

A high-density physical map of *Sinorhizobium meliloti* 1021 chromosome derived from bacterial artificial chromosome library

(genome mapping/*Rhizobium*/symbiosis)

DELPHINE CAPELA*†‡, FRÉDÉRIQUE BARLOY-HUBLER*†, MARIE-THÉRÈSE GATIUS*, JÉRÔME GOUZY‡, AND FRANCIS GALIBERT*§

*Laboratoire Recombinaisons Génétiques, Centre National de la Recherche Scientifique–UPR41, 2 Avenue du Pr Léon Bernard, 35043 Rennes Cedex, France; and †Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, Institut National de la Recherche Agronomique–Centre National de la Recherche Scientifique UMR215, Chemin de Borderouge, BP27, 31326 Castanet-Tolosan, France

Communicated by Roland Douce, University of Grenoble, Grenoble, France, May 27, 1999 (received for review March 10, 1999)

ABSTRACT As part of the European *Sinorhizobium meliloti* (strain 1021) chromosome sequencing project, four genomic bacterial artificial chromosome (BAC) libraries have been constructed, one of which was mainly used for chromosome mapping. This library consists of 1,824 clones with an average insert size of 80 kilobases and represents approximately 20-fold total genome coverage [6.8 megabases (Mbs)]. PCR screening of 384 BAC clones with 447 chromosomal markers (PCR primer pairs), consisting of 73 markers representing 118 genes (40 individual genes and 78 genes clustered in 23 operons), two markers from the *rrn* operon (three loci), four markers from insertion sequences (≈ 16 loci) and 368 sequence-tagged sites allowed the identification of 252 chromosomal BAC clones and the construction of a high-density physical map of the whole 3.7-Mb chromosome of *S. meliloti*. An average of 5.5 overlapping and colinear BAC clones per marker, correlated with a low rate of deleted or rearranged clones (0.8%) indicate a solid BAC contiguation and a correct mapping. Systematic BLASTX analysis of sequence-tagged site marker sequences allowed prediction of a biological function for a number of putative ORFs. Results are available at <http://www-recomgen.univ-rennes1.fr/meliloti>. This map, whose resolution averages one marker every 9 kilobases, should provide a valuable tool for further sequencing, functional analysis, and positional cloning.

Rhizobium meliloti, recently renamed *Sinorhizobium meliloti* (1), is a common Gram-negative soil and rhizosphere bacterium, considered as an agriculturally important nitrogen fixer. These bacteria induce nodule formation on the roots of a set of *Medicago* species, including alfalfa (*Medicago sativa*), inside which they fix atmospheric nitrogen. The formation and colonization of alfalfa root nodules by *S. meliloti* result from a series of complex interactions controlled by the exchange of molecular signals between bacteria and host plant and require spatially and temporally regulated expression of specific genes from the two symbiotic partners (2–5).

S. meliloti is being widely studied around the world because this fast-growing *Rhizobium* is amenable to genetics (6–10). Interest in *S. meliloti* has been recently fostered by the choice of *S. meliloti*–*Medicago truncatula* as a model system for the study of the *Rhizobium*–legume symbiosis by a large number of groups worldwide.

The *S. meliloti* strain 1021 genome consists of three megabase (Mb)-sized replicons: one chromosome of 3.7 Mb and two megaplasmids of 1.7 Mb and 1.4 Mb (11). Many genes involved in symbiosis are located on megaplasmids, the so-called symbiotic plasmids (pSyms) (12, 13). Attention has focused on the

two replicons, and little is known about the genetic information carried by the chromosome. Although several symbiotic genes are known to be located on the chromosome (14, 15), no detailed genetic map is available. Only low-density maps of this replicon have been obtained thus far, by using the conjugal plasmids RP4 [for *S. meliloti* strain 2011 (16)] or R68.45 [for *S. meliloti* strains 41 (17) and GR4 (18)], the Tn5-Mob conjugational transfer method and cotransduction by bacteriophage Φ M12 (15, 19), or physical restriction maps (20). To obtain a high-density chromosome map to be directly used in functional analysis, and in relation with the European chromosome sequencing project, we have created four bacterial artificial chromosome (BAC) libraries of *S. meliloti* whole genome or purified chromosome, in the pBeloBAC11 (H. Shizuya, unpublished data) and pBACE3.6 (P. De Jong, unpublished data) vectors. These BAC libraries, with genomic DNA inserts of 80 kilobases (kb) on the average, were successfully used in mapping and now provide a tool for further systematic sequencing. This paper describes the construction of a high-resolution map of the whole *S. meliloti* chromosome, on which 447 markers representing 368 sequence-tagged sites (STSs), 118 genes, *rrn* operon, and 4 insertion sequences, have been positioned by screening and ordering BAC clones. BLASTX comparisons (21) of anonymous markers from STSs allowed identification of putative ORFs with predictable biological functions.

