

Don A. Au (Am) 5  
Page 5

RESEARCH ARTICLES

served how short-wavelength excitations cause a breakdown of superfluidity in a BEC. Our results show how localized defects in a superfluid will quite generally either disperse into high-frequency ripples or end up in the form of topological defects such as solitons and vortices, and we have obtained an analytic expression for the transition between the two regimes. By varying our experimental parameters, we can create differently sized and shaped defects and can also control the number of defects created, allowing studies of a myriad of effects. Among them are soliton-soliton collisions, more extensive studies of soliton stability, soliton-sound wave collisions, vortex-soliton interactions, vortex dynamics, interaction between vortices, and the interaction between the BEC collective motion and vortices.

3355 (1998). A Mathematica notebook is available at <http://fermion.colorado.edu/~chg/Collisions>.

31. J. Javanainen, J. Ruostekoski, *Phys. Rev. A* **52**, 3033 (1995).  
 32. Y. B. Band, M. Tribbenbach, J. P. Burke, P. S. Julienne, *Phys. Rev. Lett.* **84**, 5462 (2000).  
 33. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, ed. 2, 1992).  
 34. S. E. Koonin, D. C. Merideth, *Computational Physics* (Addison-Wesley, Reading, MA, 1990).  
 35. For propagation of the Gross-Pitaevskii equation in 1D, we typically used a spatial grid with 4000 points and  $\Delta z = 0.040 \mu\text{m}$ . In 2D simulations, we typically used a  $750 \times 750$  grid with  $\Delta z = 0.21 \mu\text{m}$  and  $\Delta x = 0.057 \mu\text{m}$ . To solve the equations self-consistently, we kept track of the wave functions at previous time points and projected forward to second order. Smaller time steps and grid spacing were also used to assure convergence of the results. To mimic the nonlinear interaction strength at the center of a 3D cloud, we put in an effective condensate radius [calculated with the Thomas-Fermi approximation (39)] in the dimensions that were not treated dynamically. In all calculations, the initial condition was

the ground-state condensate wave function with all atoms in  $|1\rangle$ , obtained by propagating a Thomas-Fermi wave function in imaginary time.

36. M. R. Andrews et al., *Phys. Rev. Lett.* **79**, 553 (1997).  
 37. M. R. Andrews et al., *Science* **275**, 637 (1997).  
 38. F. Dalfovo, M. Modugno, *Phys. Rev. A* **61**, 023605 (2000).  
 39. G. Baym, C. Pethick, *Phys. Rev. Lett.* **76**, 6 (1996).  
 40. Supplementary material is available at Science Online ([www.sciencemag.org/cgi/content/full/1062527/DC1](http://www.sciencemag.org/cgi/content/full/1062527/DC1)).  
 41. Supported by a National Defense Science and Engineering Grant sponsored by the U.S. Department of Defense (C.S.), the Rowland Institute for Science, the Defense Advanced Research Projects Agency, the U.S. Air Force Office of Scientific Research, the U.S. Army Research Office OSD Multidisciplinary University Research Initiative Program, the Harvard Materials Research Science and Engineering Center (sponsored by NSF), and the Carlsberg Foundation, Denmark.

15 May 2001; accepted 11 June 2001  
 Published online 28 June 2001;  
 10.1126/science.1062527

Include this information when citing this paper.

