A portal for rhizobial genomes: RhizoGATE integrates a Sinorhizobium meliloti genome annotation update with postgenome data

Anke Becker, Melanie Barnett, Delphine Capela, Michael Dondrup, Paul-Bertram Kamp, Elizaveta Krol, Burkhard Linke, Silvia Rüberg, Kai Runte, Brenda Schroeder, et al.

Published in final edited form as:
J Biotechnol. 2009 March 10; 140(1-2): 45–50. doi:10.1016/j.jbiotec.2008.11.006.

A portal for rhizobial genomes: *Rhizo*GATE integrates a *S. meliloti* genome annotation update with postgenome data

Anke Becker[a,*], **Melanie J. Barnett**[b], **Delphine Capela**[c], **Michael Dondrup**[d], **Paul-Bertram Kamp**[e], **Elizaveta Krol**[a], **Burkhard Linke**[d], **Silvia Rüberg**[e], **Kai Runte**[d], **Brenda K. Schroeder**[f], **Stefan Weidner**[e], **Svetlana N. Yurgel**[f], **Jacques Batut**[c], **Sharon R. Long**[b], **Alfred Pühler**[e], and **Alexander Goesmann**[d]

a*Genetics and Systems Biology, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany**

b*Department of Biology, Stanford University, Stanford, CA 94305-5020, U.S.A.*

c*Laboratoire des Interactions Plantes Micro-organismes, INRA-CNRS 441-2594, Chemin de Borde Rouge, BP 52627, 31326 Castanet Tolosan cedex, France*

d*Bioinformatics Resource Facility, Center for Biotechnology (CeBiTec), Bielefeld University, 33594 Bielefeld, Germany*

e*Department of Genetics, Faculty of Biology, Bielefeld University, 33594 Bielefeld, Germany*

f*Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, U.S.A.*

Abstract

Sinorhizobium meliloti is a symbiotic soil bacterium of the alphaproteobacterial subdivision. Like other rhizobia, *S. meliloti* induces nitrogen-fixing root nodules on leguminous plants. This is an ecologically and economically important interaction, because plants engaged in symbiosis with rhizobia can grow without exogenous nitrogen fertilizers. The *S. meliloti-Medicago truncatula* (barrel medic) association is an important symbiosis model. The *S. meliloti* genome was published in 2001, and the *Medicago truncatula* genome currently is being sequenced. Many new resources and data have been made available since the original *S. meliloti* genome annotation and an update was needed. In June 2008, we submitted our annotation update to the EMBL and NCBI databases. Here we describe this new annotation and a new web-based portal *Rhizo*GATE. About 1000 annotation updates were made; these included assigning functions to 313 putative proteins, assigning EC numbers to 431 proteins, and identifying 86 new putative genes. *Rhizo*GATE incorporates the new annotion with the *S. meliloti* GenDB project, a platform that allows annotation updates in real time. Locations of transposon insertions, plasmid integrations, and array probe sequences are available in the GenDB project. *Rhizo*GATE employs the EMMA platform for management and analysis of transcriptome data and the IGetDB data warehouse to integrate a variety of heterogeneous external data sources.

*Corresponding author: Anke Becker, Phone: +49 761 203 6948, Fax: +49 761 203 2745, E-mail address: anke.becker@biologie.uni-freiburg.de.

Keywords

Rhizobiales; α-proteobacteria; symbiotic nitrogen fixation; *Medicago*; symbiosis

Introduction

Rhizobia are a diverse group of bacteria that engage in symbioses with the roots of leguminous host plants. Once established as symbionts, these bacteria reduce or "fix" dinitrogen to a form the plant can use for growth (reviewed in Prell and Poole, 2006; Jones et al., 2007). With their potential to fix up to 300 kg of molecular nitrogen/ha cultivated land, symbiotic bacteria contribute significantly to an ecologically-sound and cost-efficient nitrogen supply. Most rhizobia (order Rhizobiales) belong to families within the α-proteobacteria subdivision (Garrity et al., 2002), although symbiotic nitrogen-fixing β-proteobacteria (β-rhizobia) have also been reported (Chen et al., 2003; Rasolomampianina et al., 2005). The Rhizobiales also include the plant pathogen *Agrobacterium* and the mammalian pathogens *Bartonella* and *Brucella*. Almost 10 years ago, the first rhizobial genomes were sequenced: *Mesorhizobium loti* MAFF303099 (Phylobacteriaceae) (Kaneko et al., 2000) and *Sinorhizobium meliloti* 1021 (Rhizobiaceae) (Barnett et al., 2001; Capela et al., 2001; Finan et al. 2001; Galibert et al., 2001). Since then, more than 20 complete genome sequences from the Rhizobiales have been released. These sequences are a valuable resource that extends our knowledge of the genetic basis of the rhizobial lifestyle (MacLean et al., 2007).

The annotated *S. meliloti* 1021 genome sequence was submitted to the EMBL/GenBank database in 2001 (chromosome, AL591688; pSymA, NC_003037; pSymB, AL591985). At that time, nearly half of the predicted genes lacked a functional assignment. Hence, we wanted to exploit the wealth of newly available data to update the *S. meliloti* 1021 genome annotation. In addition, in the *S. meliloti* postgenome era, a broad range of strategies was developed to use genomic data to obtain functional information. These comprise transcriptome and proteome studies, mutagenesis, and gene fusion techniques. The number of studies applying high throughput methods is continually increasing. We integrated data from these diverse resources, and a web-based genome browser, into a portal that facilitates data management and access to the above mentioned information. Here we describe the annotation update and introduce the *Rhizo*GATE portal. These data resources are publicly available from one access site.