MATERIALS AND METHODS

Strains, Plasmids, and Media. *S. meliloti* strain 1021 (Str^t, derivative of strain 2011) was provided by S. R. Long (Department of Biological Sciences, Stanford University, Stanford, CA). *Escherichia coli* strains DH10B (GIBCO/BRL) and HB101 (Promega) were used as host strains for BAC and pUC cloning, respectively. Vectors used were pBeloBAC11 (H. Shizuya, unpublished data), pBACE3.6 (P. De Jong, unpublished data), and pUC18 (Amersham Pharmacia). Cultures of *S. meliloti* were grown in LB medium at 28°C. BAC recombinant clones were selected on LB medium containing 12.5 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol with either 50 $\mu\text{g}\cdot\text{ml}^{-1}$ 5-bromo-4-chloro-3-indolyl β -D-galactoside (X-Gal) and 25 $\mu\text{g}\cdot\text{ml}^{-1}$ isopropyl β -D-thiogalactoside (IPTG) for pBeloBAC11 or 5% sucrose for pBACE3.6. pUC recombinants were selected on LB with 50 $\mu\text{g}\cdot\text{ml}^{-1}$ ampicillin, 50 $\mu\text{g}\cdot\text{ml}^{-1}$ X-Gal, and 25 $\mu\text{g}\cdot\text{ml}^{-1}$ IPTG.

Preparation of High-Molecular-Weight DNA. *S. meliloti* strain 1021 was grown to mid-exponential phase in LB at 28°C and embedded in low melting agarose plugs as described (22).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: MB, megabase; BAC, bacterial artificial chromosome. †D.C. and F.B.-H. contributed equally to this work.

§To whom reprint requests should be addressed. E-mail: francis.galibert@univ-rennes1.fr.

The plugs were stored in 0.5 M EDTA/1% *N*-lauroyl sarcosine at 4°C.

BAC Vector Preparation. pBeloBAC11 vector was prepared as described by Woo *et al.* (23). pBACE3.6 was digested by *EcoRI* and migrated on an agarose gel to remove the pUC-link fragment before cloning.

Construction of Total Genomic BAC Libraries. Three total genomic BAC libraries were constructed as described by Brosch *et al.* (22) by using 10 µg of total genomic DNA embedded in agarose plugs partially digested with either *Sau3A* (5 units, 30 min, 37°C) or *HindIII* (10 units, 30 min, 37°C) for cloning in pBeloBAC11 or with *EcoRI* (10 units, 30 min, 37°C) for cloning in pBACE3.6. Restriction fragments between 100 and 150 kb were ligated to digested and dephosphorylated BAC vector and electroporated into *E. coli* DH10B cells (24).

Construction of Chromosomal BAC Library. Chromosomal DNA was purified by using pulse-field gel electrophoresis on a CHEF DRIII apparatus set (constant pulse of 500 sec, run of 48 h, field angle of 106°, 3 V·cm⁻¹, 1× TAE, 14°C) and excised from the gel before partial digestion with 10 units of *HindIII*. DNA fragments between 120 and 150 kb were ligated to pBeloBAC11 and electroporated into *E. coli* DH10B cells.

BAC DNA Preparation for Sequencing and Insert-Size Determination. Each BAC clone was inoculated in 50-ml Falcon polypropylene tubes containing 5 ml of LB medium supplemented with chloramphenicol (12.5 µg·ml⁻¹). BAC DNA extraction was carried out by using the alkaline lysis procedure with vigorous pipetting or vortexing during the lysis and neutralizing steps (P2 and P3 buffers) to release a larger amount of DNA (≈100–300 ng·µl⁻¹). DNA in P3 supernatant was precipitated in isopropanol, and the pellet was resuspended in 40 µl of water. For BAC end sequencing, P3 supernatant DNA was sheared through a 21GX1 needle and purified through a QiaWell column of the Qiagen (Chatsworth, CA) Miniprep kit (QiaWell 8 plasmid kit, Qiagen).

BAC End Sequencing. BAC ends were sequenced using BigDye Terminators (Perkin-Elmer Applied Biosystems) with an ABI377 automated sequencer as follows: 1 µg of DNA, 8 µl of BigDye terminator mix, 50 pmol of T7 or Sp6 primers in a total reaction volume of 25 µl. This mix was denatured at 95°C for 5 min followed by 50 thermal cycles (95°C for 45 sec, 55°C for 30 sec, 60°C for 4 min). Excess BigDye terminator was removed by using Centriflex gel filtration cartridges (Edge Biosystems, Gaithersburg, MD).

Sizing of Inserts. Insert size was determined by field inverted gel electrophoresis (FIGE) after *DraI* digestion (New England Biolabs) as described by Brosch *et al.* (22).

Chromosomal pUC Library Construction. Chromosomal DNA was separated twice from megaplasmids using the pulsed-field gel electrophoresis conditions described above. Agarose block was excised from the gel, and DNA was hydrolyzed by using 10 units of *Sau3A* restriction enzyme and 100 mg·ml⁻¹ BSA (2 hr, 37°C). Restriction fragments between 1 and 3 kb were purified with the Sephaglas Bandprep Pharmacia Kit and ligated to *Bam*HI-digested and dephosphorylated pUC18 vector (Amersham Pharmacia). After electroporation into competent *E. coli* HB101 cells, recombinant pUC clones were plated onto selective LB medium. PCR amplification of inserts with the two universal primers PU and PR enabled the selection of clones with suitable size. Another pUC library was constructed from specific chromosomal restriction fragments. Each genomic DNA plug was digested with 20 units of *PmeI* or *I-CeuI* (4 hr, 37°C), and restriction fragments were separated by using pulsed-field gel electrophoresis on a CHEF mapper Bio-Rad apparatus (ramp pulse from 34.5 sec to 181 sec, run of 30.5 hr, field angle of 120°, 1× TAE, 14°C). The 520-kb *I-CeuI*, 390-kb *I-CeuI*, and 1,040-kb *PmeI* restriction fragments were excised from the gel and hydrolyzed with 20 units of *Sau3A* (2 hr, 37°C). The resulting

DNA fragments, between 1 and 3 kb, were purified and cloned in the pUC18 vector as described above.