References and Notes

1. R. J. Donnelly, *Quantized Vortices in Helium II* (Cambridge Univ. Press, Cambridge, 1991).  
 2. R. M. Bowley, *J. Low Temp. Phys.* **87**, 137 (1992).  
 3. L. D. Landau, E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959).  
 4. A. L. Fetter, in *Bose-Einstein Condensation in Atomic Gases, Proceedings of the International School of Physics Enrico Fermi, Course CXL*, M. Inguscio, S. Stringari, C. Wieman, Eds. (International Organizations Services, Amsterdam, 1999), pp. 201-263.  
 5. For a review, see F. Dalfovo, S. Giorgini, L. P. Pitaevskii, S. Stringari, *Rev. Mod. Phys.* **71**, 463 (1999).  
 6. M. R. Matthews et al., *Phys. Rev. Lett.* **83**, 2498 (1999).  
 7. S. Burger, K. Bongs, S. Dettmer, W. Ertmer, K. Sengstock, *Phys. Rev. Lett.* **83**, 5198 (1999).  
 8. J. Denschlag et al., *Science* **287**, 97 (2000).  
 9. B. P. Anderson et al., *Phys. Rev. Lett.* **86**, 2926 (2001).  
 10. K. W. Madison, F. Chevy, W. Wohlleben, J. Dalibard, *Phys. Rev. Lett.* **84**, 806 (2000).  
 11. J. R. Abo-Shaeer, C. Raman, J. M. Vogels, W. Ketterle, *Science* **292**, 476 (2001).  
 12. C. Raman et al., *Phys. Rev. Lett.* **83**, 2502 (1999).  
 13. B. Jackson, J. F. McCann, C. S. Adams, *Phys. Rev. A* **61**, 051603(R) (2000).  
 14. L. V. Hau, S. E. Harris, Z. Dutton, C. H. Behroozi, *Nature* **397**, 594 (1999).  
 15. S. E. Harris, *Phys. Today* **50**, 36 (1997).  
 16. M. O. Scully, M. S. Zubairy, *Quantum Optics* (Cambridge Univ. Press, Cambridge, 1997).  
 17. S. E. Harris, L. V. Hau, *Phys. Rev. Lett.* **82**, 4611 (1999).  
 18. S. A. Morgan, R. J. Ballagh, K. Burnett, *Phys. Rev. A* **55**, 4338 (1997).  
 19. W. P. Reinhardt, C. W. Clark, *J. Phys. B* **30**, L785 (1997).  
 20. Th. Busch, J. R. Anglin, *Phys. Rev. Lett.* **84**, 2298 (2000).  
 21. B. B. Kadomtsev, V. I. Petviashvili, *Sov. Phys. Dokl.* **15**, 539 (1970).  
 22. C. A. Jones, S. J. Putterman, P. H. Roberts, *J. Phys. A* **19**, 2991 (1986).  
 23. C. Josseland, Y. Pomeau, *Europhys. Lett.* **30**, 43 (1995).  
 24. D. L. Feder, M. S. Pindzola, L. A. Collins, B. I. Schneider, C. W. Clark, *Phys. Rev. A* **62**, 053606 (2000).  
 25. A. V. Mamaev, M. Saffman, A. A. Zozulya, *Phys. Rev. Lett.* **76**, 2262 (1996).  
 26. L. V. Hau et al., *Phys. Rev. A* **58**, R54 (1998).  
 27. We use the definition  $\Omega_p = E_{p0} \cdot d_{13}$ ,  $\Omega_c = E_{c0} \cdot d_{23}$ , where  $E_{p0}$ ,  $E_{c0}$  are the slowly varying electric field amplitudes, and  $d_{13}$ ,  $d_{23}$  are the electric dipole matrix elements of the transitions.  
 28. S. E. Harris, J. E. Field, A. Kasapi, *Phys. Rev. A* **46**, R29 (1992).  
 29. C. Liu, Z. Dutton, C. H. Behroozi, L. V. Hau, *Nature* **409**, 490 (2001).  
 30. J. P. Burke, C. H. Greene, J. L. Bohn, *Phys. Rev. Lett.* **81**,

# The Composite Genome of the Legume Symbiont *Sinorhizobium meliloti*

Francis Galibert,<sup>1</sup> Turlough M. Finan,<sup>2</sup>  
 Sharon R. Long,<sup>3,4\*</sup> Alfred Pühler,<sup>5</sup> Pia Abola,<sup>6</sup> Frédéric Ampe,<sup>7</sup>  
 Frédérique Barloy-Hubler,<sup>1</sup> Melanie J. Barnett,<sup>3</sup> Anke Becker,<sup>5</sup>  
 Pierre Boistard,<sup>7</sup> Gordana Bothe,<sup>8</sup> Marc Boutry,<sup>9</sup> Leah Bowser,<sup>6</sup>  
 Jens Buhrmester,<sup>5</sup> Edouard Cadieu,<sup>1</sup>  
 Delphine Capela,<sup>1,7†</sup> Patrick Chain,<sup>2</sup> Alison Cowie,<sup>2</sup>  
 Ronald W. Davis,<sup>6</sup> Stéphane Dréano,<sup>1</sup>  
 Nancy A. Federspiel,<sup>6‡</sup> Robert F. Fisher,<sup>3</sup> Stéphanie Gloux,<sup>1</sup>  
 Thérèse Godrie,<sup>10</sup> André Goffeau,<sup>9</sup> Brian Golding,<sup>2</sup>  
 Jérôme Gouzy,<sup>7</sup> Mani Gurjal,<sup>6</sup> Ismael Hernandez-Lucas,<sup>2</sup>  
 Andrea Hong,<sup>3</sup> Lucas Huizar,<sup>6</sup> Richard W. Hyman,<sup>6</sup> Ted Jones,<sup>6</sup>  
 Daniel Kahn,<sup>7</sup> Michael L. Kahn,<sup>11</sup>  
 Sue Kalman,<sup>6§</sup> David H. Keating,<sup>3,4</sup> Ernö Kiss,<sup>7</sup> Caridad Komp,<sup>6</sup>  
 Valérie Lelaure,<sup>1</sup> David Masuy,<sup>9</sup> Curtis Palm,<sup>6</sup> Melicent C. Peck,<sup>3</sup>  
 Thomas M Pohl,<sup>8</sup> Daniel Portetelle,<sup>10</sup> Bénédicte Purnelle,<sup>9</sup>  
 Uwe Ramsperger,<sup>8</sup> Raymond Surzycki,<sup>6¶</sup> Patricia Thébault,<sup>7</sup>  
 Micheline Vandenbol,<sup>10</sup> Frank-J. Vorhölter,<sup>5</sup> Stefan Weidner,<sup>5</sup>  
 Derek H. Wells,<sup>3</sup> Kim Wong,<sup>2</sup> Kuo-Chen Yeh,<sup>3,4¶</sup> Jacques Batut<sup>7</sup>

