The genome of *Paenibacillus sabinae* T27 provides insight into evolution, organization and functional elucidation of *nif* and *nif*-like genes

Xinxin Li1,2†, Zhiping Deng1†, Zhanzhi Liu1†, Yongliang Yan3†, Tianshu Wang1, Jianbo Xie1, Min Lin3, Qi Cheng3† and Sanfeng Chen1,2*

**Abstract**

**Background:** Most biological nitrogen fixation is catalyzed by the molybdenum nitrogenase. This enzyme is a complex which contains the MoFe protein encoded by *nifDK* and the Fe protein encoded by *nifH*. In addition to *nifHDK*, *nifHDK*-like genes were found in some Archaea and Firmicutes, but their function is unclear.

**Results:** We sequenced the genome of *Paenibacillus sabinae* T27. A total of 4,793 open reading frames were predicted from its 5.27 Mb genome. The genome of *P. sabinae* T27 contains fifteen nitrogen fixation (*nif*) genes, including three *nifH*, one *nifD*, one *nifK*, four *nifB*, two *nifE*, two *nifN*, one *nifX* and one *nifV*. Of the 15 *nif* genes, eight *nif* genes (*nifB*, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX* and *nifV*) and two non-*nif* genes (*orf1* and *hesA*) form a complete *nif* gene cluster. In addition to the *nif* genes, there are nitrogenase-like genes, including two *nifH*-like genes and five pairs of *nifDK*-like genes. IS elements on the flanking regions of *nif* and *nif*-like genes imply that these genes might have been obtained by horizontal gene transfer. Phylogenies of the concatenated 8 *nif* gene (*nifB*, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX* and *nifV*) products suggest that *P. sabinae* T27 is closely related to *Frankia*. RT-PCR analysis showed that the complete *nif* gene cluster is organized as an operon. We demonstrated that the complete *nif* gene cluster under the control of *σ*70-dependent promoter enabled *Escherichia coli* JM109 to fix nitrogen. Also, here for the first time we demonstrated that unlike *nif* genes, the transcriptions of *nifHDK*-like genes were not regulated by ammonium and oxygen, and *nifH*-like and *nifD*-like gene could not restore the nitrogenase activity of *Klebsiella pneumonia* *nifH*− and *nifD*− mutant strains, respectively, suggesting that *nifHDK*-like genes were not involved in nitrogen fixation.

**Conclusions:** Our data and analysis reveal the contents and distribution of *nif* and *nif*-like genes and contribute to the study of evolutionary history of nitrogen fixation in *Paenibacillus*. For the first time we demonstrated that the transcriptions of *nifHDK*-like genes were not regulated by ammonium and oxygen and *nifHDK*-like genes were not involved in nitrogen fixation.

**Keywords:** *Paenibacillus sabinae* T27, *nif* gene cluster, *nif*-like gene, Genome sequence

* Correspondence: sanfeng_chen@126.com
† Equal contributors
1 Key Laboratory for Agrobiotechnology, Ministry of Agriculture, China Agricultural University, Beijing 100193, P. R. China
2 Key laboratory of Soil Microbiology, Ministry of Agriculture, China Agricultural University, Yuanmingyuan west road no.2, Haidian District, Beijing 100193, P. R. China
3 Full list of author information is available at the end of the article

© 2014 Li et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Biological nitrogen fixation, the conversion of atmospheric N₂ to NH₃, plays an important role in the global nitrogen cycle and in world agriculture [1]. Most biological nitrogen fixation is catalyzed by the molybdenum nitrogenase. This enzyme is a complex which contains the MoFe protein encoded by nifDK and the Fe protein encoded by nifH. The MoFe protein contains two metalloclusters: FeMo-co, a [Mo-7Fe-9S-C-homocitrate] cluster which serves as the active site of substrate binding and reduction and the P-cluster, a [8Fe-7S] cluster which shuttles electrons to FeMo-co [2,3]. Previous biochemical and genetic studies on Klebsiella pneumoniae carrying twenty nif genes on 24-kb region genes and Azotobacter vinelandii revealed that nifH, nifD, nifK, nifE, nifN, nifB, nifQ, nifV, nifY, nifI, nifS, nifZ, and nifM contribute to the synthesis and maturation of nitrogenase [2,3].

Contents and organization of nif genes varied significantly among N₂-fixing organisms. For example, in K. pneumoniae, twenty nif genes are co-located within a ~24 kb cluster [4], whereas in A. vinelandii the nif genes are more dispersed and distributed as two clusters in genome [5]. There is usually only one nifH gene and the nifH, nifD and nifK genes are transcribed as a single unit in many diazotrophs, such as K. pneumoniae and A. vinelandii. However, multiple nifH genes were found in a few diazotrophs. For examples, Rhizobium leguminosarum bv. phaseoli possesses three nifH genes [6] and Clostridium pasteurianum W5 has six nifH homologs [7].

Nitrogen fixation is sporadically distributed among prokaryote families: Proteobacteria, Firmicutes, Archaea, Cyanobacteria and Actinobacteria [8]. The incomplete distribution pattern and the difference in contents and organization of nif genes raise the question of origins and evolution of Mo-nitrogenase. Two conflicting hypotheses for the origin of Mo-nitrogenase have been proposed on the basis of phylogenetic examination of Mo-nitrogenase protein sequences (NifHDK) [9]. The last common ancestor (LCA) hypothesis implies that the Mo-nitrogenase had its origin in a common ancestor of the bacterial and archaeal domains. According to the LCA model gene loss has been extensive and accounts for the fact that nitrogenase is found neither in eukaryotes nor in many entire phyla of prokaryotes. The Methanogen origin hypothesis implies that nitrogen fixation originated from methanogenic archaea and subsequently was transferred into a primitive bacterium via lateral gene transfer. Recent studies based on phylogenetic analysis of NifHDK sequences supported the Methanogen origin hypothesis and implied that Mo-nitrogenase evolved in the anaerobic and hydrogenotrophic methanogens with acquisition in the bacterial domain via lateral gene transfer involving an anaerobic member of the Firmicutes [10].