Marker Design and STS BLASTX Comparisons. Three sorts of markers were used in this study: (i) *S. meliloti* gene sequences available in databases [GenBank (25) and European Molecular Biology Laboratory (26)]; (ii) pUC clones directly PCR-amplified with the two universal primers PU and PR and end-sequenced with the PU primer to generate suitable STSs; and (iii) BAC end sequences obtained with the T7 and Sp6 primers. For all marker sequences, several primers were designed to amplify fragments from 150 to 400 bp. Each STS (pUC and BAC end sequences) was systematically analyzed by BLASTX (21) comparison against nonredundant protein database from National Center for Biotechnology Information. The results are available at <http://www-recomgen.univ-rennes1.fr/meliloti>.

BAC DNA Pool Preparation and PCR Screening. Microtiter plates (4 × 96-well; 384 clones) of the *Sau3A* total genomic library were organized in 16 × 24 arrays. Individually grown bacteria were pooled in rows (x) and columns (y): 16 x pools of 24 clones and 24 y pools of 16 clones were obtained. Each pool was centrifuged (2,000 × g, 10 min) and the pellet was resuspended in 2 ml of 10 mM Tris. Working solutions were obtained by 10-fold dilutions of stock solutions in 10 mM Tris. As an initial step, 5 µl of x and y bacterial pools were tested by PCR (95°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, 30 cycles) for the presence of markers and loaded onto 3% agarose gels. Positive pools were used to determine a set of addresses corresponding to potential clones, which were subsequently validated by a second PCR analysis on individual clones.

Analysis of Positive BACs and Contig Formation. BAC contiguation was performed with the SAM version 2.5 software (C.A. Soderlund, The Sanger Centre, Cambridge, U.K.). SAM takes as input a file of BAC clones along with a set of markers for each and produces maps with one or several plausible marker orders and the alignment of clones.

RESULTS

BAC Libraries. Four different BAC libraries were constructed. Three of these have been used in this study. The library mainly used corresponds to the total genomic *Sau3A* library and consists of 1,824 clones. The two other libraries, *HindIII* chromosomal and *EcoRI* total genomic libraries, containing 1,920 and 1,440 clones respectively, were used for contig gap closure and consolidation of underrepresented regions. Size determination of 154 clones from the *Sau3A* library by pulsed-field gel electrophoresis after *DraI* cleavage indicated an average insert size of 80 kb (minimum 22 kb, maximum 150 kb). Consequently, 10-fold coverage of the whole genome requires ≈850 clones.

Marker Description. Of a set of ≈600 markers tested, 447 were mapped on the chromosome. These chromosomal markers are distributed as follows: 40 primer pairs from *S. meliloti* individual genes, 32 primer pairs designed from 78 *S. meliloti* genes belonging to 23 different operons, 6 primer pairs from repetitive sequences (two from *rrn* operon and four insertion sequences), 201 primer pairs from pUC-end sequences, and 167 primer pairs from BAC insert terminus sequences. Thirty-four additional pUC and BAC end sequences were positioned on the map by aligning sequences, but no primers were designed for them (as indicated by the symbol NA on the web site). All marker information, such as sequences, PCR primers, and chromosomal location is available on the Web site.

BAC Contig Formation. A total of 384 clones from the *Sau3A* library corresponding to ≈4.5-fold genome coverage were tested by a two-step PCR method (see Materials and Methods). Additional PCR screening were selectively performed on other BAC clones from the *Sau3A*, *HindIII*, or *EcoRI* libraries for contig gap closure and confirmation of

poorly represented regions. By using this procedure, a BAC contig of 252 clones covering the entire *S. meliloti* chromosome was constructed. This chromosomal contig could be represented by a minimal set of 48 BAC clones, with total overlap estimated to 8% of the chromosome after size determination of the BAC inserts (BAC insert sizes are available on the Web site under each BAC entry). A linear representation of the resulting physical map of the chromosome with alignment of the minimal set of clones and all marker positions is given in Fig. 1.

Establishment of the *S. meliloti* Gene Map. Mapping 447 markers on the *S. meliloti* 1021 chromosome resulted in the positioning of 118 genes (40 individual genes and 78 clustered

in 23 operons) (see Fig. 1 for localization and Web page for refs.). Only 44 of these genes had been roughly mapped on the chromosome (15, 20, 27–30), and 18 were previously assigned to this replicon but were not mapped (31–36) (Table 1). This study also allowed a new location for another 63 unassigned genes described in databases. No major contradictions with previous reports were observed regarding the relative orders of or the distances between markers. For example, we confirm (i) the Glazebrook (15) *ndvB-exoS-exoD-ntrA-pho-degP-exoR* order as well as the distance between *ntrA* and *exoD* or between *exoS* and *ndvB* (estimated to ≈ 90 and 100 kb, respectively); (ii) the Honeycutt map (20), where the *glx-ndvA-ntrA* order and the distance between *ndvA* and *ntrA* or *glx* and *ndvA* (≈ 500

