Table S1. **Geographic origin of 109 strains used in this study.**

CCBAU No.*	Geographic origin[†]	Reference
03334	Linxian	This study
83735, 83737, 83738, 83740, 83741, 83742, 83753	Aketao	Han *et al*., 2009
05557, 05561	Baoding	Li *et al*., 2011
23314, 23315, 23319	Bozhou	This study
05583, 05590, 05603, 05606, 05593	Cangxian	Li *et al*., 2011
051118, 051116, 051117, 051130, 051134, 051135, 051136, 051119	Cangxian	Zhang *et al*., 2011
05617	Hejian	Li *et al*., 2011
051142	Changli	Zhang *et al*., 2011
05701, 05702	Changli	Li *et al*., 2011
25379, 25380	Dongying	This study
83656, 83658, 83660, 83665, 83666	Farm150	Han *et al*., 2009
83654	Farm150	Han *et al*., 2009
83655	Farm150	Han *et al*., 2009
83622	Farm150	Han *et* al., 2009
83650	Farm150	Han *et al*., 2009
03328, 03329	Fenyang	This study
83635, 83639, 83640, 83641, 83644, 83646, 83643, 83647	Fukang	Han *et al*., 2009
051154, 051151, 051149	Handan	Zhang *et al*., 2011
45436	Hejian	Zhang *et al*., 2011
05618, 05631, 05641, 05636, 05638, 05642, 05644, 05646, 05647, 05640	Hejian	Li *et al*.,2011
25394, 25395, 25396, 25422, 25409, 25426	Heze	Zhang *et al*., 2011
23329, 23330	Huainan	This study
05657, 05684, 05686, 05699	Huanghua	Li *et al*.,2011
45408, 45421, 45433, 45406, 45405, 45439 , 45449	Jiaozuo	Zhang *et al*., 2011
25071	Jinan	This study
25429, 25441, 25471, 25438	Jinan	Zhang *et al*., 2011
05509	Jining	This study
25489, 25589, 25509	Jining	Zhang *et al*., 2011
03330, 03337	Linxian	This study
83619, 83621, 83631, 83680	Shihezi	Han *et al*., 2009
05564	Taian	This study
25564, 25619, 25583, 25566 , 25572	Taian	Zhang *et al*., 2011
03347, 03345	Taiyuan	This study
20016	Tianjin	Li *et al*.,2011
CCBAU107	Xinjiang	Han *et al*., 2009

*, The first two numbers represent the province where the strain was isolated, e.g. 03: Shanxi; 05: Hebei; 20: Tianjin; 23: Anhui; 25: Shandong; 45: Henan; 83: Xinjiang.

[†], Names of cities or counties.

Table S2. **Primers used in this study and GenBank accession numbers for obtained sequences.**

Genes	Primer sequence (5’-3’, forward/reverse)	Tm (°C)	GenBank Accession No. †
bacA	CAYTACRTSTTCCGYTGG / CGTYGKCCASGARTTGAC	62	KF381560 - KF381619
purL	TGACSGCCTAYGARATGATGC / GGRTTGCCGAAGTTCAGRTT	51	KF382340 - KF382399
nodM	GCCTKGARTAYCGGGGMTAYGATT / AYTGCGRCGWDAGCGGRATYT	57	KF382220 - KF382279
gcvT	CGACGACCTGATGATYRCCAA / GTYTGTTTGACGAAGGGMAGG	51	KF381860 - KF381919
cysD	CTSGAAAACGAAGCGATCCAC / CTYGGCRAARTARAGCGGYAC	55	KF381680 - KF381739
cobO	GATGGCGAAGAAGAAGAC / ASGGRTGYTTGACSAGYT	54	KF381620 - KF381679
nodE	GAAAGCGACSAGGAYCAGAT / GTGCCGTGMGCGTTGATRT	55	KF382160 - KF382219
exoA	ATCCRMARCGHATCCARAGCG / CCCAGYTGATGATGGCRAGAA	55	KF381740 - KF381799
exoY	CGGAYGGCGGMAGCRYKTTYT / CRATGACRAGRTCGGYRAKGA	59	KF381800 - KF381859
phnC	CGCCAGTTCGGCAAGA / GGCGGTGAGDTCCTT	55	KF816087 - KF816146
nodC	TGATYGAYATGGARTAYTGG / GGYGACARCCARTCGCTRTTG	55	KF381980 - KF382039
nodS	AACTATCATTTATTGCACCGTGAG / TGAAAACGGGAAAGCAAGCAG	52	KF382280 - KF382339
y4wE	TGCCGRYMSCGMATWGCTCTTC / CYCGATRCWRACGMGAAT	64	KF382640 - KF382699
nodD1	ACGCACTGATGACCGAACGCAAAC / ATCGCCAGTCGTAACGGGAGGG	60	KF382040 - KF382099
nodD2	GATATGTTTAACCCAGTTCAGTCG / TGTCATGTGCAACGCTTCCT	55	KF382100 - KF382159
rhcJ	ACACAACGAGCAGCGACACG / CCAAGAACGATGAACAAGAGGTA	59	KF382400 - KF382459
SMc00019	CADTTCCTBATHGCCATGCC / GCVGGRCANKTSAGCCADCCRTT*	66 to 50*	KF382460 - KF382519
truA	CAGGTSGCDCATSTCGAYCT / GADCGBAYCTGGTTRTGMAG*	58 to 48*	KF382580 - KF382639
thrA	GCNGGBGGYATYCCSGTBATCAAG / CGYTCGATNCGRATSACYTGSGG*	66 to 56*	KF382520 - KF382579, KF815983, KF815986, KF815990, KF815992, KF815997, KF816011, KF816014, KF816015, KF816024, KF816026, KF816027, KF816029, KF816032-KF816034, KF816038, KF816040-KF816042, KF816044, KF816046-KF816050, KF816052, KF816053, KF816056, KF816057, KF816059, KF816061, KF816063, KF816065, KF816067-KF816071, KF816073-KF816076, KF816078, KF816080-KF816084, KF816086