Firmicutes have been thought to play an important role in evolution of nitrogen fixation. Studies on evolution of nitrogen fixation in Firmicutes mainly focused on the anaerobic diazotrophic Clostridia. Although Paenibacillus is a genus of Firmicute, its nitrogen fixation traits and evolution remains unclear. It is well known that Paenibacillus is a genus of Gram-positive, facultative anaerobic, endospore-forming bacteria, originally included within the genus Bacillus and then reclassified as a separate genus in 1993 [11]. Bacteria belonging to this genus have been detected in a variety of environments such as soil, water, rhizosphere, vegetable matter, forage and insect larvae, as well as clinical samples [12]. Nitrogen-fixing Paenibacillus species have great potential for use as a bacterial fertilizer in agriculture, but genomic information of these bacteria is lacking.

Here we report the complete genome sequence of P. sabinae T27 which is a nitrogen-fixer isolated from the rhizosphere of plant Sabina squamata by our laboratory [13]. The whole genome analysis not only reveals the organization and distribution of nitrogen-fixing genes and nitrogenase-like genes, but also provides insight into the evolution of nif genes in Paenibacillus. Furthermore, we demonstrate that the complete nif gene cluster consisting of ten genes (nifB, nifH, nifD, nifK, nifE, nifN, nifX, orf1, hesA and nifV) of P. sabinae T27 is a functional unit for nitrogen fixation. Here for the first time we demonstrated that nifHDK-like genes are not involved in nitrogen fixation.

Results and discussion

General features of Paenibacillus sabinae T27 genome

The complete genome of P. sabinae T27 is composed of a single circular molecule of 5,270,569 base pairs (bp) with an average G + C content of 52.64%. The circular chromosome has a total of 4,849 putative protein-coding sequences (CDS), 26 rRNAs (8 copies of 16S-23S-5S operons and 1 copy of 16S-23S operon) and 82 tRNAs (Table 1). Among the predicted genes, 3,538 were assigned

| Table 1 General features of the genome of P. sabinae T27 |
|---------------------------------|
| Attribute                     | Value  |
| Complete genome size, bp      | 5,270,569 |
| G + C%                        | 52.64%  |
| Protein-coding sequences      | 4,849   |
| Genes with assigned function  | 3,538   |
| Genes with unknown function   | 1,311   |
| Average CDS size              | 941     |
| Percent of coding region%     | 72.96%  |
| No. of rRNAs                  | 26      |
| No. of tRNAs                  | 82      |
| Insertion sequence (IS) elements | 28     |
putative functions, covering 72.96% of the genome, and 1,311 encoded hypothetical proteins (Table 1). Twenty eight insertion sequence (IS) elements were identified in the P. sabinae T27 genome.

**Comparative genomics of P. sabinae T27**

Previous phylogeny based on nifHDK showed that Firmicutes, cyanobacteria and actinobacteria are closely related [10]. Here we compared the genomes of P. sabinae T27, Clostridium acetobutylicum ATCC 824 (a member of Firmicutes), Frankia sp. Cc13 (an actinobacterium) and Nostoc punctiforme PCC 73102 (a cyanobacterium). The four species had the core genome of 258 putative protein-coding genes (Figure 1A). There are 802 genes which are shared by P. sabinae T27 and C. acetobutylicum ATCC 824, there are 454 genes which are shared by P. sabinae T27 and Frankia sp. Cc13, and there are 553 genes which are shared by P. sabinae T27 and N. punctiforme PCC 73102. The shared genes by P. sabinae T27 and C. acetobutylicum ATCC 824 are more than those shared by P. sabinae T27 with Frankia sp. Cc13 or N. punctiforme. The results are consistent with the fact that Paenibacillus is more closely related to Clostridium than to Frankia and cyanobacteria, since Paenibacillus and Clostridium belong to the same Firmicutes. Furthermore, the genome of P. sabinae T27 was compared with those of the closely related Paenibacillus azotofixans ATCC35681 (a nitrogen-fixer) [14] and Paenibacillus polymyxa SC2 (a non-nitrogen-fixer) [15] (Figure 1B). Genome sizes of P. sabinae T27, P. azotofixans ATCC35681 and P. polymyxa SC2 are 5.27 Mb, 5.44 Mb and 6.24 Mb, respectively. Chromosome alignments showed higher level of conservation of genome architecture between P. sabinae T27 and P. polymyxa SC2 than that between P. sabinae T27 and P. azotofixans ATCC35681.

**Central metabolism**

P. sabinae T27 is a nitrogen-fixing bacterium isolated from the rhizosphere of the plant Sabina squamata [13]. The bacterium contains a wide spectrum of genes for carbon utilization and carbohydrate, amino acid and inorganic ion transport. The genome of P. sabinae T27 contains the complete set of genes for the pentose phosphate pathway (PPP) (Additional file 1: Figure S1). In addition to the metabolism of pentose, the non-oxidative PPP allows the production of intermediates necessary for nucleic acid synthesis. It contains the complete set of genes for the glycolysis pathway and allows production of acetyl-CoA. In the presence of external electron acceptors, acetyl-CoA may be completely oxidized via the citrate cycle (TCA cycle), which is encoded by the P. sabinae T27 genome (Additional file 1: Figure S1). Although the gene coding for the classical malate dehydrogenase (MDH1, EC:1.1.1.37) in TCA cycle is absent, another malate dehydrogenase (MQO, EC:1.1.5.4) gene which might be involved in pyruvate metabolism pathway metabolizing oxaloacetate to malate, is found in the genome of P. sabinae T27.