The scarcity of usable nitrogen frequently limits plant growth. A tight metabolic association with rhizobial bacteria allows legumes to obtain nitrogen compounds by bacterial reduction of dinitrogen ( $N_2$ ) to ammonium ( $NH_4^+$ ). We present here the annotated DNA sequence of the  $\alpha$ -proteobacterium *Sinorhizobium meliloti*, the symbiont of alfalfa. The tripartite 6.7-megabase (Mb) genome comprises a 3.65-Mb chromosome, and 1.35-Mb pSymA and 1.68-Mb pSymB megaplasmids. Genome sequence analysis indicates that all three elements contribute, in varying degrees, to symbiosis and reveals how this genome may have emerged during evolution. The genome sequence will be useful in understanding the dynamics of interkingdom associations and of life in soil environments.

Symbiotic nitrogen fixation is profoundly important for the environment. Most plants assimilate mineral nitrogen only from soil or

added fertilizer. An alternative source powered by photosynthesis, rhizobia-legume symbioses provide a major source of fixed



## RESEARCH ARTICLES

nitrogen. Evolution in diverse legumes of high protein content in seeds (e.g., soybean) and leaves (e.g., alfalfa) may reflect the ability of many plants in this taxon to obtain nitrogen from bacterial symbionts while growing in poor soils. Improved understanding of the rhizobia-legume symbiosis has implications for sustainable agriculture and ecosystem function. *Sinorhizobium meliloti*, the symbiont of alfalfa, is a focus of research both because of the symbiosis and because, as an  $\alpha$ -proteobacterium, it is closely related to bacterial plant and animal pathogens including *Agrobacterium* and *Brucella*. Rhizobia infect roots and induce nodules, specialized organs where bacterial endosymbionts fix nitrogen within the plant cytoplasm. The bacteria and plant exchange signals during nodule development and establish an intimate metabolic exchange of bacterial fixed nitrogen for plant carbon compounds. We understand some symbiotic mechanisms, but how

the microbe stimulates nodule organogenesis, how it invades the plant without triggering host defenses, and how and why the bacterium fixes nitrogen for the host rather than for its own metabolism attract considerable interest. Furthermore, for symbiosis to be a successful habit, the bacteria must maintain their populations in the soil and establish themselves competitively in the rhizosphere through adaptations that are little understood.

*Sinorhizobium meliloti* has been the subject of extensive genetic, biochemical, and metabolic research; this knowledge provides a solid foundation for genomic experimentation. We report here the complete and fully annotated nucleotide sequence of the *S. meliloti* strain 1021 genome (1), and an integrative analysis of the biology implied by the sequence (2). We also present the first global comparison between two rhizobial genomes, the *S. meliloti* genome and the recently reported *Me-*

*sorhizobium loti* genome (3). In addition to these two complete genomes, the 536-kb symbiotic plasmid of *Rhizobium* sp. NGR234 (4) and a 410-kb region of the chromosome of *Bradyrhizobium japonicum* (5) have been sequenced and annotated. Methods and detailed analyses of the *S. meliloti* chromosome, pSymA, and pSymB are reported concurrently (6–8).

### General Features of the Genome

Main features of the genome are listed in Table 1. The *S. meliloti* genome consists of three replicons: one large replicon of 3.65 Mb and two smaller replicons, pSymA and pSymB, of 1.35 and 1.68 Mb, respectively (Fig. 1).

Although one of the largest bacterial genomes (6.7 Mb) sequenced to date, the *S. meliloti* genome is somewhat smaller than the 7.6-Mb *M. loti* genome. We predict 6204 protein-coding genes from the *S. meliloti* genome sequence, compared with