The *S. meliloti* 1021 annotation update

Rhizobia often possess large, multipartite genomes. The *S. meliloti* genome contains a circular chromosome (3.65 Mb) and two megaplasmids, pSymA (1.35 Mb) and pSymB (1.68 Mb) (Galibert et al., 2001). The original annotation defined 6204 protein-encoding genes: 3341 on the chromosome (Capela et al., 2001), 1293 on pSymA (Barnett et al., 2001), and 1570 on pSymB (Finan et al., 2001). There was no experimentally proven function for a vast majority of the predicted genes and 40 % could not be placed in a functional category. Moreover, 8 % were orphan genes, defined as those not found in any other sequenced genome.

There are numerous changes in this annotation update (summarized in Fig. 1). The update incorporates pertinent information published from 2001 to 2008. Improved prediction tools and the availability of additional rhizobial genomes helped to identify 86 new putative genes, and adjust the start positions of 360 coding regions. There were 66 previously predicted orphan genes removed from the annotation. These were removed either because a new, more plausible candidate, was encoded on the opposite strand, or because of short length, no database match and poor *S. meliloti* codon usage patterns. It was previously suggested that short genes may be overpredicted in GC-rich genomes because stop triplets are less frequent (Skovgaard et al.,

2001;Fukuchi and Nichikawa, 2004). This updated annotation is more conservative with respect to short ORFs. About 1000 updates added new information to the annotation of gene products. Putative functions were assigned to 313 putative proteins formerly classified as hypothetical or conserved hypothetical. As a result, more than 71 % of genes now have a predicted function. The Enzyme classification (EC) numbers were changed for 160 genes and were newly assigned to 431 genes. Altogether, there have been changes in annotation in about a quarter of the genes. This number does not include minor changes such as spelling or grammar. We assigned COG (Clusters of Orthologous Groups) (Tatusov et al., 2000) and GO (Gene Ontology) (Ashburner et al., 2000) (http://www.geneontology.org/) classification terms to facilitate queries and genome comparisons by functional class and standard descriptors. Table 1 summarizes the update of the general features of the *S. meliloti* strain 1021 genome. The annotation update was released in the EMBL/GenBank/DDBJ database.

The *Rhizo*GATE portal

The *Rhizo*GATE portal (www.rhizogate.de) was constructed to integrate the *S. meliloti* genome sequence, postgenome data sets, and bioinformatics tools (Fig. 2). Organization of tools and data sets in a single portal allows easy access and exploitation of the wealth of information available for the *S. meliloti* genome. The IGetDB data repository is an integral part of this portal and allows complex queries and data mining.

The S. meliloti GenDB genome browser and annotation platform

The core of the *Rhizo*GATE portal is the *S. meliloti* GenDB project. GenDB is a genome annotation system for prokaryotic genomes (Meyer et al., 2003) (http://www.cebitec.unibielefeld.de/groups/brf/software/gendb_info/) that offers automatic identification, classification and annotation of genes using a variety of software tools. Its user interface allows expert annotation by geographically dispersed teams. Our consortium of research groups used this platform to update the 2001 *S. meliloti* genome annotation. The most important feature of the *S. meliloti* GenDB project is that it allows for continual updates, while retaining a complete annotation history. Moreover, a recent Metanor (Goesmann et al., 2005) automatic annotation is offered as an additional information source. Metanor uses a combination of standard tools like BLAST, HMMer and InterPro to assign a gene name, gene product, description, functional category, GO numbers and other attributes to each CDS. It is feasible to keep pace with the increasing amount of information and make this knowledge available to the community in real-time. In addition to the annotation changes described above, the *S. meliloti* GenDB project was enriched with information on novel genes, available mutant and gene fusion libraries, experimentally proven operons, and microarray probes (Table 2) (Fig. 3).

Recently, Mao et al. (2008) identified 20 new transcripts in a pilot study applying ultrafast transcript sequencing to *S. meliloti*. Although this study generated just 1854 non-ribosomal RNA sequences, it demonstrated that the most recent sequencing technologies will greatly facilitate gene discovery and improve genome annotation. Four of these new transcripts were independently predicted in our annotation update submitted to EMBL/GenBank/DDBJ. Since the ultrafast sequencing approach was published after submission of the update to the databases the additional 16 new transcripts could not be considered. However, these genes as well as non-coding RNA genes predicted by the Rfam database (http://www.sanger.ac.uk/Software/Rfam/) were included into the GenDB project.

Microarrays based on PCR fragments and oligonucleotide probes have been published for the *S. meliloti* genome (Rüberg et al., 2003; Krol and Becker, 2004). Positions and sequences of the probes represented on these microarrays are also denoted.

