Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501

Yongliang Yan*†, Jian Yang†, Yuetan Dou*‡, Ming Chen*, Shuzhen Ping*§, Junping Peng†, Wei Lu*, Wei Zhang*, Ziying Yao*, Hongquan Li*, Wei Liu*, Sheng He*§, Lizhao Geng*, Xiaobing Zhang†, Fan Yang†, Haiying Yu*‡, Yuhua Zhan*‡, Danhua Li*‡, Zhanglin Lin¶, Yiping Wang‖, Claudine Elmerich**††, Min Lin*§**, and Qi Jin†**

*Biotechnology Research Institute, National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China; †State Key Laboratory for Molecular Virology and Genetic Engineering, Institute of Pathogen Biology, Chinese Academy of Medical Sciences, Beijing 100176, China; ‡College of Biological Sciences, China Agricultural University, Beijing 100094, China; §National Centre for Plant Gene Research, Beijing 100101, China; ¶Department of Chemical Engineering, Tsinghua University, Beijing 100084, China; ‖College of Life Sciences, Peking University, Beijing 100871, China; and ††Département de Microbiologie, Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, Paris 75724 Cedex 15, France

Edited by Robert Haselkorn, University of Chicago, Chicago, IL, and approved March 21, 2008 (received for review February 2, 2008)

The capacity to fix nitrogen is widely distributed in phyla of *Bacteria* and *Archaea* but has long been considered to be absent from the *Pseudomonas* genus. We report here the complete genome sequencing of nitrogen-fixing root-associated *Pseudomonas stutzeri* A1501. The genome consists of a single circular chromosome with 4,567,418 bp. Comparative genomics revealed that, among 4,146 protein-encoding genes, 1,977 have orthologs in each of the five other *Pseudomonas* representative species sequenced to date. The genome contains genes involved in broad utilization of carbon sources, nitrogen fixation, denitrification, degradation of aromatic compounds, biosynthesis of polyhydroxybutyrate, multiple pathways of protection against environmental stress, and other functions that presumably give A1501 an advantage in root colonization. Genetic information on synthesis, maturation, and functioning of nitrogenase is clustered in a 49-kb island, suggesting that this property was acquired by lateral gene transfer. New genes required for the nitrogen fixation process have been identified within the *nif* island. The genome sequence offers the genetic basis for further study of the evolution of the nitrogen fixation property and identification of rhizosphere competence traits required in the interaction with host plants; moreover, it opens up new perspectives for wider application of root-associated diazotrophs in sustainable agriculture.

genome sequencing | root-associated diazotroph

The *Pseudomonas* genus belongs to the gamma subgroup of *Proteobacteria*. Until now, complete genome sequencing has been established for *Pseudomonas aeruginosa* [PAO1 (1) and PA14], *Pseudomonas fluorescens* [Pf-5 (2), PfO-1 and SBW25], *Pseudomonas putida* KT2440 (3), *Pseudomonas syringae* [DC3000, B728a (4), and 1448A], and *Pseudomonas entomophila* (5). Although phylogenetically close to *P. aeruginosa*, *Pseudomonas stutzeri* belongs to the group of nonfluorescent *Pseudomonas*. The taxonomic status and biology of this species, isolated from a large diversity of terrestrial and marine environments, have been recently reviewed (6). Some strains have attracted particular attention because of specific metabolic properties, such as denitrification, degradation of aromatic compounds, and synthesis of polyhydroxyalkanoates (6). A1501 (China General Microbiological Culture Collection Center accession no. 0351), formerly known as *Alcaligenes faecalis* A15, was isolated from rice paddy soils and has been widely used as a crop inoculant in China (7, 8). This strain belongs to genomovar 1 of *P. stutzeri* (9). Until recently, the ability to fix nitrogen was not recognized within the *Pseudomonas* genus *sensu stricto*. This property has now been reported in a few strains (6, 9, 10) and was best documented in the case of strain A1051 (11).

Biological nitrogen fixation, a key step in global nitrogen cycling, is the ATP-dependent reduction of dinitrogen to ammonia by the nitrogenase enzyme complex. The ability to fix nitrogen is widely distributed among *Bacteria* and *Archaea*. In addition to symbiotic nitrogen-fixing microorganisms, which induce differentiated structures on the host plant (root nodules of legumes and actinorhizal plants), there exist a variety of free-living nitrogen-fixing bacteria capable of association with the root system of graminaceous plants, such as *Klebsiella pneumoniae*, *Azotobacter vinelandii*, and *Azospirillum brasilense* (12). New species, such as *Azoarcus* and *Gluconacetobacter*, can also develop endophytic associations and survive within plant tissues without causing disease symptoms (12, 13).

P. stutzeri A1501 can survive in the soil, colonize the root surface, and invade the superficial layers of the root cortex (7, 8, 14). This is why it was also reported as being an endophyte (14, 15). We report here the analysis of the complete genome of A1501 and compare it with other *Pseudomonas* species. We also report clustering of the nitrogen fixation region in a putative nitrogen fixation island and the identification of several new genes required for optimal nitrogenase activity. The genome survey also revealed the presence of a number of interesting gene clusters that may be involved in rhizosphere competence and colonization of the host plant.

