

Whole-Genome Sequencing of Strain *Bacillus Alitudinis* D30202 With Inhibitory Activity Against Weed

ma xiuhua

Qinghai University

li wei

Academy of Agriculture and Forestry Sciences, Qinghai University

wang jian

Key Laboratory of Qinghai Tibet Plateau biotechnology, Ministry of Education

shen shuo (✉ fjfzss@126.com)

Academy of Agriculture and Forestry Sciences, Qinghai University

Research Article

Keywords:

Posted Date: January 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1234945/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Wild oat (*Avena fatua* L.) is among the most harmful of gramineous weeds infiltrating crops, resulting in reductions in crop quality and yield. The moderately halophilic bacterium *Bacillus altitudinis* D30202 has a good biological control effect on wild oat. In order to explore the biocontrol mechanism of *B. altitudinis* D30202, through whole-genome sequencing analysis, we explored the genome composition of *B. altitudinis* D30202, and conducted research on the prediction of herbicide secondary metabolite synthesis gene clusters to provide a reference for follow-up research on the mechanism of biological control of weeds. The whole gene sequencing results showed that the genome size of the *B. altitudinis* D30202 strain was 3,777,154 bp, the GC content was 41.32%, and there were a total of 3809 coding genes. In addition, The strain can also produce substances that may have herbicidal activity, including 4-hydroxy-3-methoxycinnamic acid and indole derivative. Online predictive analysis found that the strain has 10 secondary metabolite gene clusters. The results of whole-genome sequencing obtained in this study provide an information basis for the research and development of the molecular mechanism of weed inhibition. Genomic information reveals the interaction between several enzyme genes and wild oat, provides a reference for the study of metabolite synthesis and regulatory mechanisms, and lays the foundation for the development of new biological control agents.

Introduction

Bacillus sp., as a gram-positive stress-resistant strain of agricultural resources, has a variety of functions, including the potential to function as a biocontrol agent for corn leaf blight¹. For all kinds of active spinning, textile dyes have strong decolorization ability, and dye wastewater has strong biodegradability²⁻⁴ (Wang, 2013; Rathod, 2018; Wang, 2013; Zarka, 2021). In addition, *Bacillus* also has antibacterial activity and can promote plant growth^{5,6}. As a result, bacteria as a useful resource, have been widely applied in agriculture.

The active strain D30202 was identified as *Bacillus altitudinis* by morphological, physiological, biochemical and molecular biological analyses⁷ (Li, 2019; Rathod, 2018), and D30202 was a moderately halophilic bacterium isolated and purified from Qinghai Qarhan Salt Lake. Part of the moderately halophilic bacterium belong to the *Bacillus* sp., and it is currently the most in-depth study of biocontrol strains. It has been reported that *Paenibacillus polymyxa*, *Bacillus subtilis* and *Bacillus cereus* XG1 have certain herbicidal effects^{8,9}. And some studies have found that *B. altitudinis* has antagonistic effects on *Macrophomina phaseolina*, *Colletotrichum gloeosporioides*, *Xanthomonas oryzae* pv. *oryzicola* (Xoc), *Fusarium oxysporum* f. sp. *niveum* and *Rhizoctonia solani*¹⁰⁻¹³. However, there is no research report on the use of *B. altitudinis* for herbicidal activity at home and abroad.

Currently, weed control is largely dependent on the use of synthetic chemicals. However, the indiscriminate use of large quantities of chemical herbicides has had many harmful ecological consequences. With increasing attention given to the environmental consequences and health-related development of chemical herbicides, the concept of biological herbicides has received increasing attention, and biological herbicide technology is becoming an effective weed solution^{14,15}. Many candidate species of bacteria and fungi have been studied as potential biological herbicides, with a preliminary biopesticide research and development resource bank containing hundreds of active functional strains¹⁶. Wild oat is among the most harmful gramineous weeds

infiltrating crops; it has fast reproduction, strong stress resistance and easily causes disasters and other issues¹⁷. Microorganisms control wild oat mainly through the production of secondary metabolites and agricultural antibiotics, though there have been few studies on their mechanism. For instance, Oloquinine, an anti-grass active substance produced by *Aureobasidium pullulans* PA-2, has obvious inhibitory activity on the germ and radicle of wild oat¹⁸. The crude extract of *Trichoderma polysporum* strain HZ-31 can completely inhibit the germination of wild oat seeds and can cause metabolic changes by inactivating or inhibiting the main crop's enzymes¹⁹. In addition, The active compounds 4-hydroxy-3-methoxycinnamic acid and two indole derivatives produced by the mutant strain have inhibitory activity on the free radicals and coleoptile of the Gramineae *Digtaria sanguinalis* (L.) Scop²⁰. The eco-friendly 2-(hydroxymethyl) phenol produced by *Pseudomonas aeruginosa* (C1501) can significantly reduce the dry weight of *Amaranthus hybridus*²¹. The main sources of microbial herbicides are soil and plant pathogens, while strain D30202 is a moderately halophilic bacterium derived from the Salt Lake environment, preliminary tests have shown that its n-butanol crude extracts has certain herbicidal activity against wild oat, and the inhibitory rates on root length and shoot length at 50 mg/ml are 87% and 69%, respectively.²². Therefore, it is necessary to analyze the secondary metabolites of strain D30202 and its mechanism of action on wild oat and develop its potential as a biological herbicide. The purpose of this experiment was to determine the production of secondary metabolites that are useful for agriculture through whole-genome analysis.

Materials And Methods

Bacterial strains. The herbicide active strain D30202 was isolated from the Qarhan Salt Lake in Qinghai and is now stored in the Institute of Biotechnology, Academy of Agriculture and Forestry Sciences, Qinghai University. The raw sequencing data generated from this study have been deposited in NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>) under accession number SUB10257128.

Strains of total DNA extraction and quality inspection. The purified herbicide active strain D30202 was inoculated on improved ATCC213²³ medium and cultured in an incubator at 37 °C for 48 h. The total DNA of the strain was extracted for purity detection, with Qubit detection as the standard, Nanodrop as the auxiliary detection of DNA concentration, and 1% agarose gel electrophoresis to detect the integrity of the strain's DNA.

Whole-genome sequencing. After the samples were qualified, the library was constructed and checked, the library construction mainly uses the G-tubes method to process genomic DNA into 8-10k fragments, then repair the ends of the DNA fragments and purify the DNA fragments, the whole genome of the bacteria was sequenced by the combination of third-generation PacBio and second-generation Illumina^{24,25}. First, the genome was assembled using third-generation sequencing data after statistical analysis of the reads, and then the assembly results were corrected after subread data statistics were calculated on the second-generation data^{26,27}. Finally, genome component analysis, functional annotation, comparative genome analysis and secondary metabolite gene cluster analysis were carried out based on the corrected assembly results. and uses svg and R software drawing.