Pobigaylo et al. (2006) made a signature-tagged (STM) mini-Tn*5* mutant library in *S. meliloti* 2011, a wild type strain closely related to the reference strain *S. meliloti* 1021. Both are spontaneous streptomycin resistant derivatives of *S. meliloti* SU47 (Meade and Signer, 1977; Casse et al., 1979). A set of 412 different mini-Tn*5* transposons, each carrying two individual short sequence tags, was used for this mutagenesis approach. Each individual mutant of a set can be distinguished from every other mutant based on the tags carried by its transposon. These signature-tagged mutants can be used in large-scale genetic screens. The transposon carries a promoterless *gusA* gene, therefore mutants can be used in expression studies. Insertion sites for 5089 mutants were mapped and are marked in the GenDB project; 44 % of predicted protein-coding genes contain at least one of these mapped transposons.

In addition, plasmid integration was employed to generate defined *S. meliloti* mutants. In this approach, internal fragments, 200 to 350 bp in length, were inserted into the mobilizable suicide vector pK19mob2ΩHMB (Luo et al., 2005). These plasmids can be transferred by conjugation from the broad host range mobilizing strain *E. coli* S17-1 (Simon et al., 1983) to *S. meliloti* strains. Plasmids targeting 3207 *S. meliloti* genes are available in *E. coli* S17-1. So far, 727 plasmids from this library have been integrated into the *S. meliloti* 1021 genome. The positions of these plasmids and mutants are denoted in the GenDB project. A study of plasmid integration mutants for 90 genes encoding LysR-family transcriptional regulators has identified new regulatory genes involved in symbiosis (Luo et al., 2005). Information about this comprehensive collection of defined mutants is easily accessible via the *Rhizo*GATE portal.

A high-throughput approach was used to determine co-transcribed genes in *S. meliloti* (Krol et al., unpublished). This approach combined polar plasmid integrations in the first gene of putative operons with transcriptional reporter gene fusions in downstream genes. These reporter gene fusions were derived from the mini-Tn*5* mutant library (Pobigaylo et al., 2006). Significant reduction of reporter gene expression caused by plasmid interruption of transcription indicates an operon structure. This study confirmed co-transcription from 75 operons comprising 225 genes. These data, along with published co-transcription data were included into the *S. melioti* GenDB project.

Recently, Cowie et al. (2006) generated a library of 6,298 plasmids carrying randomly generated *S. meliloti* DNA fragments fused to flanking pairs of reporter genes, *gfp+*, *lacZ* and *gusA*, *rfp*. These plasmids were integrated into the *S. meliloti* genome. Positional information and orientation of the fragments is given in the GenDB project.

In addition to the *S. meliloti* 1021 genome, published annotations were included for ten more Rhizobiales genomes (Table 3). These genomes are presented in one GenDB project, which facilitates searches on all genomes.

The EMMA platform for management and analysis of transcriptome data

We implemented the open source platform EMMA (Dondrup et al., 2003, 2008) (http://www.cebitec.uni-bielefeld.de/groups/brf/software/emma_info/) within *Rhizo*GATE to store and analyze transcriptome data from *S. meliloti*. EMMA provides access to gene expression data sets and processing tools including normalizations to reduce noise and systematic errors, statistical tests for the identification of significantly regulated genes, and clustering algorithms. It allows import of data from spotted microarrays, Affymetrix GeneChips, and NimbleGene DNA chips. This platform connects to the genomic sequences in GenDB so that expression data is linked to the most recent annotation. The *S. meliloti* EMMA project currently holds 277 hybridizations arranged in 22 experiments. These transcriptome data sets can be queried by using the IGetDB data warehouse (Table 4) or the EMMA web interface accessible via the *Rhizo*GATE portal.

The IGetDB data warehouse

IGetDB (http://www.cebitec.uni-bielefeld.de/groups/brf/software/igetdb_info/) is an online platform for data mining that stores and integrates data from various sources. Currently, it holds annotation, mutant (Pobigaylo et al., 2006, 2008), gene fusion library (Cowie et al., 2006), and ORFeome data (Schroeder et al, 2005), as well as transcriptome and proteome data (Table 4). Annotation data and positions of mutants are derived automatically from the GenDB project. Published mutant phenotypes from high-throughput studies (Pobigaylo et al., 2006, 2008) and from the NodMutDB mutant data collection (Mao et al., 2005) as well as symbiotic phenotypes of about 250 mTn*5* or plasmid insertion mutants (Capela et al., unpublished data) are also accessible via the IGetDB interface. Furthermore, this interface links to expression data of the gene fusion library that was assayed in complex medium and minimal medium with either glucose or succinate as sole carbon source (Cowie et al., 2006). Also, *Rhizo*GATE provides a link to information about the Affymetrix SymbiosisChip, including gene and intergenic region sequences that were used for GeneChip probe set design. Published expression data applying this GeneChip were included in the data warehouse.

Implications and summary

For decades, *S. meliloti* has been the focus of numerous research groups from around the world. This research has led to many important discoveries, some of which are applicable to prokaryotes in general, making *S. meliloti* a model bacterium in its own right. For example, a long sought after step in cobalamin biosynthesis was first elucidated in *S. meliloti* (Campbell et al., 2006). Recently, substrates were experimentally determined for many members of the *S. meliloti* ABC transporter (Mauchline et al., 2006) and short-chain dehydrogenase/reductase (Jacob et al., 2008) families. Together with its model plant host, *Medicago truncatula*, *S. meliloti* is one of the most important symbiosis model systems.

