Comparative genomic analysis of *Bacillus mycoides* Gnyt1 strain

Yang Xiaomei (✉ yangxiaomeirz@126.com)

Gansu Agricultural University

Tuo Yao

Gansu Agricultural University

Research article

Keywords: Comparative genomic analyses, *Bacillus mycoides*, *Bacillus thuringiensis*, Functional gene analysis, Strain evolution

Posted Date: March 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15751/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

This study focused on the comparative genomic analyses of five carrying *Bacillus* strains which might give us insights on the difference and similarity in the genomic evolutionary kinships and contexts of *Bacillus mycoides* Gnyt1. Comparative genomic analyses of the carrying *Bacillus* strains genomes with emphasis on their antimicrobial resistance genes and virulence factors were performed. Most nitrogen-fixing genes and the phosphate-dissolving genes of the comparative genomic strain existed and the gene type was significantly different. And the action to nitrogen-fixing gene and phosphate-dissolving gene are dependent on different genes. Sharing some adherence, similarity, and the *Bacillus mycoides* strain Gnyt1 developed evolution is consistent with other *Bacillus*. Related functions absorb different gene-specific acquisitions, pigments, proteases, regulatory, secretory systems and functional factors associated with nitrogen-fixing phosphorus. Sharing most 16S rDNA analysis, Mauve-based genome-wide, Gene family analysis and phylogenetic tree, the *Bacillus* strain Gnyt1 and *Bacillus mycoides* strain AH621 differ in composition gene expression and expression cause conditions and factors.

1. Introduction

Bacillus mycoides Gnyt1 is a significant plant growth promoting rhizobacteria (PGPR), which could swarm by thousands of biological of the rhizosphere, the nutrient-enriched zone in the immediate vicinity of the soil-root interface and bacterial species (Bais et al., 2006; Mendes et al., 2013). In reported, PGPR strains can promote plant growth of promoting nutrient availability and improving plants to adapt more and more environmental stresses, etc (Kumar et al., 2014; Glickmann et al., 1995). *Bacillus* and *Pseudomonas* are often interrelated with the plant growth and soil quality, *Bacillus mycoides* has an important role in soil remediation and plant growth (Ahmed et al., 2014; Aeron et al., 2011; Orhan et al., 2006). Absorbing soil nutrients and water by root, which affecting morphological development lead to plant grow. Plant root play an important role on nutrient availability (Kapulnik et al., 1985).

The use of comparative genomics can more directly determine the evolutionary relationship of strains. However, the infections caused by *Bacillus mycoides* comparative genomic have also been incidentally reported (G. Lombardi et al., 2002; S. Ladhani et al., 1998). The Gnyt1 is part of *Bacillus mycoides* is a normal isolate from the rhizosphere of the dominant plant *polygonum viviparum* of the alpine grassland where was suffering from Tianzhu in Gansu province, China, in 2014 (Liu T et al., 2017). The performance of the strain has been previously studied and genome-wide sequencing (Li JH, 2017), the complete genomes of five strains had been sequenced and annotated in reported study.

In this study, we also demonstrate the strain sequence and compare its similar *Bacillus* genome with the three major representative strains (CM000719, CM000743, and CM000749) that are economically considerable species of plant microorganism. By comparing the four strains genomic analysis, our aim was to compare to obtain the information of the target genome based on the gene sequence and the genomic map, our comparisons revealed not only conserved components of the Gnyt1 core genome but also unique genes for each pathogen. In particular, this study focused on the comparative genome

analyses of four *Bacillus* strains which might give us insights on the similarity and difference in the genomic contexts of *Bacillus mycoides* Gnyt1 gene. These data will provide a foundation for the molecular mechanisms and functional analyses of the Gnyt1 relevant to their Nitrogen-fixing phosphorus mechanism, etc.

2. Material And Methods

2.1. Nucleotide accession numbers

The complete sequence of *Bacillus mycoides* Gnyt1 strain genome, preserved in the microbiology laboratory of Gansu Agricultural University and plasmid are available in GenBank Database under accession number CM000719, CM000743, and CM000749, respectively.