Sucrose is the common carbon source used for isolation of P. sabinae T27 [13]. The genome of the bacterium has the sucrose-6-phosphate hydrolase and alphaglucosidase for metabolizing sucrose to glucose and fructose. Transporter systems are an important element for bacteria to communicate with their environment. The genome of P. sabinae T27 contains an extensive set of transporters.
of 247 transport related genes. Of the 247 transport related genes, 64 are involved in carbohydrate transport, 66 encode components of amino acid transporters and 107 encode components of inorganic ion transporters. Importantly, Fe (iron), molybdenum, sulfate and NH₄⁺ are related to nitrogen fixation and nitrogen metabolism.

**Nitrogen fixation and nitrogenase-like genes**

One of the most distinct features of *P. sabinae* T27 is its ability to fix nitrogen. The genome of *P. sabinae* T27 contains fifteen *nif* genes, including four *nifB*, three *nifH*, one *nifD*, one *nifK*, two *nifE*, two *nifN*, one *nifX* and one *nifV*. Of the 15 *nif* genes, eight *nif* genes (*nifB, nifH, nifD, nifK, nifE, nifN, nifX* and *nifV*) and two non-*nif* genes (*orf1 and hesA*) which are located between *nifX* and *nifV* form a complete *nif* gene cluster, the four *nif* genes (*nifE, nifN, nifB and nifH*) are clustered together and the other three *nif* genes (two *nifB* and one *nifH*) are scattered at different locations (Figure 2). In addition to the *nif* genes, there are nitrogenase-like genes, including two *nifH*-like and five pairs of *nifDK*-like genes. Our results are consistent with the reports that *nifDK*-like genes existed in Archaea and Firmicutes [8]. Interestingly, genome of *P. sabinae* T27 does not contain transcription regulatory gene *nifA* which is found in almost all of Gram-negative diazotrophs, such as in *K. pneumoniae, A. vinelandii* and *Pseudomonas stutzeri* A1501 [16]. The lack of *nifA* suggests that there may be a different regulation mechanism of nitrogen fixation in *P. sabinae* T27.

**The content and organization of the complete *nif* gene cluster**

Bioinformatics analysis revealed that the ten genes *nif*BHDKENXorf1hesAnifV in within the complete *nif* gene cluster are organized as an operon within an 11 kb region. The gene designated as *hesA* is also found in *Frankia* [17] and cyanobacteria [18]. The *orf1*, whose predicted product is a hypothetical protein, is also found in several *N₂*-fixing *Paenibacillus* species [19]. The predicted product of HesA shares ~ 45% identity with the putative molybdenum cofactor biosynthesis protein HesA. HesA is a member of the ThIF-MoeB-HesA family and contains an N-terminal nucleotide binding domain and a C-terminal MoeZ/MoeB-like domain. The gene content and organization of the complete *nif* gene cluster is unique to *Paenibacillus* [19,20]. Although *Paenibacillus* and *Clostridium* are the members of the Firmicute, their *nif* gene content and organization varied greatly. For example, *nifN-B* fusion gene was found in the *nif* gene clusters of the three species of *Clostridia: C. acetobutylicum, C. beijerinckii*, and *C. pasteurianum*. Also, there are two genes *nifI1* and *nifI2* located between *nifH* and *nifDK* in *C. acetobutylicum*. Although *nifI1* and *nifI2* are not essential for nitrogen fixation, but serve a regulatory function [22]. Actually, the *nif* gene content and organization of *Clostridium* spp. are more similar to those of *Methanosarcina acetivorans* and *Methanococcus maripaudis*, since two genes *nifI1* and *nifI2* also exist between *nifH* and *nifDK* in these archaea.

**IS may play important roles in the evolution of the *nif* and *nif*-like genes**

As described above, twenty eight insertion sequence (IS) elements, belonging to six transposase families were identified in *P. sabinae* T27 chromosome. IS elements were found to be located on the flanking region of the

![Figure 2 Chromosome map and distribution of nif genes in P. sabinae T27](http://www.biomedcentral.com/1471-2164/15/723)
complete nif gene cluster, other nif genes and nif-like genes (Figure 3). It is generally accepted that IS abundance correlates positively with the frequency of horizontal gene transfer (HGT) [23]. IS elements can mediate the transfer of genetic information (such as antibiotic resistance and new metabolic capabilities) between genomes or between replicons of the same genome and they can also induce duplications, deletions, and rearrangements of genetic information [24]. The existence of transposase in the flanking region of the complete nif gene cluster suggests that the nif cluster might be acquired in P. sabinae T27 by HGT event from other diazotrophs and the additional nifBHEN genes and nifHDK-like genes might be horizontally transferred or duplicated. The nif genes acquired by HGT were also reported in several diazotrophs. For example, A sequence reminiscent of a transposase gene located just upstream the nif cluster in Herbaspirillum seropedicae is an indicative of HGT event [25]. It was generally recognized that variations of G + C contents between nif cluster and genome are indicative of HGT. For example, G + C content of the nif gene cluster was higher than the average of the entire genome (66.8% vs. 63.8%) in P. stutzeri A1501 [16]. However, we found that the G + C contents of the complete nif gene cluster of P. sabinae T27 is as same as the average of the entire genome (52.64% vs. 52.63%), suggesting that the complete nif gene cluster of P. sabinae T27 has undergone a longer time of evolution.

Evolution of nif and nif-like genes of P. sabinae T27
To further evaluate the evolution of nitrogen fixation in P. sabinae T27, we reconstructed the phylogenies based on the concatenation of the NifBHDKENXV sequences (Figure 4). Notably, the Nif protein sequences of P. sabinae T27 used for the phylogenetic trees were from the complete nif cluster. The phylogenetic tree showed that Paenibacillus and Frankia are sister groups, suggesting that P. sabinae T27 may originate from a common ancestor with Frankia. Also, we constructed the phylogenetic trees based on the HesA and Orf1 sequences which are contained within the complete nif gene cluster. The HesA phylogenetic tree revealed supported that Paenibacillus and Frankia are sister groups (Additional file 2: Figure S5). The Orf1 phylogenetic tree showed that P. sabinae T27 is closely related to Clostridium (Additional file 3: Figure S6). IS element on the flanking region of the complete nif cluster suggested that the complete nif cluster may have been acquired in P. sabinae T27 by HGT. Interestingly, these data revealed that although Paenibacillus and Clostridium are the members of the Firmicutes, their nif genes are not very closely related.