To facilitate research on this important model organism, we established the *Rhizo*GATE portal as a single access site for the growing volume of *S. meliloti* genome and postgenome data and included additional Rhizobiales genomes. We intend to add further value to this platform by integrating additional publicly available transcriptome, proteome, metabolome, and mutant datasets into the IGetDB data repository, and implementing more analysis tools.

About one quarter of soil fixed nitrogen results from rhizobial symbioses (Madigan et al. 1997; Zahran, 1999; Werner and Newton, 2005). This amount is insufficient to sustain food production, so use of chemical nitrogen fertilizers is necessary. Unfortunately, manufacture of these fertilizers requires vast amounts of energy, mostly in the form of non-renewable fossil fuels. Therefore, gains in understanding rhizobial symbiosis have the potential to help improve nutrition, climate, and world economies. We hope that this portal will be one small step in aiding such research.

Acknowledgements

The work was funded by grants 031U213D and 0313805A from Bundesministerium für Bildung und Forschung, Germany in the framework of the GenoMik and GenoMik Plus Competence Networks. M.J.B was supported by National Institutes of Health Grant GM30962 (to S.R.L.), B.K.S. and S.N.Y. were supported by grants MCB-0131376 and MCB-0131376 from the U.S. National Science Foundation, awarded to Michael L. Kahn. We are very grateful to many colleagues for contributing to the annotation update by sharing their knowledge. The authors further wish to thank the BRF system administrator team for expert technical support.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE,

Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet 2000;25:25–29.

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Sue Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh K-C, Davis RW, Federspiel NA, Long SR. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc. Natl. Acad. Sci. USA 2001;98:9883–9888. [PubMed: 11481432]

Barnett MJ, Toman CJ, Fisher RF, Long SR. A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc. Natl. Acad. Sci. U.S.A 2004;47:16636–16641. [PubMed: 15542588]

Becker, A. Functional Genomics of Rhizobia. In: Pawlowski, K., editor. Prokaryotic Symbionts in Plants (Microbiology Monographs). Germany: Springer, Heidelberg; 2008. in press

Becker A, Bergès H, Krol E, Bruand C, Rüberg S, Capela D, Lauber E, Meilhoc E, Ampe A, de Bruijn FJ, Fourment J, Francez-Charlot A, Kahn D, Küster H, Liebe C, Pühler A, Weidner S, Batut J. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol. Plant-Microbe Interact 2004;17:292–203. [PubMed: 15000396]

Bobik C, Meilhoc E, Batut J. FixJ : a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. J. Bacteriol 2006;188:4890–4902. [PubMed: 16788198]

Campbell GR, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC. *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. Proc. Natl. Acad. Sci. U.S.A 2006;103:4634–4639. [PubMed: 16537439]

Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, Boistard P, Becker A, Boutry M, Cadieu E, Dréano S, Gloux S, Godire T, Goffeau A, Kahn D, Lelaure V, Masuy D, Pohl T, Portelle D, Pühler A, Purnelle B, Ramsperger U, Thébault P, Vandenbol M, Weidner S, Galibert F. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti*. Proc. Nat. Acad. Sci. USA 2001;98:9877–9882. [PubMed: 11481430]

Capela D, Carrere S, Batut J. Transcriptome-based identification of the *Sinorhizobium meliloti* NodD1 regulon. Appl. Environ. Microbiol 2005;71:4010–4013.

Capela D, Filipe C, Bobik C, Batut J, Bruand C. *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. Mol. Plant-Microbe Interact 2006;19:363–372. [PubMed: 16610739]

Casse FC, Boucher C, Julio J-S, Michel M, Dénarié J. Identification and characterization of large plasmids in *Rhizobium meliloti* using agarose gel electrophoresis. J. Gen. Microbiol 1979;113:229–242.

Chao T-C, Becker A, Buhrmester J, Pühler A, Weidner S. The *Sinorhizobium meliloti fur* gene regulates in dependence of Mn(II) the transcription of the *sitABCD* operon encoding a metal type transporter. J. Bacteriol 2004;186:3609–3620. [PubMed: 15150249]

Chao T-C, Buhrmester J, Hansmeier N, Pühler A, Weidner S. Role of the regulatory gene *rirA* in the transcriptional response of *Sinorhizobium meliloti* to iron limitation. Appl. Environ. Microbiol 2005;71:5969–5982. [PubMed: 16204511]

Chen WM, James EK, Prescott AR, Kierans M, Sprent JI. Nodulation of Mimosa spp. by the beta-proteobacterium *Ralstonia taiwanensis*. Mol. Plant-Microbe Interact 2003;16:1051–1061. [PubMed: 14651338]

Cowie A, Cheng J, Sibley CD, Fong Y, Zaheer R, Patten CL, Morton RM, Golding GB, Finan TM. An integrated approach to functional genomics: construction of a novel reporter gene fusion library of *Sinorhizobium meliloti*. Appl. Environ. Microbiol 2006;72:7156–7167. [PubMed: 16963549]

Djordjevic MA, Chen HC, Natera S, Van Noorden G, Menzel C, Taylor S, Renard C, Geiger O, Weiller GF. the *Sinorhizobium meliloti* sequencing consortium. A global analysis of protein expression profiles in *Sinorhizobium meliloti*:discovery of new genes for nodule occupancy and stress adaptation. Mol. Plant-Microbe Interact 2003;16:508–524. [PubMed: 12795377]