Results

General Features of the Genome and Comparative Genomics. The *P. stutzeri* A1501 genome is composed of a single circular chromosome of 4,567,418 bp, encoding 4,146 probable proteins, 59 tRNA genes, and four rRNA operons (Fig. 1, Table 1). No intact (or nearly intact) prophages were found in the genome of *P. stutzeri* A1501, but it carries six genes encoding site-specific recombinases of the phage integrase family. We also identified

Author contributions: Y.Y. and J.Y. contributed equally to this work; Y.W., C.E., M.L., and Q.J. designed research; Y.Y., J.Y., Y.D., Z.Y., W. Liu, L.G., and D.L. performed research; M.C., S.P., J.P., W. Lu, W.Z., H.L., S.H., X.Z., F.Y., H.Y., Y.Z., C.E., and L.M. analyzed data; and Y.Y., J.Y., Z.L., C.E., and M.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The annotated genome sequence reported in this paper has been deposited in the GenBank database (accession no. CP000304). Raw microarray data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE6572).

**To whom correspondence may be addressed. E-mail: elmerich@pasteur.fr, linmin57@vip.163.com, or zdsys@sina.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801093105/DCSupplemental.

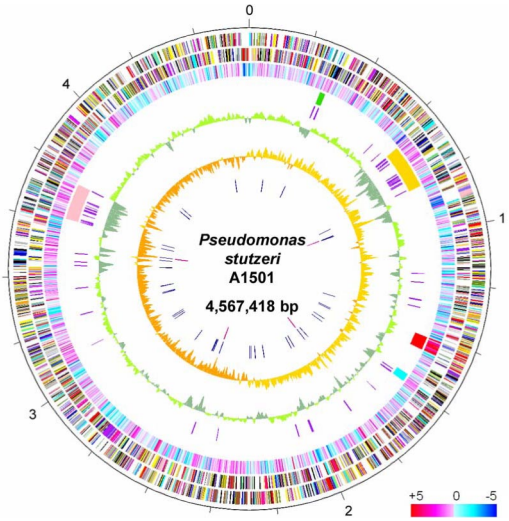

Fig. 1. Circular representation of the *P. stutzeri* A1501 genome and gene expression profiles in nitrogen fixation conditions in comparison with nitrogen excess conditions. Outer scale is marked in 200 kb. (From the outer to the inner concentric circle) Circles 1 and 2, genes encoded by leading and lagging strands, respectively. Coding sequences are color-coded by COG categories. Circle 3, $\log_2$ ratio of the expression level under nitrogen fixation conditions versus ammonia excess conditions for each gene; values are color-coded as shown in the right bottom bar. Circle 4, location of the genomic islands. Green, PST0244 to PST0260; yellow, PST0573 to PST0711; blue, PST1463 to PST1490; pink, PST3342 to PST3471; red, *nif* island PST1302 to PST1359. Circle 5, transposase, recombinase, and integrase genes. Circles 6 and 7, G + C content and GC skew (G − C/G + C), respectively, with a 10-kb window size. Circles 8 and 9, distribution of tRNA genes and *rrn* operons, respectively.

a total of 42 copies of repeat sequences (covering 1.35% of the genome) grouped into 10 types encoding 57 transposases. Four distinct regions with atypical GC content and flanked by tRNA genes (Fig. 1) had general characteristics of typical genomic islands (16), suggesting recent transfer of genetic material into A1501. A fifth region containing the *nif* genes could also be considered a genomic island, as discussed later.

The size of the genome was much smaller than those reported for other *Pseudomonas* species sequenced to date. Deduced translation products of A1501 were compared with the predicted proteomes of the five other representative *Pseudomonas* species to identify interspecies ortholog pairs (Table 1), and orthologous gene synteny was found to be severely disordered [supporting information (SI) Fig. S1]. A1501 exhibited the highest overall similarity to *P. aeruginosa* compared with other species, because 66.8% (2, 770) of the A1501 genes had counterparts in the *P. aeruginosa* PAO1 genome. This result is consistent with phylogenetic analysis based on 12 housekeeping proteins in the 11 sequenced *Pseudomonas* strains (Fig. S2). Based on genome comparison, the number of genes previously recognized as being part of the *Pseudomonas* core genome may be limited to 1,997. Indeed, most of the genes related to virulence and pathogenicity in *P. aeruginosa* PAO1 were absent in A1501, such as type III/VI secretion systems, synthesis of both types of quorum sensing molecules, alginate polymer synthesis, siderophores, and antibiotic biosynthesis pathways (1, 17).

Nitrogen Fixation. All genetic information specific to nitrogen fixation was clustered in the A1501 genome in a 49-kb nitrogen fixation region (PST1302–PST1359) consisting of 59 genes. The G + C content of this region was higher than the average of the entire genome (66.8% vs. 63.8%) (Fig. 1). This favored the hypothesis that the *nif* region in *P. stutzeri* constitutes a nitrogen fixation island, as discussed in other nitrogen fixers (16, 18, 19). For example, *nif* genes are part of an island in *Wolinella succinogenes* (16, 19), and the existence of several genomic islands in *Rhizobium leguminosarum* was based on differences in G + C content (18). The *nif* region in A1501 is flanked by genes coding for cobalamin synthesis (cobalamin 5′-phosphate synthase, PST1301) on one side and glutathione peroxidase (PST1360) on the other (Fig. S3). High gene conservation in the flanking regions was noted in other *Pseudomonas* species (unable to fix nitrogen), raising the possibility that there exists an insertion hotspot flanked by cobalamin 5′-phosphate synthase and glutathione peroxidase in *Pseudomonas* species (Table S1) in which the nitrogen fixation island in A1501 was acquired from lateral gene transfer.