The complete genome sequence revealed that there are three nifH, one nifD, two nifH-like, five pairs of nifDK-like genes in P. sabinae T27. Here we constructed phylogenetic trees with real NifH/NifD/NifK and NifH/NifD/NifK-like sequences (Figure 5) and the phylogenetic tree revealed that NifH/NifD/NifK-like sequences are clearly divergent from conventional nitrogenase. All NifH-like, NifD-like and NifK-like sequences are clustered together by themselves, suggesting that they may have been resulted from duplication.

As described above, in addition to the ten genes nifBHDKENXorf1hesAnifV within the complete nif gene cluster, three nifB, two nifH, one nifE and one nifN genes exist in the genome of P. sabinae T27. Here we further constructed NifB, NifH and the concatenated NifEN phylogenetic trees (Additional files 4, 5, 6: Figures S2-S4) and phylogenetic analysis revealed that these multiple nifB, nifH and nifEN are clustered with their own corresponding genes within the complete nif gene cluster,
suggesting that they may result from duplication of nifB, nifH, nifE and nifN, respectively, of the complete nif gene cluster.

**Characterization of multiple nitrogenase-like genes**

The nifHDK are structural genes of Mo-nitrogenase, with the nifD and nifK genes encoding the α and β subunits, respectively, of the molybdenum iron protein (dinitrogenase) and the nifH the γ subunit of the iron protein (dinitrogenase reductase). The genome of *P. sabinae* T27 contains two nifH-like, five nifD-like and five nifK-like genes. Conserved residues in alignments of NifH-like sequences (Figure 6) with NifH sequences show that 4Fe-4S iron sulfur cluster-ligating cysteines

![Figure 4](image_url) Maximum-likelihood tree based on NifBHDKENXV protein sequences of *P. sabinae* T27 and the representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a neighbor-joining analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.

![Figure 5](image_url) Maximum-likelihood phylogenetic tree of NifHDK/NifHDK-like sequences. NifHDK/NifHDK-like sequences were derived from *P. sabinae* T27 and the representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a neighbor-joining analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.
and the P-loop/MgATP binding motif are invariant, suggesting that these proteins may function analogously to dinitrogenase reductase. Conversely, NifD/NifK-like sequences are highly diverged from both the nitrogenase subunits. FeMoco-ligating residues at $\alpha$Cys275 and $\alpha$His442 of NifD (Figures 7, 8) are not conserved in NifD/Nifk-like sequences, although several—but not all—conserved cysteines involved with P-cluster coordination are found in NifD/Nifk-like sequences.

Expressions of nifHDK and nifHDK-like genes in N$_2$-fixing and non-N$_2$-fixing conditions

It is generally recognized that nif genes are expressed in N$_2$-fixing conditions (the microaerobic or anaerobic and without ammonium or limited ammonium). In order to examine whether the transcription of nifHDK and nifHDK-like genes is regulated by ammonium and oxygen in N$_2$-fixing organisms, suggesting that the nifHDK genes of P. sabinae T27 are involved in nitrogen fixation. In contrast to nifHDK, nifHDK-like genes of P. sabinae T27 were not significantly differently expressed in N$_2$-fixing and non-N$_2$-fixing conditions, suggesting that these nif-like genes did not function in nitrogen fixation.

Functional analysis of nifH/nifH-like and nifD/nifD-like genes in nitrogen fixation

To further comparatively study the functions of the nif and nif-like gene of P. sabinae T27, K. pneumonia nifH mutant strain 1795 and nifD mutant strain I$\alpha$423P [26], both of which have no or very low nitrogenase activity, were complemented with the nifH/nifH-like and nifD/nifD-like genes of P. sabinae T27 under the control of K. pneumonia nifH promoter, respectively. As shown in Figure 10A, the complementary strains carrying nifH-like1 or nifH-like2 of P. sabinae T27 could not resume the nitrogenase activity of K. pneumonia nifH mutant strain 1795, while the nifH1 from the complete nif cluster of P. sabinae T27 could restore to nearly 50% of the
wild-type strain M5al. The data are consistent with our previous report that the three copies of *nifH* could restore nitrogenase activity of *K. pneumonia nifH*− mutant strain 1795 [27]. Likewise, *nifD* of *P. sabinae* T27 could resumed the nitrogenase activity of *K. pneumonia nifD*− mutant strain Iα423P, although *K. pneumonia* enabled *K. pneumonia* nifD− mutant strain Iα423P to have higher nitrogenase activity than *P. sabinae* T27 did (Figure 10B). In contrast, none of *nifD*-like1, *nifD*-like2, *nifD*-like3, *nifD*-like4 and *nifD*-like5 could restore the nitrogenase activity of *K. pneumonia nifD*− mutant. These data suggest that *nif*-like genes may be not involved in nitrogen fixation.

The complete *nif* gene cluster is organized as an operon. Bioinformatics analysis revealed that the ten genes *nifBHDKENXorf1hesAnifV* within the complete *nif* gene cluster are organized as an operon. Here RT-PCR experiments using primers designed to span across intergenic regions indicated that the nine genes within the *nif* cluster are organized in a single operon (Additional file 7: Figure S7). Single operon *nif* clusters have been reported.
in gram-positive prokaryotes and in the archaea, e.g. *Heliobacterium chlorum* [28] and *Methanococcus maripaludis* [29]. However, in contrast to these *nif* clusters *P. sabinae* T27 does not contain the negative regulatory genes *nifI1* and *nifI2* (homologues of *glnB*), which are involved in post-translational regulation of nitrogenase activity in response to fixed nitrogen [30].