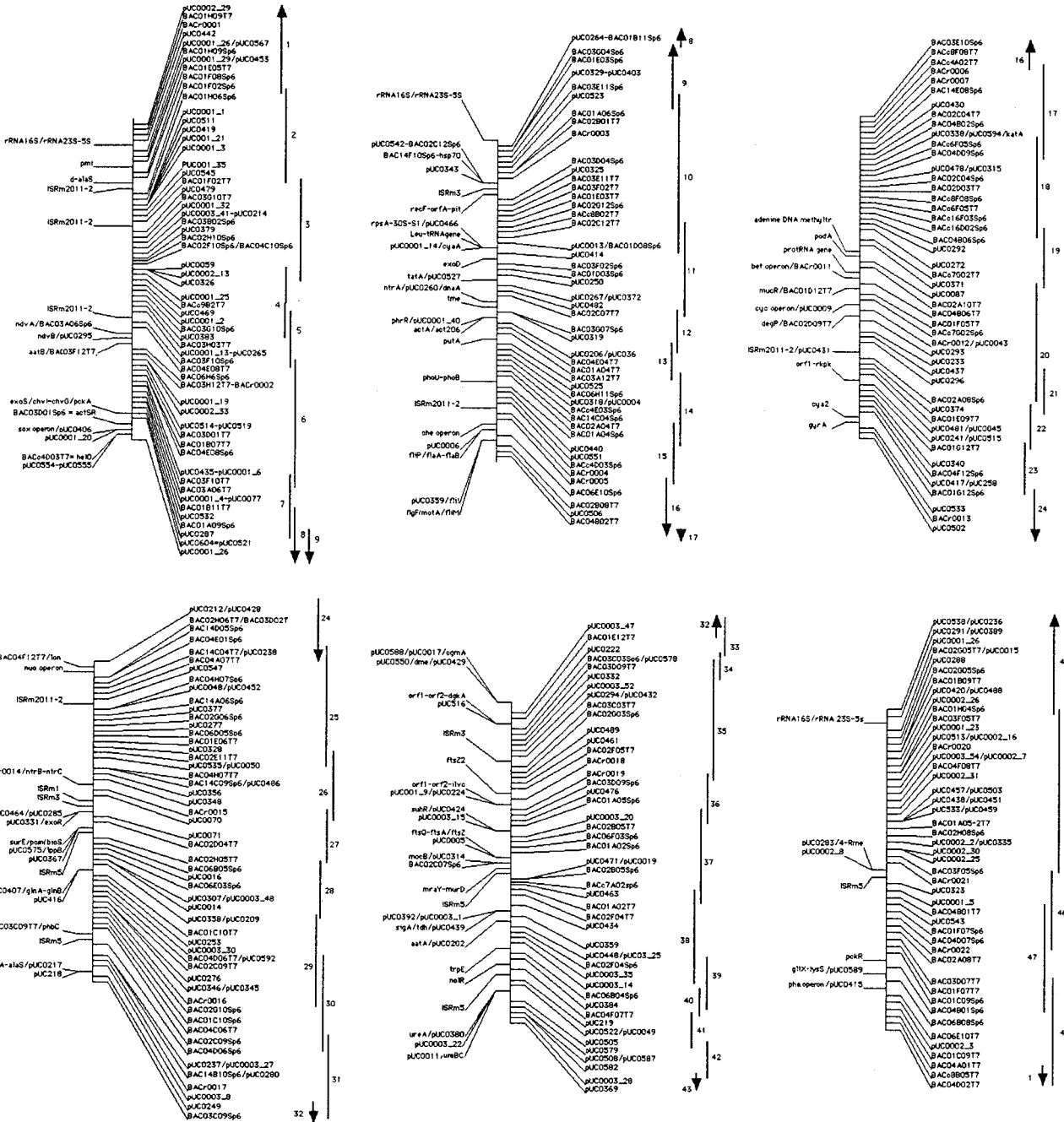


FIG. 1. High-density map of the *S. meliloti* 1021 chromosome, represented for convenience in six linear contiguous parts of ≈ 600 kb. *Left* presents the gene marker positions, whereas the *Right* shows the location of STS markers. Markers with aligned sequences are positioned on the same line. Spacing between each marker is estimated to ≈ 9 kb. On the *Right* side, the 48 BAC clones forming the minimal set of overlapping clones needed to cover the whole chromosome are shown. For convenience, genes belonging to the same operon are designated by their generic name followed by operon (for example *cyc operon* represents *cycHJKL*).