The nitrogenase complex comprised the MoFe-protein encoded by *nifDK* and the Fe protein encoded by *nifH*. Full assembly of the complex required the products of a dozen other *nif* genes extremely conserved in free-living and symbiotic diazotrophs, in particular for processing of nitrogenase metalloclusters and catalytic stability (*nifMZ*, *nifW*, and *nifUS*) and for synthesis of a specific molybdenum cofactor (FeMo-co) bound to the MoFe protein (*nifB*, *nifQ*, *nifENX*, *nifV*, and *nifH*) (20). Although these genes are common to most systems, their organization, content, and regulation differ between genera.

We compared the organization of A1501 *nif* genes with those of four other well characterized nitrogen-fixing bacteria relatively phylogenetically close to *Pseudomonas*, *A. vinelandii* AvOP (21), *K. pneumoniae* M5a1 (22), *A. brasilense* Sp7 (12), and *Azoarcus* sp. BH72 (13). The general organization of the *nif* genes in A1501 showed a high degree of similarity to that of *A. vinelandii*, except that the *nif* genes were not contiguous in *A. vinelandii*, but were distributed into two portions of the genome (Fig. 2). A first group of genes (from PST1302 to PST1323 in A1501) contained *nifB* and *nifQ* involved in synthesis of FeMo-co and *nifLA*. NifL and NifA were the negative/positive regulatory proteins for *nif* operon expression (23, 24). Four additional genes associated with *nifB* and *nifQ* (Fig. 2) were also conserved in *A. vinelandii* (21). The *rnfABCDGEF* operon, which encodes subunits of a membrane-bound protein complex involved in electron transport to nitrogenase (25), was located next

Table 1. General features of *Pseudomonas* genomes

Species	Size, Mb	CDS, no.	G + C content, %	Coding density, %	tRNA, no.	rRNA operon, no.	Plasmid, no.	Oholog shared with A1501, no.	Average similarity of orthologs, %
Pst	4.56	4,146	63.8	89.8	59	4	—	—	—
Pae	6.26	5,570	66.6	89.4	63	4	—	2,770	72.2
Pfl	7.07	6,144	63.3	88.8	71	5	—	2,674	70.6
Psy	6.39	5,763	58.4	86.8	63	5	2	2,472	70
Ppu	6.18	5,420	61.6	87.7	74	7	—	2,604	69.9
Pen	5.88	5,169	64.2	89.1	78	7	—	2,531	70.4

Pst, *P. stutzeri* A1501; *Pae*, *P. aeruginosa* PAO1; *Pfl*, *P. fluorescens* Pf-5; *Psy*, *P. syringae* pv. *tomato* DC3000; *Ppu*, *P. putida* KT2440; *Pen*, *P. entomophila* L48.

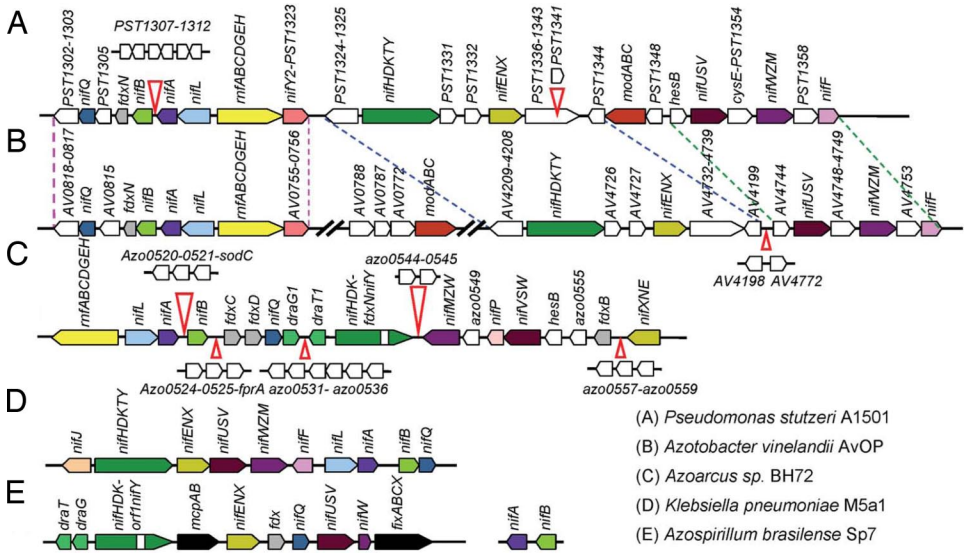

Fig. 2. Schematic representation of the *P. stutzeri* A1501 *nif* gene cluster compared with syntenic *nif* clusters of *A. vinelandii* AvOP, *K. pneumoniae* M5a1, *Azoarcus* sp. BH72, and *A. brasilense* Sp7. Connecting lines indicate corresponding orthologous gene clusters with high homology at the protein level. Identically colored segments indicate orthologs. Open boxes represent additional genes of unknown function outside or inside *nif* gene clusters. Red wedges indicate location of additional genes. *fdxN* between *nifB* (PST1306) and PST1305 corresponds to PST1305a.