The complete *nif* gene cluster of *P. sabinae* T27 has a σ70-dependent promoter

Almost all of the *nif* genes in Gram-negative nitrogen-fixing bacteria, such as *K. pneumoniae* and *A. vinelandii*, are transcribed from σ70 promoters (−24/-12) whose expression depends on activator NifA [31]. However, the presumed promoter regions for the *nif* genes of *P. sabinae* T27 have sequences which are similar to the *E. coli* σ70-dependent promoter. A σ54-dependent −24/-12 promoter sequence was not observed upstream of the *nif* cluster. Downstream of *nifV*, a potential transcriptional termination site was identified, containing two potential stem loops followed by a T-rich region (Additional file 8: Figure S8A). These findings indicate that the *nif* genes in *P. sabinae* T27 are organized as a single operon containing 9 genes, which is transcribed from an rpoD-dependent promoter.

To analyze the σ70-dependency of the *nifB* promoter, electrophoretic mobility shift assays (EMSA) were carried out using either *E. coli* σ70-RNAP (RNA polymerase) or σ70 from *P. sabinae* T27, which was overexpressed and purified from *E. coli* (Additional file 8: Figure S8B). EMSA experiments revealed that both purified σ70 from *P. sabinae* T27 and *E. coli* σ70-RNAP holoenzyme bind to the 45 bp *nifB* promoter fragment. Competition experiments with non-labelled *nifB* DNA indicated that the *E. coli* RNAP holoenzyme binds more tightly to this DNA fragment, since higher concentrations of competitor were apparently required to dissociate the *E. coli* σ70-RNAP
These results are consistent with the ability of $\sigma^A$ (or $\sigma^{70}$) of *Bacillus subtilis* to bind to promoters independent of core RNAP [32,33].

The complete *nif* gene cluster of *P. sabinae* T27 enables *E. coli* to fix nitrogen

We further cloned the 12-kb full-length *nif* gene cluster consisting of its own *nif* promoter and the contiguous nine genes *nifBHDKENXorf1hesAnifV* into the wide-host plasmid pVK100 and then transformed this into *E. coli* JM109, yielding the recombinant *E. coli* strain 27 (Additional file 9: Figure S9). To determine whether the *Paenibacillus nif* gene cluster functions in *E. coli*, we employed two independent methods to assess nitrogenase activity: firstly, reduction of the alternative substrate acetylene to ethylene, which can be readily quantified by gas chromatography [34,35] and secondly, a $^{15}\text{N}_2$ enrichment assay to directly measure the incorporation of this tracer into organic nitrogen [36]. When grown anaerobically in nitrogen-deficient medium, *P. sabinae* T27 exhibits both acetylene reduction and $^{15}\text{N}_2$ incorporation (Additional file 9: Figure S9). The recombinant *E. coli* strain 27, which expresses the *nif* genes from the native promoter showed approximately 10% of the specific activity for acetylene residue production for the native promoter.
reduction when compared with *Paenibacillus* and was competent to assimilate $^{15}$N$_2$. The results demonstrated that the complete *nif* gene cluster is a functional unit.

**Conclusions**

In this study, we uncovered the contents and organization of *nif* and *nif*-like genes of *P. sabinae* T27 by completing its genome sequence. The genome of *P. sabinae* T27 contains fifteen nitrogen fixation (*nif*) genes, including three *nifH*, one *nifD*, one *nifK*, four *nifB*, two *nifE*, two *nifN*, one *nifX* and one *nifV*. Of the 15 *nif* genes, eight *nif* genes (*nifB*, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX* and *nifV*) and two non-*nif* genes (orf1 and hesA) form a complete *nif* gene cluster. Phylogenetic analysis suggests that the complete *nif* cluster of *P. sabinae* T27 was originated from a common ancestor with *Frankia*. Multiple *nifB*, *nifH*, *nifE*, *nifN* may result from duplication. The complete *nif* gene cluster is organized in an operon as a functional unit for nitrogen fixation. The complete *nif* gene cluster under the control of its $\sigma^{70}$-dependent promoter enabled *Escherichia coli* JM109 to fix nitrogen. *P. sabinae* T27 contains two *nifH*-like genes and five pairs of *nifDK*-like genes. Unlike *nif* genes, the transcriptions of *nifHDK*-like genes were not regulated by ammonium and oxygen and *nifHDK*-like genes were not involved in nitrogen fixation.

**Methods**

**Strains and media**

Strains used in this study is listed in Additional file 10: Table S1. *P. sabinae* T27 and the recombinant *E. coli* strains were routinely grown in LD medium (per liter contains: 2.5 g NaCl, 5 g yeast and 10 g tryptone) at 30°C with shaking. When appropriate, antibiotics were added in the following concentrations: 100 $\mu$g/ml ampicillin, and 20 $\mu$g/ml tetracycline for maintenance of plasmids.

Nitrogen-free and nitrogen-deficient media were used in this study. Nitrogen-free medium contains: 100 $\mu$g/ml g NaCl, 5 $\mu$g/ml Na$_2$HPO$_4$, 3.4 $\mu$g/ml KH$_2$PO$_4$, 1.6 mg CaCl$_2$• 2H$_2$O, 30 $\mu$g/ml MgSO$_4$• 7H$_2$O, 0.3 $\mu$g/ml MnSO$_4$, 36 $\mu$g/ml Ferric citrate, 7.6 $\mu$g/ml Na$_2$MoO$_4$• 2H$_2$O, 10 $\mu$g/ml p-aminobenzoic acid, 5 $\mu$g/ml biotin and 4 $\mu$g/ml glucose as carbon source. Nitrogen-deficient medium contained 2 mM glutamate as nitrogen source in nitrogen-free medium [20].

**Phylogenetic analysis**

Maximum-likelihood (ML) phylogenetic trees were constructed using PhyML (version 3.0) [37] software and multiple alignment of amino acid sequences were carried out by ClustalW (version 2.1) [38].