Genome composition and gene function annotation analysis. The coding genes were predicted by NCBI²⁸ or Prokka²⁹, the rRNA was predicted by RNAmmer³⁰, the tRNA regions and the secondary structures of the tRNAs were predicted by tRNAscan-SE³¹, the repeat sequences of the bacterial genome were predicted by RepeatMasker³², the tandem repeat sequences of the bacterial genome were predicted by TRF³³, and the gene islands of the genome were predicted using the online tool IslandViewer4 with four different genome island

prediction methods³⁴. Finally, prophage prediction was carried out with Phage_Finder³⁵. The functional databases that provide annotations are Nr (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/>), Swiss-Prot(<https://www.uniprot.org/>), KEGG (<https://www.genome.jp/kegg/>) and COG(<https://www.ncbi.nlm.nih.gov/COG/>). The functional annotation information is obtained by comparing the data with Blast³⁶ and Diamond (<https://github.com/bbuchfink/diamond>). Advanced analysis uses CARD³⁷ and VFDB (<http://www.mgc.ac.cn/VFs/main.htm>) annotations to analyze the types and quantities of virulence factors and resistance genes. In addition, CAZy (<http://www.cazy.org/>) annotations were used to analyze carbohydrate types.

Comparative genome analysis. To determine the similarity between the genes in strain D30202 and other species, analyzed the genes in the single-copy gene family in the whole genome of D30202 and the reference species, the complete whole-genome reference sequences of *B. altitudinis* 6ww6, *Bacillus amyloliquefaciens* FZB42, *Bacillus velezensis* S3-1 and other model strains were downloaded from the NCBI Genome database, and comparative genomic analysis was carried out with the whole-genome sequence of strain D30202, including core gene, specific gene, gene family and evolutionary tree analyses³⁸.

Prediction and analysis of the secondary metabolism gene cluster. The secondary metabolic gene cluster was predicted and analyzed using antiSMASH software online³⁹. Combined with the results of NCBI BLAST comparison analysis, the predicted secondary metabolite synthesis gene cluster was analyzed.

Results

Summary of the sequencing and assembly of strain D30202. Using the combination of third-generation PacBio and second-generation Illumina technology, the whole genome of strain D30202 was sequenced. The sequencing results showed that the whole genome size of the strain was 3,777,154 bp, the average GC content was 41.32%, the total coding genes were 3809, including 3 prophages, and the numbers of tRNAs and rRNAs were 82 and 24, respectively. In addition, the numbers of tandem and interspersed repeats were 63 and 29, respectively (Fig. S1). Their total lengths were 7803 bp and 1957 bp, respectively, and the sequence ratios were both 0.05%. The analysis found that the whole-genome size and average GC content of strain D30202 were within range of *Bacillus* species (Table 1). whole-genome size:1.13–9.5 Mp, GC content: 32.8–70.7%

Table 1 Assembly and component information for the *B. altitudinis* D30202 genomes

| Assembly information | <i>B. altitudinis</i> D30202 |
|--------------------------------|------------------------------|
| Genome size(bp) | 3,777,154 |
| GC content(%) | 41.32% |
| Component analysis information | |
| CDS | 3809 |
| GC content(%) | 41.82% |
| Prophage Number | 3 |
| tRNA Number | 82 |
| 16SrRNA Number | 8 |
| 5SrRNA Number | 8 |
| 23SrRNA Number | 8 |
| Tandem repeat | 63 |
| Interpersed repeat | 29 |

Annotation analysis of basic functions of the strain D30202 genome. Using the Nr, Swiss-Prot, KEGG and COG functional annotation databases, the protein sequences of predicted genes were compared in each functional database by BLAST and Diamond, and a total of 3809 genes were obtained (Fig. S2). Among them, the numbers of annotated genes from Nr, Swissport, COG and KEGG analyses were 3805, 3125, 2774 and 2280, respectively, the number of unannotated genes was 4, and there were 2064 genes annotated by four databases simultaneously. Based on the assembled genome sequence, combined with the prediction results of coding genes and COG annotation, the genome circle map (Fig. S3) and COG annotation functional classification map were drawn (Fig. 1). Among the 3809 genes predicted, there were 3354 COG functional classification genes, including 596 genetic information category genes and 549 cell-related categories, with 54 genes related to defense mechanisms and 171 genes related to signal transduction mechanisms. Moreover, there were 1493 genes related to the metabolic category, only 425 genes predicted by routine function and 291 genes with unknown function.

KEGG metabolic pathway prediction. From the analysis of 134 metabolic pathways annotated by KEGG (Fig. 2), it was found that strain D30202 had a complete herbicide active substance 4-hydroxy-3-methoxycinnamic acid phenylalanine metabolic pathway and herbicide active substance indole derivative tryptophan metabolic pathway (Fig. 3, Fig. S4). At present, Comparison found that strain D30202 has multiple pathways to produce indole acetic acid, which is the downstream substance of indole derivatives. After a detailed search and comparison, the production of 2-(hydroxymethyl) phenol was not found, but through the correlation path search, the upstream precursor 2-hydroxyphenylacetic acid of 2-(hydroxymethyl) phenol was found (Fig. S5). The enzymes and marker genes used in the biosynthesis pathway of 4-hydroxy-3-methoxycinnamic acid were obtained by comparing DIAMOND with the KEGG database. The main enzymes involved are phenylalanine ammonia-lyase ([EC:4.3.1.24], [EC:4.3.1.25]) that catalyzes the direct removal of ammonia on L-phenylalanine to produce trans-cinnamic acid, and the trans-cinnamic acid 4-monooxygenase ([EC:1.14.14.91]) can oxidize DADPH to NADP⁺ and water under the action of oxygen, at the same time, it can hydroxylate the 4-position of the

benzene ring of trans-Cinnamic acid to form 4-hydroxycinnamate. The p-coumarate 3-hydroxylase ([EC:1.14.13.-]) can catalyze the hydroxylation of 3-C of 4-coumarate to generate caffeate (3,4-dihydroxy-trans-cinnamate), Besides, there is also caffeic acid 3-O-methyltransferase([EC:2.1.1.68]), its function is catalyzes the conversion of caffeate to 4-hydroxy-3-methoxycinnamic acid. The marker gene D30202_01823 encodes lyases, D30202_03503 encodes oxidoreductases, D30202_00958 and D30202_01056 encode hydroxylases, D30202_00280, D30202_01060, D30202_01249, D30202_03642, D30202_03556, D30202_03413, D30202_02231, and D30202_02018 encode methyltransferases, and the enzymes encoded by these marker genes promote the production of herbicidal active substances.

Advanced functional analysis of strain D30202 genome

CAZy annotation and classification. CAZy is a resource database that is capable of synthesizing or decomposing complex carbohydrates and sugar complexes. The annotation results show that there are a total of 567 gene encoding enzyme system families in the genome of strain D30202 (Fig. S6, Table S1), including 4 types of auxiliary activity (AA) enzymes 4, 18 types of carbohydrate-binding modules (CBMs) 88, 10 types of carbohydrate esterases (CEs) 42, 51 types of glycoside hydrolysis (GHs) enzyme family proteins 192, 22 types of glycosyl transferase (GT) family proteins 238, and 3 types of polysaccharide lyases (PLs) 3. Wild oat, as an annual oat crop, is rich in protein, carbohydrates and crude fiber. Strain D30202 encodes glycoside hydrolyzed enzymes (family 18, YaaH), endoglucanase ([EC: 3.2.1.4]), type III pantothenate kinase ([EC: 2.7.1.33]), cellulase (EglA) and other enzyme genes, prompting studies to further understand the biological control mechanism of strain D30202.