2.2 Genome alignment and functional analysis

The genome of Gnyt1 strain and reported four strains were using the progressive Mauve algorithm in Mauve(A. E. Darling et al., 2010)by BLAST Generator(N. F. Alikhan et al., 2011)and phylogenetic tree construction based on 16s rDNA. Enrichment analyses and comparison of GO terms, COG categories. etc annotated to the all chromosome were performed, the Gnyt1 strain resistance genes and factors acquired were firstly reported, Finally, based on MUMmer’s genome-wide sequence alignment(W. F. Liu et al., 2018).

3. Results And Discussion

3.1 General genomic features

The whole genome sequence of the strain indicates the total length of sequence after genomic sequence splicing is 5,597,907bp and GC content of 35.57%(Table 1). The strain Gnyt1 ORF prediction is 5,876, the average of ORF is 792.17bp, some plasmids were identified in the genome of *Bacillus* strain Gnyt1 which has 4,654,818bp in ORF total length.

Table 1 General genome features of the *Bacillus mycoides* Gnyt1 strain

Sample	Property	Value
Gnyt1	Seq Length(bp)	5,597,907
	GC Content(%)	35.57
	ORF number	5,876
	Longest ORF (bp)	15,033
	Avg. ORF length (bp)	792.17
	Avg. GC %	36.13
	Coding Region (bp)	4,654,818
	% of genome	83.15

3.2 Comparative gene strain performance

Comparative strains CM000719, CM000743 and CM000749

3.3 Strain genome map construction

The *Bacillus mycoides* Gnyt1 genome map (Fig.1) which genomic size is 5,597.907bp. From inside to outside, the first circle represents the scale(Li H et al., 2009), the second circle represents GC Skew, the third circle represents the GC content, the fourth and seventh laps represent the COG to which each CDS belongs, the fifth and sixth circles represent the location of CDS, tRNA and rRNA on the genome.

Fig. 1 *Bacillus mycoides* Gnyt1 genome circle

3.4 Phylogenetic analysis by 16S rDNA analysis

Each branch represents a species, and the length of the branches represents the evolutionary distance between the two species, in the degree of difference between species(Giardine B et al., 2005). The number (bootstrap value) marked on the node represents the credibility of all the species in the branch to form a branch. Generally, the higher the value, the better the credibility. If the value is greater than 70, the branch is considered to be the branch. More reliable. The ruler represents the degree of difference between the two species(Wang Y et al., 2017).

Annotate the fasta files of 15 strains and analyze it with software MEGA7 to obtain 16s, a phylogenetic tree diagram of rDNA, molecular characterization based on 16S rDNA homology of partial sequences with the available sequences in NCBI(NCBI Resource Coordinators, 2018) database website (<http://www.ncbi.nlm.nih.gov/blast>)(Altschul, S. F. et al., 1997)confirmed the preliminary identification as *Bacillus*(Bodour et al., 2003).The sequences were analyzed correspondly of different *Bacillus* reported

from different parts of the word. Strain Gnyt1 showed 95% homology with *Bacillus mycoides* strain AH621 (CM000719.1) and *Bacillus mycoides* strain AH603(CM000737.1). Similarly, strain Gnyt1 showed 90% homology with *Bacillus thuringiensis* IBL 4222(CM000759.1). In order to track the evolutionary relationship of the tested isolates and find the closest relationship to their sequence in NCBI, using Neighbor-Joining method constructed the phylogenetic tree by mathematical averages (UPGMA) among 16S rDNA sequences of this strains and corresponding sequences of not the same *Bacillus* isolates (Fig.2).

Fig 2. Phylogenetic tree reconstruction based on 16s rDNA of strain *Bacillus mycoides* Gnyt1

3.5 based genome-wide sequence alignment

To clarify the level of genomic disagreement, multiple alignments of four *Bacillus* strains genome sequences were conducted in Mauve software, the results are as follows(Fig 3), in the figure, from top to bottom, they are *Bacillus mycoides* Gnyt1 strain, *Bacillus myciodes* strain AH621, *Bacillus pseudomycoides* strain Rock14 and *Bacillus subblis* strain ATCC19217. The local collinear block (LCB) identified by the Mauve alignment revealed high sequence similarity of the four genomes, indicating that the genomic structure is very conserved in terms of gene identity and sequence. The chromosome of *Bacillus mycoides* Gnyt1 strain showed 62% coverage and 80% identity with that of *Bacillus myciodes* strain AH621 strain and *Bacillus myciodes* strain AH603 strain. This is consistent with the study of other genes from *Lactobacillus sakei* were found in both genomes(Lara Eisenbach et al., 2019; J. Y. Liu et al., 2018; Chaillou S. et al., 2005; Chaillou S et al.,2009).

