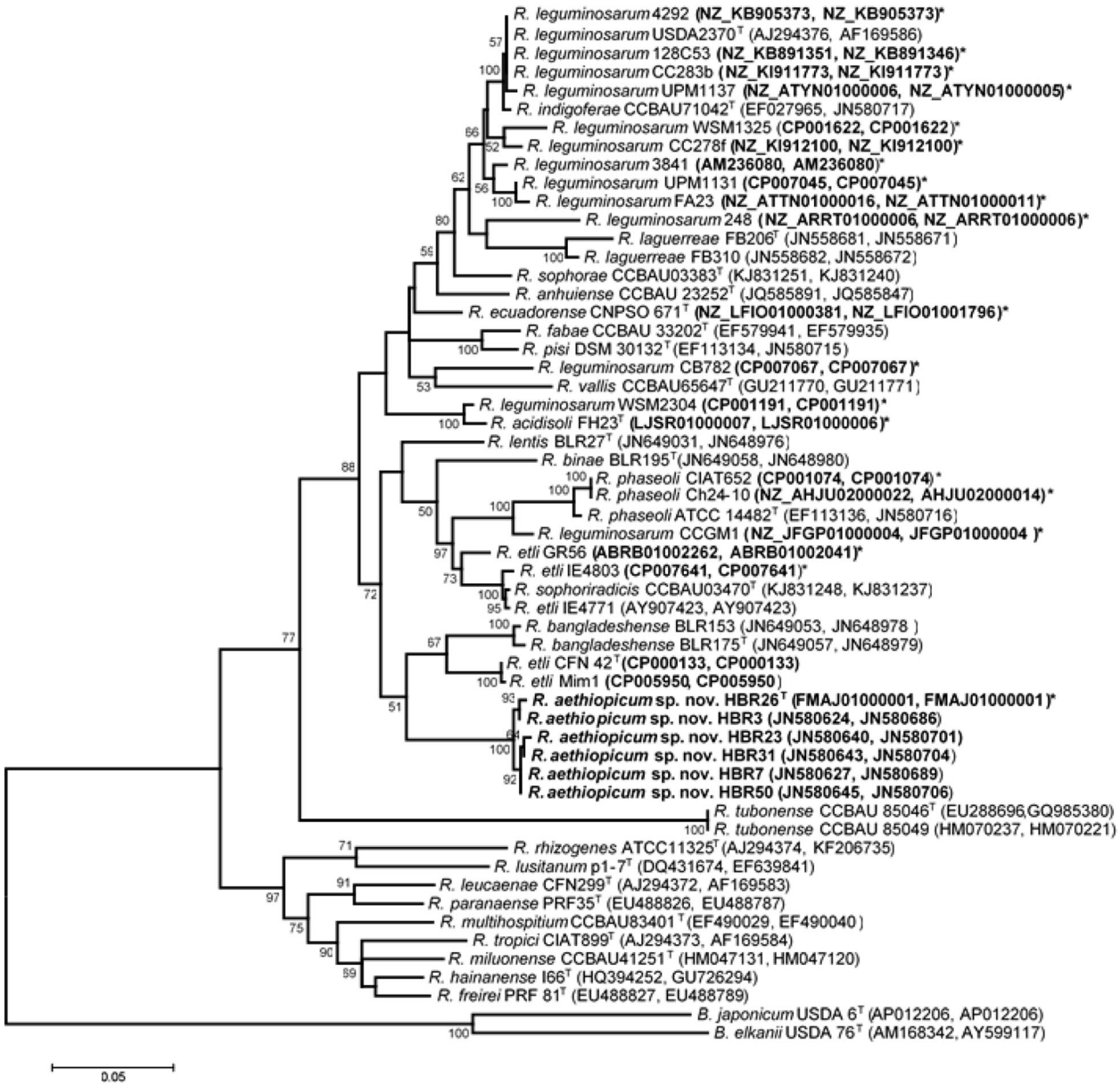

Fig. 2 Maximum Likelihood phylogenetic tree reconstructed based on *recA-glnII* concatenated nucleotide sequences, showing the relationships between *Rhizobium aethiopicum* sp. nov. (in bold) and recognized species of the genus *Rhizobium*. The tree was constructed by using Tamura-Nei model using MEGA version 7. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3397). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 32.0253% sites). Bootstrap values (100 replicates) are indicated at the branching points. Reference type strains are indicated with superscript 'T'. Bar, % estimated substitutions. GenBank accession numbers of the sequences (*recA, glnII* in order) are listed in parentheses next to strains codes. The accession numbers of whole genome sequenced strains are indicated with bold*. Abbreviations: B, *Bradyrhizobium*; R, *Rhizobium*; sp., species

DOE JGI. A summary of the genome project information is listed in Table 2.

Growth conditions and genomic DNA preparation

First HBR26^T (=HAMBI 3550^T=LMG 29711^T) was grown aerobically on YEM agar plates at 28 °C. A pure colony was transferred into 3 ml YEM broth medium and the cell culture was grown for four days in a shaker incubator (200 rpm) at 28 °C. One ml was used to inoculate 150 ml YEM broth, and cells were grown on a shaker (200 rpm) again at 28 °C until the culture reached late-logarithmic phase. DNA was isolated from cell pellets collected in a 60 ml following the CTAB bacterial genomic DNA isolation protocol Version Number 3 provided by the DOE JGI [41].

Table 1 Classification and general features of *Rhizobium aethiopicum* sp. nov. HBR26^T [63]

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [64]
		Phylum *Proteobacteria*	TAS [65]
		Class *Alphaproteobacteria*	TAS [66]
	Classification	Order *Rhizobiales*	TAS [67]
		Family *Rhizobiaceae*	TAS [68]
		Genus *Rhizobium*	TAS [68, 69]
		Species *R. aethiopicum* sp. nov.	IDA
		Type strain HBR26^T	IDA
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	28 °C	IDA
	pH range; Optimum	5–10; 7	IDA
	Carbon source	Varied (see Additional file 2: Table S2)	IDA
MIGS-6	Habitat	Soil, root nodule, on host	TAS [24]
MIGS-6.3	Salinity	Non-halophile	IDA
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free living, symbiotic	IDA
MIGS-14	Pathogenicity	Non-pathogenic	NAS
MIGS-4	Geographic location	Central Ethiopia	TAS [24]
MIGS-5	Sample collection	September, 2007	TAS [24]
MIGS-4.1	Latitude	8° 35′ 49.80″	TAS [24]
MIGS-4.2	Longitude	39° 22′ 49.27″	TAS [24]
MIGS-4.4	Altitude	1661	TAS [24]

Evidence codes – *IDA* Inferred from Direct Assay, *TAS* Traceable Author Statement (i.e., a direct report exists in the literature), *NAS* Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [70]

Genome sequencing and assembly

The genome was sequenced at the DOE JGI using a combination of Illumina HiSeq 2500 and Illumina HiSeq 2500-1 TB technologies [42]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 9,310,748 reads totaling 1405.9 Mbp. Methods used for library construction and sequencing can be found at the DOE JGI website [43]. In order to discard artifacts from Illumina sequencing and library preparation, all raw Illumina sequence data was passed through the program DUK at DOE JGI [43]. Filtered Illumina reads were assembled using Velvet (version 1.2.07) [44] and then from Velvet contigs, 1–3 kb simulated paired-end reads were constructed using wgsim (version 0.3.0) (https://github.com/lh3/wgsim). Allpaths–LG (version r46652) [45] was used to assemble Illumina reads with a simulated read. The final assembly was based on 1,290.5 Mbp of Illumina data, which provides 258.1× input read coverage of the genome. The draft genome is 6.6 Mbp in size and contains 64 contigs in 62 scaffolds.