Virulence factor analysis. The genome of strain D30202 was compared with the VFDB database. A total of 287 virulence factor genes were found in the genome of strain D30202, and the genes had 12 functions (Table S2). The largest gene was 907 bp, and the smallest gene was 29 bp.

Resistance gene analysis. The predicted gene sequences were compared to the database maintained by Resistance Gene Identifier (RGI), and the resistance mechanisms, toxin classifications, types of resistant antibiotics and resistance-related genes were studied. Strain D30202 encodes 44 kinds of genes (321 total genes) labeled as having resistance mechanisms, toxin classifications, types of antibiotic tolerance and resistance-related genes, accounting for 8.4% of the total coding genes, mainly including efflux pump inhibitors, transporters, fluoridequinolones, pyrazinamides, tetracyclines, aminocoumarins, enzymes, cycloaliphatic peptides, and isoniazid antibiotic-related resistance genes, of which there are approximately 20 special genes(Table S3). The *GyrA*, *gyrB* and *parE* genes are involved in fluoroquinolone resistance, the *pncA* and *rpsA* genes are involved in pyrazinamide resistance, the *EF-Tu* genes are involved in elfamycin resistance, the *fusA*, *kasA*, *fabI*, *fabG*, *ndh* and *gidB* genes are involved in antibiotic resistance, and the *pgsA*, *cls*, *liaS*, *mprF*, *liaR* and *gshF* genes are involved in daptomycin resistance. *katG* is an isoniazid resistance gene. The *murA* gene is involved in transferase. And the function of all genes is to alteration the antibiotic target.

Comparative genomics analysis. Comparing the whole-genome information of strain D30202 with those of the other three sequenced *Bacillus* species(Table 2), the whole-genome size of strain D30202 was smaller than those of the other three strains, and the GC content and number of tRNAs were lower than those of strains FZB42 and S3-1. The number of CDSs was close to that of strain 6ww6 but was significantly higher than those of strains FZB42 and S3-1. The number of rRNAs was the same as that of 6ww6 and lower than those of the other two strains.

Table 2 Comparison of genome characteristics of strain D30202 with other strains

| Items | <i>B.altitudinis</i> | <i>B.altitudinis</i> | <i>B.amyloliquefaciens</i> | <i>B.velezensis</i> |
|-----------------|----------------------|-------------------------------|----------------------------|---------------------|
| | D30202 | 2. <i>altitudinis</i> 6ww6 | FZB42 | S3-1 |
| Genome size(bp) | 3777154 | 3804206 | 3918596 | 3929772 |
| GC content(%) | 41.32 | 41.30 | 46.50 | 46.50 |
| CDS | 3809 | 3834 | 3734 | 3748 |
| Number of tRNAs | 82 | 81 | 88 | 85 |
| Number of rRNAs | 24 | 24 | 29 | 27 |

Based on the Core-Pan analysis of the whole-genome sequences of 4 strains, 2564 core genes of strain D30202 and 3 model strains and a total of 518 pan genes were identified(Fig. 4A, Fig. 4B). The specific gene numbers of strains FZB42, S3-1, 6ww6 and D30202 were 203, 148, 212 and 216, respectively, and the number of specific genes of strain D30202 was the highest among the four strains. Gene Family analysis showed that the numbers of Gene Families of the four strains FZB42, S3-1, 6ww6 and D30202 were 3704, 3637, 3806 and 3808, respectively; the number of Gene Families of strain D30202 was the largest, with a total of 2564. The average nucleotide identity (ANI) analysis heat graph (Fig. 5A), homologous gene family Venn graph(Fig. 5B) and phylogenetic tree are shown(Fig. 5C). The genome of strain D30202 had the highest homology with the model strain 6ww6, which was clustered in the same branch and had a genetic distance of approximately 0.01 relative to the other two model strains.

Prediction and analysis of secondary metabolite gene clusters. Through antiSMASH online prediction and NCBI BLAST comparison analysis, it was found that strain D30202 encoded 10 secondary metabolite synthesis gene clusters (Fig. S7, Table 3), of which 2 gene clusters could find identified gene clusters with high similarity, the other 4 gene clusters could find identified gene clusters with low similarity, and 4 gene clusters could not find identified gene clusters. Table 3 shows that the secondary metabolites encoded by strain D30202 are lichenysin, fengycin and bacillibactin in the nonribosomal pathway (NRP), carotenoids in the terpene biosynthesis pathway and bacilysin in other pathways. In addition, there is RiPP:Bottromycin A2 in the Bottromycin pathway. In the comparison with known gene clusters, the similarity of lichenysin- and bacilysin-coding gene clusters was 85%, that of carotenoid-coding gene clusters was 50%, that of fengycin- and bacillibactin-coding gene clusters was 53%. The similarity of Bottromycin A2-coding gene cluster is the lowest, at 6%.

Table 3 Gene clusters of secondary metabolite of *B.altitudinis* D30202

| Type | Region | | Most similar (known)cluster | | | |
|---------------------|---------|---------|-----------------------------|------------------|------------|--|
| | From | To | Products | Pathways | Similarity | Resources |
| NRPS | 362336 | 444093 | lichenysin | NRP | 85% | <i>Bacillus licheniformis</i> DSM 13 |
| RRE-containing | 897947 | 918852 | – | – | – | |
| terpene,siderophore | 1067858 | 1096576 | carotenoid | terpene | 50% | <i>Halobacillus halophilus</i> DSM 2266 |
| betalactone | 1822796 | 1850340 | fengycin | NRP | 53% | <i>Bacillus velezensis</i> FZB42 |
| terpene | 1919077 | 1940954 | – | – | – | |
| T3PKS | 1980564 | 2021661 | – | – | – | |
| betalactone | 2479683 | 2512087 | Bottromycin A2 | RiPP:Bottromycin | 6% | <i>Streptomyces</i> sp. BC16019 |
| other | 3419395 | 3460816 | bacillysin | other | 85% | <i>Bacillus velezensis</i> FZB42 |
| RiPP-like | 3631941 | 3642285 | – | – | – | |
| NRPS | 3710873 | 3760583 | bacillibactin | NRP | 53% | <i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168 |

Note: “–” represents None.