Table 1. Database genes mapped on the *S. meliloti* chromosome

Gene	Description	Ref.
<i>aatA</i> and <i>aatB</i>	Aspartate aminotransferase A and B	
<i>actA-phrR-act206</i> [†]	Genes required for acid tolerance	53
<i>actS-actR</i>	Two-component regulatory system involved in acid tolerance	
<i>adenine DNA methyltransferase betICBA</i> [†]	Glycine betaine synthesis	31
<i>cgmA</i>	Unknown	
<i>cheYLAWRBY2-orf1-orf2-orf9</i> *	Chemotaxis operon	29
<i>chvI-chvG</i> *	Two-component regulatory system	27
<i>cyaA</i> and <i>cya2</i>	Adenylcyclase A and 2	
<i>cycHJKL</i> *	Cytochrome <i>c</i> -type biogenesis operon	54
<i>degP</i> *	Protease	15
<i>delta-ALA synthetase</i>	δ -aminolevulinic acid synthase	
<i>dgkA</i>	Diacylglycerol kinase	
<i>dme</i> *	NAD-malice enzyme	30
<i>dnaA</i>	Chromosomal replication initiator protein	
<i>exoD</i> *	Nodule invasion protein ExoD	15
<i>exoR</i> *	Probable transcriptional regulator	15
<i>exoS</i> *	Histidine kinase sensory protein ExoS	15
<i>flaA-flaB</i> *	Flagellin genes	29
<i>fliI-flgF-motA-fliM</i> *	Flagellae operon	
<i>fliP</i> *	Protein involved in production of functional flagella	28
<i>ftsQ-ftsA-ftsZ</i> *	Cell division proteins	55
<i>ftsZ2</i> *	Cell division protein	56
<i>glnA-glnB</i>	Glutamine synthetase I; PII regulatory protein	
<i>gltX-lysS</i> *	Glutamyl-tRNA synthetase; lysyl-tRNA synthetase	20
<i>gyrA</i>	DNA gyrase subunit A	
<i>helO</i>	Putative helicase	
<i>hsp70</i>	Heat shock protein	
<i>ilvC</i> [†]	Acetohydroxy acid isomeroreductase	57
<i>katA</i> [†]	Catalase	32
<i>lon</i>	Protease	
<i>leu-tRNA gene</i>	Transfer RNA-leucine	
<i>mucR</i>	Transcriptional regulatory protein	
<i>mraY-murD</i>	UDP- <i>N</i> -acetylmuramoyl-L-alanyl-D-glutamate synthetase	
<i>ndvA</i> *	Protein involved in the production of β -(1-2)-glucan	20
<i>ndvB</i> *	Protein involved in the production of β -(1-2)-glucan	15
<i>nolR</i> [†]	Nodulation protein NolR	36
<i>ntrA</i> *	RNA polymerase σ -54 factor	15, 20
<i>ntrB-ntrC</i> [†]	Two-component regulatory proteins	34
<i>nuoABC</i>	NADH-ubiquinone oxidoreductase	
<i>pckA</i> *	Phosphoenolpyruvate carboxykinase	27
<i>pckR</i>	Putative regulator of <i>pckA</i> expression	
<i>phaABCDEFGF</i>	K ⁺ efflux system involved in pH adaptation	
<i>phbC</i>	Poly- β -hydroxybutyrate synthase	
<i>phoU-phoB</i> *	Phosphate regulatory proteins	15
<i>pmi</i>	Phosphoisomerase	
<i>podA</i> *	Pyruvate orthophosphate dikinase	33
<i>pro-tRNA</i> [†]	Transfer RNA-proline	58
<i>putA</i> [†]	Proline dehydrogenase	37
<i>recA-alaS</i>	RecA protein; alanyl-tRNA synthetase	
<i>recF-orfA-pit</i> *	Phosphate transport protein	59
<i>rkpK-orfI</i> [†]	UDP-glucose dehydrogenase; UDP-glucuronic acid epimerase	60
<i>rpsA</i> [†]	30S ribosomal protein S1	61
<i>sigA-tdh</i> [†]	σ -factor; threonine dehydrogenase	35
<i>soxADB-glxABCD-glnT</i>	Sarcosine oxidase; glutamine synthetase III operon	
<i>surE-pcm-bioS-lppB</i>	Possible survival operon	
<i>suhR</i>	<i>rpoH</i> suppressor	
<i>tatA</i>	Tyrosine aminotransferase	
<i>tme</i> *	NADP-dependent malic enzyme	30
<i>trpE</i>	Anthranilate synthase	
<i>ureABC</i>	Urease	
<i>4_Rme</i>	Putative σ -54 dependent transcriptional activator	

*Genes previously mapped on the chromosome (located) as described in the respective references.

[†]Genes previously assigned to the chromosome (not located) as described in the respective references.

kb each) are identical here; and (iii) the *pckA-ntrA-tme-podA-dme* positions in the Osteras map (37). The location of the chemotaxis region linked to the *his-39* marker (19) as de-

scribed by Sourjik *et al.* (38) is also in agreement with our results, even if we found a slight discrepancy in this 45-kb chromosomal region. Actually, the tiling path and validity of

our BAC contig enabled us to determine without ambiguity the order *che-fljP-flaA-flaB-fljI-motA-fljM* instead of *che-fljM-motA-fljI-fljP-flaA-flaB*. This may reflect an effective local difference between the two strains used in these studies (strains 1021 and 41).

Validity of the Map. The 252 BAC clones represent a total of ≈ 20 Mb, i.e., 5.5-fold chromosome coverage. No chromosomal region is represented by a unique clone, and 10 very short regions (6% of markers) are covered by two BAC clones only. Statistical analysis showed an average of 5.5 colinear clones per marker and an average spacing between marker of nine kb, assuming a uniform distribution of markers and considering operon structures or aligned sequences as a unique locus. In view of this coverage rate, the resolution of the map and the colinear distribution of the markers (only three BAC clones were found to be rearranged), the contig formation can be considered as valid even though local rearrangements (< 9 kb) cannot be excluded.