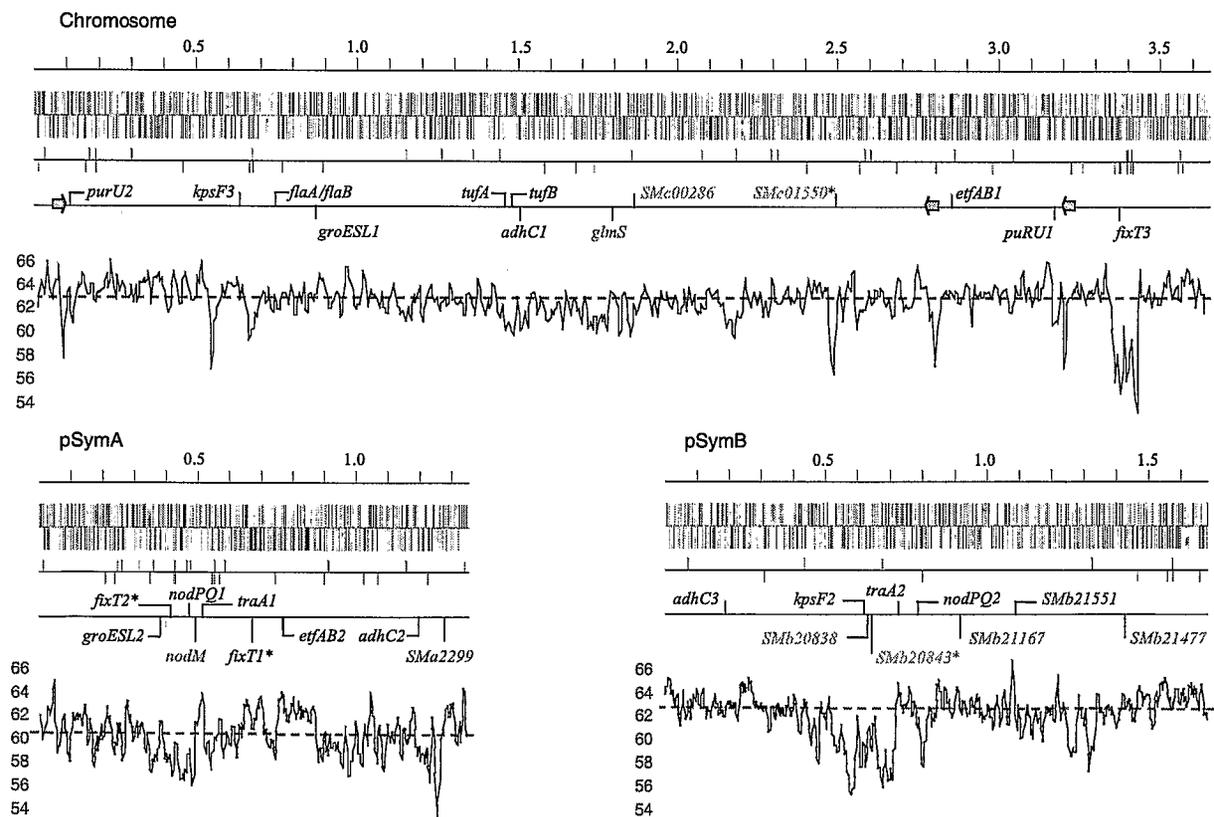


Fig. 1. Linear representation of the *S. meliloti* genome (strain 1021). Each replicon is drawn to scale. First line: Coordinates relative to the sequence on Web site (2) (in megabases). Second line: Distribution of genes according to direction of transcription (+ strand above) and functional categories (blue, small molecule metabolism; green, macromolecule metabolism; orange, structural elements; yellow, cell processes; red, elements of external origin; pink, not classified regulators; gray, conserved hypothetical and unknown/hypothetical). Third line:

Distribution of IS and phage-related sequences. Fourth line: Recently duplicated genes (those with at least 90% nucleotide identity over their entire length) including *rtn* operons (green arrows). Duplications of differently named genes are matched by color. Because of space constraints, the *Sma0753/Sma0758* duplication at 0.41 Mb is not shown. Loci where clusters of genes are reiterated are indicated by an asterisk. Fifth line: GC% variation along the replicons with mean value as a red dotted line.

## RESEARCH ARTICLES

6752 for *M. loti* (3). A function could be postulated for 59.7% of *S. meliloti* genes on the basis of database comparisons, whereas 8.2% of the *S. meliloti* gene products had no database match. The proportion of orphan genes was significantly higher on the megaplasmids than on the chromosome, with 11.5% on pSymA and 12.3% on pSymB (Table 1).

Contrary to expectations (9–11), the genome of *S. meliloti* is not highly reiterated. A limited number of genes appear to be recently duplicated, including several symbiotic genes (Fig. 1). However, the *S. meliloti* genome contains many ancient duplications, because 42% (2589) of *S. meliloti* genes belong to 548 paralogous families, ranging from 2 to 134 genes per family (2, 12). This high level of paralogy suggests that genome size has been little constrained during *S. meliloti* evolution, facilitating the acquisition of new adaptive functions for life in the soil and for symbiosis. This is illustrated by the rich set of transport and regulatory functions (see below).

Insertion sequence (IS) elements and phage sequences compose 2.2% of the *S. meliloti* genome, but their distribution varies (Table 1). Overall abundance is higher on pSymA, especially near symbiotic genes (Fig. 1), a feature similar to other rhizobial

symbiotic plasmids and regions (3–5, 13). This provides additional evidence that symbiotic regions are prone to DNA rearrangements (3). Twenty-one types of IS are identified on the *S. meliloti* genome: four are chromosome-specific, four are pSymA-specific, and one is pSymB-specific [see (2) for additional data on IS].

**Replication, transfer, and maintenance of pSym megaplasmids.** The unusual size of the megaplasmids raises the question of whether they are plasmids or chromosomes. pSymA and pSymB share plasmid features with *Rhizobium* sp. pNGR234a and *Agrobacterium* Ti and Ri plasmids: *rep-ABC* genes were identified by sequence similarity, and a linked putative origin of replication was inferred from GC skew analysis. pSymA contained putative conjugative transfer genes (*traACDG*) and a putative *oriT* sequence, but lacked the *traI-RMBF* and *trbDJKLFH* genes found on other rhizobial plasmids. Transfer experiments are required to determine whether pSymA is a transferable plasmid. pSymB

lacks transfer genes, except for a paralog of the pSymA *traA* and *oriT*. Its lower G + C content (60.4%) compared with the other two replicons (Table 1) and its strikingly distinct codon usage [see Web site (2)] suggest an alien origin for pSymA.