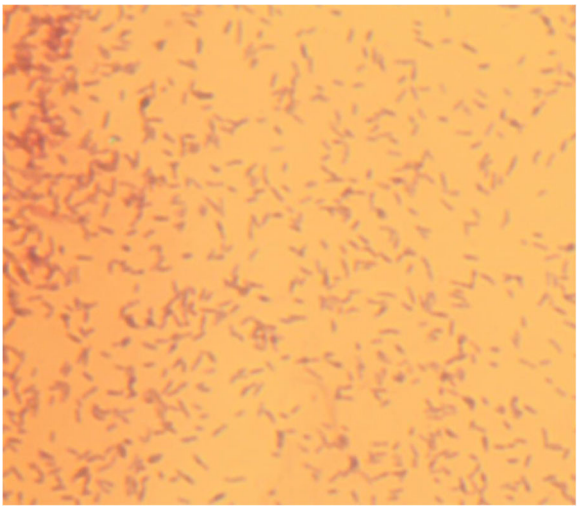

Fig. 3 Gram stain of *Rhizobium aethiopicum* sp. nov. strain HBR26^T

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Illumina std shotgun library
MIGS 29	Sequencing platforms	Illumina HiSeq 2500, Illumina HiSeq 2500-1 TB
MIGS 31.2	Fold coverage	258.1×
MIGS 30	Assemblers	Velvet (version 1.2.07), Allpaths–LG (version r46652)
MIGS 32	Gene calling method	Prodigal
	Locus Tag	ATF61
	Genbank ID	FMAJ00000000
	Genbank Date of Release	03-AUG-2016
	GOLD ID	Gp0108286
	BIOPROJECT	PRJNA303274
MIGS 13	Source Material Identifier	HBR26
	Project relevance	Symbiotic N_2 fixation, agriculture

Genome annotation

Genes were predicted using Prodigal [46] and using the DOE JGJ annotation pipeline [47]. The identified protein-coding genes were translated and functionally annotated by comparing the sequences with the NCBI non-redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNA genes were found using tRNAScanSE tool [48] and ribosomal RNA genes were identified by searches against models of the ribosomal RNA genes at the SILVA database [49]. Other non–coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL [50]. Additional analysis was accomplished using the IMG tool [51]. The same tool was also used for manual functional annotation of the predicted genes and for examining the genome sequence.

Genome properties

The genome of HBR26^T is arranged in 62 scaffolds and consists of 6,557,588 bp, with a 61% G + C content. In total 6307 genes were predicted, of these 6221 were protein-coding genes and 86 were RNA genes. Five rRNAs identified including one 16S rRNA, two 5S rRNA, and two 23S rRNA genes. There were 52 tRNA genes and 29 other (miscRNA) RNA genes. The statistics and properties of the genome are summarized in Table 3. The majority of the protein-coding genes, 5054 (80.13%) were assigned with putative functions (Table 3), and of these 4578 genes (72.59%) were assigned to COG functional categories (Table 4). The remaining genes were annotated as hypothetical proteins (1167 genes, 18.5%).

Table 3 Genome statistics

Attribute	Value	% of total
Genome size (bp)	6,557,588	100
DNA coding (bp)	5,707,275	87.03
DNA G + C (bp)	4,004,707	61.07
DNA scaffolds	62	100
Total genes	6307	100
Protein coding genes	6221	98.64
RNA genes	86	1.36
Pseudo genes	not determined	
Genes in internal clusters	962	15.25%
Genes with function prediction	5054	80.13%
Genes assigned to COGs	4578	72.59%
Genes with Pfam domains	5315	84.27%
Genes with signal peptides	530	8.40%
Genes with transmembrane helices	1406	22.29%
CRISPR repeats	0	

Insights from the genome sequence

Genome wide comparative analysis

Based on *recA-glnII* concatenated sequence comparisons, the proposed type strain HBR26^T and strains included in *R. aethiopicum* sp. nov., HBR23, HBR3, HBR31, HBR7, and HBR50 were closely related to each other (99–100% sequence identity). Nevertheless, these strains were only distantly related to the closest reference strains *R. etli* CFN42^T (94%) and *R. bangladeshense* BLR175^T (93%). In order to further resolve the taxonomy of the novel group, genomic comparative analyses were done between HBR26^T and several relatively close reference strains presented in the Fig. 2. For this the genomes of a number strains, such as *R. etli* CFN42^T, *Rhizobium etli* IE4771, *Rhizobium etli* Mim1, *Rhizobium etli* IE4803, *Rhizobium phaseoli* Ch24-10, *Rhizobium phaseoli* CIAT652, *Rhizobium acidisoli* FH23, *Rhizobium ecuadorense* PSO671^T, and *Rhizobium leguminosarum* CB782, CCGM1, WSM2304, PM1131, WSM1325, 4292, 3841, and UPM1137 were retrieved from the DOE JGI genome portal (Tables 5 and 6). ANI was computed from protein-coding genes of the genomes using the MiSI program implemented in the IMG database [35]. For a pair of genome sequences, the system calculates ANI by averaging the nucleotide identity of orthologous genes identified as bidirectional best hits and also calculates Alignment Fraction of orthologous genes [35]. In addition, partially sequenced genome reads from *R. bangladeshense* BLR175^T, *Rhizobium lentis* BLR27^T, *Rhizobium binae* BLR 95^T, *Rhizobium anhuiense* CCBAU23252^T, *R. pisi* DSM30132^T and *Rhizobium fabae* CCBAU33202^T were used for calculation of additional ANI with the JSpecies program using default parameters as previously used [52, 53]. Table 5 shows the ANI values obtained between HBR26^T and reference strains (numbers above the diagonal). The numbers below the diagonal show pairwise orthologous genes identified as bidirectional best hits between genomes. AF was >0.68 in all ANI calculations among whole or draft genomes but the AF value was <0.6 in all ANI calculations with partially sequenced genome reads. The ANI values obtained between HBR26^T and references strains varied between 87.4 and 91.8%, which was below 96%, the value of relatedness recommended for species delineation [35]. The closest strains were *R. bangladeshense* LR175^T and *R. etli* CFN42^T with ANI values 91.8 and 90.2%, respectively. This result is in agreement with the *recA-glnII* concatenated analysis (Fig. 2), confirming that that HBR26^T is distantly related to the *R. etli* and *R. bangladeshense* species but belongs to the novel *Rhizobium* species. The ANI between *R. etli* IE4803 and *R. etli* IE4771 was 97.7%. However, ANI