Localization of *S. meliloti* Repetitive Elements. The *Sau3A* BAC library was also analyzed for repetitive sequence content such as ribosomal RNA (*rnm*) operons and insertion sequences (IS). Concerning *rnm* operons (16S-23S-5S), we found three loci in *S. meliloti* chromosome, each containing a *I-CeuI* restriction site as described (20). We have also tested the presence of insertion sequence elements *ISRm1*, *ISRm2011-2*, *ISRm3* (*ISRm3-2*), and *ISRm5* to determine the extent of their distribution and to point out putative rearrangement, deletion, or duplication sites. The interest of their location also lies in their frequent association with nodulation and nitrogen-fixation genes. At least two copies of *ISRm1* (39) were found, one only of which is on the chromosome near the *ntrB-ntrC* operon on BAC27. *ISRm2011-2*, the only member of the IS630-Tc1/IS3 retroposon superfamily found in *S. meliloti* (40, 41), was more widespread within the genome, with at least seven copies on the chromosome (BAC01, BAC03, BAC05, BAC18, BAC21, BAC22, and BAC25). *ISRm3* (*ISRm3-2*) (42) was found in eight to nine copies on the genome, which is in keeping with previous studies, and we were able to identify three of these loci on the chromosome: near *hsp70-recF* site (BAC10), between *ISRm1* and *exoR* (BAC27), and between *ilvC* and *ftsZ2* (BAC35). Finally, *ISRm5* (43) was also found in 8–10 copies with five chromosomal locations: in *surE-pcm-bioS-lppB* operon (BAC27), near *phbC* (BAC30), *mraY* (BAC37), *nolR* (BAC40), and *4-Rme* genes (BAC46) (Fig. 1).

STS Marker Identification. The sequence of 201 pUC inserts and 167 BAC ends ($\approx 160,000$ bases) was compared with the complete nonredundant protein database from National Center for Biotechnology Information using the BLASTX program (21). Significant nucleotide similarities ($P < 1e^{-10}$) with known sequences were found for 54% of the query sequences, which could reflect the high proportion of putative coding region in the chromosome. Among those, we noticed 21% of sequences with strong similarities to *Rhizobiaceae* family proteins, in the following proportions: 34 sequences from *S. meliloti*, 13 from *Bradyrhizobium*, 1 from *Azorhizobium*, 12 from *Rhizobium* sp., and 17 from *Agrobacterium*. In addition, most of the sequences have similarities ($P < 1e^{-4}$) with bacterial proteins: 10% of *Bacillaceae*, 7% of *Pseudomonas* and *Sphingomonas*, 20% of *Cocci*, 15% of *Rhizobacter* and *Rhodobacter*, 5% with cyanobacteria (*Synechocystis* sp.), 7% of pathogenic bacteria (*Ralstonia*, *Yersinia*), and 6% of *Mycobacterium*. In parallel, several categories could readily be discerned: chemotaxis mechanism (*mcpA* from *Agrobacterium*, *Y4FA* from *Rhizobium* sp. *NGR234*), genes belonging to regulatory systems (*regR-regS* and *ragA* from *B. japonicum*, *rfpB* from *Xanthomonas campestris*, *frzZ* from *Mycobacterium xanthus* or *regX* from *R. sphaeroides*) and also a high number of ABC-transporter systems (*Y4OS*, *Y4TQ*, *Y4WD*, *Y4MO* from *Rhizobium* sp. *NGR234*, *YHES* from *E. coli*). Moreover, we observed putative *S. meliloti* multicopy genes located both on

the chromosome and on one of the megaplasmids. These include a third copy of *fixT* (44), a second copy of *glnII* (GSII) for which we also found a pSymb copy, and a second *syrm-like* (45) copy, with a weaker sequence similarity. BLASTX results are available at the Web site under each BAC entry. For each STS marker (pUC or BAC ends), the two best hits, if any, are given together with a synthetic description of the hit (BLAST score, expected value, percentage of identity, commentary on the hits as retrieved from the databases). Hypertext links allow access to full BLASTX results, marker DNA sequences, and database sequence entries.

DISCUSSION

Generating a high-density physical map of the *S. meliloti* chromosome served two major goals: (i) positioning enough genes and STS markers to construct a map directly useful for functional studies of this agriculturally important bacterium and (ii) preparing the European sequencing project of the chromosome, started on 1 December 1998. At the present time, 14 complete eubacterial genomes have been sequenced. Two experimental strategies have been used thus far: a direct approach, with shotgun sequencing of the whole genome, and a two-step approach implying construction of a contig of recombinant clones followed by shotgun sequencing. The latter, more traditional approach was used for sequencing the genomes of *E. coli*, *Bacillus subtilis*, and *Synechocystis* sp. (4.6, 4.2, and 3.5 Mb in size, respectively) (46–48). The former and more recent approach has been applied thus far to genomes of a much smaller size (≈ 2 Mb). Although the direct shotgun approach clearly has many advantages, we have decided to follow the two-step strategy. The reasons for this are four-fold: (i) the size of the chromosome (3.7 Mb); (ii) the difficulty to prepare chromosomal DNA in pure form and sufficient quantity; (iii) the genome organization in three replicons, which led to a worldwide sharing of the sequencing effort with S. R. Long at Stanford University that will sequence the 1.4-Mb replicon, T. Finan and B. Driscoll at Hamilton University (Ontario, Canada) associated with A. Pühler at Bielefeld University (Germany), which will sequence the 1.7-Mb replicon; and (iv) the known existence of quite a number of nucleic acid sequence repeats belonging to different families (49). The interest in using recombinant BAC clones lies in the low level of chimerism or rearrangements, in the stability of the clones, which is convenient for keeping large insert DNA at the quality level and in quantities suitable for genome analysis, and in the low copy number of cloned genes in the heterologous bacterial host, which can reduce their toxic and deleterious effects.