No essential gene could be predicted on pSymA, consistent with previous data (14). However, essential genes are present on pSymB, including the arginine tRNA, *Arg-tRNA<sub>CCG</sub>*; the *minCDE* cell division genes that may also be essential; and two candidate genes for asparagine synthesis (*asn*), one of which should be required for growth in minimal medium (8). Therefore pSymA is clearly plasmidlike, whereas pSymB has several chromosomal features.

**Transport functions.** Genes encoding transport systems constitute the largest (12%) class of genes in the *S. meliloti* genome. Most of these are ABC transporters (Fig. 2), as is the case in other bacterial genomes. Their relative abundance is particularly high (17.4%) on pSymB, where almost all are predicted to be import sys-

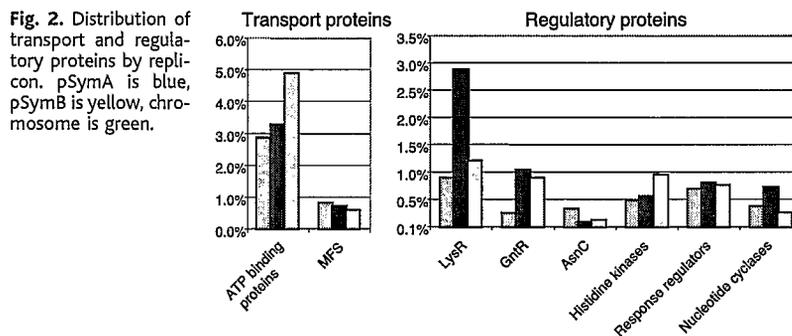


Fig. 2. Distribution of transport and regulatory proteins by replicon. pSymA is blue, pSymB is yellow, chromosome is green.

<sup>1</sup>UMR6061-CNRS, Laboratoire de Génétique et Développement, Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes cedex, France. <sup>2</sup>Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1. <sup>3</sup>Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA. <sup>4</sup>Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA. <sup>5</sup>Universität Bielefeld, Biologie VI (Genetik), Universitätsstrasse 25, D-33615 Bielefeld, Germany. <sup>6</sup>Stanford Center for DNA Sequencing and Technology, Stanford, CA 94305, USA. <sup>7</sup>Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, UMR215-CNRS-Institut National de la Recherche Agronomique (INRA), Chemin de Borde Rouge, BP 27, F-31326 Castanet Tolosan Cedex, France. <sup>8</sup>GATC Biotech AG, Jakob-Stadler-Platz GmbH 7, D-78467 Konstanz, Germany. <sup>9</sup>Unité de Biochimie physiologique, Université Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium. <sup>10</sup>Unité de Biologie Animale et Microbienne, Faculté des Sciences Agronomiques de Gembloux, Avenue Maréchal Juin 6, B-5030 Gembloux, Belgium. <sup>11</sup>Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, USA.

\*To whom correspondence should be addressed.  
 †Present address: Institut Curie, 26 rue d'Ulm, 75005 Paris, France.  
 ‡Present address: Exelixis, Inc., 170 Harbor Way, Post Office Box 511, South San Francisco, CA 94083-0511, USA.  
 §Present address: Incyte Genomics, 3160 Porter Drive, Palo Alto, CA 94304, USA.  
 ||Present address: Département de Biologie Moléculaire Sciences 2, Université de Genève, Geneva, Switzerland 1211.  
 ¶Present address: Institute of BioAgricultural Sciences, Academia Sinica, Nankang, Taipei, Taiwan 11529.

Table 1. General features of the *S. meliloti* strain 1021 genome.

Feature	Chromosome	pSymA	pSymB	Genome
Length (bp)	3,654,135	1,354,226	1,683,333	6,691,694
G + C ratio	62.7%	60.4%	62.4%	62.1%
Protein-coding regions	85.8%	83.2%	88.6%	85.9%
Transfer RNAs	51	2	1	54
tmRNA*	1	0	0	1
Ribosomal RNA operons	3	0	0	3
Protein-coding genes	3341	1293	1570	6204
Average length of protein-coding genes (pb)	938	871	950	927
Genes with functional assignment	59%	56.5%	64.4%	59.7%
Orphan genes (% of total protein-coding genes)	5%	11.5%	12.3	8.2%
Regulatory genes (% of total protein-coding genes)	7.2%	10.4%	10.5%	8.7%
Insertional and phage sequences (% of replicon size)	2.2%	3.6%	0.9%	2.2%
RIME elements	185	6	27	218
Palindromes A, B, and C	253	0	5	258

\*tmRNA derives its name from the presence of two separate domains, one that functions as a tRNA, and another that serves as an mRNA.

## RESEARCH ARTICLES

tems (8). Thus, pSymB plays a prominent role in importing small molecules. Rht transporters (hydroxylated amino acid efflux proteins) are unexpectedly abundant (12 members) in *S. meliloti*. No phosphoenolpyruvate sugar phosphotransferase (PTS) transport system was found, implying that sugars are transported and subsequently phosphorylated by cytoplasmic sugar kinases that are encoded by the chromosome and pSymB.