Dominguez-Ferreras A, Perez-Arnedo R, Becker A, Olivares J, Soto MJ, Sanjuan J. Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. J. Bacteriol 2006;188:7617–7625. [PubMed: 16916894]

Dondrup M, Goesmann A, Bartels D, Kalinowski J, Krause L, Linke B, Rupp O, Sczyrba A, Pühler A, Meyer F. EMMA: a platform for consistent storage and efficient analysis of microarray data. J. Biotechnol 2003;106:135–146. [PubMed: 14651856]

Dondrup M, Albaum SP, Griebel JT, Henckel K, Jünemann S, Kleindt CK, Küster H, Linke B, Mertens D, Mittard-Runte V, Neuweger H, Runte KJ, Tauch A, Tille F, Pühler A, Goesmann A. EMMA 2 - A MAGE-compliant system for the collaborative analysis and integration of microarray data. BMC Bioinformatics, submitted. 2008

Finan TM, Weidner S, Chain P, Becker A, Wong K, Cowie A, Buhrmester J, Vorhölter F-J, Golding B, Pühler A. The complete sequence of the *Sinorhizobium meliloti* pSymB megaplasmid. Proc. Nat. Acad. Sci. USA 2001;98:9889–9894. [PubMed: 11481431]

Fukuchi S, Nishikawa K. Estimation of the number of authentic orphan genes in bacterial genomes. DNA Research 2004;11:219–231. [PubMed: 15500248]

Galibert F, Finan TM, Long SR, Pühler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dréano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating D, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl T, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thébault P, Vandenbol M, Vorhölter F-J, Weidner S, Wells DH, Wong K, Yeh K-C, Batut J. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 2001;293:668–672. [PubMed: 11474104]

Garrity, GM.; Johnson, KL.; Bell, JA.; Searles, DB. Bergey's Manual of Systematic Bacteriology Release 3.0. Vol. Second Edition. New York: Springer-Verlag; 2002. Taxonomic Outline of the Procaryotes.

Gibson KE, Barnett MJ, Toman CJ, Long SR, Walker GC. The symbiosis regulator CbrA modulates a complex regulatory network affecting the flagellar apparatus and cell envelope proteins. J. Bacteriol 2007;189:3591–3602. [PubMed: 17237174]

Goesmann A, Linke B, Bartels D, Dondrup M, Krause L, Neuweger H, Oehm S, Paczian T, Wilke M, Meyer F. BRIGEP—the BRIDGE-based genome– transcriptome–proteome browser. Nucleic Acids Res 2005;33:W710–W716. [PubMed: 15980569]

Hoang HH, Becker A, González JE. The LuxR homolog ExpR, in combination with the Sin quorum sensing system, plays a central role in *Sinorhizobium meliloti* gene expression. J. Bacteriol 2004;186:5460–5472. [PubMed: 15292148]

Jacob AI, Adham SA, Capstick DS, Clark SR, Spence T, Charles TC. Mutational analysis of the *Sinorhizobium meliloti* short-chain dehydrogenase/reductase family reveals substantial contribution to symbiosis and catabolic diversity. Mol. Plant-Microbe Interact 2008;21:979–987. [PubMed: 18533838]

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. Nat. Rev. Microbiol 2007;5:619–633. [PubMed: 17632573]

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Research 2000;7:331–338. [PubMed: 11214968]

Krol E, Becker A. Global transcriptional analysis of phosphate stress responses in *Sinorhizobium meliloti* strains 1021 and 2011. Mol. Gen. Genomics 2004;272:1–17.

Luo L, Yao S-Y, Becker A, Rüberg S, Yu Q-Q, Zhu J-B, Cheng HP. Identification of two new *Sinorhizobium meliloti* LysR-like transcriptional regulators required in nodulation. J. Bacteriol 2005;187:4562–4572. [PubMed: 15968067]

MacLean AM, Finan TM, Sadowsky MJ. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiol 2007;144:615–622. [PubMed: 17556525]

Madigan, MT.; Martinko, JM.; Parker, J.; Madigan, MT. Brock Biology of Microorganisms. Vol. Eighth edition. Upper Saddle River, NJ: Prentice Hall; p. 986

Mao C, Qiu J, Wang C, Charles TC, Sobral BW. NodMutDB: a database for genes and mutants involved in symbiosis. Bioinformatics 2005;21:2927–2929. [PubMed: 15817696]

Mao C, Evans C, Jensen RV, Sobral BW. Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. BMC Microbiol 2008;8:72. [PubMed: 18454850]

Mauchline TH, Fowler JE, East AK, Sartor AL, Zaheer R, Hosie AH, Poole PS, Finan TM. Mapping the Sinorhizobium meliloti 1021 solute-binding protein-dependent transportome. Proc. Natl. Acad. Sci. USA 2006;103:17933–17938. [PubMed: 17101990]

Meade HM, Signer ER. Genetic mapping of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. U.S.A 1977;74:2076–2078. [PubMed: 266730]

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A. GenDB - an open source genome annotation system for prokaryote genomes. Nucl. Acids Res 2003;31:2187–2195. [PubMed: 12682369]