Statistics on clone size based on a large number of measurements indicated an average BAC insert size of 80 kb, which is less than the majority of eukaryotic BAC libraries constructed to date (50) but quite comparable to other prokaryotic libraries (22). The BAC insert size seems to depend strongly on the origin of DNA and particularly on G/C content.

Library screening was successfully performed with markers from various types of sequence (database coding genes, STS, repetitive elements) and provides substantial information (new gene localization, identification of putative ORFs). The relative order of markers is consistent with previous *S. meliloti* chromosome-mapping reports (15, 20, 27–30) as far as comparisons can be done. However, the previous genetic linkage maps (16–18, 51) are difficult to correlate with our work, because no auxotrophic markers or restriction sites were used here. However, when available, GenBank reports helped us to confirm the position of genetic loci.

Whenever possible, we tended to refine and more accurately position genetic loci in the high-resolution map by aligning marker sequences and BLASTX results comparisons. However the relative order of a low number of markers cannot be deduced from the BAC contig, which is indicated on Fig. 1 by

a unique anchorage site or on the Web site in the BAC information pages. This represents the unique limitation found for the BAC contig approach and will be resolved by total sequencing.

Markers with multiple loci such as insertion sequences were mapped at once. The presence of these mobile elements could explain the multilocation of genes such as *fixT*, *glnII*, or *syrM* and account for genetic transfers between megaplasmids and chromosome or between diverse *Rhizobium* species or *Agrobacterium-Rhizobium* as described by Deng *et al.* (52). They could also help to understand genome redundancy and the three-replicon structure and could allow explanation of closed relationships between symbiotic bacteria genomes and possible DNA exchanges between bacteria and plant, but evidence of this is still needed.

The present high-density map of *S. meliloti* (strain 1021) chromosome will certainly be of interest in comparative genomics, for instance, between the three *S. meliloti* replicons.

We thank Alain Billault and Catherine Soravito de Francesky (Centre d'Étude du Polymorphisme Humain, Paris, France) for their involvement in the construction of the BAC libraries. We are also particularly grateful to Dominique Lavenier (Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France) for computing assistance. We thank Jacques Batut (UMR215, Institut National de la Recherche Agronomique-Centre National de la Recherche Scientifique, Toulouse, France) and Jean-Claude Chuat (UPR41, Centre National de la Recherche Scientifique, Rennes) for valuable discussions and careful reading of the manuscript. This work was supported in part by Centre National de la Recherche Scientifique (genome program) and Institut National de la Recherche Agronomique grants.