**Regulatory proteins.** Regulatory genes make up a substantial fraction (8.7%) of the *S. meliloti* genome, especially the megaplasmids (Table 1). The LysR family (86 members) predominates, particularly on pSymA (Fig. 2). GntR regulators are more frequently found on megaplasmids, whereas the AsnC family is more common on the chromosome. Thus, each replicon has a distinct regulatory gene profile.

With only seven members,  $\sigma$ 54-dependent transcriptional regulators constitute a small family in *S. meliloti*. A single "quorum-sensing" system (SMc00168, SMc00170) was found. We identified 36 response regulators and 37 histidine kinases, but no serine-threonine kinases. Thus far, *S. meliloti* encodes the most nucleotide cyclases (26 members) of any bacterial genome [see (7) for a detailed analysis]. We identified 14 putative RNA polymerase sigma factor genes, most belonging to the extracytoplasmic function (ECF) subfamily. Similarly to *Caulobacter crescentus* (15) and *M. loti* (3), *S. meliloti* lacks a *rpoS*.

**Bacterial adhesion and surface structural elements.** How rhizobia adhere to plant root hairs is poorly understood. We identified one

putative adhesin (SMc01708), two agglutinin-like genes (SMc00638 and SMc00639), and an ABC transporter (SMa0950 to SMa0953) resembling the *attA1A2BC* attachment genes of *A. tumefaciens* (16). Two previously unknown pili were postulated: a type T pilus system similar to the *virB*-encoded type IV system of *Agrobacterium* [see (6)], and one strikingly similar to the *Caulobacter crescentus* pilus (17), encoded by two sets of homologous genes (*pilA/cpa*) located on the chromosome and pSymA. *Sinorhizobium meliloti* lacks a type III secretion system, unlike *Rhizobium* NGR234 (4), *M. loti* (3), and *B. japonicum* (5). Therefore, use of type III secretion systems to infect plant cells is not a universal strategy among rhizobia and instead may play a role in host-specificity (18).

*Sinorhizobium meliloti* surface polysaccharides, including exopolysaccharides (EPSs), lipopolysaccharides (LPSs), capsular polysaccharides (CPSs), and cyclic  $\beta$ -glucans, encoded mainly by the chromosome and pSymB, are crucial for successful plant infection, possibly by suppressing plant defense responses (19). As many as 12% of pSymB genes may be involved in polysaccharide biosynthesis. We identified two new loci on the chromosome (7) and nine on pSymB (8). It will be interesting to find out the roles that these surface modification genes play in the interaction between the bacterium, the plants with which it comes in contact, and different soil environments.

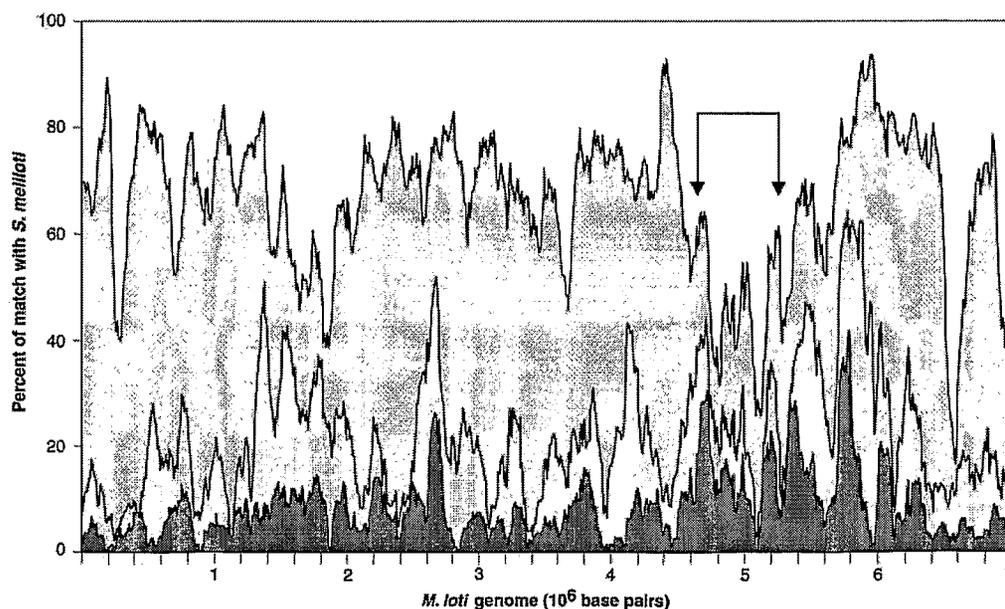
**Nodulation.** In *S. meliloti*, nodulation genes required for the synthesis and export of Nod factors are located on pSymA. Our analysis sheds new light on the possible

origin of these genes in *S. meliloti*.