Discussion

Farmland weeds are among the main reasons for the reduction in crop yield. Herbicides are widely used in farmland as the most important means to control weeds. In previous studies, it was found that large-scale weeding mainly depends on chemical herbicides. At the same time, chemical herbicides can cause some environmental pollution problems, increase weed resistance to herbicides, and change weed population structure⁴⁰. Biological herbicides are presently being used as a new weed control method because of their wide herbicide spectrum, environmental protection and long persistence^{41,42}. The discovery of natural herbicides has great potential for the development of new pesticides. Approximately 400 strains of rhizosphere bacteria from five species of weeds and wheat were isolated, and the strains producing phytotoxic substances were screened. It was found that they were phytotoxic to wild oat and little seed canary grass (*Phalaris minor* Retz.), which confirmed their potential as natural herbicides⁴³. Rust fungus (*Puccinia spegazzinii*) can control *Mikania micrantha* invading Papua New Guinea and can be used as a traditional biological control agent⁴⁴. Microbial herbicides control weeds mainly through the production of secondary metabolites. To study the weed inhibition mechanism of herbicides and tap the gene cluster resources of secondary metabolites, it is difficult for traditional experiments and identification methods to comprehensively analyze the weed inhibition substances of *B. altitudinis*, nor can we fully identify all of its weed inhibition genes. Therefore, to more completely strain D30202, this study used a combination of third-generation PacBio and second-generation Illumina and biological analysis

methods to sequence the whole genome of the bacteria to explore the complete genetic information of the organism and reveal its genomic characteristics, metabolic pathways, and gene cluster types.

B. altitudinis D30202 is a moderately halophilic bacterium. Sequencing revealed that its whole genome size is 3,777,154 bp, which is smaller to the genome sizes of *B. altitudinis* 6ww6. Nevertheless, through Core-Pan and Gene Families analysis, it is found that the number of genes between strain 30202 and the model strain is different, indicating that it may produce a different secondary metabolite gene cluster from the model strain, and the secondary metabolite prediction found that strain 30202 contains inhibition of protein synthesis Bottromycin A2⁵⁰. Through the analysis of the assembled data, the genomic composition of the CDSs and RNAs and the types and numbers of repetitive sequences were revealed. According to the Nr annotation results (Fig. S8), strain D30202 was determined to be *B. altitudinis*, which was consistent with the previous strain identification results. Comparing genomes, it was found that the genome of strain D30202 and the model strain 6ww6 were clustered to the same branch, and the homology reached 100% but had a certain genetic distance from the other two model strains. KEGG annotation results are predicted to be consistent with the results of previous studies^{20,21}. Based on the compounds produced by related metabolic pathways, it is speculated that the herbicidal activity of strain D30202 against wild oat may be caused by these active compounds. On this basis, the coding genes were predicted, and it was found that among the 3809 coding genes predicted for strain D30202, there were a total of 567 gene-coding protein domains, with the most predominant being GT family proteins, and the least predominant being PLs, with 238 and 3, respectively. In addition, it was also found that there were 287 virulence factors with 12 functions encoded in the genome. Compared with most *Bacillus* species, this strain had more virulence factors, and the genes related to drug resistance reached 8.4% of the total coding genes.

It was found that *Bacillus* had the ability to inhibit grass and to promote bacteriostasis, growth and biodegradation; however, no related reports on grass inhibition by halophilic bacteria were found especially from the halophilic bacteria in the salt lake environment⁴⁵⁻⁴⁷. The crude extracts of ethyl acetate and n-butanol fermentation broth have obvious inhibitory effects on the root and shoot growth of wild oat, and further research has been performed to clarify their mechanisms of action. In this study, a complete synthesis pathway of 4-hydroxy-3 methoxycinnamic acid and indole derivatives was found, but whether there are more herbicide active substances synthesized will require further experiments and analyses for verification. Through online antiSMASH prediction and NCBI BLAST comparison analysis, it was found that strain D30202 encoded 10 synthetic gene clusters of secondary metabolites. In the comparison with known gene clusters, the similarity of lichenysin-coding gene clusters to *Bacillus licheniformis* DSM13 and *Bacillus velezensis* FZB42 was 85%, while the similarity between bottromycin A2-coding gene clusters and *Streptomyces sp.* BC16019 was the lowest, at only 6%. At present, these secondary metabolite gene clusters have been determined to have antibacterial functions, though there has been no report on their prevention and control of weeds. Further experiments may be required for verification^{48,49}.

Through systematic and comprehensive analyses of the genome data of *B. altitudinis* D30202, the taxonomic status of the species and the number of gene distributions were determined. The prediction of secondary metabolite gene clusters is for the rapid and effective separation of herbicide active substances. And to study the related active substances and mechanisms of weed control, the related genes and synthetic pathways were analyzed in detail. In general, this study provides a reference for the interaction mechanism between strain D30202 and wild oat and provides a convenient source for searches for related biosynthesis and secondary

metabolite regulation genes at later stages. In addition, the analysis of related enzymes and genes confirmed its potential as a biological control agent.

Declarations

Acknowledgements

This work was financially supported by International Cooperation Project of Science and Technology Department of Qinghai Province (No. 2021-HZ-804). We would like to thank Guangzhou Genedenovo Biotechnology Co., Ltd for assisting in sequencing and analysis.

Author contributions

S.S., W.L. and J.W. conceptualized the study; S.S. and X.M. designed and performed the experimental work; S.S., X.M. performed the data analysis; X.M. drafted the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Javandira, C. Aini L Q. Sugiharto A N. et al. THE POTENCY OF *Bacillus* sp. AND *Pseudomonas* sp. AS BIOLOGICAL CONTROL AGENTS AGAINST CORN LEAF BLIGHT DISEASE CAUSED BY *Pantoea* sp. *AGRIVITA Journal of Agricultural Science*. **35**(2), 103–109(2013).
2. Rathod, M. G. & A. P. Pathak. Efficient Decolorization of Textile Dyes by Alkaline Protease Producing Bacterial Consortia. *Indian Journal of Geo-Marine Sciences*. **47**(7), 1468–1477(2018).