to *nifLA* in AvOP, BH72, and A1501 (11). A second group (contiguous to group 1 in the case of A1501) contained a set of genes (*nifHDKTY*, *nifNEX*, *nifUSV*, and *nifWZMF*) in the same order in *A. vinelandii*, *K. pneumoniae*, and A1501. These genes were contiguous in *K. pneumoniae*, whereas additional coding sequences (CDSs) were found interspaced between or within the *nif* operons both in *A. vinelandii* and A1501. In A1501, very few nucleotides were present in the noncoding regions between genes, and translational overlapping between start and stop codons of some of the genes was observed (Fig. S4). For example, the 10 genes from PST1349 (*hesB*-like gene) to PST1358 were likely to be part of the same 7.5-kb transcriptional unit, suggesting a superoperon with translational coupling. Although nitrogenase on-off switching was reported in A1501 (11), *draT* and *draG* genes known to control the nitrogenase ADP ribosylation process were not found in A1501, suggesting an inactivation mechanism different from that reported in *Azoarcus* and *A. brasilense* (12, 13).

Transcriptomic analysis of the whole A1501 genome was implemented by using genome-wide microarrays. Significant changes were observed in expression of a number of genes. Among these, 255 genes, including most of the genes within the nitrogen fixation island, showed increased transcription (ratio $\geq$2) under nitrogen fixation conditions compared with nitrogen excess conditions (Fig. 1 and Table S2). This suggested that some of the CDSs interspaced between *nif* genes might be involved in nitrogen fixation. We further chose 16 uncharacterized genes within the *nif* region. Up-regulation under nitrogen-fixing conditions was confirmed by quantitative real-time RT-PCR (using *nifH* as a control), whereas mutant construction by gene disruption (enabling transcription of downstream genes) led to a significant decrease in nitrogenase activity for all except two of them in PST1308 and PST1331 (Table 2 and Fig. S4). This suggests that the corresponding gene products, although not essential, are required for full nitrogenase activity under the assay conditions used.

Rhizosphere Competence Traits and Root Surface Colonization. *P. stutzeri* A1051 is a soil bacterium, but it can colonize rice roots in an aquatic environment (7, 8, 14). It can also colonize wheat (14). Bacteria are found in the root tip mucilage, in root hair, at the emergence of lateral roots and, in some cases, in the first layers of the root cortex (8, 14). Analysis of the genome led to identification of genes coding for different properties that presumably permit A1501 to adapt to the highly competitive environment in the rhizosphere (26).

Some specific metabolic capacity may contribute to fitness in diverse environment and the genome carries a wide spectrum of genes for carbon and nitrogen source utilization, including a large number of transporter genes. Transport systems are key components in communication of bacteria with their environment. A1501 possesses >300 transporter-encoding genes, including at least 138 members of the ATP binding cassette (ABC) transporter superfamily (Fig. 3). This number is in the range of that observed in *P. aeruginosa* and *P. putida.* However, 87 genes (including 33 genes for ABC-type transporters) present in A1501 were not found in other *Pseudomonas* genomes. We found many transport systems for carbohydrates, amino acids and dicarboxylic acids, which are the major components of root exudates (Fig. 3). Acquisition of iron is an important trait for rhizosphere competition (27). A1501 does not carries pathways for the synthesis of siderophores, and it contains fewer TonB-dependent siderophore receptors than other *Pseudomononas* species. (Table S3). A1501 should get iron from siderophores produced by other soil bacteria in the rhizosphere, a feature also reported in *Azoarcus* (13).

Pathways for aromatic degradation through the protocatechuate (*pca*) and catechol (*ben*) routes found in the A1501 genome are similar to those found in other *Pseudomonas*, such as *P. entomophila* (5) and *P. putida* (28). Using *in vivo* expression technology, one of the *pca* gene was reported to be up-regulated in the rhizosopere of the *P. stuzeri* A15 (14). In general, *P. stutzeri* strains are not reported to accumulate polyhydroxyalkanoates (PHA) granules, except for one strain, *P. stutzeri* 1317 (6, 29). The A1501 genome contains *phaCABR* genes involved in the three-step polyhydroxybutyrate biosynthetic pathway, with *phaR* coding for a transcriptional regulator, but it contains a single phasin gene (*phaP*), in contrast to *Ralstonia*, which carries many (30).

Denitrification is another general property of many *Pseudomonas* species. The organization of genes involved in nitrate metabolim in A1501 is close to that of the *P. stutzeri* Zobell strain (31), with a super cluster containing *nor*, *nir*, and *nos* genes

Table 2. Identification of 16 uncharacterized genes within the nitrogen fixation island with increased expression under nitrogen fixation conditions and nitrogenase activity of corresponding mutant strains

