

*Databases and ontologies***NodMutDB: a database for genes and mutants involved in symbiosis**Chunhong Mao^{1,*}, Jing Qiu¹, Chunxia Wang¹, Trevor C. Charles² and Bruno W. S. Sobral¹¹Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA and ²Department of Biology, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Received on February 15, 2005; accepted on March 31, 2005

Advance Access publication April 7, 2005

ABSTRACT

Summary: Functional genomics research is producing large amounts of data on the functions of individual genes related to symbiosis. We have developed a relational database, NodMutDB (Nodulation Mutant Database), to provide a comprehensive resource for depositing, organizing and retrieving information on symbiosis-related genes, mutants and published literature. NodMutDB brings together new studies and existing mutant-based literature to facilitate our understanding of how genes function in symbiotic processes in both Rhizobia and their host plants.

Availability: <http://nodmutdb.vbi.vt.edu>**Contact:** cmao@vbi.vt.edu**Supplementary information:** Database schema and data curation model are available at <http://nodmutdb.vbi.vt.edu>**1 INTRODUCTION**

Rhizobia-legume symbiosis is an attractive research field because of its importance in agriculture. Rhizobia interact with host legume plants in soil to develop root nodules, which convert atmospheric nitrogen into ammonia, a form of nitrogen used by plants as nutrients. Much research effort has been devoted to study the interaction between Rhizobia and legumes, with special focus on two model species, *Sinorhizobium meliloti* and its host *Medicago truncatula*. The *S.meliloti*1021 genome sequence and the initial annotation of the genome were completed and published in 2001 (Barnett *et al.*, 2001; Capela *et al.*, 2001; Finan *et al.*, 2001; Galibert *et al.*, 2001). The *M.truncatula* genome is in the process of being sequenced and annotated (<http://medicago.org>). Pre- and post-genomic research on *S.meliloti* and *M.truncatula* activities have rapidly generated vast amounts of information about genes, gene functions and regulation in bacteria and plants. Forward and reverse genetic studies involving bacteria and plant mutants have also provided valuable information about gene functions and their roles in symbiosis. For example, the isolation of *S.meliloti* mutants with loss of nodulation phenotype led to the discovery of a cluster of Nod genes. These genes are induced by plant flavonoid signals to produce Nod factors (NFs), which then trigger a series of plant responses resulting in the formation of the nitrogen fixing root nodules [see Luyten and Vanderleyden (2000) for a review]. Screening of *M.truncatula* non-nodulating mutants

led to the identification of a Nod factor receptor gene, NORK (Sagan *et al.*, 1995; Endre *et al.*, 2002; Esseling *et al.*, 2004) and other genes such as NFP, DMI1, DMI3, NSP1 and NSP2 that are involved in nodulation at various stages [see Mitra *et al.* (2004) for a review]. These studies thus established the bi-directional signaling circuitry for nodule development. We believe that the integration of the rapidly growing genomic and genetic information with published literature will undoubtedly improve our understanding of the mechanisms of symbiotic processes and facilitate further research. As of February 2005, more than 1900 published research articles on *S.meliloti* and *M.truncatula* are available on PubMed (NCBI, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>). There is a need for the symbiosis-research community to have a comprehensive and high quality database for depositing, organizing and retrieving information on symbiosis-related genes, mutants and published literature. We have developed such a relational database called NodMutDB (Nodulation Mutant Database) with special emphasis on quality control of data input and on the ability of researchers to use the database to search for general patterns.

2 NodMutDB DESIGN AND DESCRIPTION

NodMutDB is a comprehensive relational database (implemented in Oracle 9i) for organizing, storing and retrieving normalized information on symbiosis-related genes, mutants and publications. Data are collected through literature review and public database searches by a team of curators. The database can be easily accessed and searched through a user-friendly Web interface (written in PHP; Fig. 1). The curated information in the database includes (1) gene information: symbol, name, id, product, function, location, mutant collection, COG classification, physical properties and external links to GenBank, SwissProt/UniProt and other public databases; (2) mutant information: parental background, mutant strain/line, mutagenesis methods, mutation locations, mutant class, e.g., Nod⁻ for defective in nodulation, and Fix⁻ for defective in nitrogen fixation, phenotypes at different nodule developmental stages, cell growth, gene regulations, phenotypic effects on counter partners, genes affected, allelism and related articles; (3) published literature on *S.meliloti* and its host plants. In order to help to understand the interaction between Rhizobia and the plant host, we place special emphasis on accurately describing the symbiotic phenotype of the mutants including (1) how the mutant bacteria affect nodulation or nitrogen fixation of

*To whom correspondence should be addressed.