Table 4 Number of genes associated with general COG functional categories

Code	Value	%age	Description
J	221	4.24	Translation, ribosomal structure and biogenesis
A	0	00	RNA processing and modification
K	467	8.96	Transcription
L	123	2.36	Replication, recombination and repair
B	2	0.04	Chromatin structure and dynamics
D	41	0.79	Cell cycle control, Cell division, chromosome partitioning
V	115	2.21	Defense mechanisms
T	252	4.83	Signal transduction mechanisms
M	274	5.25	Cell wall/membrane biogenesis
N	85	1.63	Cell motility
U	106	2.03	Intracellular trafficking and secretion
O	189	3.62	Posttranslational modification, protein turnover, chaperones
C	267	5.12	Energy production and conversion
G	557	10.68	Carbohydrate transport and metabolism
E	557	10.68	Amino acid transport and metabolism
F	108	2.07	Nucleotide transport and metabolism
H	239	4.58	Coenzyme transport and metabolism
I	209	4.01	Lipid transport and metabolism
P	274	5.25	Inorganic ion transport and metabolism
Q	145	2.78	Secondary metabolites biosynthesis, transport and catabolism
R	566	10.83	General function prediction only
S	363	6.96	Function unknown
-	1729	27.41	Not in COGs

The total is based on the total number of protein coding genes in the genome

values between these strains and the type strain *R. etli* CFN42^T (= < 90.2%) was much below the cutoff value of strains of the same species. Several *R. leguminosarum* strains included in Table 5 may represent species other than *R. leguminosarum* (ANI < 96% each other). The genome of *R. leguminosarum* CCGM1 showed a significantly higher degree of similarity with *R. phaseoli* Ch24-10 (97.2% ANI) and CIAT652 (97.2% ANI), and could thus be classified as *R. phaseoli*. *R. leguminosarum* WSM2304 showed 96.6% genomic relatedness with *R. acidisoli* FH23^T. Accordingly, we suggest the classification of WSM2304 under *R. acidisoli* species. The ANI value between *R. fabae* CCBAU33202^T and *R. pisi* DSMZ30132^T was 96.6%. This value corroborates the relationship between the two strains as reported previously [24], which is also shown in the *recA-glnII* based phylogenetic tree in Fig. 2, suggesting that *R. fabae* CCBAU33202^T and *R. pisi* DSMZ30132^T might represent one and the same species.

Table 6 shows the genome statistics and functional category comparison between HBR26^T and close reference rhizobial strains. The draft genome of HBR26^T (6.6 Mbp) is about the same size as that of *R. phaseoli* Ch24-10 (6.6 Mbp) and slightly greater than *R. etli* CFN42^T (6.5 Mbp) and *R. phaseoli* CIAT652 (6.4 Mbp). However, strain HBR26^T has smaller genome size compared to *R. leguminosarum* CCGM1 (6.8 Mbp), *R. etli* IE4803 (6.9 Mbp), *R. acidisoli* FH23 (7.3Mbp), *R. ecuadorense* CNPSO671 (6.9Mbp) and all other *R. leguminosarum* (6.8-7.9 Mbp) symbiovar *viciae* and *trifolii* reference strains (Table 6). Though the gene content of strain HBR26^T (6307) is only greater than of CIAT652 (6132), it has got the highest percentage of genes assigned to Pfam (84.3%), TIGRfam (24.8%), and KEGG (29.7%). HBR26^T also has the highest percentage of genes assigned to COG (72.6%) and KOG (18.1%) functional categories, with the exceptions *R. leguminosarum* UPM1131 (72.9%), and WSM2304 (18.6%), respectively.

In Fig. 4 the Venn diagram plotted in the OrthoVenn program shows overlapping orthologous protein clusters between the genomes of HBR26^T and other common bean-nodulating references *R. etli* CFN42^T, and *R. phaseoli* Ch24-10, CIAT652 and CCGM1. The orthologous clusters were identified with default parameters, 1e-5 e-value cutoff for all protein similarity comparisons and 1.5 inflation value for the generation of orthologous clusters [54]. In total the strains formed 6534 protein clusters, 6462 orthologous clusters (at least containing two strains) and 4273 single-copy gene clusters. All five strains shared in common 4385 orthologous protein clusters. On a pairwise basis, HB26^T shares 32, 42 and 44 proteins with CCGM1, Ch24-10, and CIAT652, respectively. Strain HBR26^T shares the most with CFN42^T with 164 orthologous group. This result is in agreement with *recA-glnII* phylogenetic and ANI analysis, supporting that HBR26^T is more closely related to CFN42^T compared to the other bean-nodulating strains. The genome of HBR26^T contains the highest number of genome-specific proteins of the five strains with 665 singletons followed by CFN 42^T, CIAT652, Ch24-10 and CCGM1 with 568, 549 and 516 singletons, respectively.

Comparative analysis of accessary genes: emphasis on symbiotic genes

Genes which are not essentially present in all bacterial strains are known as accessory genes. These genes are contained by mobile elements such as plasmids, genomic islands, transposons or phages and thus can be gained or lost among bacterial strains through horizontal gene transfer mechanisms. Accessory genes in the genome of HBR26^T were searched by assembling against the reference genome *R. etli* CFN42^T using the Genome Gene Best Homologs package from program IMG-ER [55].

Table 5 ANI Genomic comparison between *R. aethiopicum* sp. nov. HBR26^T and other members of *Rhizobium* species