Gene ID	Probable function (genes)	Involvement in nitrogen fixation in other systems	Nitrogenase activity, %	Up-regulation factor in microarrays*	Up-regulation factor in real time RT-PCR†
PST1302	Glutaredoxin-related protein	Unknown	76	++++	23.95 ± 4.99
PST1303	Thiosulfate sulfurtransferase	Unknown	27	++++	15.08 ± 2.32
PST1305	Arsenate-reductase-related protein	Unknown	52	++++	28.46 ± 9.56
PST1308	LysR, transcriptional factor	Unknown	91	+	2.52 ± 0.54
PST1312	Thiopurine S-methyltransferase (*tmpA*)	Unknown	60	+	2.82 ± 0.13
PST1325	Unknown	Unknown	24	+++	27.86 ± 7.92
PST1326	Nitrogenase Fe protein (*nifH*)‡	Yes	0	++++	132.5 ± 18.3
PST1331	Unknown	Unknown	86	+++	10.09 ± 1.14
PST1332	Unknown	Unknown	60	+	3.69 ± 0.78
PST1336	Unknown	Unknown	39	++	3.71 ± 1.23
PST1337	Unknown	Unknown	72	++++	15.28 ± 1.70
PST1342	Unknown	Unknown	57	+	3.31 ± 0.23
PST1343	Unknown	Unknown	32	+	4.01 ± 0.21
PST1344	Unknown	Unknown	66	++	18.91 ± 2.23
PST1349	Fe-S cluster assembly protein (*hesB*)	Yes	2	++++	221.51 ± 24.40
PST1353	Serine acetyltransferase (*cysE*-like)§	Yes	19	++++	4.42 ± 0.12
PST1354	Unknown	Unknown	68	+++	19.31 ± 1.29

All uncharacterized genes have counterparts in *A. vinelandii*, except for PST1312. Nitrogenase activity of mutants (enabling transcription of downstream genes) is expressed as percentage of wild-type A1501 activity.
*+, 2–4; ++, 4–8, ++; +++, 8-16; ++++, >16.
†Values are mean ± SD.
‡*nifH* was used as control.
§Also designated *nifP* or *iscA*.

(Table S4), except that A1501 does not harbor the regulatory gene *nirR*. A recent report revealed that a mutant of *narG*, coding for a component of nitrate reductase, was more competitive than the wild-type strain in colonization of rice and wheat (32).

Analysis of the genome revealed that A1501 has a number of genes reported to play a role in osmotolerance (33). In particular, it carries genes coding periplasmic glucan biosynthesis proteins MgoG and MgoH found in most aerobic bacteria. Except for *betA* and *betB*, other genes of the glycine-betaine pathway found

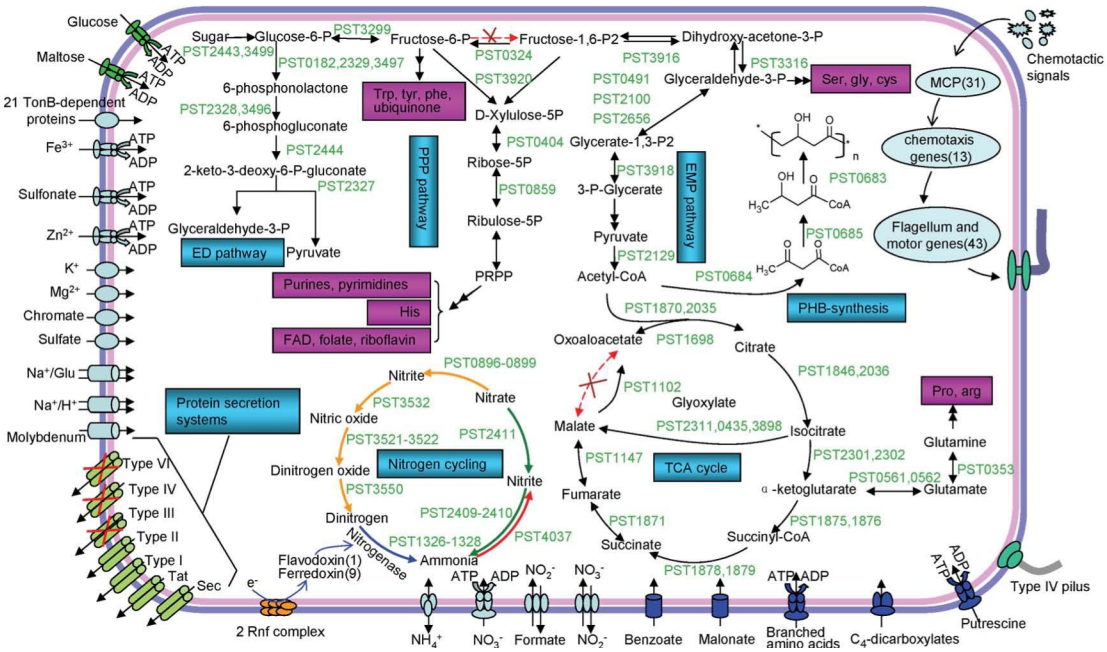

Fig. 3. Schematic overview of metabolic pathways and transport systems in *P. stutzeri* A1501. Gene identifiers are shown by numbers in green. Predicted transporters are grouped by substrate specificity: light blue, inorganic ions; emerald green, sugar; dark blue, organic sources. Reactions for which no candidate enzyme was confidently predicted are indicated by dashed arrows. Pathways that involve multiple reactions are shown by double arrows. Final biosynthetic products are indicated with pink boxes. Arrows indicate direction of transport. Crosses indicate pathways or reactions that are apparently not present in strain A1501.

in other *Pseudomonas* species are missing in A1501. The genome carries an *ectABC* cluster responsible for synthesis and accumulation of ectoine in the moderate halophile *Halomonas* but (except for *ectB*) absent in other *Pseudomonas*. The genome contains a number of genes that can play a role in detoxification of reactive oxygen species commonly found in *Pseudomonas* genomes. It includes five catalases, two superoxide dismutases, four peroxidases, four hydroperoxide reductases, and nine glutathione *S*-transferases (34) (Table S5).

