Supporting Information

Tian et al. 10.1073/pnas.1120436109



Fig. S1. Genetic diversity and biogeography of rhizobia nodulating soybean. Red points indicate sampling sites. The species in red belongs to *Sinorhizobium*. Relative abundance of each species is also indicated. *In Xinjiang (1); 98.6% of *Bradyrhizobium* strains were isolated from soybean nodules in soils with low values of electronic conductivity (EC) (77–178 μ S/cm), and 69.4% of them were collected from a sampling site of near-neutral soil (pH 7.3; EC, 77 μ S/cm), whereas 93.0% of *S. fredii* strains were isolated from nodules in alkaline soils (pH 7.6–8.04), and 66.2% of them were found at a site with relatively high value of soil EC (1,090 μ S/cm). [†]In Huang-Huai-Hai (2, 3); *Bradyrhizobium* dominated (89.9%) soybean nodules in soils with pH 6.91–7.73, whereas *Sinorhizobium* was the major (85.5%) microsymbiont of soybean in soils with higher pH values (pH \geq 7.81).

Han LL, et al. (2009) Unique community structure and biogeography of soybean rhizobia in the saline-alkaline soils of Xinjiang, China. *Plant Soil* 324:291–305.
Li QQ, et al. (2011) Diversity and biogeography of rhizobia isolated from root nodules of *Glycine max* grown in Hebei Province, China. *Microb Ecol* 61:917–931.
Zhang YM, et al. (2011) Biodiversity and biogeography of rhizobia associated with soybean plants grown in the North China Plain. *Appl Environ Microbiol* 77:6331–6342.



Fig. S2. Neighbor-joining tree of *dnaK-rpoB*. (Scale bar: a 5% nucleotide sequence divergence.) A total of 1,000 bootstrap replicates were made, and bootstrap values are indicated at each node.



Fig. S3. Neighbor-joining tree of nodC. (Scale bar: a 5% nucleotide sequence divergence.) A total of 1,000 bootstrap replicates were made, and bootstrap values are indicated at each node. Branches in gray boxes are those of bootstrap values above 70%.



Fig. S4. Hierarchical clustering of rhizobia based on heat map of 887 functional genes. Approximately unbiased probability and bootstrap probability (au/bp) values of >70% were indicated at each node. α , the branch of α -proteobacteria; β , the branch of β -proteobacteria.



Fig. S5. Species tree of rhizobia. (*A*) Supertree based on 295 trees of core proteins of rhizobia. (*B*) ML tree based on 295 core genes shared by rhizobial genomes. (Scale bar: a 10% nucleotide sequence divergence.) A total of 100 bootstrap replicates were made for both trees, and bootstrap values are indicated at each node. To construct the supertree of 295 core proteins shared by 26 representative rhizobia, the following pipeline was applied: (*i*) conserved blocks from multiple alignment of test protein were selected by using Gblocks (1); (*ii*) PTP integrated in PAUP* 4.0b10 (2) was used to examine whether a tree from the original data was no better than random (results with a PTP score of P = 0.01 were accepted for further analysis); (*iii*) best-fit models of protein evolution were obtained with ProtTest3 (3); (*iv*) the neighbor-joining tree with 100 bootstrap replicates was produced for each protein by using PHYLIP (4); and (*v*) 295 protein trees were then analyzed with the most similar supertree analysis (MSSA) method integrated in Clann to produce the supertree, and 100 bootstrap replicates were used to assess the support for internal branches on the supertree (5).

- 1. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.
- 2. Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Sinauer, Sunderland, MA).
- 3. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. Bioinformatics 27:1164-1165.
- 4. Retief JD (2000) Phylogenetic analysis using PHYLIP. Methods Mol Biol 132:243-258.
- 5. Creevey CJ, McInerney JO (2009) Trees from trees: Construction of phylogenetic supertrees using clann. Methods Mol Biol 537:139-161.



Fig. S6. Hierarchical clustering of rhizobia based on heat map 66,150 pan genes. Approximately unbiased probability and bootstrap probability (au/bp) values of >70% were indicated at each node.

Other Supporting Information Files

Table S1 (DOC) Table S2 (DOC) Table S3 (DOC) Table S4 (DOC) Table S5 (DOC) Table S6 (DOC) Dataset S1 (XLS)