	Genome Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	*R. aethiopicum* HBR26^T		**90.2**	89.6	89.1	89.0	89.0	88.9	88.1	87.8	87.7	87.5	88.5	87.3	88.2	87.6	**91.8**	**90.2**	87.4	89.4	87.7	87.7	88.0	88.6
2	*R. etli* CFN 42^T	5003		98.7	90.1	90.3	90.3	90.2	87.6	88.1	87.5	87.8	89.6	87.2	88.2	87.5	89.8	89.3	87.7	88.9	87.4	87.5	90.2	88.8
3	*R. etli* Mim1	4876	5410		89.57	89.73	89.78	89.7	87.6	87.9	87.4	87.1	89.63	87.1	87.4	87.4	89.5	89.9	88.3	88.6	87.3	87.4	89.6	88.1
4	*R. etli* IE4803	5127	4979	4885		92.5	92.4	92.5	88.2	88.7	88	87.9	97.7	87.9	88.6	88.1	88.8	89.8	89.5	89	87.9	87.7	89.0	89.7
5	*R. leguminosarum* CCGM1	4943	5060	4882	5132		97.2	97.3	88.2	88.8	87.8	87.8	91.8	87.6	88.5	87.8	88.3	89.3	90.3	89.4	88	87.8	89.0	89.8
6	*R. phaseoli* Ch24-10	4923	5120	4831	5076	5697		98.3	88	88.7	87.8	88.8	91.8	87.6	88.4	87.8	87.5	89.6	91	89.3	87.9	87.7	89.0	89.7
7	*R. phaseoli* CIAT 652	4767	4764	4667	4834	5274	5205		88.1	88.8	87.9	87.8	91.9	87.7	88.6	87.9	88.9	89.8	91	89.1	88	87.8	89.0	89.8
8	*R. leguminosarum* CB782	4736	4735	4762	4674	4822	4701	4521		90.9	89.0	89.1	88.1	89.0	89.2	89.0	88.2	88.6	86.3	89.6	88.3	88.2	90.9	90.7
9	*R. leguminosarum* WSM2304	4540	4650	4687	4617	4591	4770	4515	4999		89.1	89.2	88.6	89.1	89.3	89.3	88.5	89	87.2	90.2	88.4	88.4	96.6	92.6
10	*R. leguminosarum* UPM1131	4583	4576	4572	4585	4708	4512	4334	4859	4844		93.1	87.9	94.0	93.0	94.8	88.8	89.7	88.7	91.5	89.1	89	89.3	89.4
11	*R. leguminosarum* UPM1137	4532	4653	4661	4616	4616	4653	4463	4954	4950	5489		87.8	93.3	98.5	93.3	88.6	89.8	88.6	92	89.3	89	89.3	89.4
12	*R. etli* IE4771	4975	4889	5038	5621	5071	4830	4744	4705	4612	4650	4643		87.8	87.9	87.8	88.5	89.3	89.7	88.8	87.7	87.6	88.8	88.9
13	*R. leguminosarum* WSM1325	4559	4608	4580	4528	4640	4561	4361	4904	4899	5345	5436	4568		93.3	94.4	88.6	89.4	87.1	91.5	88.6	88.7	89.0	89.4
14	*R. leguminosarum* 4292	4775	4897	4697	4872	4957	4895	4690	4959	4922	5281	6230	4711	5405		93.2	88.3	89.5	87.4	91.8	89	88.8	89.5	90.0
15	*R. leguminosarum* 3841	4659	4657	4635	4627	4611	4574	4375	4981	4916	5499	5565	4637	5647	5536		88.4	89.5	87.7	91.5	88.8	89	89.2	89.2
16	*R. bangladeshense* BLR 175^T	40	41	43	40	43	39	39	40	37	43	42	37	42	40	39		89.2	86.9	85.6	87.2	89.1	88.5	88.6
17	*R. lentis* BLR 27^T	55	63	55	53	58	56	53	58	57	53	56	51	52	70	69	22		87.7	77.8	88.9	87.7	89.1	89.6
18	*R. binae* BLR 195^T	79	88	81	86	86	82	78	90	90	97	97	85	100	84	97	33	27		80.0	88.1	88.5	87.2	87.1
19	*R. anhuiense* CCBAU 23252^T	51	52	56	56	60	55	54	54	54	54	54	53	55	55	52	25	22	25		89.2	90.0	90.0	89.7
20	*R. pisi* DSMZ 30132^T	1479	1480	1487	1478	1510	1507	1424	1552	1513	1530	1640	1484	1537	1581	1564	39	44	79	48		97.5	88.4	88.5
21	*R. fabae* CCBAU 33202^T	2058	2088	2158	2079	2156	2141	1988	2162	2105	2139	2237	2088	2147	2222	2184	46	52	95	58	1467		88.2	88.3
22	*R. acidisoli* FH23^T	4745	5003	4876	4803	5098	4959	4686	4991	5399	4858	4954	4781	4918	5140	4933	42	65	89	54	1514	2129		92.6
23	*R. ecuadorense PSO671*^{T}	4630	4728	4619	4784	4906	4820	4803	4838	4848	4593	4778	4564	4674	5013	4808	41	55	84	56	1451	2058	5023	

Gray shade indicates ANI calculated using partially sequenced genomes. The conting fatsa files of the reads were obtained from Professor J.P.W. Young, the University of York and read data are also deposited at NCBI database under Bioproject accession number PRJEB7125 or PRJEB7987; number below the diagonal are pairwise orthologous genes identified as bidirectional best hits between genomes; AF was >0.68 in all ANI calculation among whole or draft genome but AF value was <0.6 in all ANI calculation with partially sequenced genome reads. Numbers above the diagonal are ANI between genomes. Reference type strains are indicated with superscript 'T'; R, *Rhizobium*

Additional file 3: Table S3 shows homologous *repABC* (plasmid replication genes) and symbiotic genes found in the genome of HBR26^T. The result revealed that HBR26^T carries five different *repABC* genes homologous to the genes found in five of the *R. etli* CFN42^T [56] plasmids 42b, 42c, 42d, 42e, 42f, suggesting that HBR26^T may have five additional replicons other than the chromosome. The *repABC* genes corresponding to the symbiotic plasmid 42d showed high sequence similarity between other common bean nodulating strains CFN42^T, CIAT652, IE4803, Ch24-10, 4292 and CCGM1 (identity ranging 99–100%). This implies that bean- nodulating strains and HBR26^T may share common symbiotic plasmids. The HBR26^T *repABC* genes homologous to 42b, 42c, 42e and 42f also showed sequence similarity in the ranges 86–89%, 84–93%, 92–94%, 84–93%, respectively, with strains CIAT652, CFN42^T, IE4803, Ch24-10, 4292, CCGM1 and *R. etli* sv. *mimosae* Mim1.

The symbiosis between rhizobia and legume plants is initiated when plant exudates known as flavonoids trigger expression of the rhizobial nodulation genes that code for the synthesis of LCO Nod factors. The backbone of this LCO is encoded by the common *nodABC* accessory genes. There are also additional genes (*nol*, *noe*) which code for the substituent groups that decorate the LCO core [57]. The symbiosis between rhizobia and legumes results in the formation of specialized organs on plant roots known as nodules in which rhizobia differentiate into N_2-fixing bacteroids [58]. Like most symbiotic rhizobia, the genome of HBR26^T carries the

Table 6 Genome statistics of *R. aethiopicum* sp. nov. HBR26^T and reference rhizobial strains

Status	Genome Name	IMG Genome ID	GenBank Accession number	Quality	Host Name	Genome Size (Mbp)	Gene	Scaf-fold[a]	GC %	CDS %	RNA[a]	COG %	KOG %	Pfam %	TIGR-fam %	KEGG %
Draft	*R. aethiopicum* BR26^T	2615840624	PRJNA303274	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.6	6307	62	0.61	98.64	86	72.6	18.1	84.3	24.8	29.7
Finished	*R. etli* CFN 42^T	2623620267	CP000133	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.5	6345	7	0.61	98.5	95	69.9	17.5	82.0	24.2	29.0
Finished	*R. etli* Mim1	2565956559	CP005950	High	*Mimosa affinis*[b]	7.2	7006	7	0.61	97.82	153	70.2	17.8	80.2	24.3	28.7
Finished	*R. etli* IE4803	2630968325	CP007641	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	7.0	6708	5	0.61	98.57	96	71.3	17.6	83.0	24.6	29.0
P. Draft	*R. leguminosarum* CCGM1	2609460209	JFGP00000000	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.9	6711	55	0.61	98.63	92	69.2	17.1	81.0	23.8	28.1
P. Draft	*R. phaseoli* Ch24-10	2548876814	AHJU00000000	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.6	6593	352	0.61	98.82	78	67.6	16.6	81.0	23.7	28.2
Finished	*R. phaseoli* CIAT 652	642555152	CP001074	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.5	6132	4	0.61	99.02	60	70.7	17.7	81.2	25.2	29.0
Finished	*R. leguminosarum* CB782	2510065076	CP007067	High	*Trifolium semipilosum* (http://plants.usda.gov/core/profile?symbol=TRSE7)	6.7	6559	4	0.61	98.67	87	72.5	18.4	83.2	24.5	28.8
Finished	*R. leguminosarum* WSM2304	643348569	CP001191	High	*T. polymorphum* (http://plants.usda.gov/java/ClassificationServlet?source=display&classid=TRPO6)	6.9	6643	5	0.61	99.07	62	70.9	18.6	83.1	23.9	28.5
P. Draft	*R. leguminosarum* UPM1131	2513237084	CP007045	High	*Pisum sativum* (http://plants.usda.gov/core/profile?symbol=PISA6)	7.2	6951	41	0.61	98.83	81	72.9	17.9	83.5	23.7	27.9
P. Draft	*R. leguminosarum* UPM1137	2513237085	ATYN00000000	High	*P. sativum* (http://plants.usda.gov/core/profile?symbol=PISA6)	7.7	7462	49	0.61	99.04	72	71.0	17.6	81.9	22.4	28.1
Finished	*R. etli* IE4771	2585427632	CP006986	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	7.1	6894	6	0.61	98.23	122	71.3	17.9	81.1	24.3	28.9
Finished	*R. leguminosarum* WSM1325	644736401	CP001622	High	*Trifolium* (http://www.theplantlist.org/tpl1.1/record/ild-8146)	7.4	7292	6	0.61	99.18	60	68.7	17.5	81.6	22.4	26.9
P. Draft	*R. leguminosarum* 4292	2516653085	AQZR01000000	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	7.3	7193	5	0.61	98.83	84	71.8	17.9	83.2	23.2	28.5
Finished	*R. leguminosarum* 3841	2623620212	AM236080	High	*P. sativum* (http://plants.usda.gov/core/profile?symbol=PISA6)	7.8	7447	7	0.61	98.74	94	71.7	17.7	82.7	22.5	27.3