3. Wang, Z. W. & J. S. Liang. Decolorization of Reactive Black 5 by a newly isolated bacterium *Bacillus* sp. YZU1. *International Biodeterioration & Biodegradation*. **76**, 41–48(2013).
4. Zarka, B., K. Maryam, C. G. Abbas, M. Ghulam & S. Saba. Evaluation of the Potential Role of *Bacillus altitudinis* MT422188 in Nickel Bioremediation from Contaminated Industrial Effluents. *Sustainability*. **13**(13), 7353–7353(2021).
5. El-Esawi, M. A. et al. *Bacillus firmus* (SW5) augments salt tolerance in soybean (*Glycine max* L.) by modulating root system architecture, antioxidant defense systems and stress-responsive genes expression. *Plant Physiology and Biochemistry*. **132**, 375–384(2018).
6. Feli, F., B. U. Mardhiah, N. Yuana, L. et al. The potentials of secondary metabolites from *Bacillus cereus* SN7 and *Vagococcus fluvialis* CT21 against fish pathogenic bacteria. *Microbial Pathogenesis*. **158**(12), 105062–(2021).
7. Li, W. & S. Shen. Screening and identification of herbicidal active strains. *Journal of Qinghai University*. **37**(06), 11–19 (2019).
8. Wei, L., S. Shuo & G. Qingyun. Identification and Biological Activity of Three Strains Isolated from Distiller's Grains of Highland Barley Wine. *Chinese Journal of Biological Control*. **32**, 544–552(2016).
9. Juan, Y. & Z. Ming-yue. Isolation, identification, and herbicidal activity of *Bacillus cereus* XG1. *Journal of Agricultural University of He Bei*. **39**(02), 81–86(2016).
10. Lamsal, K., Y. S. Lee, S. K. Yun & W. K. Sang. Evaluation of Rhizobacterial Isolates for Their Antagonistic Effects against Various Phytopathogenic Fungi. *The Korean Journal of Mycology*. **44**(1), 36–47(2016).
11. Goswami, M. & S. Deka. Isolation of a novel rhizobacteria having multiple plant growth promoting traits and antifungal activity against certain phytopathogens. *Microbiological Research*. **240**, 126516(2020).
12. Meng-sang, H. & Y. Rui-huan. *Bacillus altitudinis* 181-7, a potential biocontrol agent against bacterial leaf streak of rice. *Acta Phytopathologica Sinica*. **51**(06), 962–974(2021).
13. Satyaprasad, K., Udayini V. Reddy M S. et al. Effect of *Bacillus cereus*, a plant growth promoting rhizobacterium (PGPR) on *Fusarium* root and stalk rot pathogen of sorghum(2011).
14. El-Metwally, I. M. & M. A. El-Wakeel. Comparison of safe weed control methods with chemical herbicide in potato field. *Springer Berlin Heidelberg*. **43**(1), 16(2011).
15. Guan, X., Chen X. Qiu C. et al. Effects of long-term herbicide application on the crops in soybean-peanut rotations in the red soil upland of Southern China. *Field Crops Research*. **248**(2020).
16. Chutia, M., J. J. Mahanta, N. Bhattacharyya, M. Bhuyan, P. Boruah & T. C. Sarma Microbial Herbicides for Weed Management: Prospects, *Progress and Constraints*. *Plant Pathology Journal*. **6**(3), 210(2007).
17. Wei, L. & S. Shuo Bio-herbicidal potential of wheat rhizosphere bacteria on *Avena fatua* L. *grass*. *Bioengineered*. **12**(1), 516–526(2021).
18. Xinya, z. Genome sequencing of *Aureobasidium pullulans* PA-2 and isolation and identification of herbicidal active substances. *Qinghai University*.(2020).
19. Haixia, Z. & M. Yongqiang. Effect of *Trichoderma polysporum* HZ-31 infection on physiological mechanism of *Avena fatua* L. *Acta Agriculturae Zhejiangensis* **33**, 490–496(2021).
20. Li-hui, Z., Z. Jin-lin, L. Ying-chao, C. et al. Isolation and Structural Speculation of Herbicide-Active Compounds from the Metabolites of *Pythium aphanidermatum*. *Journal of Integrative Agriculture*. **12**(06), 1026–1032(2013).

21. Adetunji, C. O., J. K. Oloke, O. M. Bello, M. Pradeep & R. S. Jolly. Isolation, structural elucidation and bioherbicidal activity of an eco-friendly bioactive 2-(hydroxymethyl) phenol, from *Pseudomonas aeruginosa* (C1501) and its ecotoxicological evaluation on soil. *Environmental Technology & Innovation*. **13**, 304–317(2019).
22. Weijia, L. Screening of Strains with Inhibitory Activity against Weed and its Preliminary Inhibitory Mechanisms. *Qinghai University*.(2019).
23. Qianfu, W. Studies on Isolation and Polyphasic Taxonomy of Halophiles. *Lanzhou University*, (2007).
24. Chen, S., Y. Zhou, Y. Chen & J. Gu. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. **34**(17), i884-i890 (2018).
25. Chin, C.-S., P. Peluso & F. J. Sedlazeck. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods: Techniques for life scientists and chemists*. **13**(12), 1050–1054(2016).
26. Mayumi, K., H. Sumitaka, S. Kengo, T. Atsushi, F. Asao & S. Yasubumi. Assembly results with PacBio subreads and Illumina MiSeq reads.(2014).
27. Volant, S., P. Lechat, P. Woringer, L. et al. SHAMAN: a user-friendly website for metataxonomic analysis from raw reads to statistical analysis. *BMC Bioinformatics*. **21**(1), 59–64(2020).
28. Tatiana, T. & D. Michael. NCBI prokaryotic genome annotation pipeline. *Nucleic acids research*. **44**(14), 6614–6624(2016).
29. Torsten, S. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)* **30**(14), 2068–2069(2014).
30. Karin, L., H. Peter, R. E. Andreas, S. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*. **35**(9), 3100–3108(2007).
31. Lowe, T. M. & P. P. Chan. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*. **44**(W1), W54-W57(2016).
32. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*. **25**(1), 4.10.1-4.10.14(2004).
33. G, B. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*. **27**(2), 573–80(1999).
34. Bertelli, Claire, Brinkman, Fiona, S. & L. Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics*. **34**(13), 2161–2167(2018).
35. Fouts, D. E. Phage_Finder: Automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Research*. **34**, 5839–5851(2006).
36. Altschul, S., W. Gish, W. Miller, E. Myers & D. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*. **215**(3), 403–410(1990).
37. Jia, B., A. R. Raphenya, B. Alcock, N. Waglechner & A. G. McArthur. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*. **45**(D1), D566-D573(2016).
38. Riley, R. W., A. Salamov, I. Grigoriev & D. Hibbett. Phylogeny and comparative genome analysis of a Basidiomycete fungi. Ernest Orlando Lawrence Berkeley National Laboratory, *Berkeley, CA (United States)*. (2016).