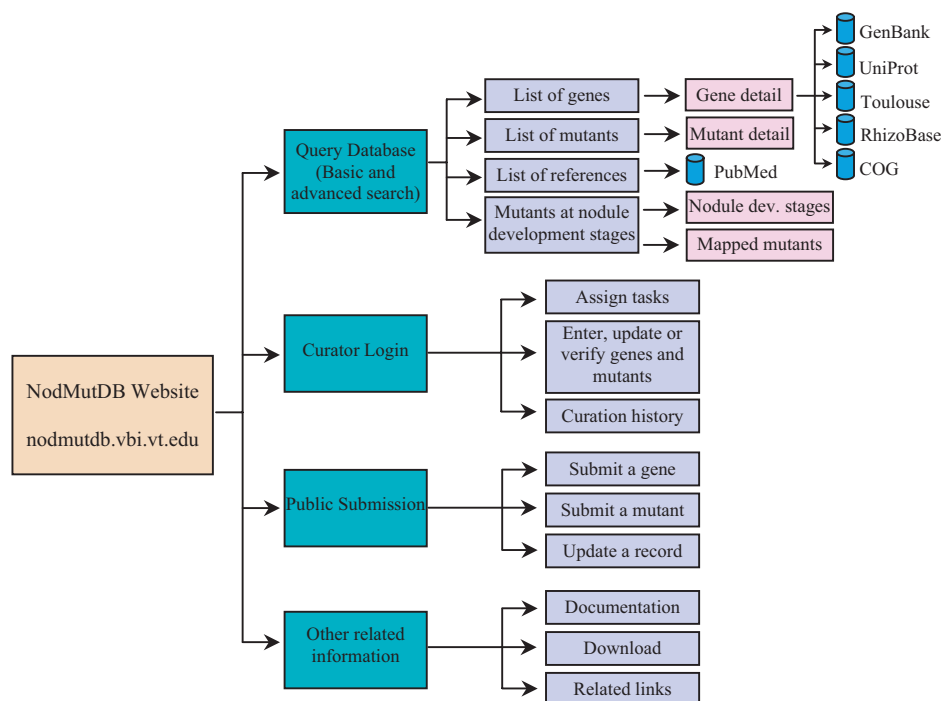


Fig. 1. NodMutDB Web interface.

the wild-type or a mutant plant; (2) how well the mutant plant grows and whether it makes functional nodules when inoculated with Nod factors, a wild-type or a mutant rhizobial strain; (3) which nodule development stages are affected by the mutants and under what conditions. Users can also submit new genes and mutants to the database, which will be validated by our curators before releasing to the public.

3 DATA ENTRY AND CURATION

Publications related to genes and mutants involved in symbiosis of *S.meliloti* and *M.truncatula* were collected by a team of curators through exhaustive PubMed, Google and journal-specific searches and literature reviews and citations. The criteria to include a gene in the NodMutDB database, of which only one must be met, are: (1) experimental evidence showing the involvement of the gene in symbiosis signal transduction pathways or nodule development and metabolic functions; (2) the mutants of the gene displaying defective symbiotic phenotype with a measurable effect on nodule formation or function; (3) marker genes for nodulation such as nodulins; (4) if one of the genes in an operon has been demonstrated to be involved in symbiosis, then other genes of the same operon are also included in NodMutDB, so as to help to understand the function of that operon. Candidate genes and their information were manually entered into NodMutDB by our curators via the NodMutDB Web interface. Each gene was then assigned to a curator to search and curate mutants being studied for that gene. Mutants and their related publications were entered into NodMutDB. The information was reviewed, updated and verified by a senior curator to ensure quality and accuracy. The curation status of each gene and mutant is available on the Web <http://nodmutdb.vt.edu>. The status includes date/time/who submitted, entered, updated or verified the information.

We welcome the public submission of mutants/genes that affect nodulation and nitrogen fixation to facilitate data collection. At the same time, we also want to ensure the quality and accuracy of the information. Researchers may submit their mutant/gene information through our Web interface. Once we receive the public submissions, our curators will validate and enter the information into the database for publishing on the website. The public submitters will be acknowledged for their contribution to NodMutDB.