Fig 3. Global alignment of chromosomes of the strain *Bacillus mycoides* Gnyt1

3.6 Gene family analysis

In the process of genome evolution, a gene produces two or more copies through gene duplication. The sum of these genes is called a gene family(X. Y. Xu et al.,2019; Z. Zhou et al., 2019). Usually, each gene within a gene family has a similar biochemical function.

Firstly, the protein sequence of the reference genome is downloaded and screened according to the length of the protein sequence to remove sequences with a sequence length of less than 50 amino acids. Combine all the protein sequences to be analyzed into one file, build the database based on the data set, and use the data set as a query for all-VS-all blastp analysis. The threshold of the series alignment is set to $1e-10$. The results of the sequence alignment were processed using orthomcl (version 2.0.8) software(D. Fischer et al., 2011), in which the length of the sequence alignment was set to 70%, and the gene family was clustered by MCL. (Inflation) is set to 1.5. Finally, the results obtained by clustering are collated and counted using a self-made Perl script(Fig 4).

To reveal the evolutionary relationships and find the analysis of different genes family by six strains, As shown in Fig.4, Most species have undergone genome-wide replication and large fragment replication during evolution, so most genes are in the form of gene families, which are large and small, as small as two, as large as a few hundred(Y. Wang et al., 2019; H. Zhang et al., 2018; T. Shi, 2016). There are 3322 genes described as the same gene family in six strains, respectively. By contrast, the research strain's role of strain gene family more CM00715.1 and CM000719.1. Interestingly, a high number of genes were grouped in *Bacillus mycoides* strain and *Bacillus cereus* strain, this also corresponds to the main gene function of *Bacillus*(Jean-Philippe Rasigade et al., 2018; Y. Feng et al., 2018; Y. Y. Lin et al., 2017; Y. M. Hu et al., 2018).

Fig 4. Gene family analysis of the strain *Bacillus mycoides* Gnyt1

3.7 Construction of phylogenetic tree based on single copy genes

Proteins from 14 strains were clustered using orthoMCL software, and phylogenetic tree construction was performed using PhyML. The best amino acid substitution model was selected using ProtTest (version 3.2) (Darriba D et al., 2011). A phylogenetic tree(Crisuolo A, 2011)was constructed using the Maximum Likelihood method in PhyML 3.1 software, in which the best model (HKY8 model) obtained by ProtTest was used for the amino acid substitution model. After the completion of the tree, verify the reliability of the phylogenetic tree branch (bootstrap, 1000 replications).

The protein sequences were analyzed correspondly of different *Bacillus* reported from different parts of the world in NCBI (Fig.5). Strain Gnyt1 showed 100% homology with *Bacillus mycoides* strain AH621 (CM000719.1) and *Bacillus mycoides* strain AH603(CM000737.1). Similarly, strain Gnyt1 showed 90% homology with *Bacillus cereus* AH1273(CM000741.1).

Fig 5. Phylogenetic tree of single copy genes

4. Conclusion

Through comparative analyses of the five *Bacillus* strains carrying different genomes, with emphasis on their similarities and differences were identified. Sharing most 16S rDNA analysis, Mauve-based genome-wide, Gene family analysis and phylogenetic tree, the *Bacillus* strain Gnyt1 and *Bacillus mycoides* strain AH621 differ in composition gene expression and expression cause conditions and factors. The nitrogen-fixing gene and the phosphate-dissolving gene of the comparative genomic strain existed and the gene type was significantly different. And the action to nitrogen-fixing gene and phosphate-dissolving gene are dependent on different genes. Sharing some adherence, similarity, and the *Bacillus mycoides* strain Gnyt1 developed evolution is consistent with other *Bacillus*. Related functions absorb different gene-specific

acquisitions, pigments, proteases, regulatory, secretory systems and functional factors associated with nitrogen-fixing phosphorus.

Declarations

Availability of data and materials

All data generated and analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

Xiaomei Yang analyzes the difference and similarity in the genomic evolutionary kinships and contexts of *Bacillus mycoides* Gnyt1. Tuo Yao designed the research and project outline and finalized the manual. Xiaomei Yang and Tuo Yao were genetically analyzed. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31660688) and the Provincial Department of Organizations "Development of Key Technologies and New Products for Microbial Fertilizers" (LYRC2019–113).