- De Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrujuan, G., Neyra, M., Collins, M. D., Dreyfus, B., Kersters, K. & Gillis, M. (1994) *Int. J. Syst. Bacteriol.* **44**, 715–733.
- Denarié, J., Debéllé, F. & Rosenberg, C. (1992) *Annu. Rev. Microbiol.* **46**, 497–531.
- Denarié, J. & Cullimore, J. (1993) *Cell* **74**, 951–954.
- Long, S. R. (1996) *Plant Cell* **8**, 1885–1898.
- Fisher, R. F. & Long, S. R. (1992) *Nature (London)* **357**, 655–660.
- Glazebrook, J. & Walker, G. C. (1991) *Methods Enzymol.* **204**, 398–418.
- Long, S. R. (1989) *Annu. Rev. Genet.* **23**, 483–506.
- Martin, M. O. & Long, S. R. (1984) *J. Bacteriol.* **159**, 125–129.
- Souppène, E., Foussard, M., Boistard, P., Truchet, G. & Batut, J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3759–3763.
- Ruvkun, G. B. & Ausubel, F. M. (1981) *Nature (London)* **289**, 85–88.
- Sobral, B. W., Honeycutt, R. J., Atherly, A. G. & McClelland, M. (1991) *J. Bacteriol.* **173**, 5173–5180.
- Julliot, J. S., Dusha, I., Renalier, M. H., Terzaghi, B., Garnerone, A. M. & Boistard, P. (1984) *Mol. Gen. Genet.* **193**, 17–26.
- Batut, J., Terzaghi, B., Ghérardi, M., Huguet, M., Terzaghi, E., Garnerone, A. M., Boistard, P. & Huguet, T. (1985) *Mol. Gen. Genet.* **199**, 232–239.
- Forrai, T., Vincze, E., Banfalvi, Z., Kiss, G. B., Randhawa, G. S. & Kondorosi, A. (1983) *J. Bacteriol.* **153**, 635–643.
- Glazebrook, J., Meiri, G. & Walker, G. C. (1992) *Mol. Plant Microbe Interact.* **5**, 223–227.
- Meade, H. M. & Signer, E. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2076–2078.
- Kondorosi, A., Kiss, G. B., Forrai, T., Vincze, E. & Banfalvi, Z. (1977) *Nature (London)* **268**, 525–526.
- Casadesus, J. & Olivares, J. (1979) *Mol. Gen. Genet.* **174**, 203–209.
- Klein, S., Lohman, K., Clover, R., Walker, G. C. & Signer, E. R. (1992) *J. Bacteriol.* **174**, 324–326.
- Honeycutt, R. J., McClelland, M. & Sobral, B. W. (1993) *J. Bacteriol.* **175**, 6945–6952.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B. G. & Cole, S. T. (1998) *Infect. Immun.* **66**, 2221–2229.
- Woo, S. S., Jiang, J., Gill, B. S., Paterson, A. H. & Wing, R. A. (1994) *Nucleic Acids Res.* **22**, 4922–4931.
- Sheng, Y., Mancino, V. & Birren, B. (1995) *Nucleic Acids Res.* **23**, 1990–1996.
- Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette, B. F. F., Rapp, B. A. & Wheeler, D. L. (1999) *Nucleic Acids Res.* **27**, 12–17.
- Stoesser, G., Tuli, M. A., Lopez, R. & Sterk, P. (1999) *Nucleic Acids Res.* **27**, 18–24.
- Osteras, M., Stanley, J. & Finan, T. M. (1995) *J. Bacteriol.* **177**, 5485–5494.
- Finan, T. M., Gough, C. & Truchet, G. (1995) *Gene* **152**, 65–67.
- Ziegler, R. J., Peirce, C. & Bergman, K. (1986) *J. Bacteriol.* **168**, 785–790.
- Driscoll, B. T. & Finan, T. M. (1997) *Microbiology* **143**, 489–498.
- Pocard, J. A., Vincent, N., Boncompagni, E., Smith, L. T., Poggi, M. C. & Le Rudulier, D. (1997) *Microbiology* **143**, 1369–1379.
- Herouart, D., Sigaud, S., Moreau, S., Frendo, P., Touati, D. & Puppo, A. (1996) *J. Bacteriol.* **178**, 6802–6809.
- Jimenez-Zurdo, J. I., Garcia-Rodriguez, F. M. & Toro, N. (1997) *Mol. Microbiol.* **23**, 85–93.
- Szeto, W. W., Nixon, B. T., Ronson, C. W. & Ausubel, F. M. (1987) *J. Bacteriol.* **169**, 1423–1432.
- Rushing, B. G. & Long, S. R. (1995) *J. Bacteriol.* **177**, 6952–6957.
- Cren, M., Kondorosi, A. & Kondorosi, E. (1995) *Mol. Microbiol.* **15**, 733–747.
- Osteras, M., Driscoll, B. T. & Finan, T. M. (1997) *Microbiology* **143**, 1639–1648.
- Sourjik, V., Sterr, W., Platzer, J., Bos, I., Haslbeck, M. & Schmitt, R. (1998) *Gene* **223**, 283–290.
- Wheatcroft, R. & Watson, R. J. (1988) *J. Gen. Microbiol.* **134**, 113–121.
- Selbitschka, W., Arnold, W., Jording, D., Kosier, B., Toro, N. & Puhler, A. (1995) *Gene* **163**, 59–64.
- Martinez-Abarca, F., Zekri, S. & Toro, N. (1998) *Mol. Microbiol.* **28**, 1295–1306.
- Wheatcroft, R. & Laberge, S. (1991) *J. Bacteriol.* **173**, 2530–2538.
- Laberge, S., Middleton, A. T. & Wheatcroft, R. (1995) *J. Bacteriol.* **177**, 3133–3142.
- Foussard, M., Garnerone, A. M., Ni, F., Souppène, E., Boistard, P. & Batut, J. (1997) *Mol. Microbiol.* **25**, 27–37.
- Barnett, M. J. & Long, S. R. (1990) *J. Bacteriol.* **172**, 3695–3700.
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277**, 1453–1474.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature (London)* **390**, 249–256.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiyama, M., Sasamoto, S., *et al.* (1996) *DNA Res.* **3**, 185–209.
- Flores, M., Gonzalez, V., Brom, S., Martinez, E., Pinero, D., Romero, D., Davila, G. & Palacios, R. (1987) *J. Bacteriol.* **169**, 5782–5788.
- Schibler, L., Vaiman, D., Oustry, A., Guinec, N., Dangy-Caye, A. L., Billault, A. & Cribiu, E. P. (1998) *Mamm. Genome* **9**, 119–124.
- Kondorosi, A., Vincze, E., Johnston, A. W. B. & Beringer, J. E. (1980) *Mol. Gen. Genet.* **178**, 403–408.
- Deng, W., Gordon, M. P. & Nester, E. W. (1995) *J. Bacteriol.* **177**, 2554–2559.
- Tiwari, R. P., Reeve, W. G., Dilworth, M. J. & Glenn, A. R. (1996) *Microbiology* **142**, 601–610.
- Kereszt, A., Slaska-Kiss, K., Putnoky, P., Banfalvi, Z. & Kondorosi, A. (1995) *Mol. Gen. Genet.* **247**, 39–47.
- Margolin, W., Corbo, J. C. & Long, S. R. (1991) *J. Bacteriol.* **173**, 5822–5830.
- Margolin, W. & Long, S. R. (1994) *J. Bacteriol.* **176**, 2033–2043.
- Aguilar, O. M. & Grasso, D. H. (1991) *J. Bacteriol.* **173**, 7756–7764.
- Papp, I., Dorgai, L., Papp, P., Jonas, E., Olasz, F. & Orosz, L. (1993) *Mol. Gen. Genet.* **240**, 258–264.
- Bardin, S. D., Voegelé, R. T. & Finan, T. M. (1998) *J. Bacteriol.* **180**, 4219–4226.
- Kereszt, A., Kiss, E., Reuhs, B. L., Carlson, R. W., Kondorosi, A. & Putnoky, P. (1998) *J. Bacteriol.* **180**, 5426–5431.
- Schnier, J., Thamm, S., Lurz, R., Hussain, A., Faist, G. & Dobrinski, B. (1988) *Nucleic Acids Res.* **16**, 3075–3089.