4 WEB INTERFACE

We developed a user-friendly Web interface for easy viewing and retrieving of the NodMutDB database. Figure 1 illustrates the navigation of the Webpages. Users can query all genes, mutants and their references in the database as well as search for specific ones using either a basic or an advanced search engine. For example, the user may query *S.meliloti* mutants in the 'Nod⁻' mutant class that affected the root hair curling developmental stage. A summary of information for the matched genes, mutants or references will be displayed with each item (gene, mutant, reference) linking to a more detailed information page and other public databases such as GenBank, UniProt/SwissProt, Toulouse *Sinorhizobium meliloti* 1021 genome sequencing projects (<http://sequence.toulouse.inra.fr/meliloti.html>), RhizoBase (<http://www.kazusa.or.jp/rhizobase/Sinorhizobium/index.html>), COG (clusters of orthologous groups, <http://www.ncbi.nlm.nih.gov/COG/>) and PubMed (Fig. 1). We also mapped *M.truncatula* mutants to the nodule developmental timeline to assess the temporal effects and upstream/downstream relationship of the mutants during the symbiosis process. We provide a simple diagram of the nodule development with *M.truncatula* mutants mapped at the different stages.

Table 1. NodMutDB data curation status

	<i>S.meliloti</i>	<i>M.truncatula</i>
Number of genes	234	53
Number of mutants	664	48
References on mutant study	201	30
Total references collected	1709	254

Each development stage is defined and explained based on previously published reviews. Each mutant gene group is linked to a more detailed information page for the mutant alleles and their symbiotic phenotypes.

5 CURRENT STATUS AND FUTURE PLANS

The current data curation status for the collected genes, mutants and published articles in the database is summarized in Table 1. At the time of writing this manuscript, we completed curating collected paper references on *S.meliloti* and *M.truncatula* published by then. Since NodMutDB is an ongoing effort, we continue to curate and consistently update the database as new studies are published. We also plan to include other legume species such as alfalfa, lotus, soybean and pea and Rhizobia species such as *Bradyrhizobium japonicum* and *Mesorhizobium loti*. As we incorporate more plant and bacteria species in the database, we will provide a mechanism to compare genes, mutants and legume–rhizobia interactions between different species. As more species are added to the database and the amount of publications concerning genes and mutants increases, data curation will become more challenging. We encourage the plant-microbe symbiosis research community to be involved to facilitate

the curation process as we provide the foundation. We hope that NodMutDB will become a community-based database.

ACKNOWLEDGEMENT

We thank Drs. Pedro Mendes and Allan Dickerman for helpful discussions. We also thank Dr. Daniel Kahn at INRA-CNRS, France for making NodMutDB gene/mutant links available at the international *S.meliloti* genome project Toulouse website. The NodMutDB project is supported by funding from the Virginia Bioinformatics Institute and the Ontario Genomics Institute.

REFERENCES

- Barnett, M.J. *et al.* (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc. Natl Acad. Sci. USA*, **98**, 9883–9888.
- Capela, D. *et al.* (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl Acad. Sci. USA*, **98**, 9877–9882.
- Endre, G. *et al.* (2002) A receptor kinase gene regulating symbiotic nodule development. *Nature*, **417**, 962–966.
- Esseling, J.J. *et al.* (2004) A nonsymbiotic root hair tip growth phenotype in NORK-mutated legumes: implications for nodulation factor-induced signaling and formation of a multifaceted root hair pocket for bacteria. *Plant Cell*, **16**, 933–944.
- Finan, T.M. *et al.* (2001) The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *Proc. Natl Acad. Sci. USA*, **98**, 9889–9894.
- Galibert, F. *et al.* (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, **293**, 668–672.
- Luyten, E. and Vanderleyden, J. (2000) Survey of genes identified in *Sinorhizobium meliloti* spp., necessary for the development of an efficient symbiosis. *Eur. J. Soil Biol.*, **36**, 1–26.
- Mitra, R.M. *et al.* (2004) Six non-nodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. *Proc. Natl Acad. Sci. USA*, **101**, 10217–10222.
- Sagan, M. *et al.* (1995) Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after γ -ray mutagenesis. *Plant Sci.*, **111**, 63–71.