References

1. Ahmed, E.A., Hassan, E.A., El Tobgy, K.M., Ramadan, E.M. 2014. Evaluation of rhizobacteria of some medicinal plants for plant growth promotion and biological control. *Ann. Agric. Sci.* 59, 273-280.
2. Aeron, A., Dubey, R.C., Maheshwari, D.K., Pandey, P., Bajpai, V.K., Kang, S.C. 2011. Multifarious activity of bioformulated *Pseudomonas fluorescens* PS1 and biocontrol of *Sclerotinia sclerotiorum* in Indian rapeseed (*Brassica campestris* L.). *Eur. J. Plant Pathol.* 131, 81-93.

3. A.E. Darling, B. Mau, N.T. Perna. 2010. ProgressiveMauve: multiple genome alignment with gene gain; loss and rearrangement, PLoS One 5 e11147. ☐
4. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped Blast and Psi-Blast: a new generation of protein database search programs. Nucl. Acids Res. 25 , 3389-3402.
5. Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S., & Vivanco, J. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual Review of Plant Biology*, 57, 233-266.
6. Bodour, A.A., Drees, K.P., Maier, R.M. 2003. Distribution of biosurfactant-producing bacteria in undisturbed and contaminated arid southwestern soils. Appl. Environ. Microbiol. 69 , 3280-3287.
7. Criscuolo A. 2011. More PhyML: improving the phylogenetic tree space exploration with PhyML 3. Molecular phylogenetics and evolution, 61(3), 944-94.
8. Chaillou S., Champomier-Verges, M.C. Cornet, M. Crutz-Le Coq, A.M. Dudez, et al. 2005. The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. Nat. Biotechnol. 23 , 1527-1533.
9. Chaillou S, Daty M, Baraige, F Dudez A.M., Anglade, et al. 2009. Intraspecies genomic diversity and natural population structure of the meat-borne lactic acid bacterium *Lactobacillus sakei*. Appl. Environ. Microbiol. 75, 970-980.
10. Fischer, S.M. Thomas and C. Beierkuhnlein. 2011. Modelling climatic suitability and dispersal for disease vectors: the example of a phlebotomine sandfly in Europe. Procedia Environmental Sciences. 7, 164-169.
11. Darriba D, Taboada GL, Doallo R. 2011. Posada D: ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics . 27(8), 1164-1165.
12. Glickmann, E., Dessaux, Y. 1995. A critical examination of the specificity of the salkowski reagent for indolic compounds produced by phytopathogenic bacteria. Appl. Environ. Microbiol. 61, 793-796.
13. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15, 1451-5.
14. Lombardi, F. Luzzaro, J.D. Docquier, M.L. Riccio, M. Perilli, et al.. 2002. Nosocomial infections caused by multidrug-resistant isolates of *Pseudomonas putida* producing VIM-1 metallo- β -lactamase, J. Clin. Microbiol. 40, 4051-4055.
15. Zhang, J.J. Jin, L.F. Jin, Z.F. Li, G.Y. Xu, et al. 2002. Identification and analysis of the chloride channel gene family members in tobacco(*Nicotiana tabacum*). Gene. 676, 56-64.
16. J.Y. Liu, L. Yang, L. Li, B. Li, D.Q. Chen, et al. 2002. Comparative genomic analyses of two novel qnrVC6 carrying multidrug-resistant *Pseudomonas. spp* strains. Microbial Pathogenesis. 123, 269-274.
17. Jean-Philippe Rasigade, Florian Hollandt, Thierry Wirth. 2002. Genes under positive selection in the core genome of pathogenic *Bacillus cereus* group members. Infection, Genetics and Evolution, 65, 55-64.