Pobigaylo N, Wetter D, Szymczak S, Schiller U, Kurtz S, Meyer F, Nattkemper TW, Becker A. Construction of a large sequence signature-tagged miniTn*5* transposon library and its application to mutagenesis of *Sinorhizobium meliloti*. Appl. Environ. Microbiol 2006;72:4329–4337. [PubMed: 16751548]

Pobigaylo N, Szymczak S, Nattkemper TW, Becker A. Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. Mol. Plant-Microbe Interact 2008;21:219–231. [PubMed: 18184066]

Prell J, Poole P. Metabolic changes of rhizobia in legume nodules. Trends Microbiol 2006;14:161–168. [PubMed: 16520035]

Rasolomampianina R, Bailly X, Fetiarison R, Rabevohitra R, Bena G, Ramaroson L, Raherimandimby M, Moulin L, De Lajudie P, Dreyfus B, Avarre JC. Nitrogen-fixing nodules from rose wood legume trees (Dalbergia spp.) endemic to Madagascar host seven different genera belonging to alpha- and beta-Proteobacteria. Mol. Ecol 2005;14:4135–4146. [PubMed: 16262864]

Rossbach S, Mai DJ, Carter EL, Sauviac L, Capela D, Bruand C, de Bruijn FJ. Response of *Sinorhizobium meliloti* to elevated concentrations of cadmium and zinc. Appl. Environ. Microbiol 2008;74:4218–4221. [PubMed: 18469129]

Rüberg S, Tian ZX, Krol E, Linke B, Meyer F, Wang Y, Pühler A, Weidner S, Becker A. Construction and validation of a DNA microarray for genome-wide expression profiling in *Sinorhizobium meliloti*. J. Biotechnol 2003;106:255–268. [PubMed: 14651866]

Sauviac L, Philippe H, Phok K, Bruand C. An extracytoplasmic function sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*. J. Bacteriol 2007;189:4204–4216. [PubMed: 17400745]

Schroeder BK, House BL, Mortimer MW, Yurgel SN, Maloney SC, Ward KL, Kahn ML. Development of a functional genomics platform for *Sinorhizobium meliloti*: construction of an ORFeome. Appl. Environ. Microbiol 2005;71:5858–5864. [PubMed: 16204497]

Simon R, Priefer U, Pühler A. A broad host range mobilisation system for *in vivo* genetic engineering: transposon mutagenesis in gram-negative bacteria. Biotechnology 1983;1:784–791.

Skovgaard M, Jensen LJ, Brunak S, Ussery D, Krogh A. On the total number of genes and their length distribution in complete microbial genomes. Trends in Genet 2001;17:425–428. [PubMed: 11485798]

Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucl. Acids Res 2000;28:33–36. [PubMed: 10592175]

Wells DH, Chen EJ, Fisher RF, Long SR. ExoR is genetically coupled to the ExoS-ChvI two-component system and located in the periplasm of *Sinorhizobium meliloti*. Mol. Microbiol 2007;64:647–664. [PubMed: 17462014]

Werner, D.; Newton, WE., editors. Nitrogen fixation in agriculture, forestry, ecology, and the environment. The Netherlands: Springer, Dordrecht; 2005. p. 347

Zahran HH. *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiol. Mol. Biol. Rev 1999;63:968–989. [PubMed: 10585971]

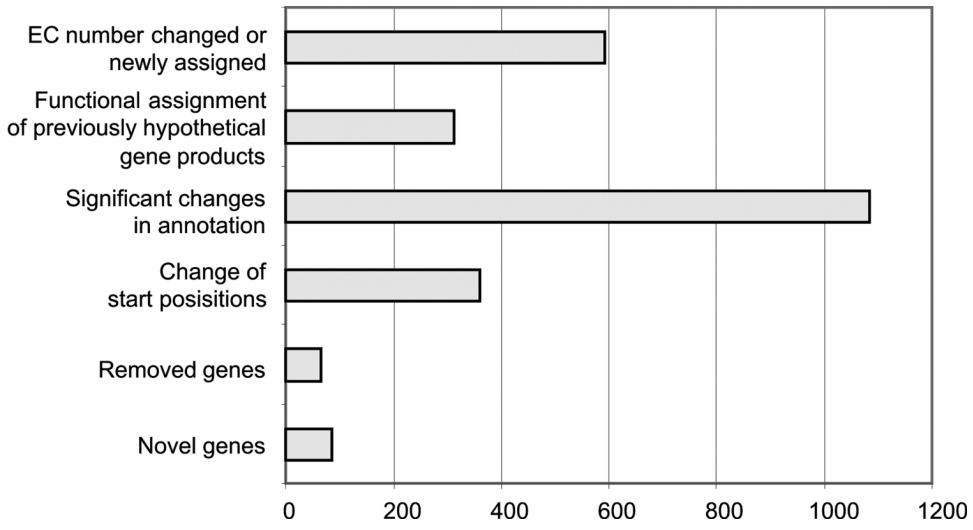

Fig. 1.
Comparison of the annotation update to the original annotation (Galibert et al., 2001).