We found two highly conserved duplications of *nod* genes in the genome (Fig. 1). *nodM* is 99% identical at the nucleotide sequence level to *gImS*, encoding D-glucosamine synthetase, which suggests that *nodM* emerged recently from duplication of the housekeeping chromosomal *gImS*. Each megaplasmid carries a copy of *nodPQ*, which is 99% conserved at the nucleotide level (9) and is involved in the activation of sulfate to 3'-phosphoadenosine 5'-phosphosulfate for sulfation of Nod factors in *S. meliloti*. Vestiges of an IS element next to the pSymA copy of *nodPQ* (20) suggest that this copy arose from transposition of an ancestral pSymB copy. In addition to these duplications, we discovered that *nodG* is a paralog of the housekeeping chromosomal *fabG*. Overall, sequence analysis suggests that the *S. meliloti nod* genes have two distinct origins: horizontal gene transfer, mediated by import of pSymA from an unknown bacterium, and resident gene duplication.

**Nitrogen fixation and nitrogen metabolism.** Nitrogen metabolism is a prominent feature encoded by the *S. meliloti* genome, particularly pSymA. Whereas nitrogenase synthesis and activity require up to 20 *nif* genes in *Klebsiella pneumoniae*, only nine *nif* genes are found in the *S. meliloti* genome (*nifA*, *nifB*, *nifHDKE*, *nifX*, *nifN*, and *nifS*). Except for a likely *nifS* ortholog on the chromosome and a possible *nifV* gene (SMc02546), all of these genes are located on pSymA. Although *nifQ*, *nifZ*, and *nifW* are found in *Rhizobium* sp. NGR234 (4) and *M. loti* (3), no homologs were found in *S. meliloti*.

**Fig. 3.** Comparison of *M. loti* and *S. meliloti* predicted proteins. The *M. loti* genome from bp 1 to 7 Mb is distributed along the x axis. In any given window along the x axis, the proportion within that window that has a significant match [see (27)] in the *S. meliloti* genome is displayed, and the color indicates the location of the match: blue for pSymA, yellow for pSymB, and green for the chromosome. White represents the proportion that has no global match to *S. meliloti*. Arrows indicate the *M. loti* symbiotic island.



## RESEARCH ARTICLES

Besides nitrogen fixation genes, pSymA carries glutamate dehydrogenase (*gdhA*), a full subset of genes necessary for denitrification (*nos*, *nor*, and *nap*), and nitrate transport genes. The chromosome bears the known *ntrBC*, *glnB*, *glnA*, and *glnT* genes; an alanine dehydrogenase (*ald*); the ammonium transporter *amtB*; the regulatory proteins *ntrXY*, *glnE*, *glnK*, and *glnD*; the GOGAT glutamate synthase system *gltBD*; and three previously unknown glutamine synthetase homologs. pSymB encodes a nitrate reductase (*narB*), two nitrate transporters (*nrtA*, SMb20436), and a single glutamine synthetase *glnI*.

**Energy metabolism in relation to symbiosis.** *Sinorhizobium meliloti* is an aerobic bacterium that must generate high levels of energy to support nitrogen fixation in the low-oxygen environment of the nodule. A previously characterized cytochrome c oxidase of the *cbb<sub>3</sub>* type with high affinity for oxygen is encoded by two sets of duplicated *fixNOQP* genes on pSymA. Analysis revealed an additional, less-conserved copy of the *fixNOQP* cluster. Both pSymA and the chromosome carry a large NADH-ubiquinone dehydrogenase gene cluster that may enhance energy metabolism in symbiosis, possibly along with the *fixNOQP*-encoded *cbb<sub>3</sub>* oxidase. pSymA also encodes two formate dehydrogenases (6).

**Comparison of the *S. meliloti* genome to other rhizobial genomes.** We compared (21) the predicted protein content of the *S. meliloti* genome with that of the recently sequenced *M. loti* genome (Fig. 3). Several conclusions emerged from this comparison: (i) Thirty-five percent of *M. loti* genes have no ortholog in *S. meliloti*; (ii) the genetic information carried by pSymA or pSymB in *S. meliloti* is dispersed in the *M. loti* genome; (iii) the *M. loti* MAFF303099 symbiotic island contains, besides nodulation and nitrogen fixation genes, genes that have no ortholog in the *S. meliloti* genome. Similarly, a high proportion (54%) of the 536-kb *Rhizobium* sp. NGR234 symbiotic plasmid genes have no ortholog in *S. meliloti* and those which do are distributed over the three *S. meliloti* replicons [see figure on Web site (2)]. Altogether these observations indicate that rhizobia, despite their taxonomic relatedness and symbiotic habit, differ significantly in gene content and organization. It is not known whether different isolates of a particular species will likewise show a high degree of genetic diversity. Further work will be needed to deter-

mine whether conserved and varying genes relate to adaptations for particular plant rhizospheres, for other environmental conditions, or for other adaptations not yet defined.