Chemotaxis and bacterial surface components (such as flagella, pili, and surface polysaccharides) are of importance in adhesion to a surface and biofilm formation (12, 26). A1501 has a complex chemosensory protein system by which the bacterium can sense and respond to the presence of certain substances, with 31 genes coding for methyl accepting receptors widespread over the genome, a situation close to that of *P. putida* KT2440 (3). It also carries 13 genes for signal excitation and adaptation and methyl removal proteins (*che*) and clusters containing *flg* and *fli* genes responsible for flagella biogenesis and motility. Indeed, an *rpoN* mutant that was devoid of flagella displayed less surface colonization than the wild type (12). A set of 28 genes responsible for type IV pili biogenesis are present in A1501, including genes known to be required for natural DNA transformation in other *P. stutzeri* strains (*pilA*, *comL*, *pilE*, *pilC*, *pilT*, and *pilU*) (35). However, type IV pili are also required for twitching motility, and *pilA* was found to be essential for root surface colonization and for infection of plant tissue in *Azoarcus* (13).

Alginate is the major exopolysaccharide found in fluorescent *Pseudomonas*. However, except for *algA* (phosphomannose isomerase) and *algI* (acetylase), all of the other genes involved in alginate biosynthesis are missing in A1501, whereas most of the regulatory genes known to control its biosynthesis (36) are present in A1501 (Table S6). This suggests that the ability to synthesize alginate was lost in strain A1501. We noted a cluster of genes for cellulose biosynthesis in A1501 sharing some similarities with that of *P. putida* KT2440 that may play a role in attachment to the plant surface. There also exists a gene (PST2494) encoding a hydrolytic enzyme resembling an endoglucanase, which is known to play a role in endophytic colonization in *Azoarcus* (13); however, no gene coding for pectinolytic activity, which is often found in pathogens, was detected in the A1501 genome. We have not identified genes possibly involved in indole-3-acetic acid (IAA) synthesis. But the genome contains *acdS*, which codes for a deaminase of the precursor of ethylene, a phytohormone known to inhibit root elongation. This gene is a good candidate for involvement in plant growth promotion by preventing ethylene formation (12, 37).

Discussion

The *P. stutzeri* A1501 genome is much smaller than that of fluorescent *Pseudomonas* species sequenced to date (Table 1). This might be due to reductive evolution and could reflect a more stringent association with its plant host, although smaller size is not specific to A1501 and was reported for other *P. stutzeri* strains that do not necessarily have the same lifestyle (38). In addition, *P. putida* KT2440, which is also a rhizospheric strain, has a much larger genome (3, 39) (Table 1). We identified a set of 487 genes common to other *Pseudomonas* and absent in the A1501 genome, suggesting loss of functions in A1501, as noted for the alginate biosynthesis gene cluster and for fluorescent siderophore production.

Lateral gene transfer appears to be a mechanism for continuous adaptation of soil bacteria to environmental changes (16). The A1501 genome contains four typical genomic islands (Fig. 1), suggesting recent transfer. Some genes associated with these genomic islands encode conserved proteins sharing similarity with transport and heavy metal resistance proteins, whose functions may be related to better adaptation to the rhizosphere. We also noted that genes for type I restriction-modification (R-M) systems (Table S7) were present in three of the genomic islands but nowhere else in the genome, suggesting recent acquisition of R-M by lateral gene transfer.

The presence of *nif* genes in this particular strain of *P. stutzeri*, when most other *Pseudomonas* species lack this property, raises the question of their origin. The nitrogen fixation process is believed to have played a crucial role in early cellular evolution, in particular when the geochemical reserves of fixed nitrogen in the biosphere became depleted. Because nitrogen fixation is maintained in *Bacteria* and *Archaea*, it is often hypothesized that *nif* genes may originate from the last common ancestor, even though other scenarios are proposed (40, 41). It can be presumed that *nif* genes were lost in most lineages of *Bacteria* and *Archaea* and that lateral gene transfer played a role in recent acquisition of *nif* genes in some lineages (16, 42). The rareness of nitrogen fixation properties in true *Pseudomonas* suggests that *nif* genes were lost during evolution in most (if not all) *Pseudomonas* species. Packing of the *nif* genes in A1501, and the difference in G + C content from the rest of the genome, favor the hypothesis that the 49-kb *nif* region is a genomic island. The *K. pneumoniae* 24-kb *nif* gene cluster comprises the tightest organization of *nif* genes identified to date (22). In all other systems studied, additional genes associated with core *nif* genes have been identified, but in most cases, their role in nitrogen fixation remains hypothetical (20). From the data reported here, it is likely that many of the associated genes play a role in some steps of the nitrogen fixation process. We presume that many of these genes, which do not exist in other systems except for *Azotobacter*, might encode functions for adapting the nitrogen fixation process, such as electron transport chains to nitrogenase and oxygen protection mechanisms. Because the physiology of nitrogen fixation is more restrictive in *K. pneumoniae* (anaerobiosis) than in *P. stutzeri* (microaerobiosis) and *A. vinelandii* (aerobiosis), it is tempting to speculate that there exists lateral transfer of the *nif* gene from a common ancestor and subsequent acquisition of extra genes so as to adapt to the physiology of the respective hosts.