**Genome sequencing, genome annotation and analysis**

Genomic DNA of *P. sabinae* T27 was isolated according to [13]. Genome sequencing was performed by Tianjin Research Center for Functional Genomics and Biochip in China. The genome *P. sabinae* T27 was sequenced by using a hybrid sequencing approach that incorporates 454 pyrosequencing with Illumina Genome Analyzer. Sequencing by both methods was performed according to manufacturer’s instructions, Roche and Illumina.

The rRNA genes were identified with RNAmmer [39]. Transfer RNA (tRNA) genes were identified by the program tRNAscan-SE [40]. Genes coding for proteins with known functions were annotated by searches against KEGG Genes, Pfam, and SWISSPROT [41]. The complete sequence has been assigned GenBank accession no. CP004078.

**Construction of recombinant plasmid for expression of the complete nif cluster in E. coli**

Genomic DNA of *P. sabinae* T27 was used as template for cloning *nif* genes. A 12 kb Xho I-Xho I DNA fragment containing the complete *nif* gene cluster (a 310 bp promoter region and the contiguous ten genes *nifHDKXorfAhesAnifV*) was PCR amplified with primers T-up and T-down (Additional file 11: Table S2). The PCR product was ligated to Xho I site of pVK100, yielding plasmid pKY100-27. Then the plasmid was transferred to *E. coli* JM109, yielding the recombinant *E. coli* 27 strain.

**Construction of plasmids for complementation studies**

In order to determine the function of *nifH1/nifH*-like and *nifD/nifD*-like genes, overlap PCR was performed to fuse the coding regions of *nifH1*, *nifH*-like1, *nifH*-like2, *nifD*, *nifD*-like1, *nifD*-like2, *nifD*-like3, *nifD*-like4 and *nifD*-like5 of *P. sabinae* T27 with the *nifH* promoter of *K. pneumoniae*. The primers used in fusion were listed in Additional file 11: Table S2. The amplified PCR products were cloned to pVK100. The recombinant pVK100 were transformed to *K. pneumoniae nifH1* mutant or *K. pneumoniae nifD* mutant for complementation.

**Transcription start site identification**

The 5′-RACE method was used to determine the transcription start site (TSS) using the SMARTer RACE cDNA Amplification Kit (Clontech). Gene-specific primers are listed in Additional file 11: Table S2. The PCR product was cloned into the pMD18-T Vector and then sequenced.

**Overexpression and purification of $\sigma^{70}$ from P. sabinae T27 in E. coli**

A 1134 bp DNA fragment carrying the *rpoD* gene (encoding $\sigma^{70}$ of *P. sabinae* T27) was PCR amplified with primers sigma A-F and sigma A-R (Additional file 11: Table S2). The PCR product was ligated to the pET-28b expression vector, yielding plasmid pET28-$\sigma^{70}$. *E. coli* strain BL21 (DE3) was transformed with expression plasmid pET28-$\sigma^{70}$ and utilized for protein expression. The bacterial cells were grown in LB medium to the end of
log phase and then a final concentration of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) was added to the culture and the cells were harvested after incubation for another 4 h at 16°C. The cells were then harvested and disrupted by sonication on ice. The protein was purified from the supernatant with Ni²⁺-NTA agarose (Qiagen) according to the manufacturer’s instructions.

Electrophoretic mobility shift assay (EMSA)
For the electrophoretic mobility shift assay (EMSA), a 50 bp nif promoter fragment (from −47 to +3 relative to the transcription start site of nifB in \textit{P. sabiniae} T27) was synthesized by Sangon Biotech Co., Ltd (Shanghai). To do this, two DNA fragments corresponding to the sequences of the first strand (5’- GAGAAGTGAATGCGTCTGA TTTGCTCCTGTCCTAAA-GTAATATTAT-3’) and the complementary DNA strand (5’- ATATAATACATC TTAGAGAC-AGGACAAATACGTAATTCACCTTC TCC-3’) were synthesized. The two strands were annealed and then labeled with digoxin using the DIG Gel Shift Kit (Roche). The binding shift experiment of \textit{E. coli} σ⁷⁰-RNAP (RNA polymerase) (Epicentre) or σ²⁹ of \textit{P. sabiniae} T27 to the nif promoter was carried out using a gel shift kit (Roche). At the same time, a scrambled 39 bp DNA fragment formed by annealing the following complementary oligonucleotides (5’- GTACGGAGTATCAGCTCCGTA GCATGCAATCTCCTGG-3’) and (5’-CCAGAGGATT TGCACTACGGAGCTGGATCTCCGTA -3’) were used to assay non-specific binding.

RT-PCR and qRT-PCR analysis
For RT-PCR, \textit{P. sabiniae} T27 was grown in N₂-fixing conditions (without NH₄Cl and O₂). For qRT-PCR, \textit{P. sabiniae} T27 was grown in N₂-fixing conditions (without NH₄Cl and O₂) for nitrogenase activity assays was performed according to Wang et al’s reports [20].

Additional files

Additional file 1: Figure S1. Schematic overview of metabolic pathways and transport systems in \textit{P. sabiniae} T27. Predicted transporters are grouped by energy specificity: red, ATP-dependent transporters; deep pink, symporters; light pink, ion channels; yellow, transporter family. Arrows indicate direction of transport. Final biosynthetic products are indicated with orange boxes. Crosses indicate pathways or reactions that are apparently not present in \textit{P. sabiniae} T27.

Additional file 2: Figure S5. Maximum-likelihood tree based on complete HesA protein sequences showing relationships between HesA protein of \textit{P. sabiniae} T27 and HesA proteins from representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a neighbor-joining analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.

Additional file 3: Figure S6. Maximum-likelihood tree based on ORF1 sequences showing relationships between ORF1 of \textit{P. sabiniae} T27 and ORF1 from representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a neighbor-joining analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.

Additional file 4: Figure S2. Maximum-likelihood tree based on complete NifF protein sequences showing relationships between NifF proteins of \textit{P. sabiniae} T27 and NifF proteins from representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a neighbor-joining analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.