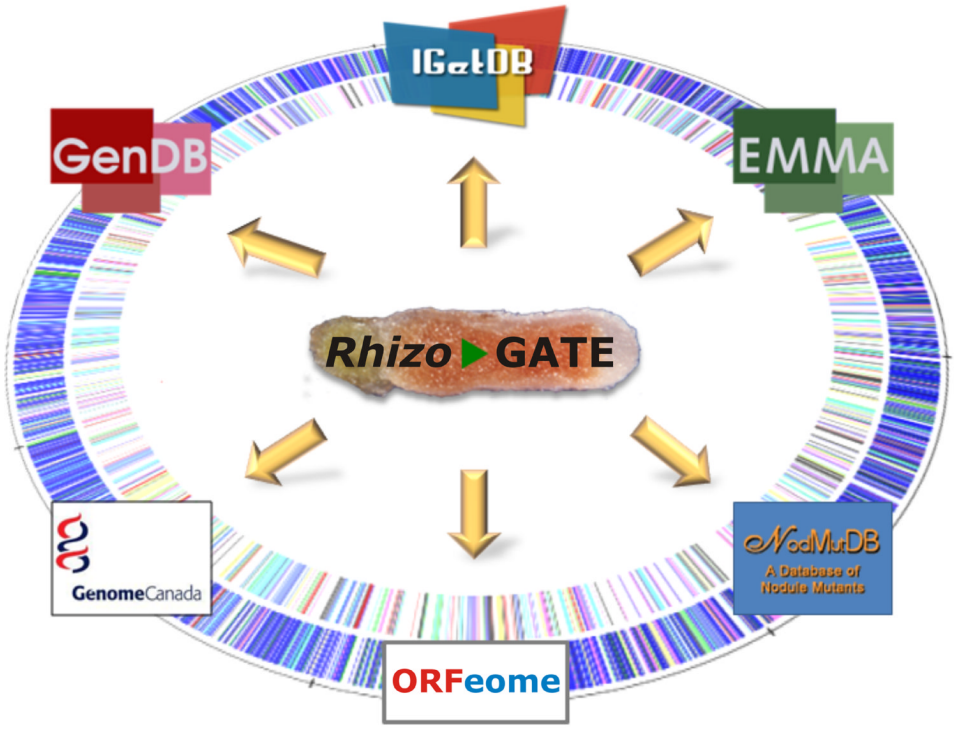

Fig. 2.

The *Rhizo*GATE portal (www.rhizogate.de) provides access to genome (GenDB, Bielefeld University) and transcriptome (EMMA, Bielefeld University) data, includes a data warehouse (IGetDB, Bielefeld University), and links to *S. meliloti* gene fusion (McMaster University, GenomeCanada) and ORFeome library data (Washington State University), and to the NodMutDB mutant database (Virginia Bioinformatics Institute).