Table 6 Genome statistics of *R. aethiopicum* sp. nov. HBR26^{T} and reference rhizobial strains *(Continued)*

P. Draft	*R.acidisoli* FH23^{T}	2648501703	LJSR00000000	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	7.3	7111	104	0.61	98.83	83	69.6	17.4	81.6	22.9	27.6
P. Draft	*R. ecuadorense* CNPSO 671^{T}	2648501138	LFIO00000000	High	*P. vulgaris* (http://plants.usda.gov/core/profile?symbol=PHVU)	6.9	6668	139	0.61	98.85	77	71.2	17.8	82.3	24.2	29.1

P. draft, permanent draft; [a] number of scaffolds or number of RNA; [b]broad host range, including plants of *M. affinis* (http://www.theplantlist.org/tpl1.1/record/ild-15931), *Leucaena leucocephala* (http://plants.usda.gov/core/profile?symbol=LELEL2), *Calliandra grandiflora* (http://www.theplantlist.org/tpl1.1/record/ild-20119), *Acaciella angustissima* (http://www.theplantlist.org/tpl1.1/record/ild-28474) as well as *P. vulgaris* [71]. Reference type strains are indicated with superscript 'T'; R, *Rhizobium*

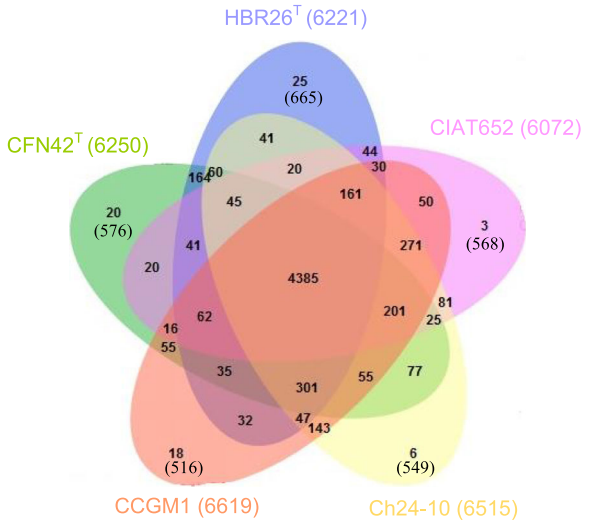

Fig. 4 Venn diagram plotted by OrthoVenn program shows shared orthologous protein clusters among the genomes of bean-nodulating rhizobial strains (in the center): *Rhizobium etli* CFN42^T, *Rhizobium phaseoli* Ch24-10, *Rhizobium phaseoli* CIAT652, *Rhizobium leguminosarum* CCGM1 and *Rhizobium aethiopicum* type strain HBR26^T. The number of protein clusters comprising multiple protein families is indicated for each genome and also the number of singletons i.e., protein with no orthologous of each strain are shown in parenthesis. The total number of protein sequences of each genome are indicated in parentheses next to strains codes

symbiotic genes encoding for the synthesis of LCO structures, substituent groups and genes coding for nitrogen fixation (Additional file 3: Table S3). Several of the nodulation and nitrogen-fixing genes are located on the scaffolds Ga0061105_135 and Ga0061105_130, 141, 144 and 150. The first scaffold contains the main nodulation genes except *nodA,* while the other scaffolds encompass many of the nitrogen-fixing genes (Additional file 3: Table S3).

The genomes of HBR26^T, *R. etli* CFN42^T, *R. phaseoli* Ch24-10 and CIAT652 were aligned using the progressive Mauve alignment tool [59], using default parameters. The genomic features were visualized using the Artemis Comparison Tool [60, 61]. The Mauve alignment in Fig. 5 shows the presence of a similar *nodBCSIJD* module organization between the genome of HBR26^T and the genomes of other bean-nodulating rhizobial strains CFN42^T, CIAT652, and Ch24-10. The *nodDIJSCB* genes are flanked by transposase genes and hypothetical protein-coding genes. A similar arrangement of the *nod* genes was also found in the genomes of CCGM1 and IE4803, which are also micro-symbionts of common bean (data not shown).

All HBR26^T, CFN42^T, Ch24-10, CIAT652, and CCGM1 genomes carry additional *nodZ, noeI* and *nolE* genes adjacent to the *nodBCSIJD* region. Similarly, in the genomes of clover and faba bean nodulating *R. leguminosarum* WSM2304, UPM1131 and 3841 the nodulation genes *nodD, nodB, nodC, nodI,* and *nodJ* are also clustered in the same region. In the latter case, this region contains additional *nodA, nodL, nodE,* and *nodF* genes as well. The *nodA* and *nolL* genes of HBR26^T which are located in the scaffolds Ga0061105_134 and Ga0061105_130, respectively, are very similar to the corresponding gene sequences of bean-nodulating rhizobial strains CFN42^T, Ch24-10, CCGM1, CIAT652 and IE4803 (99–100% similarity). Its *nodB* gene is also homologous with CFN42^T, CIAT652, and IE4803. The highest identity (100%) is with *nodB* of IE4803 followed by CFN42^T (98%) and CIAT652 (97%). *nodC* of HBR26^T shares 97% similarity with *nodC* of CIAT652, CFN42^T and, CCGM1. All *nodS, nodI* and *nodJ* genes of HBR26^T share high identity with those of CIAT652 (99%), CFN42^T (98%), CCGM1 (98%) and Ch24-10 (98%).

The nitrogenase complex, an enzyme responsible for nitrogen fixation in diazotrophs, consists of two components known as dinitrogenase and dinitrogenase reductase [62]. The *nif* genes are required for the synthesis and functioning of the nitrogenase complex [62]. Many of these genes in the genome of HBR26^T are harbored in four different scaffolds Ga0061105_130, Ga0061105_150, Ga0061105_144, and Ga0061105_141. The first scaffold contains the *nifA-nifB-nifT-nifZ-nifW* genes, and the second scaffold includes the *nifE, nifN* and *nifX* genes. The nitrogen-fixing genes *nifH, nifU* and *nifQ* are retained in the scaffold Ga0061105_141. An additional *nifH* gene, *fixG* and *fixH* genes are found in the scaffold Ga0061105_144 and a *nifK* gene is located in the scaffold Ga0061105_162. The dinitrogenase component of the nitrogenase complex is a product of *nifD* and *nifK* genes and the dinitrogenase reductase is coded by *nifH* [62]. However, the *nifD* gene is missing in the draft genome of HBR26^T. This gene is important to enable the nitrogenase enzyme complex functional. On the other hand, the strain HBR26^T makes effective nitrogen-fixing symbiosis with common bean plants. Thus, the reason behind the absence of *nifD* in the genome of HBR26^T is probably because our data is a draft genome and probably *nifD* was missed during sequencing. It is also possible that *nifD* sequence was truncated when the library was constructed.