18. Kumar, A., Prasad, S., Singh, S.K. 2014. Screening of free living rhizobacteria associated with wheat rhizosphere for plant growth promoting traits. *Afr. J. Agric. Res.* 13, 1094-1100.
19. Kapulnik, Y., Okon, Y., Henis, Y. 1985. Changes in root morphology of wheat caused by *Azospirillum* inoculation. *Can. J. Microbiol.* 31, 881-887.
20. Liu T, Yao T, Chen JG, Ma WB, Liu H, Ma CY, Jiang YM. 1985. Identification and study on the effects of plant growth promoting rhizobacteria of *Carex enervis*. *Acta Prataculturae Sinica*, 12, 130-139.
21. Li JH. 2017. The characteristics of a fine plant growth-promoting rhizobacteria *Bacillus mycoides* Gnyt1 and its whole genome sequencing analysis. Gansu agricultural university, Doctoral dissertation.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-9.
23. Lara Eisenbach, Andreas J. Geissler, Matthias A, Ehrmann, Rudi F. Vogel. 2019. Comparative genomics of *Lactobacillus sakei* supports the development of starter strain combinations. *Microbiological Research*. 221, 1-9.
24. Mendes, R., Garbeva, P., & Raaijmakers, J. M.. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, 37, 634-663.
25. N.F. Alikhan, N.K. Petty, N.L. Ben Zakour, S.A. Beatson. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons, *BMC Genom.* 12 ,402.
26. NCBI Resource Coordinators. 2018. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 46, D8-13.
27. Orhan, E., Esitken, A., Ercisli, S., Turan, M., Sahin, F. 2006. Effects of plant growth promoting rhizobacteria (PGPR) on yield, growth and nutrient contents in organically growing raspberry. *Sci. Hortic.* 111 ,38-43.
28. Ladhani, Z.A. Bhutta. 1998. Neonatal *Pseudomonas putida* infection presenting as staphylococcal scalded skin syndrome, *Eur. J. Clin. Microbiol. Infect. Dis.* 17, 642-644.
29. Shi. 2016. Impact of gene family evolutionary histories on phylogenetic species tree inference by gene tree parsimony. *Molecular Phylogenetics and Evolution*, 96, 9-16.
30. W.F. Liu, S.Y. Wu, Q. Lin, S.H Gao, F. Ding, et al. 2018. A Reference-based Genome Assembly and Annotation Tool for New Genomes and Upgrade of Known Genomes. *Genomics, Proteomics & Bioinformatics*. 26, 9-12.
31. Wang Y, Song F, Zhu J, Zhang S, Yang Y, et al. 2017. GSA: genome sequence archive. *Genomics Proteomics Bioinformatics* 15, 14-8.
32. X.Y. Xu, Y.P. Zhou, B. Wang, L. Ding, Y. Wang, et al. 2019. Genome-wide identification and characterization of laccase gene family in *Citrus sinensis*. *Gene*, 689, 114-123.
33. Wang, Y.F. Ding, X.M. Wang, H.J. Chen, H.B. Cao. 2019. Analysis of PpGLV gene family suggests that PpGLV4 peptide coordinates auxin and ethylene signaling in peach. *Scientia Horticulturae*, 246, 12-

20.

34. Feng, S. Liu, Y. Jiao, Y. L. Wang, M. Wang, et al. 2018. Gene cloning and expression of the Lasparaginase from *Bacillus cereus* BDRD-ST26 in *Bacillus subtilis* WB600. *Journal of Bioscience and Bioengineering*, 19 , 1-12.
35. Y.Y. Lin, P.T. Chen. 2017. Development of polyhydroxybutyrate biosynthesis in *Bacillus subtilis* with combination of PHB-associated genes derived from *Ralstonia eutropha* and *Bacillus megaterium*. *Journal of the Taiwan Institute of Chemical Engineers*, 79, 110-115.
36. Y.M. Hu, F. Nan, Sarah Wanjiku Maina, J. Guo, J. Guo, et al. 2018. Clone of plipastatin biosynthetic gene cluster by transformation-associated recombination technique and high efficient expression in model organism *Bacillus subtilis*. *Journal of Biotechnology*, 288, 1-8.
37. Zhou, J.L. Zhu, X.C. Zhang. 2019. Genome-wide identification and characterization of the Dof gene family in cassava (*Manihot esculenta*). 687, 298-307.

Figures

Figure 1

Bacillus mycoides Gnyt1 genome circle

Figure 2

Phylogenetic tree reconstruction based on 16s rDNA of strain *Bacillus mycoides* Gnyt1

Figure 3

Global alignment of chromosomes of the strain *Bacillus mycoides* Gnyt1

Figure 4

Gene family analysis of the strain *Bacillus mycoides* Gnyt1

Figure 5

Phylogenetic tree of single copy genes