*, Primers and PCR conditions for *SMc00019*, *truA* and *thrA* were described in Zhang *et al.* (2012).

†, 109 sequences were obtained for *thrA* and 60 sequences were obtained for each of the other 18 genes.

Table S3. Summary of 60 representative strains used in this study .

CCBAU No.*	Genotype of Core genes	Genotype of symbiosis-related genes			Geographic origin†	Reference
		Chromosome	pSymB	pSymA		
S. fredii (38 strains in total)						
83635	4	1	2	2	**Aketao**	Han *et al*., 2009
83640, 83643	4	1	2	2	**Fukang**	Han *et al*., 2009
83735, 83753	4	1	7	2	**Aketao**	Han *et al*., 2009
03334	4	1	2	11	Linxian	This study
83655	4	1	1	12	**Farm150**	Han *et al*., 2009
83666	4	1	1	12	**Farm150**	Han *et al*., 2009
20016	4	2	2	2	Tianjin	Li *et al*.,2011
23330	5	3	6	2	Huainan	This study
03347	5	3	6	2	Taiyuan	This study
051118	5	3	6	2	Cangxian	Zhang *et al*., 2011
05557, 05561	5	3	6	2	Baoding	Li *et al*., 2011
25379	5	3	6	2	Dongying	This study
25409	5	3	6	2	Heze	Zhang *et al*., 2011
25566	5	3	6	2	Taian	Zhang *et al*., 2011
45436	5	3	6	2	Hejian	Zhang *et al*., 2011
03328	3	4	1	10	Fenyang	This study
03329	3	4	1	9	Fenyang	This study
03330	3	4	1	15	Linxian	This study
051154	2	1	6	2	Handan	Zhang *et al*., 2011
45433	1	1	1	1	Jiaozuo	Zhang *et al*., 2011
051149	1	1	1	2	Handan	Zhang *et al*., 2011
05564	1	1	1	2	Taian	This study
05657	1	1	4	5	Huanghua	Li *et al*., 2011
23315	1	1	3	2	Bozhou	This study
25380	1	1	6	2	Dongying	This study
25422	1	1	4	2	Heze	Zhang *et al*., 2011
25438	1	1	3	2	Jinan	Zhang *et al*., 2011
25509	1	1	3	3	Jining	Zhang *et al*., 2011
03337	1	1	1	2	Linxian	This study
83619	1	5	1	2	**Shihezi**	Han *et al*., 2009
05583	1	5	1	2	Cangxian	Li *et al*., 2011
83622	1	5	1	2	**Farm150**	Han *et al*., 2009
45406	6	5	1	4	Jiaozuo	Zhang *et al*., 2011
05593	7	5	1	2	Cangxian	Li *et al*., 2011
051142	1	6	5	2	Changli	Zhang *et al*., 2011
Sinorhizobium sp. II (1 strain)						
05617	8	9	9	1	Hejian	Li *et al*., 2011
Sinorhizobium sp. I (3 strains)						
05702	9	7	8	8	Changli	Li *et al*., 2011
05701	10	7	8	8	Changli	Li *et al*., 2011
25426	10	7	8	8	Heze	Zhang *et al*., 2011
Sinorhizobium sp. III (9 strains)						
051119	11	8	10	13	Cangxian	Zhang *et al*., 2011
051151	11	8	10	6	Handan	Zhang *et al*., 2011
051135	11	8	10	14	Cangxian	Zhang *et al*., 2011
051116	12	8	10	7	Cangxian	Zhang *et al*., 2011
05590					Cangxian	Li *et al*., 2011
051130	12	8	10	2	Cangxian	Zhang *et al*., 2011
05618, 05631, 05641	12	8	10	13	Hejian	Li *et al*., 2011
Sinorhizobium sojae (9 strains)						
05603, 05606	13	10	11	2	Cangxian	Li *et al*., 2011
05636, 05638, 05642, 05647	13	10	11	2	Hejian	Li *et al*., 2011
05684, 05686, 05699	13	10	11	2	Huanghua	Li *et al*., 2011