The genes *nifB, nifT, nifZ, nifE, nifN, nifX, fixG, fixH, nifW, nifQ, nifK* and *nifH* all share high identity with homologous genes found in CFN42^T (98–100%), Ch24-10 (98–100%), CCGM1 (98–100%), 4292 (96–99%) or in IE4803 (92–100%). In our previous study, we identified rhizobial strains belong to *R. phaseoli*, *R. etli* and *R. leguminosarum* from root nodules of common bean plants growing in the soils of Ethiopia [24]. Thus, the close

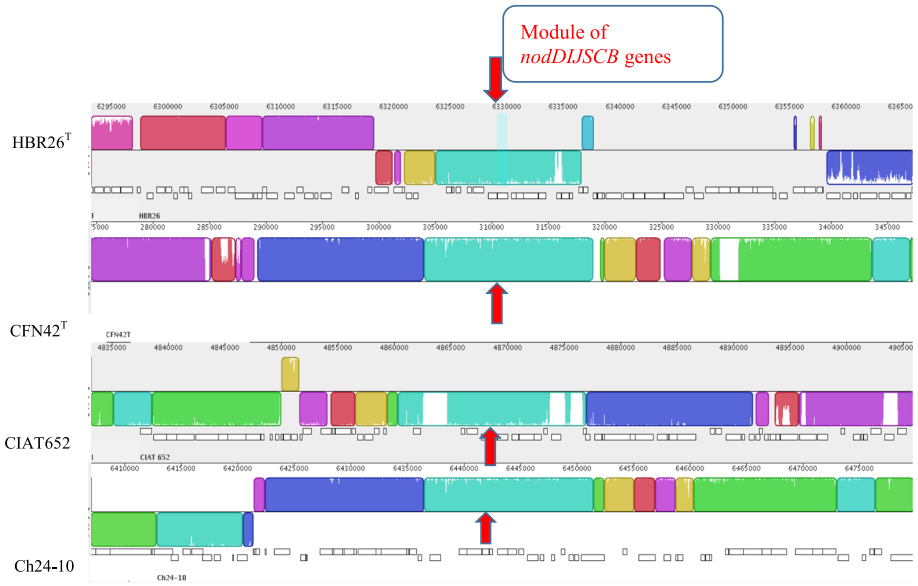

Fig. 5 Mauve alignment comparing the genome of *Rhizobium aethiopicum* type strain HBR26^T with the genome of *Rhizobium etli* CFN42^T, *Rhizobium phaseoli* CIAT652 and *Rhizobium phaseoli* Ch24-10. The module of *nodDIJSCB* genes are indicated by the *arrows*

similarity of the *nod*, *nif* and *fix* genes between HBR26^T and bean-nodulating *R. etli, R. phaseoli* and *R. leguminosarum* strains suggests that those genes might be shared between these rhizobial species through horizontal gene transfer mechanisms.

Conclusion

This study presents the genome sequence for the *R. aethiopicum* sp. nov. strain HBR26^T. The result from phylogenetic analyses of multilocus sequences of core genes showed a novel species within the genus *Rhizobium*. This result was further supported by ANI calculation, in which the genome of the type strain HBR26^T exhibited < 91.8% identity when compared with the genomes of close *Rhizobium* species. This value is much lower than the 96% ANI limit for delineating a species. The data confirms that *R. aethiopicum* sp. nov. should be considered as a new *Rhizobium* species. Thus, on the basis of phylogenetic, comparative genomic analyses and ANI results and by including phenotypic characteristics, we formally propose the creation of *R. aethiopicum* sp. nov. that contains the strain HBR26^T (= HAMBI 3550^T=LMG 29711^T). The strains included in this species are effective nitrogen-fixing rhizobia in symbiosis with common bean plants. The genome of the type strain HBR26^T carries five plasmid replication *repABC* genes homologous to the genes found in five of the *R. etli* CFN42^T plasmids, suggesting that HBR26^T may have five additional replicons other than the chromosome. The organization of *nodBCSIJD* genes is similar between the genomes of HBR26^T and other bean-nodulating rhizobial species. The symbiotic genes necessary for nodulation and for nitrogen fixation share high sequence similarity between bean-nodulating strains, such as *R. etli, R. phaseoli* and *R. leguminosarum*, which suggests that these genes might be shared between bean-nodulating rhizobial species through horizontal gene transfer mechanisms.

Description of *Rhizobium aethiopicum* sp. nov.

Rhizobium aethiopicum (ae.thi.o'pic.um. L. neut. adj. *aethiopicum*, pertaining to Ethiopia). Fast-growing, forming moist, raised and smooth colonies 3–5 mm in diameter within 3–4 days on YEM agar plates under optimal growth conditions, at 28 °C and pH7. The strains are able to grow between 15 °C and 30 °C. The organisms require no or trace amounts of NaCl for growth and are only able to grow at NaCl concentrations of 0–0.5% and at pH values in the range 5–10. No growth occurred at pH4, at temperature 4 °C and at 37 °C, and 1–5% NaCl. Cells are Gram-negative rod-shaped and 1.0–2.4 μM in length. Oxidation of the following substrates as carbon sources in Biolog GN2 microplates was recorded positive; dextrin, glycogen, N-acetyl-D-glucosamine, adonitol, L-arabinose, D-arabitol, D-cellobiose, I-erythritol, D-fructose, L-fucose, D-galactose, α-D-glucose, α-D-lactose, lactulose, maltose, D-mannitol, D-mannose, D-melibiose, β-methyl-D-glucoside, D-psicose, D-raffinose, L-rhamnose, D-sorbitol, sucrose, D-trehalose, turanose, xylitol, pyruvic acid methyl ester, succinic acid mono-methyl-ester, β-hydroxybutyric acid, γ-hydroxybutyric acid, itaconic acid, α-keto butyric acid, α-keto glutaric acid, D,L-lactic acid, succinic acid, bromo-succinic acid, succinamic acid, L-alaninamide,

D-alanine, L-alanine, L-alanyl-glycine, L-asparagine, L-aspartic acid, L-glutamic acid, glycyl-L-glutamic acid, L-histidine, hydroxy-L-proline, L-ornithine, L-proline, D,L-carnitine, γ-amino butyric acid, urocanic acid, nosine, uridine, thymidine, glycerol, α-d-glucose-1-phosphate and D-glucose-6-phosphate. However, the oxidation was negative for the following substrates: α-cyclodextrin, Tween 40, Tween 80, N-acetyl-D-galactosamine, gentiobiose, acetic acid, D-galacturonic acid, D-gluconic acid, D-glucosaminic acid, D-glucuronic acid, p-hydroxy phenylacetic acid, α-keto valeric acid, propionic acid, D-saccharic acid, glucuronamide, L-phenylalanine, L-pyroglutamic acid, D-serine, phenyethyl-amine, putrescine, and 2-aminoethanol. The type strain HBR26^T (= HAMBI 3550^T =LMG 29711^T) was isolated from root nodules of common bean plants growing in Ethiopia. The genome size of the type strain is 6.6 Mbp and the G + C content of the genome is 61%. The genome sequence of the type strain is deposited at DOE JGI genome portal under IMG genome/Taxon ID: 2615840624 [39] and also available at European Nucleotide Archive [40] under accession numbers FMAJ01000001-FMAJ01000062. The type strain has been deposited in the HAMBI (HAMBI 3550^T) and LMG (LMG 29711^T) culture collections.