We have identified, in the genome of A1501, a large number of genes that might be responsible for survival in the soil and colonization of the host plant. Except for a few of them (*narG*, *pcA*, and *rpoN*) identified in refs. 12 and 14, their actual involvement in colonization remains to be established. Recent work in the rhizospheric *P. pudtia* KT2440 identified 91 genes up-regulated in the presence of maize exudates (39), and 31 of them are conserved in A1501 (including regulators, transport systems, stress response, and chemotaxis) (data not shown). Present knowledge of the genome should help in elucidating regulatory mechanisms that control the association with the host plant, thus enabling better comprehension of mechanisms of nitrogen transfer. An understanding of the mechanisms by which root-associated diazotrophs improve productivity of nonleguminous crops is an important issue. Our research is aimed at improving associative nitrogen fixation for sustainable agriculture and for reducing crop chemical fertilizer requirements. Finally, the availability of the A1501 genome sequence provides insights into the evolution of nitrogen-fixing organisms and opens up new perspectives for molecular interaction between the microbe and host plant.

Methods

Genome Sequencing. The genome sequence was determined by using whole-genome shotgun strategy as described in ref. 43. Plasmid libraries with one small insert (2–4 kb) and one large insert (4–6 kb) were constructed in an alkaline-phosphatase-treated pUC19 vector after random mechanical shearing (ultrasonic nebulization) of genomic DNA. Finally, 50,408 valid reads and 30.7 Mb of total length were established, providing 7-fold genome coverage.

For a detailed description of genome assembly, annotation and comparisons, see *SI Methods*.

Microarray. The microarray used in this study featured 3,988 of the 4,146 CDSs identified in *P. stutzeri A1501*. Full details are in *SI Methods*.

Quantitative Real-Time RT-PCR. Quantitative real-time PCR experiments were performed with three independent RNA preparations, using the 7000 Sequence Detection system (Applied Biosystems) and SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's recommendations. First-strand cDNAs were synthesized from 5 μg of total RNA in a 20-μl reaction volume, using the SuperScript first-strand synthesis system for RT-PCR (Invitrogen).

Construction of Mutants. Uncharacterized genes within the *nif* island were inactivated by homologous suicide plasmid integration (44), using pK18mob as a vector. DNA fragments of ≈200 bp were amplified by using total DNA of A1501 as template and oligonucleotide primers designed to generate amplicons for the creation of mutations, enabling the transcription of downstream genes. Amplicons were ligated into pK18mob, and the resulting plasmids were introduced into *P. stutzeri* A1501; recombination at the correct location was checked by PCR. Nitrogenase activity was assayed with bacterial suspensions incubated at an OD_{600} of 0.1 in N-free minimal K medium at 30°C under an argon atmosphere containing 0.5% oxygen and 10% acetylene according to the protocol described in ref. 11. Nitrogenase activity is expressed as nanomoles of ethylene per min per (mg protein).

ACKNOWLEDGMENTS. We thank Prof. Jiliang Tang for helpful discussions, Dr. Kornelia Smalla for critical reading of the manuscript, and Dr. Valérie Karplus and Mrs. Jerri Bram for text editing. This work was supported by the Ministry of Science and Technology of China National Basic Research Program Grants 2007CB707805, 2007CB109203, and 2001CB108904; National High-Tech Program Grants 2007AA021304 and 2006AA020101; National Natural Science Foundation of China Grants 30470047 and 30200007; and Sichuan Habio Bioengineering Company.