Additional file 5: Figure S3. Maximum-likelihood tree based on complete NifG protein sequences showing relationships between NifG-like proteins of \textit{P. sabiniae} T27 and NifG proteins from representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a maximum-likelihood analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.2 substitutions per amino acid position.

Additional file 6: Figure S4. Maximum-likelihood tree based on complete NifE protein sequences showing relationships between NifE proteins of \textit{P. sabiniae} T27 and NifE proteins from representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on a maximum-likelihood analysis of 100 resampled datasets; only values at or above 50% are given, Bar 0.1 substitutions per amino acid position.

Nitrogenase activity assays by acetylene reduction method
For nitrogenase activity assays, \textit{P. sabiniae} T27 and the recombinant \textit{E. coli} 27 strain were grown in 5 mL of LD media (supplemented with antibiotics when necessary) in 50 mL flasks shaken at 250 rpm for 16 h at 30°C. Nitrogenase activity assays was performed according to Wang et al’s reports [20].

\(^{15}\)N\(_2\) incorporation assay
\textit{P. sabiniae} T27 and the recombinant \textit{E. coli} strain were grown overnight in LD medium. The cultures were collected and resuspended in 70 ml nitrogen-deficient medium containing 2 mM glutamate as nitrogen source to an OD\(_{600}\) of 0.4 in a 120 ml serum bottle. \(^{15}\)N\(_2\) incorporation assay was performed according to Wang et al’s report [20].
Wrote paper: SC. All authors helped to draft and approve the final draft.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

Analyzed the data: ZD, XL, ZL, JX. Conducted experiments: XL, ZD, ZL. Designed experiments and wrote paper: SC. All authors helped to draft and approved the final manuscript.

Acknowledgements

This work was supported by funds from the National “973” Project (Grant No. 2010CB126504).

Author details

1Key Laboratory for Agrobacteriology, Ministry of Agriculture, China Agricultural University, Beijing 100193, P. R. China. 2Key laboratory of Soil Microbiology, Ministry of Agriculture, China Agricultural University, Yuanmingyuan west road no.2, Haidian District, Beijing 100193, P. R. China. 3Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, P. R. China.

Received: 7 November 2013 Accepted: 31 July 2014 Published: 27 August 2014

References

1. Falkowski PG. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. Nature 1997, 387(6638):272–275.
2. Rubio LM, Ludden PW. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annu Rev Microbiol 2008, 62(933–111.
3. Hu Y, Ribbe MW. Biosynthesis of nitrogenase FeMoCo. Coord Chem Rev 2011, 255(9–10):1219–1224.
4. Arnold W, Rump A, Klipp W, Pfafer F, Pühler A: Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. J Mol Biol 1988, 203(3):715–738.
5. Setubal JC, dos Santos P, Goldman BS, Eresy W, Espin G, Rubio LM, Valla S, Almeida NF, Balabasunamian D, Crones L, Curati L, Zijin G, Gudse Y, Goodner B, Burnis KH, Hernandez JA, Houmieu K, Imperial J, Kennedy C, Larson TJ, Latreille P, Ligot LS, Li J, Maerk M, Miller LM, Norton S, Caroli IP, Paulsen I, Rafts EC, Roemer R, et al. Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J Bacteriol 2009, 191(4):4534–4545.
6. Michels J, D’Hooghe I, Vennet C, Pelemans H, Vanderleyden J. Characterization of the Rhizobium leguminosarum biovar phaseoli nif gene, a positive regulator of nif gene expression. Arch Microbiol 1994, 161(4):404–408.
7. Wang SZ, Chen JS, Johnson JL: The presence of five nifH-like sequences in Clostridium pasteurianum: sequence divergence and transcription properties. Nucleic Acids Res 1988, 16(43):454–454.
8. Dos Santos P, Fang Z, Morgan J, Setubal JC, Dixon R: Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics 2012, 13(1):162.
9. Raymon JD, Steffert JL, Staples CR, Bankshipt RE: The natural history of nitrogen fixation. Mol Biol Evol 2004, 21(3):541–554.
10. Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW: A late methanogen origin for molybdenum-dependent nitrogenase. Geobiology 2011, 9(3):221–232.
11. Ash C, Priest F, Collins MD: Molecular identification of RNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Antonie Van Leeuwenhoek 1993, 64(3):253–260.
12. Benedicti A, Campos S, Ambrosini A, De Souza R, Granada C, Costa P, Arruda L, Moreira F, Vargas LK, Weiss Y, Tippoo E, Farco H, Souza EM, Pedrosa FO, Passaglia LM: Genome sequence of the diazotrophic Gram-positive rhizobacterium Paenibacillus nriograndis SBSR1(1). J Bacteriol 2011, 193(22):6391–6392.
13. Ma Y, Xia Z, Liu X, Chen S: Paenibacillus sp. nov., a nitrogen-fixing species isolated from the rhizosphere soils of shrubs. Int J Syst Evol Microbiol 2007, 57(1):1–6.
14. Sellin L, Van Etas D, Penido EGC: Bacillus azotofixans sp. nov., a nitrogen-fixing species from Brazilian soils and grass roots. Int J Syst Bacteriol 1984, 34:451–456.
15. Ma M, Wang C, Ding Y, Li L, Shen D, Jiang X, Guan D, Cao F, Chen H, Feng R, Wang X, Ge Y, Yao L, Bing X, Yang X, Li J, Du B: Complete genome sequence of Paenibacillus polymyxa SC2, a strain of plant growth-promoting rhizobacterium with broad-spectrum antimicrobial activity. J Bacteriol 2011, 193(3):311–312.
16. Yang Y, Yang J, Dou Y, Chen M, Peng S, Peng J, Lu W, Zhang W, Yao Z, Li H, Liu W, He S, Geng L, Zhang X, Yang F, Yu H, Zhan Y, Li D, Lin Z, Wang Y, Elmichen C, Lin M, Qin O: Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated Pseudomonas alcaligenes A1501. Proc Natl Acad Sci U S A 2008, 105(21):7564–7569.
17. Oh CJ, Kim KB, Kim J, Kim WU, Lee H, An CS: Organization of nif gene cluster in Frankia sp. Eu01 strain, a symbiont of Elaeagnus umbellata. Arch Microbiol 2012, 194(1):29–34.
18. Welsh EA, Liberton M, Stockel J, Loh T, Elvitigala T, Wang C, Wollam A, Fulton RS, Clifton SW, Jacobs JM, Aurora R, Ghost BK, Sherman LA, Smith RD, Wilson RK, Pakrasi HB: Functional response to nitrogen-fixing rhizobacteria in the marine nitrogen cycle. Proc Natl Acad Sci U S A 2008, 105(21):15094–15099.
19. Xie JB, Du Z, Bai L, Tian C, Zhang Y, Xie J, Wang T, Liu X, Chen X, Cheng Q, Chen S, Li J: Comparative genomic analysis of N-fixing and non-N-fixing Paenibacillus spp: organization, evolution and expression of the nitrogen fixation genes. PLoS Genet 2014, 10(3):e1004231.
20. Wang L, Zhang L, Liu Z, Zhao D, Liu X, Zheng B, Xie J, Hong Y, Li P, Chen S, Dixon R, Li J: A minimal nitrogen fixation gene cluster from Paenibacillus sp. WLY78 enables expression of active nitrogenase in Escherichia coli. *PloS Genet* 2013, 9(1):1033865.