Additional files

Additional file 1: Table S1. Phenotypic characteristics of *Rhizobium aethiopicum* sp. nov. strains. (DOCX 21 kb)

Additional file 2: Table S2. Carbon sources utilization response between *Rhizobium aethiopicum* sp. nov. strains and *Rhizobium etli* CFN42^T. (DOCX 25 kb)

Additional file 3: Table S3. Rhizobium aethiopicum sp.nov. strain HBR26T repABC and symbiotic genes homologous to genes found in the symbiotic plasmid 42d of Rhizobium etli CFN 42T. (XLSX 45 kb)

Abbreviations

AF: Alignment fraction; ANI: Average nucleotide identity values; CTAB: Cetyl Trimethyl Ammonium Bromide; DDH: DNA-DNA Hybridization; DOE: Department of energy; GOLD: Genomes online database; IMG: Integrated microbial genomes; IMG-ER: Integrated microbial genomes – expert review; JGI: Joint Genome Institute; LCO: Lipochito-Oligosaccharide; MIGS: Minimum information about a genome sequence; MiSI: Microbial species identifier; MLSA: Multilocus sequence analysis; N_2: Dinitrogen; R2A: Reasoner's 2A Agar; YEM: Yeast Extract Mannitol

Acknowledgements

All microbiological lab work, data analyses, and manuscript preparation were supported by the SOILMAN project funded by Academy of Finland, University of Helsinki. Sequencing was performed by DOE JGI and the sequencing project was a part of the DOE JGI 2014 Genomic Encyclopedia of Type Strains, Phase III the genomes of soil and plant-associated and newly described type strains. We would like to thank Professor J.P.W. Young (University of York, UK) for providing Conting fasta files of additional reference genome reads. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported under Contract No. DE-AC02-05CH11231.

Authors' contributions

AAA, KL and WBW planned the genome sequencing project. AAA isolated the described strains and performed cultivation, microbiological laboratory experiments, phenotypic characterization, DNA extraction, PCR, 16S rRNA gene, *recA* and *glnII* gene sequences analyses. AAA prepared phylogenetic trees, figures, genomic data analysis and wrote the manuscript. TW and NCK participated in the genome sequencing, assembly and genome annotation. KL and WBW conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Environmental Sciences, University of Helsinki, Viikinkaari 2a, Helsinki, Finland. [2]DOE Joint Genome Institute, Walnut Creek, USA. [3]Department of Microbiology, University of Georgia, Biological Sciences BuildingAthens, USA.

Received: 5 August 2016 Accepted: 24 December 2016
Published online: 26 January 2017

References

1. Franche C, Lindström K, Elmerich C. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. Plant Soil. 2009;321:35–59.
2. Gepts P, Debouck D. Origin, domestication and evolution of the common bean (*Phaseolus vulgaris* L.). In: van Schoonhoven A, Voysest O, editors. Common beans, research for crop improvement. Wallingford, UK, Cali-Colombia: CAB; 1991. p. 7–53.
3. Kaplan L. Archeology and domestication in American Phaseolus (Beans). Econ Bot. 1965;19:558–68.
4. Singh SP. Production and utilization. In: Singh SP, editor. Common bean improvement in the twenty-first century. Boston: Kluwer Academic Publishers; 1999. p. 1–24.
5. Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ, Hirota MM. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. Biol Conserv. 2009;142:1141–53.
6. Assefa T, Assefa H, Kimani P. Development of improved haricot bean germplasm for the mid and low-altitude sub-humid agroecologies of Ethiopia. In: Ali K, Ahmed S, Beniwal S, Kenneni G, Rajandra SM, Makkouk K, editors. Workshop on food and forage legumes of Ethiopia. Aleppo: ICARDA; 2006. p. 87–94.
7. Martínez-Romero E, Segovia L, Mercante FM, Franco AA, Graham P, Pardo MA. *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees. Int J Syst Bacteriol. 1991;41:417–26.
8. Segovia L, Young JPW, Martinez-Romero E. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strain as *Rhizobium etli* sp. nov. Int J Syst Bacteriol. 1993;43:374–7.
9. Amarger N, Macheret V, Laguerre G. *Rhizobium gallicum* sp. nov. and *Rhizobium giardinii* sp. nov. from *Phaseolus vulgaris* nodules. Int J Syst Bacteriol. 1997;47:996–1006.
10. Ribeiro RA, Rogel MA, Lòpez-Lòpez A, Ormen˜o-Orrillo E, Barcellos FG, Martínez J, Thompson FL, Martínez-Romero E, Hungria M. Reclassification of *Rhizobium tropici* type A strains as *Rhizobium leucaenae* sp. nov. Int J Syst Evol Microbiol. 2012;62:1179–84.
11. Valverde A, Igual JM, Peix A, Cervantes E, Velázquez E. *Rhizobium lusitanum* sp. nov. a bacterium that nodulates *Phaseolus vulgaris*. Int J Syst Evol Microbiol. 2006;56:2631–7.
12. Wang F, Wang ET, Wu LJ, Sui XH, LiJr Y, Chen WX. *Rhizobium vallis* sp. nov., isolated from nodules of three leguminous species. Int J Syst Evol Microbiol. 2011;61:2582–8.
13. Ribeiro RA, Martins TB, Ormeño-Orrillo E, et al. *Rhizobium ecuadorense* sp. nov., an indigenous N_2-fixing symbiont of the Ecuadorian common bean (*Phaseolus vulgaris* L.) genetic pool. Int J Syst Evol Microbiol. 2015;65:3162–9.
14. Lopez-Lopez A, Rogel-Hernandez MA, Barois I, et al. *Rhizobium grahamii* sp. nov., from nodules of *Dalea leporina*, *Leucaena leucocephala* and *Clitoria ternatea*, and *Rhizobium mesoamericanum* sp. nov., from nodules of *Phaseolus vulgaris*, siratro, cowpea and *Mimosa pudica*. Int J Syst Evol Microbiol. 2012;62:2264–71.