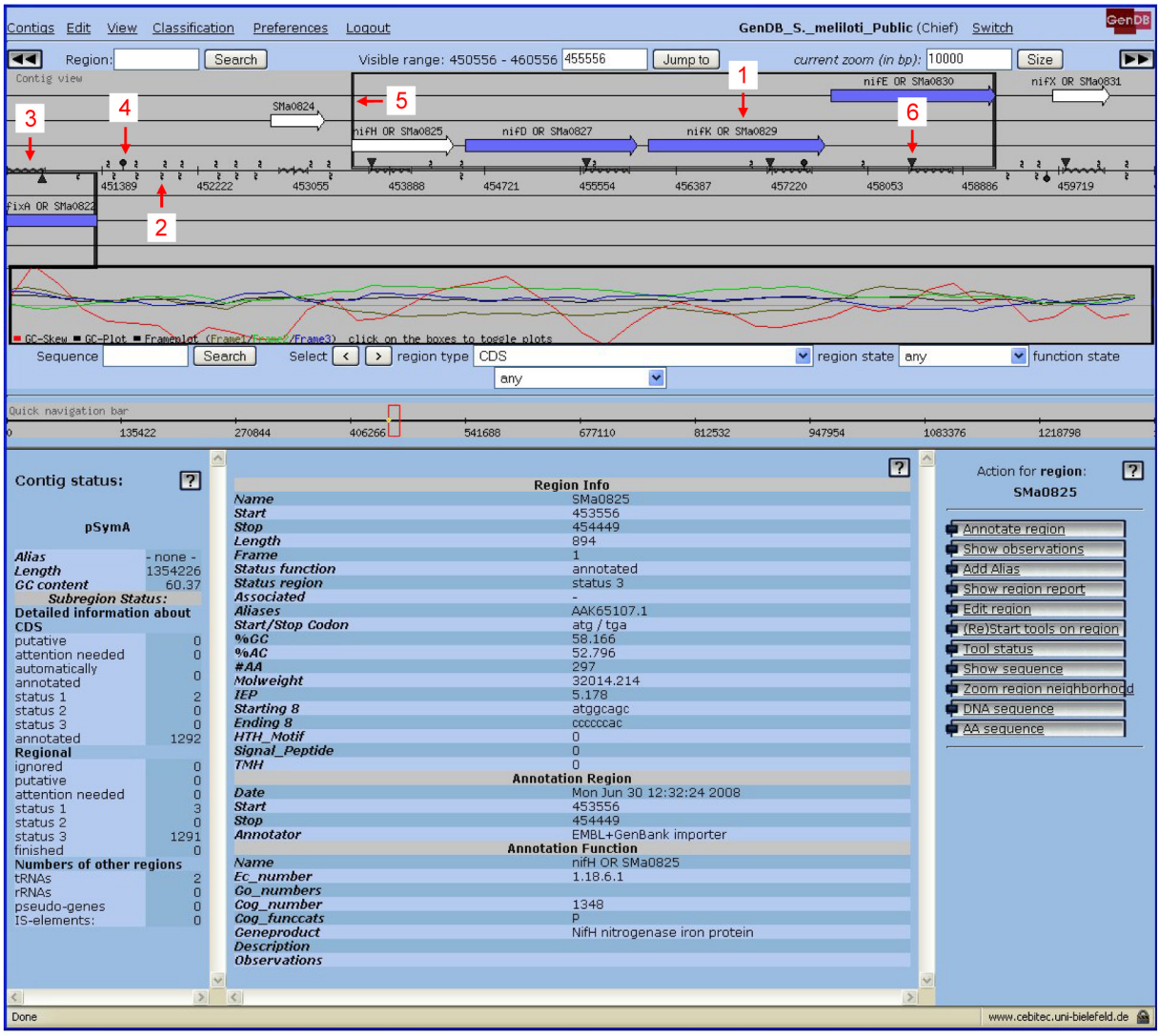

Fig. 3. Main contig view of the GenDB *Rhizo*GATE project

1, coding region; 2, microarray oligonucleotide probe; 3, microarray PCR fragment probe; 4, transposon insertion; 5, operon; 6, integrated plasmid

Table 1

General features of the *S. meliloti* strain 1021 genome

	Chromosome 2001	update	pSymA 2001	update	pSymB 2001	update	Genome 2001	update
Length (bp)	3,654,135	3,654,135	1,354,226	1,354,226	1,683,333	1,683,333	6,691,694	6,691,694
G+C ratio	62.7 %	62.7 %	60.4 %	60.4 %	62.4 %	62.4 %	62.1 %	62.1 %
tRNAs	51	51	2	2	1	1	54	54
rRNA operons	3	3	0	0	0	0	3	3
Protein-coding genes	3341	3351	1293	1291	1570	1583	6204	6225
Genes with functional assignment	59 %	74.5 %	56.5 %	60.6 %	64.4 %	74.7 %	59.7 %	71.6 %

* % of total protein-coding genes

Table 2

Data types accessible via the *S. meliloti* GenDB project

Data type	Reference
Genome sequence, coding regions, RNA genes, repeat motifs (insertion elements, transposons, RIME elements)	Galibert et al., 2001
mTn*5*-STM mutants	Pobigaylo et al., 2006
Plasmid integration mutants	Rüberg and Becker; Capela et al.; unpublished
Inserts of gene fusion plasmids	Cowie et al., 2006
Operons	Krol et al., unpublished
Sm6kPCR microarray probes	Rüberg et al., 2003
Sm6kOligo microarray probes	Krol and Becker, 2004
Sm14kOLI microarray probes	Becker, 2008
Observations (Blast matches, InterPro, Pfam, TMHMM, SignalP, helix-turn-helix, TIGRFAM, Priam)	

Table 3

Rhizobiales genomes in the GenDB project

Organism	Accession numbers
Sinorhizobium meliloti 1021	AL591688, NC_003037, AL591985
Sinorhizobium medicae WSM419	NC_009636, NC_009620, NC_009621, NC_009622
Rhizobium etli CFN42	NC_007761, NC_007762, NC_007763, NC_007764, NC_007765, NC_007766, NC_004041
Rhizobium etli CIAT 652	NC_010994, NC_010998, NC_010996, NC_010997
Rhizobium leguminosarum bv. viciae 3841	NC_008380, NC_008381, NC_008384, NC_008378, NC_008382, NC_008383, NC_008379
Rhizobium sp. NGR234	NC_000914
Mesorhizobium loti MAFF303099	NC_002678, NC_002679, NC_002682
Mesorhizobium sp. BNC1	NC_008254, NC_008242, NC_008243, NC_008244
Bradyrhizobium japonicum USDA 110	NC_004463
Bradyrhizobium sp. BTAi1	NC_009485, NC_009475
Bradyrhizobium sp. ORS278	NC_009445

Table 4

Data sources accessible via the IGetDB repository

Data type	Reference
GenDB region information: e.g. coding regions, mutants, microrray probes, operons Transcriptome data	Galibert et al., 2001; Rüberg et al., 2003; Krol and Becker, 2004; Pobigaylo et al., 2006; Becker, 2008 Barnett et al., 2004; Becker et al., 2004; Chao et al., 2004, 2005; Capela et al., 2005 ; Hoang et al., 2004; Krol and Becker, 2004; Bobik et al., 2006; Capela et al., 2006; Dominguez-Ferreras et al., 2006; Gibson, 2007; Sauviac et al., 2007; Wells et al., 2007 ; Rossbach et al., 2008 Djordjevic et al., 2003
Proteome data	Djordjevic et al., 2003
Gene fusion library: expression information	Cowie et al. 2006
ORFeome: clone information	Schroeder et al., 2005
mTn*5*-STM mutants: phenotypes	Pobigaylo et al., 2006; Pobigaylo et al. 2008 ; Capela et al. unpublished
Plasmid integration mutants: phenotypes	Luo et al. 2005 ; Capela et al., unpublished