39. Kai, B., W. Thomas, C. M. G., L. Xiaowen, S. C. J. et al. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Narnia*. **45**(W1), W36-W41(2017).
40. Mekky, M. S., A. M. A. Hassanien, E. M. Kamel & A. E. A. Ismail. Allelopathic effect of *Ocimum basilicum* L. extracts on weeds and some crops and its possible use as new crude bio-herbicide. *Annals of Agricultural Sciences*. **64**(2),(2019).
41. Kang, S.-M. & S. Bilal. Effect of Ammonia and Indole-3-acetic Acid Producing Endophytic *Klebsiella pneumoniae* YNA12 as a Bio-Herbicide for Weed Inhibition: *Special Reference with Evening Primroses. Plants*. **9**(6), 761(2020).
42. Park, J.-M., R. Radhakrishnan, S.-M. Kang & I.-J. Lee. IAA Producing *Enterobacter* sp. I-3 as a Potent Bio-herbicide Candidate for Weed Control: A Special Reference with Lettuce Growth Inhibition. *Indian Journal of Microbiology*. **55**(2), 207–212(2015).
43. Tasawar, A., Z. Z. Ahmad, N. Muhammad, A. M. S., S. E. Mohamed & A. E. Mohamed. Field Performance of Allelopathic Bacteria for Biological Weed Control in Wheat: Innovative, Sustainable and Eco-Friendly Approach for Enhanced Crop Production. *Sustainability*. **12**(21), 8936–8936(2020).
44. M.D., D. Assessing the potential of the rust fungus *Puccinia spegazzinii* as a classical biological control agent for the invasive weed *Mikania micrantha* in Papua New Guinea. *Biological Control*. **67**(2), 253–261(2013).
45. Ap, A., B. Aas & C. Mk. Synergistic adsorption and biodegradation of heavy crude oil by a novel hybrid matrix containing immobilized *Bacillus licheniformis*: Aqueous phase and soil bioremediation. *Ecotoxicology and Environmental Safety*. **222**(2021).
46. Jimtha, J. C., G. E. Mallikarjunaswamy & N. Noushad. Probiotic Rhizospheric *Bacillus* sp. from *Zingiber officinale* Rosc. Displays Antifungal Activity against soft rot pathogen *Pythium* sp. *Current Plant Biology*. **27**, 100217(2021).
47. Salwan, R., V. Sharma, R. Saini & M. Pandey. Identification of plant beneficial *Bacillus* spp. for Resilient agricultural ecosystem. *Current Research in Microbial Sciences*. **2**(2021).
48. Jakline, N., E. A. O, M. M. S, R. M. A & A. R. K. The biosurfactants iturin, lichenysin and surfactin, from vaginally isolated lactobacilli, prevent biofilm formation by pathogenic *Candida*. *FEMS microbiology letters*. **367**(15),(2020).
49. Catherine, N., V. H. Quynh, G. Annika, C. Simon, N. T. T. Thi & M. Jacques. Bacilysin within the *Bacillus subtilis* group: Gene prevalence versus antagonistic activity against gram-negative foodborne pathogens. *Journal of Biotechnology*. **327**(2020).
50. NOBUO TANAKA et al. Inhibition of Protein Synthesis by Bottromycin A2 and Its Hydrazide[J]. *The Journal of Biochemistry*. **60**(4), 405–410(2008).

Figures

COG function classification

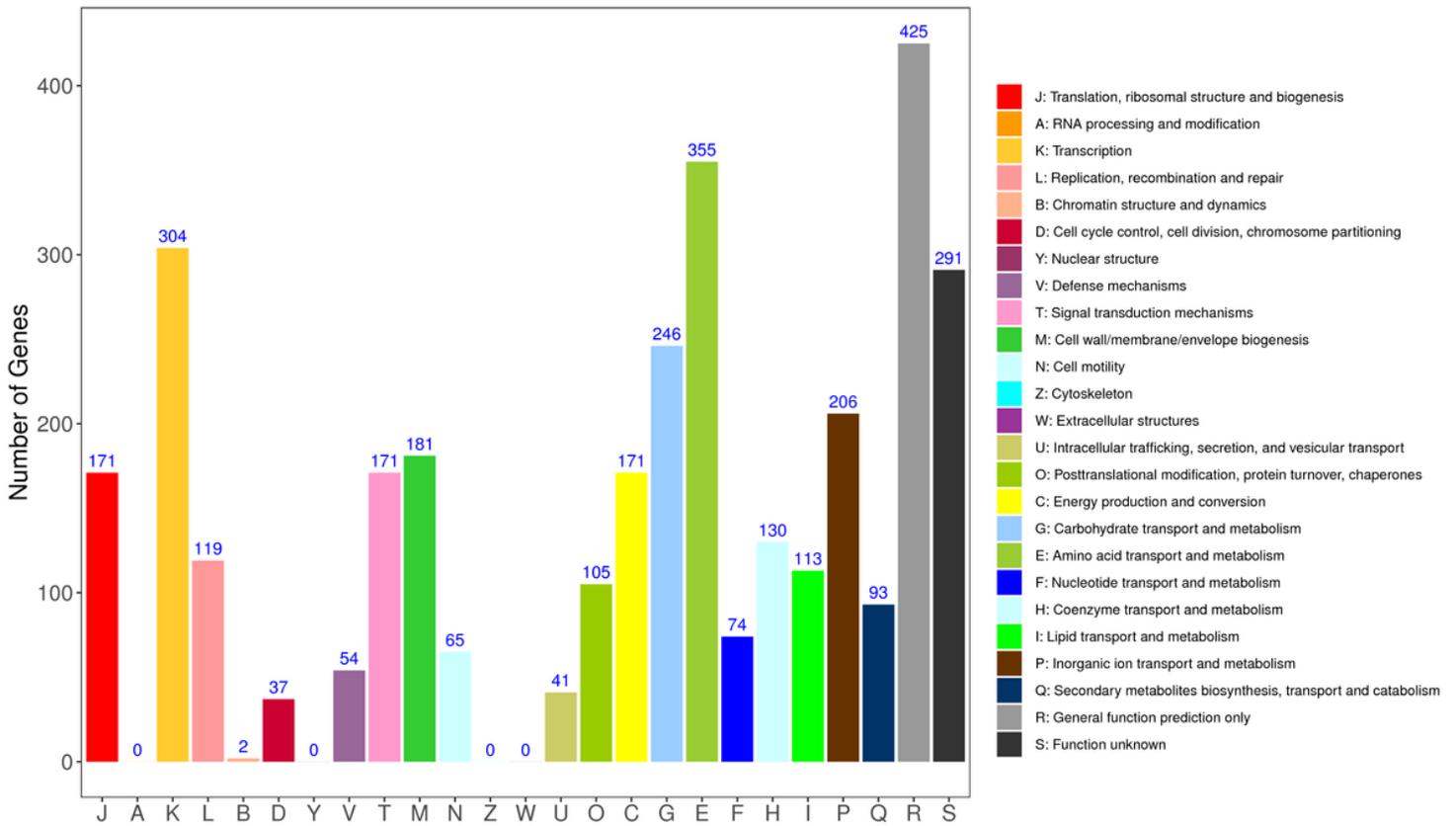


Figure 1

COG annotation functional classification map of the *B. altitudinis* D30202. The horizontal axis represents the function categories, which are displayed in different colors letters, and the vertical axis represents the number of predicted genes included in each function category. The number on the bar graph represents the number of genes on the annotation.