15. Dall'Agnol RF, Ribeiro RA, Ormeno-Orrillo E, et al. *Rhizobium freirei* sp. nov., a symbiont of *Phaseolus vulgaris* that is very effective at fixing nitrogen. Int J Syst Evol Microbiol. 2013;63:4167–73.
16. Mnasri B, Liu TY, Saidi S, et al. *Rhizobium azibense* sp. nov., a nitrogen fixing bacterium isolated from root nodules of *Phaseolus vulgaris*. Int J Syst Evol Microbiol. 2014;64:1501–6.
17. Román-Ponce B, Zhang JY, Vásquez-Murrieta MS, et al. *Rhizobium acidisoli* sp. nov., isolated from root nodules of *Phaseolus vulgaris* in acid soils. Int J Syst Evol Microbiol. 2016;66:398–406.
18. Mnasri B, Moncef M, Gisèle L, Mohamed A, Ridha M. Salt-tolerant rhizobia isolated from a Tunisian oasis that are highly effective for symbiotic N_2-fixation with *Phaseolus vulgaris* constitute a novel *biovar* (bv. mediterranense) of *Sinorhizobium meliloti*. Arch Microbiol. 2007;187(1):79–85.
19. Sadowsky MJ, Cregan PB, Keyser HH. Nodulation and nitrogen fixation efficiency of *Rhizobium fredii* with *Phaseolus vulgaris* genotypes. Appl Environ Microbiol. 1988;54:1907–10.
20. Mhamdi R, Laguerre G, Aouani ME, et al. Different species and symbiotic genotypes of field rhizobia can nodulate *Phaseolus vulgaris* in Tunisian soils. FEMS Microbiol Ecol. 2012;41:77–84.
21. Toledo I, Lloret L, Martinez-Romero E. *Sinorhizobium americanus* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. Syst Appl Microbiol. 2003;26:54–64.
22. Talbi C, Delgado MJ, Girard L, Ramírez-Trujillo A, Caballero-Mellado J, Bedmar EJ. *Burkholderia phymatum* strains capable of nodulating *Phaseolus vulgaris* are present in Moroccan soils. Appl Environ Microbiol. 2010;76:4587–91.
23. Amann RI, Lin CH, Key R, Montgomery L, Stahl DA. Diversity among fibrobacter strains: towards a phylogenetic classification. Syst Appl Microbiol. 1992;15:23–32.
24. Aserse AA, Räsänen LA, Assefa F, Hailemariam A, Lindström K. Phylogeny and genetic diversity of native rhizobia nodulating common bean (*Phaseolus vulgaris* L.) in Ethiopia. Syst Appl Microbiol. 2012;35(2):120–31.
25. Martens M, Dawyndt P, Coopman R, Gillis M, De Vos P, Willems A. Advantages of Multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer*. Int J Syst Evol Microbiol. 2008;58:200–14.
26. Germano MG, Menna P, Mostasso FL, Hungria M. RFLP analysis of the RNA operon of a Brazilian collection of bradyrhizobial strains from thirty three legume species. Int J Syst Evol Microbiol. 2006;56:217–29.
27. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, Nesme X, Rosselló-Mora, Swings R, et al. Report of the ad hoc committee for the reevaluation of the species definition in bacteriology. Int J Syst Evol Microbiol. 2002;52:1043–7.
28. Vinuesa P, Leon-Barrios M, Silva C, Willems A, Jarabo-Lorenzo A, Perez-Galdona R, Werner D, Martinez-Romero E. *Bradyrhizobium canariense* sp. nov. , an acid-tolerant endosymbiont that nodulates genistoid legumes (Papilionoideae: Genisteae) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. genistearum, *Bradyrhizobium* genospecies a and *Bradyrhizobium* genospecies b. Int J Syst Evol Microbiol. 2005;55:569–75.
29. Willems A, Munive A, de Lajudie P, Gillis M. In most *Bradyrhizobium* groups sequence comparison of 16S–23S rDNA internal transcribed spacer regions corroborates DNA–DNA hybridizations. Syst Appl Microbiol. 2003;26:203–10.
30. Rosselló-Mora R. DNA-DNA re-association methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E, editor. Molecular identification, systematics and population structure of prokaryotes. Berlin: Springer; 2006. p. 23–50.
31. Mousavi SA, Willems A, Nesme X, de Lajudie P, Lindström K. Revised phylogeny of *Rhizobiaceae*: Proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations. Syst Appl Microbiol. 2015;38(2):84–90.
32. Mousavi SA, Osterman J, Wahlberg N, Nesme X, Lavire C, Vial L, Paulin L, de Lajudiee P, Lindström K. Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. Syst Appl Microbiol. 2014;37(3):208–15.
33. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102:2567–72.
34. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014;64:346–51.
35. Varghese JN, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. Microbial species delineation using whole genome sequences. Nucleic Acids Res. 2015;43(14):6761–71.
36. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier PE. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol. 2014;64:384–91.
37. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, De Vos P, Vandamme P, Eisen JA, Garrity G, Hugenholtz P, Kyrpides NC. Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains. Stand Genomic Sci. 2015;10:26.
38. Biolog GN2 manufacturer's instructions http://www.ecologiemicrobiennelyon.fr/IMG/pdf/GN2-1.pdf).
39. DOE Joint Genome Institute. http://www.jgi.doe.gov/.
40. www.ebi.ac.uk/ena/data/view/Taxon:1138170 (European Nucleotide Archive).
41. CTAB DNA extraction protocol [http://jgi.doe.gov/collaborate-with-jgi/pmooverview/protocols-sample-preparation-information/].
42. Bennett S. Solexa Ltd. Pharmacogenomics. 2004;5(4):433–8.
43. Mingkun L, Copeland A, Han J. DUK, 2011. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4623924/pdf/40793_2015_Article_77.pdf. Accessed 26 Oct 2015.
44. Zerbino D, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
45. Gnerre S, MacCallum I. High–quality draft assemblies of mammalian genomes from massively parallel sequence data. PNAS. 2011;108(4):1513–8.
46. Hyatt D, Chen GL, Lacascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 2010;11:119.
47. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genomic Sci. 2015;10:86.
48. Lowe TM, Eddy SR. tRNAscan–SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
49. Pruesse E, Quast C, Knittel, Fuchs B, Ludwig W, Peplies J, Glckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:2188–7196.
50. INFERNAL. Inference of RNA alignments. http://infernal.janelia.org.
51. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42(1):D560–7.
52. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
53. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106:19126–31.
54. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2015;43(W1):W78–84. doi:10.1093/nar/gkv487.
55. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG-ER: a system for microbial genome annotation expert review and curation. Bioinformatics. 2009;25:2271–8.
56. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci U S A. 2006;103:3834–9.
57. Perret X, Staehelin C, Broughton WJ. Molecular basis of symbiotic promiscuity. Microbiol Mol Biol Rev. 2000;64:180–201.
58. Terpolilli JJ, Hood GA, Poole PS. What determines the efficiency of N_2-fixing *Rhizobium*-legume symbioses? Adv Microb Physiol. 2012;60:325–89.
59. Darling AE, et al. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;87(12):4576-9.
60. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinf. 2009;10:421. doi:10.1186/1471-2105-10-421.
61. Carver T, et al. BamView: viewing mapped read alignment data in the context of the reference sequence. Bioinformatics. 2010;26:676–7.
62. Rubio LM, Ludden PW. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annu Rev Microbiol. 2008;62:93–111.
63. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7.

64. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. PNAS USA. 1990; 87:4576–9.
65. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology, vol. Volume 2, Part B. Secondth ed. New York: Springer; 2005.
66. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer - Verlag; 2005.
67. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology, vol. Volume 2, Part C. Secondth ed. New York: Springer; 2005. p. 324.
68. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.
69. Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, R. rhizogenes, *R. rubi*, *R. undicola* and *R. vitis*. Int J Syst Evol Microbiol. 2001; 51:89–103.
70. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
71. Wang ET, Rogel MA, García-de los Santos A, Martínez-Romero J, Cevallos MA, Martínez-Romero E. *Rhizobium etli* bv. *mimosae*, a novel *biovar* isolated from *Mimosa affinis*. Int J Syst Bacteriol. 1999;49:1479–91.