1. Stover CK, *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406:959–964.
2. Paulsen IT, *et al.* (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat biotechnol* 23:873–878.
3. Nelson KE, *et al.* (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* 4:799–808.
4. Feil H, *et al.* (2005) Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv *tomato* DC3000. *Proc Natl Acad Sci USA* 102:11064–11069.
5. Vodovar N, *et al.* (2006) Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. *Nat biotechnol* 24:673–679.
6. Lalucat J, Bennasar A, Bosch R, Garcia-Valdes E, Palleroni NJ (2006) Biology of *Pseudomonas stutzeri*. *Microbiol Mol Biol Rev* 70:510–547.
7. Qiu Y, Zhou S, Mo X, You C, Wang D (1981) Investigation of dinitrogen fixation bacteria isolated from rice rhizosphere. *Chinese Sc Bull (Kexue Tongbao)* 26:383–384.
8. You C, Lin M, Fang X, Song W (1995) Attachment of *Alcaligenes* to rice roots. *Soil Biol Biochem* 27:463–466.
9. Vermeiren H, *et al.* (1999) The rice inoculant strain *Alcaligenes faecalis* A15 is a nitrogen-fixing *Pseudomonas stutzeri*. *Syst Appl Microbiol* 22:215–224.
10. Hatayama K, Kawai S, Shoun H, Ueda Y, Nakamura A (2005) *Pseudomonas azotifigens* sp. *nov.*, a novel nitrogen-fixing bacterium isolated from a compost pile. *Int J Syst Evol Microbiol* 55:1539–1544.
11. Desnoues N, *et al.* (2003) Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology* 149:2251–2262.
12. Elmerich C, Newton WE (2007) *Associative and Endophytic Nitrogen fixing Bacteria and Cyanobacterial Associations—Introduction* (Springer, Dordrecht, The Netherlands).
13. Krause A, *et al.* (2006) Complete genome of the mutualistic, N(2)-fixing grass endophyte *Azoarcus* sp. *strain* BH72. *Nat biotechnol* 24:1385–1391.
14. Rediers H, *et al.* (2003) Development and application of a *dapB*-based *in vivo* expression technology system to study colonization of rice by the endophytic nitrogen-fixing bacterium *Pseudomonas stutzeri* A15. *Appl Environ Microbiol* 69:6864–6874.
15. You CB, Zhou F (1989) Non-nodular endorhizospheric nitrogen fixation in wetland rice. *Can J Microbiol* 35:403–408.
16. Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2:414–424.
17. Haas D, Keel C (2003) Regulation of antibiotic production in root-colonizing *Peudomonas* spp. and relevance for biological control of plant disease. *Annu Rev Phytopathol* 41:117–153.
18. Young JP, *et al.* (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 7:R34.
19. Baar C, *et al.* (2003) Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc Natl Acad Sci USA* 100:11690–11695.
20. Klipp W, Masepohl B, Gallon JR, Newton WE (2004) *Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria* (Kluwer Academic, Dordrecht, The Netherlands).
21. Jacobson MR, *et al.* (1989) Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *J Bacteriol* 171:1017–1027.
22. Arnold W, Rump A, Klipp W, Priefer UB, Puhler A (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *J Mol Biol* 203:715–738.
23. Martínez-Argudo I, Little R, Shearer N, Johnson P, Dixon R (2004) The NifL-NifA System: A multidomain transcriptional regulatory complex that integrates environmental signals. *J Bacteriol* 186:601–610.
24. Xie Z, *et al.* (2006) Interaction between NifL and NifA in the nitrogen-fixing *Pseudomonas stutzeri* A1501. *Microbiology* 152:3535–3542.
25. Jeong HS, Jouanneau Y (2000) Enhanced nitrogenase activity in strains of *Rhodobacter capsulatus* that overexpress the *rnf* genes. *J Bacteriol* 182:1208–1214.
26. Lugtenberg BJ, Dekkers LC (1999) What makes *Pseudomonas* bacteria rhizosphere competent? *Environ Microbiol* 1:9–13.
27. Neilands JB (1995) Siderophores: Structure and function of microbial iron transport compounds. *J Biol Chem* 270:26723–26726.
28. Jimenez JI, Minambres B, Garcia JL, Diaz E (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ Microbiol* 4:824–841.
29. Xi J, *et al.* (2000) Hyperproduction of polyesters consisting of medium-chain-length hydroxyalkanoate monomers by strain *Pseudomonas stutzeri* 1317. *Antonie Van Leeuwenhoek* 78:43–49.
30. Pohlmann A, *et al.* (2006) Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nat biotechnol* 24:1257–1262.
31. Zumft WG (1997) Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* 61:533–616.
32. Rediers H, Vanderleyden J, De Mot R (2007) Nitrate respiration in *Pseudomonas stutzeri* A15 and its involvement in rice and wheat root colonization. *Microbiol Res*, 10.1016/j.micres.2007.03.003.
33. Wood JM (1999) Osmosensing by bacteria: Signals and membrane-based sensors. *Microbiol Molec Biol Rev* 63:230–262.
34. Vuilleumier S, Pagni M (2002) The elusive roles of bacterial glutathione *S*-transferases: New lessons from genomes. *Appl Microbiol Biotechnol* 58:138–146.
35. Sikorski J, Teschner N, Wackernagel W (2002) Highly different levels of natural transformation are associated with genomic subgroups within a local population of *Pseudomonas stutzeri* from soil. *Appl Environ Microbiol* 68:865–873.
36. Chitnis CE, Ohman DE (1993) Genetic analysis of the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* shows evidence of an operonic structure. *Mol Microbiol* 8:583–593.
37. Glick BR (2005) Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *FEMS Microbiol Lett* 251:1–7.
38. Ginard M, Lalucat J, Tummler B, Romling U (1997) Genome organization of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. *Int J Syst Bacteriol* 47:132–143.
39. Matilla MA, *et al.* (2007) Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. *Genome Biol* 8:R179.
40. Fani R, Gallo R, Lio P (2000) Molecular evolution of nitrogen fixation: The evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J Mol Evol* 51:1–11.
41. Raymond J, Siefert JL, Staples CR, Blankenship RE (2004) The natural history of nitrogen fixation. *Mol Biol Evol* 21:541–554.
42. Kechris KJ, Lin JC, Bickel PJ, Glazer AN (2006) Quantitative exploration of the occurrence of lateral gene transfer by using nitrogen fixation genes as a case study. *Proc Natl Acad Sci USA* 103:9584–9589.
43. Jin Q, *et al.* (2002) Genome sequence of *Shigella flexneri* 2a: Insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res* 30:4432–4441.
44. Windgassen M, Urban A, Jaeger KE (2000) Rapid gene inactivation in *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* 193:201–205.