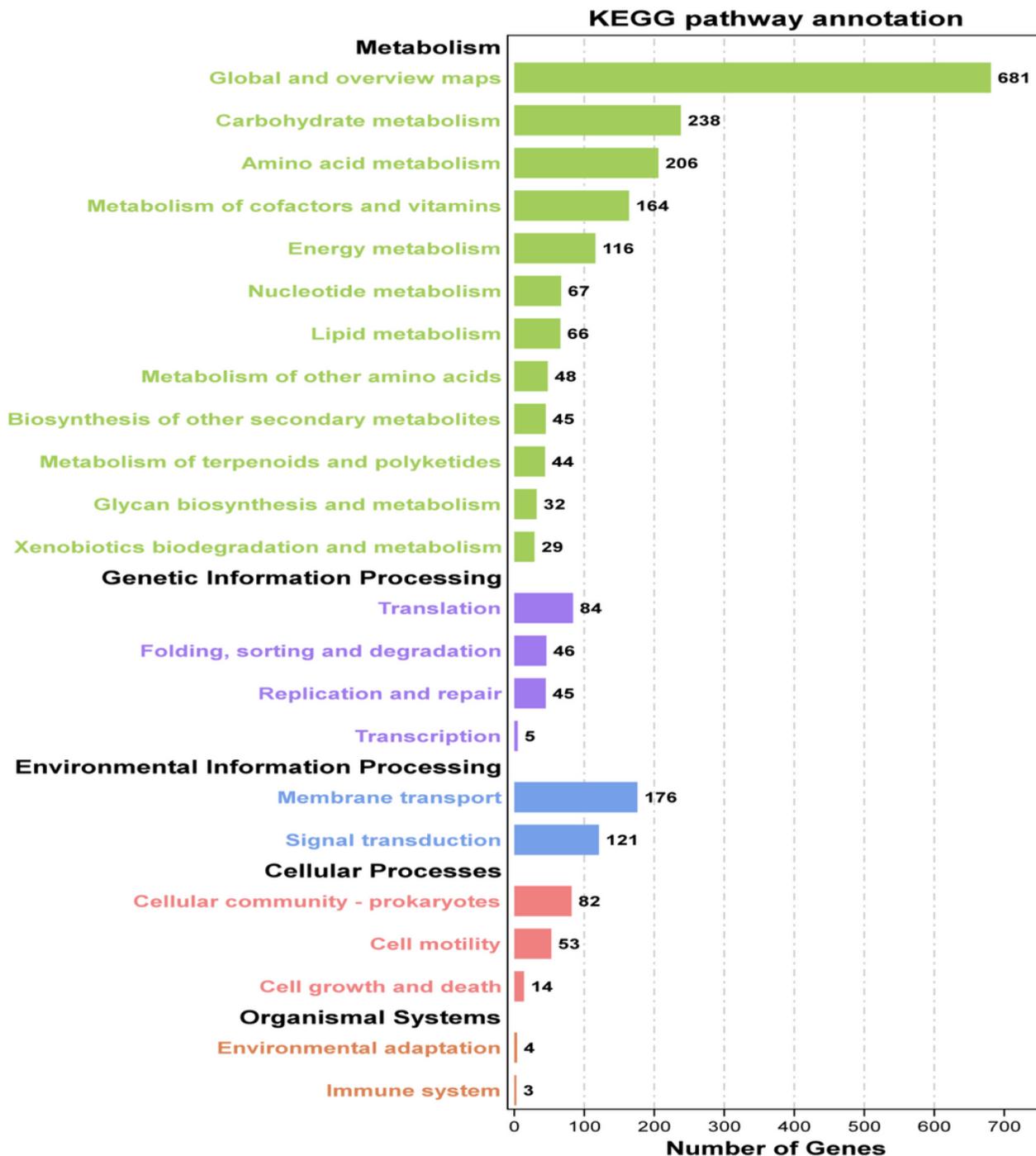


Figure 2

KEGG pathway annotation result statistical map of the *B. altitudinis* D30202, the horizontal axis represents the number of genes, the vertical axis represents the functional classification of the KEGG database, the black font is the level 1 classification, and the color font is the level 2 classification. The number on the bar graph represents the number of genes on the annotation.