21. Bothe H, Schmitz O, Yates MG, Newton WE: Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiol Mol Biol Rev* 2010, 74(4):529–551.

22. Leigh JA: Nitrogen fixation in methanogenic: the archival perspective. *Curr Issues Mol Biol* 2000, 2(4):125–131.

23. Touchon M, Rocha EP: Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* 2007, 24(4):969–981.

24. Francino MP: An adaptive radiation model for the origin of new gene functions. *Nat Genet* 2005, 37(8):573–577.

25. Pedrosa FO, Monteiro RA, Wassem R, Cruz LM, Ayub RA, Colauto NB, Fernandez MA, Fungaro MH, Grisard EC, Hungria M, Madeira HM, Nodari RO, Osaku CA, Petzl-Erler ML, Tenerzi H, Vieira LG, Weiss VA, Pereira LF, Almeida MI, Alves LR, Marín A, Araujo LM, Balsanelli E, Baura VA, Chubatsu LS, Facoro H, Favetti A, Friedermann G, Glienke C, et al: *Genome of Herbaspirillum seropedicae strain SmR1*, a specialized diazotrophic endophyte of tropical grasses. *PloS Genet* 2011, 7(5):e1001964.

26. Guo Q, Peng T, Chang T, Zhang G, Jiang W, Li Y, Li J: Mutagenesis at α-423Ile of MoFe protein reduces the catalytic activity of nitrogenase in *Klebsiella oxytoca*. *Chin Sci Bull* 2014, 59:849–856.

27. Hong Y, Ma Y, Wu L, Maki M, Qin W, Chen S: Characterization and analysis of *nif* genes from *Pseudomonas sp.* T27. *Microbiol Res* 2012, 167(10):596–601.

28. Enkh-Amgalan J, Kawasaki H, Oh-oka H, Seki T: Cloning and characterization of a novel gene involved in nitrogen fixation in *Helobacterium chlorum*: a possible regulatory gene. *Arch Microbiol* 2006, 186(4):327–337.

29. Kesseler PS, Blank C, Leigh JA: The *nif* gene operon of the methanogenic archaeon *Methanococcus maripaludis*. *J Bacteriol* 1998, 180(6):1504–1511.

30. Doddsworth JA, Leigh JA: Regulation of nitrogenase by 2-oxoglutarate-reversible, direct binding of a PHI-like nitrogen sensor protein to dinitrogenase. *Proc Natl Acad Sci U S A* 2006, 103(26):9779–9784.

31. Dixon R, Kahn D: Genetic regulation of biological nitrogen fixation. *Nature Rev Microbiol* 2004, 2(8):621–631.

32. Yamada M, Kubo M, Miyake T, Sakaguchi R, Higo Y, Imanaka T: Promoter sequence analysis in *Bacillus* and *Escherichia*: construction of strong promoters in *E. coli*. *Gene* 1991, 99(1):109–114.

33. Jammer H, Lansen TS, Krogh A, Savild HH, Bunka S, Knudsen S: Sigma A recognition sites in the *Bacillus subtilis* genome. *Microbiology* 2001, 147(Pt 9):2417–2424.

34. Dilkworth MJ: Acetylene reduction by nitrogen-fixing preparations from *Clostridium pasteurianum*. *Biochim Biophys Acta* 1966, 127(2):285–294.

35. Schrollhorn R, Burris RH: Acetylene as a competitive inhibitor of N₂ fixation. *Proc Natl Acad Sci U S A* 1967, 58(2):213–216.

36. Montoya JP, Voss M, Kahler P, Capone DG: A simple, high-precision, high-sensitivity tracer assay for N(inf2) fixation. *Appl Environ Microbiol* 1996, 62(3):986–993.

37. Guindon S, Duftayard J, Lefort V, Ansinsima M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of *PhyML*. *Syst Biol* 2010, 59:307–321.

38. Thompson JD, Gibson T, Higgins DG: Multiple sequence alignment using *ClustalW* and *ClustalX*. *Curr Protoc Bioinformatics* 2002, 0.2.3.2:3.1–3.2.22.

39. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW: *RNAmmer*: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007, 35(9):3100–3108.

40. Lowe TM, Eddy SR: *tRNAscan-SE*: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, 25(1):955–964.

41. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gusekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A: The Pfam protein families database. *Nucleic Acids Res* 2010, 38(Database issue):D211–D222.