Note: Core genes: *SMc00019*, *thrA*, *truA*; Chromosome: *bacA, purL, nodM, gcvT, cysD, cobO, nodE*; pSymB: *exoA, exoY, phnC*; pSymA: *nodC, nodS, y4wE, nodD1, nodD2, rhcJ*. *. The first two numbers represent the province where the strain was isolated, e.g. 03: Shanxi; 05: Hebei; 20: Tianjin; 23: Anhui; 25: Shandong; 45: Henan; 83: Xinjiang. †, Names of cities or counties, sampling sites in bold are in Xinjiang and the other sites are in Huang-Huai-Hai region.

Table S4. Analysis of nucleotide polymorphism for the five species based upon the combined sequences of core genes or symbiosis-related genes.

Species	Segregating sites	h (Number of haplotypes)	Hd (Haplotype diversity)	Pi (nucleotide diversity)
SMc00019-truA-thrA				
S. fredii	74	12	0.829	0.01286
Sinorhizobium sp. I	32	3	1.000	0.01669
Sinorhizobium sp. III	32	4	0.694	0.01213
S. sojae	2	3	0.417	0.00035
bacA-purL-nodM-gcvT-cysD-cobO-nodE				
S. fredii	112	16	0.893	0.00846
Sinorhizobium sp. I	5	2	0.667	0.00114
Sinorhizobium sp. III	38	4	0.750	0.00708
S. sojae	7	2	0.222	0.00053
exoA-exoY-phnC				
S. fredii	73	12	0.811	0.02164
Sinorhizobium sp. I	0	1	0.000	0.00000
Sinorhizobium sp. III	7	4	0.583	0.00143
S. sojae	0	1	0.000	0.00000
nodC-nodS-y4wE-nodD1-nodD2-rhcJ				
S. fredii	23	10	0.461	0.00112
Sinorhizobium sp. I	0	1	0.000	0.00000
Sinorhizobium sp. III	11	5	0.806	0.00183
S. sojae	0	1	0.000	0.00000

Note: Strain number is 38 for *S. fredii,* 3 for *Sinorhizobium* sp. I, 1 for *Sinorhizobium* sp. II, 9 for *Sinorhizobium* sp. III and 9 for *S. sojae,* respectively.

Table S5. Breakpoints detected by Recombination Detection Program (RDP).

Genes (bp)	breakpoints (bp)
SMc00019-thrA-truA (1278 bp)	11-969 318-578 570-975 607-1238 689-900 738-975 751-975 751-977 973-1274 979-1249 979-1263
bacA-purL-nodM-gcvT-cysD-cobO-nodE (2925 bp)	3-413 318-578 461-2487 464-2545 806-1188 806-1209
exoA-exoY-phnC (1088 bp)	629-1088 697-1015 697-1023
nodC-nodS-y4wE-nodD1-nodD2-rhcJ (2879 bp)	-