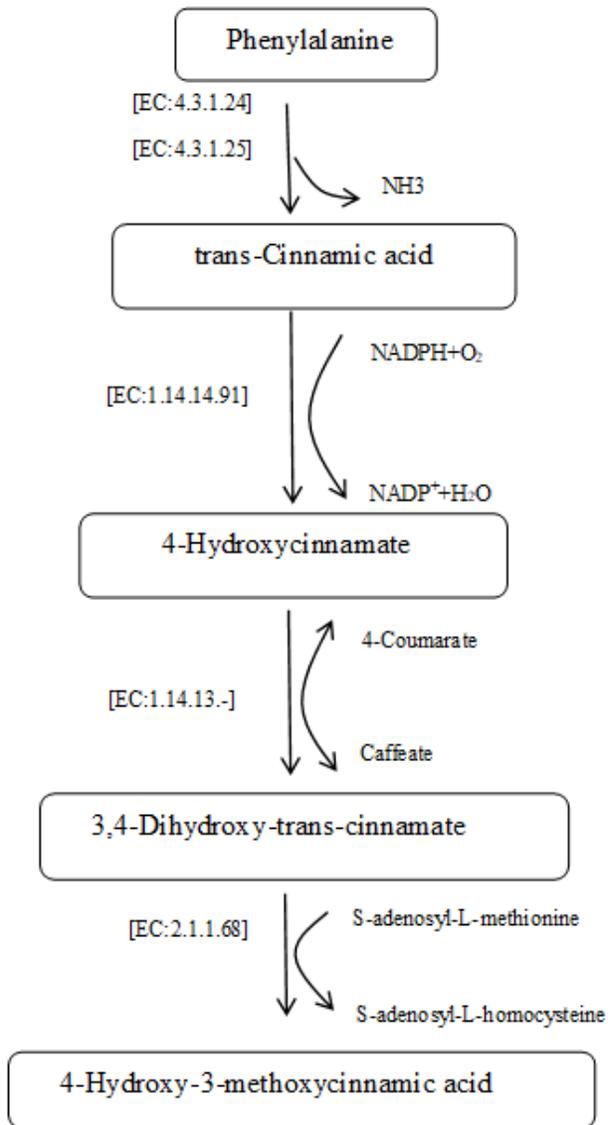


Figure 3

Metabolic pathway prediction of 4-hydroxy-3-methoxycinnamic acid

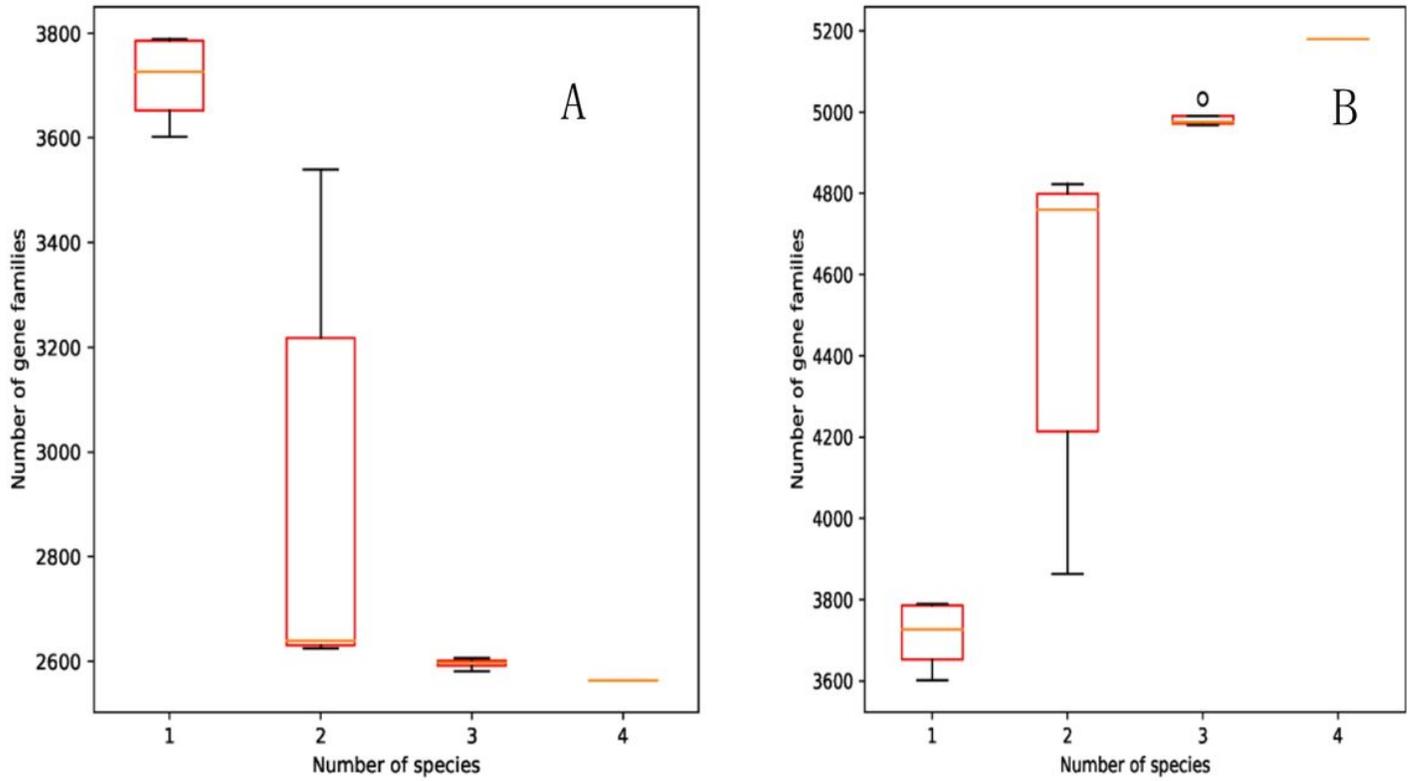


Figure 4

Core-Pan analysis of the whole genome sequence of 4 strains. **A** is Core gene box plot, **B** is Pan Gene Box Plot.. Note: The horizontal axis represents the number of fitted genomes, and the vertical axis represents the number of core/pan genes predicted by the fitting analysis.

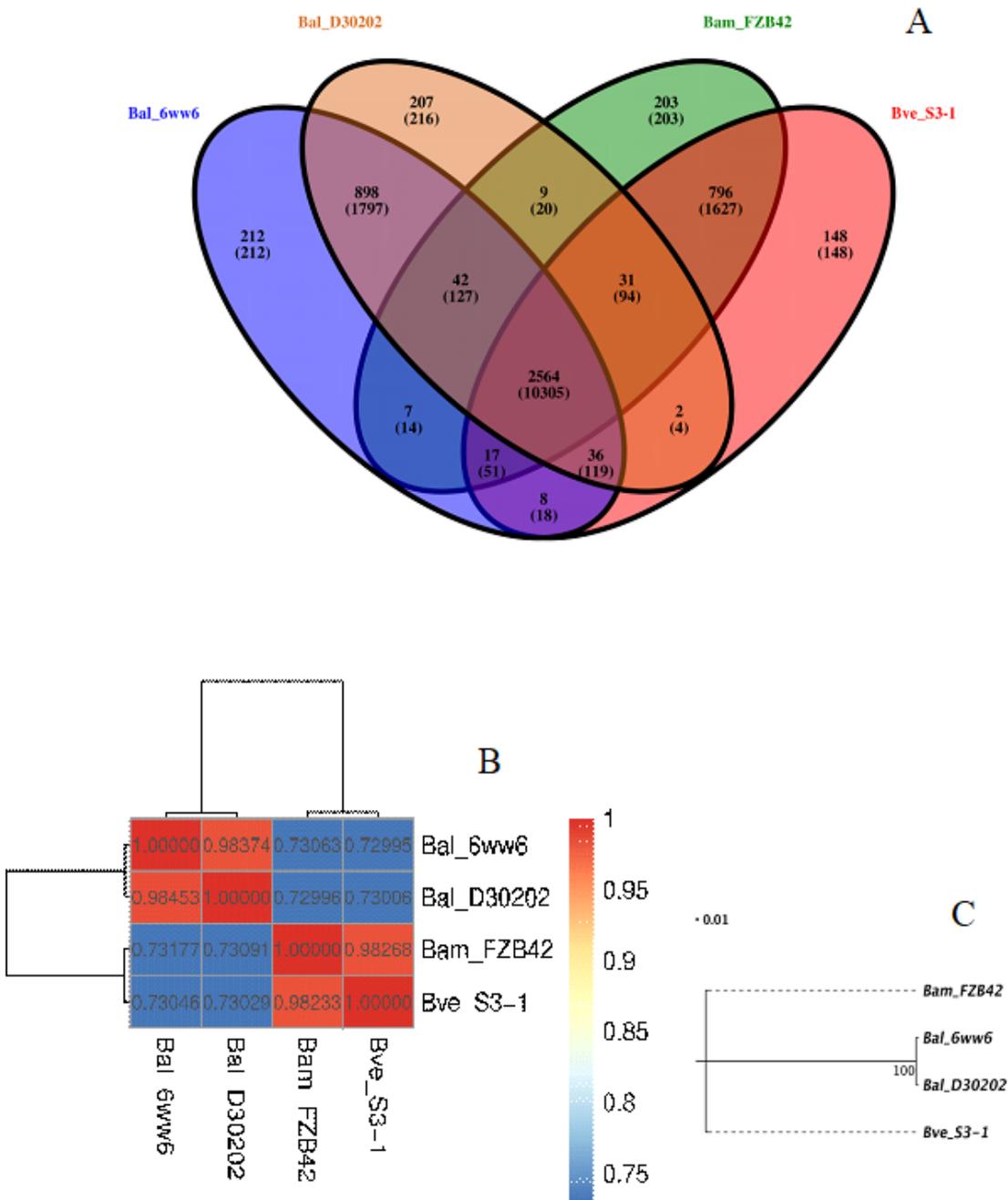


Figure 5

Gene Family analysis of the whole genome sequence of 4 strains. **A** is Venn graph of homologous gene family. Each ellipse represents a genome, the number above each region represents the number of gene families in the species in this region, and the number in parentheses below represents the total number of genes in the gene family in the species in this region. **B** is ANI analysis heat graph, **C** is phylogenetic tree based on characteristic genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure.pptx](#)