### Conclusion and Perspectives

Determination of the *S. meliloti* 1021 genome sequence shows that it has a composite architecture, consisting of three replicons with distinctive structural and functional features. We interpret this as a consequence of its recent emergence. Both structural and gene function analyses are consistent with the hypothesis that the two megaplasmids were acquired separately by an ancestor whose genome consisted of a single chromosome. pSymA was acquired more recently, in evolutionary terms, as indicated by its distinctive GC% and codon usage, its paucity of *Rhizobium*-specific intergenic mosaic elements (RIMEs) and ABC elements, and the specificity of its IS content. pSymB acquisition probably preceded that of pSymA or may have resulted from a chromosomal excision event. However, distinct features of pSymB, including gene specialization, low abundance of IS elements, and a high proportion of orphan genes (Table 1), argue against a chromosomal origin for pSymB.

It is tempting to speculate how acquisition of the megaplasmids by the ancestral rhizobium widened its metabolic capacities and environmental adaptability. The chromosome of *S. meliloti* is that of a typical aerobic, heterotrophic bacterium. Acquisition of pSymB considerably extended the metabolic capabilities of the microbe by allowing it to metabolize a large variety of small compounds encountered in the soil or in the plant rhizosphere. An increased capacity in synthesizing polysaccharides may also have significantly improved the colonization potential of these microbes. Finally, acquisition of pSymA led to the emergence of nodulation, as well as the bacterium's capacity to colonize the low-oxygen environment of the nodule. pSymA also expanded the capacity to metabolize nitrogen compounds under a variety of chemical forms, including molecular dinitrogen. Such speculation may offer new perspectives for microbial evolution and for identifying the origins of the rhizobium-legume symbiosis. The complete *S. meliloti* genome sequence and its detailed annotation creates opportunities for an expanded analysis of symbiotic nitrogen fixation by allowing researchers to focus on specific metabolic

and regulatory circuits. Functional analyses of the *S. meliloti* genome will lead to further insights in understanding this and other rhizobium-legume symbioses.

### References and Notes

- Sequence accession numbers: chromosome AL591688 (EMBL, CON entry); pSymA accession: AE006469 (GenBank); pSymB accession number AL591985 (EMBL).
- Annotation, tools, and supplementary data are available at <http://sequence.toulouse.inra.fr/meliloti.html>
- T. Kaneko *et al.*, *DNA Res.* **7**, 331 (2000).
- C. Freiberg *et al.*, *Nature* **387**, 394 (1997).
- M. Gottfert *et al.*, *J. Bacteriol.* **183**, 1405 (2001).
- M. J. Barnett *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
- D. Capela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
- T. Finan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
- J. S. Schwedock, S. R. Long, *Genetics* **132**, 899 (1992).
- M. H. Renalier *et al.*, *J. Bacteriol.* **169**, 2239 (1987).
- M. Flores *et al.*, *J. Bacteriol.* **169**, 5782 (1987).
- Genes are considered paralogs if their products match over more than 80% of their length with more than 30% amino acid sequence identity.
- V. Viprey, A. Rosenthal, W. J. Broughton, X. Perret, *Genome Biol.* **1**, 1 (2000).
- I. J. Oresnik, S. L. Liu, C. K. Yost, M. F. Hynes, *J. Bacteriol.* **182**, 3582 (2000).
- W. C. Nierman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4136 (2001).
- A. G. Matthysse, H. A. Yarnall, N. Young, *J. Bacteriol.* **178**, 5302 (1996).
- J. M. Skerker, L. Shapiro, *EMBO J.* **19**, 3223 (2000).
- X. Perret, C. Staehelin, W. J. Broughton, *Microbiol. Mol. Biol. Rev.* **64**, 180 (2000).
- V. Viprey, X. Perret, W. J. Broughton, *Subcell. Biochem.* **33**, 437 (2000).
- J. S. Schwedock, S. R. Long, *Mol. Plant. Microbe Interact.* **11**, 1119 (1994).
- BlastP comparisons of the *M. loti* and *S. meliloti* proteins were run using the following parameters: expect value threshold = 0.1; gap opening penalty = 9; gap extension penalty = 2. Genes are considered possible orthologs if they match over 80% of either the query or subject protein length with an expect of less than 1e-6.
- We are grateful to Thomas Schiex [Biométrie et Intelligence artificielle, Institut National de la Recherche Agronomique (INRA), Toulouse, France] for the Frame D program. D.C. was supported by a CNRS (BDI) doctoral fellowship and E.K. by an INRA postdoctoral fellowship. We also thank Lion Bioscience AG (Heidelberg, Germany) and IIT Biotech GmbH (Bielefeld, Germany) for their participation. This work has been supported by grants from the European Union (MELILO BIO4-CT98-0109) to F.G., A.P., A.G., D.P., T.P., and J.B., the Natural Sciences and Engineering Research Council of Canada (NSERC) Research, Strategic and Genomics to T.M.F. and B.G., the Bundesministerium für Forschung und Technologie (0311752) to A.P., the CNRS (Genome Research Program) and INRA (AIP98/P00206) to F.G. and J.B., the U.S. Department of Energy, Energy Biosciences Program (DE-FG03-90ER20010) to S.R.L. and (DE-FG03-99ER20225) to M.K.; and the U.S. NIH (GM30962) and the Howard Hughes Medical Institute to S.R.L.

30 May 2001; accepted 5 June 2001