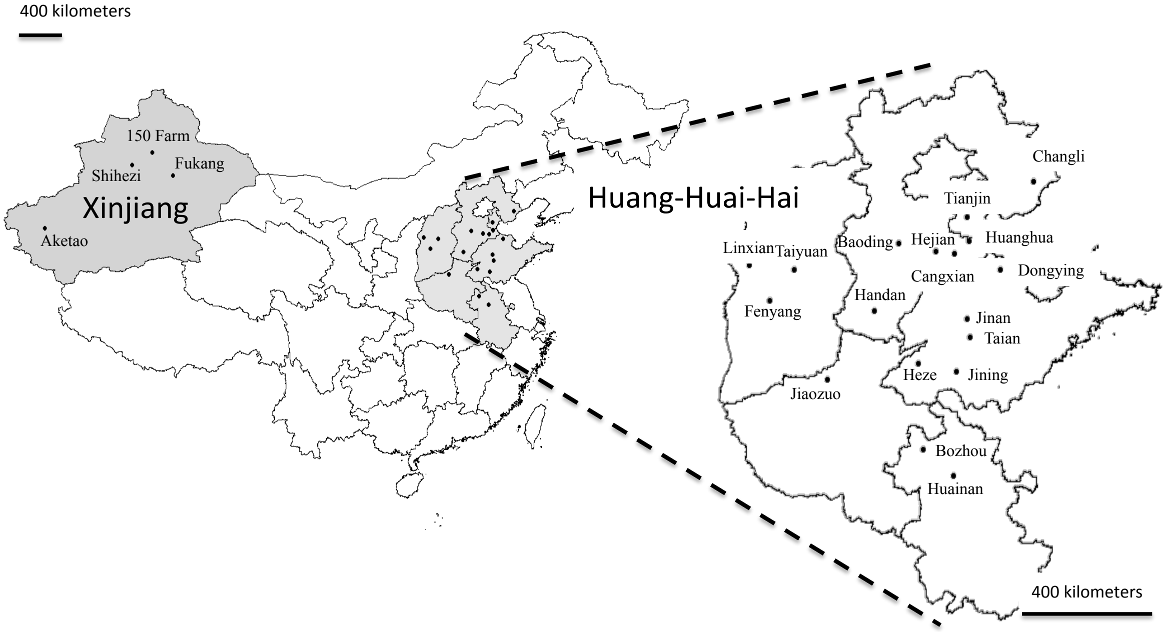
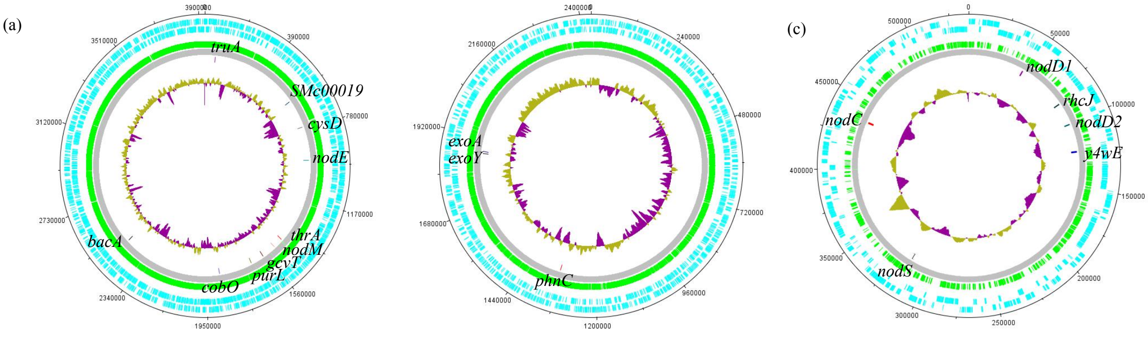


Fig. S1 Sampling sites of soybean rhizobia used in this study (black dots). Grey areas indicate two eco-regions (Xinjiang and Huang-Huai-Hai) where sampling sites were distributed. The two maps were created by using DIVA-GIS software (www.diva-gis.org) and the sampling information of soybean rhizobia firstly reported in this study and those described in earlier publications (Han et al. 2009, Li et al. 2011 and Zhang et al. 2011).



Gene	Complete name or Function	Gene	Complete name or Function
truA	RNA pseudouridine synthase A	*exoA*	glucosyltransferase protein
thrA	homoserine dehydrogenase	*exoY*	galactosyltransferase protein
SMc00019	conserved hypothetical protein	*phnC*	phosphonates import ATP-binding protein
bacA	putative transport protein, essential for bacteroid development protein	*y4wE*	histidinol-phosphate aminotransferase
purL	probable phosphoribosylformyl glycinamidine synthetase II protein	*nodC*	N-acetylglucosaminyltransferase
nodM	putative glucosamine synthase	*nodS*	N-methyl transferase
gcvT	glycine cleavage system T protein	*nodD1*	transcription regulator
cysD	sulfate adenylyltransferase	*nodD2*	nod box-dependent transcription activator
cobO	cob(I)yrinic acid a,c-diamide adenosyltransferase	*rhcJ*	Type III secretion component
nodE	beta ketoacyl ACP synthase		

Fig. S2. Location of the genetic markers used in this study on the genome of *Sinorhizobium* sp. NGR234. (a): three housekeeping genes and seven symbiosis-related genes on chromosome; (b): three symbiosis-related genes on pSymB; (c): six symbiosis-related genes on pSymA.

References

Han LL, Wang ET, Han TX, Liu J, Sui XH, Chen WF, Chen WX. 2009. Unique community structure and biogeography of soybean rhizobia in the saline-alkaline soils of Xinjiang, China. Plant Soil 324:291–305.

Li QQ, Wang ET, Zhang YZ, Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX. 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, Li Jr. Y, Chen WF, Wang ET, Tian CF, Li QQ, Sui XH, Chen WX. 2011. Biodiversity and biogeography of rhizobia associated with soybean plants grown in the North China Plain. Appl. Environ. Microbiol. 77:6331–6342.

Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX. 2012. Robust Markers Reflecting Phylogeny and Taxonomy of Rhizobia. PLoS One. 7, e44936.