Whole-Genome Sequencing of Strain Bacillus Altitudinis 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 (✉ fjzss@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 blight\(^1\). For all kinds of active spinning, textile dyes have strong decolorization ability, and dye wastewater has strong biodegradability\(^2\)\(^-\)\(^4\). 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 analyses\(^7\)\(^\{\}\)\(^{Li, 2019 \#1487; Rathod, 2018 \#930}\), 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 strains\(^16\). Wild oat is among the most harmful gramineous weeds.
infiltrating crops; it has fast reproduction, strong stress resistance and easily causes disasters and other issues. 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 oat. 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 enzymes. 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 *Digitaria sanguinalis* (L.) Scop. The eco-friendly 2-(hydroxymethyl) phenol produced by *Pseudomonas aeruginosa* (C1501) can significantly reduce the dry weight of *Amaranthus hybridus*. 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. 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 ATCC213 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. 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. 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 NCBI or Prokka, the rRNA was predicted by RNAmmer, the tRNA regions and the secondary structures of the tRNAs were predicted by tRNAscan-SE, the repeat sequences of the bacterial genome were predicted by RepeatMasker, the tandem repeat sequences of the bacterial genome were predicted by TRF, and the gene islands of the genome were predicted using the online tool IslandViewer4 with four different genome island
prediction methods. Finally, prephage prediction was carried out with Phage_Finder. 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 Blast and Diamond (https://github.com/bbuchfn/diamond). Advanced analysis uses CARD 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 analyses.

**Prediction and analysis of the secondary metabolism gene cluster.** The secondary metabolic gene cluster was predicted and analyzed using antiSMASH software online. 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**

<table>
<thead>
<tr>
<th>Description</th>
<th>B. altitudinis D30202</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size (bp)</td>
<td>3,777,154</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>41.32%</td>
</tr>
</tbody>
</table>

**Component analysis information**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>3809</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>41.82%</td>
</tr>
<tr>
<td>Prophage Number</td>
<td>3</td>
</tr>
<tr>
<td>tRNA Number</td>
<td>82</td>
</tr>
<tr>
<td>16SrRNA Number</td>
<td>8</td>
</tr>
<tr>
<td>5SrRNA Number</td>
<td>8</td>
</tr>
<tr>
<td>23SrRNA Number</td>
<td>8</td>
</tr>
<tr>
<td>Tandem repeat</td>
<td>63</td>
</tr>
<tr>
<td>Interpersed repeat</td>
<td>29</td>
</tr>
</tbody>
</table>

**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-monoxygenase ([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 ber. 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, fluoroquinolones, 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, fabl, 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

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

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 bacillysin in other pathways. In addition, there is RiPP:Botromycin A2 in the Botromycin pathway. In the comparison with known gene clusters, the similarity of lichenysin- and bacillysin-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 Botromycin A2-coding gene cluster is the lowest, at 6%.

Table 3  Gene clusters of secondary metabolite of B. altitudinis D30202
<table>
<thead>
<tr>
<th>Type</th>
<th>Region</th>
<th>Most similar (known)cluster</th>
<th>Products</th>
<th>Pathways</th>
<th>Similarity</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRPS</td>
<td>From 362336</td>
<td>To 444093</td>
<td>Products</td>
<td>Pathways</td>
<td>Similarity</td>
<td>Bacillus licheniformis DSM 13</td>
</tr>
<tr>
<td></td>
<td>From 1067858</td>
<td>To 1096576</td>
<td>lichenysin</td>
<td>NRP</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>RRE-containing</td>
<td>From 897947</td>
<td>To 918852</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>terpene,siderophore</td>
<td>From 1067858</td>
<td>To 1096576</td>
<td>carotenoid</td>
<td>terpene</td>
<td>50%</td>
<td>Halobacillus halophilus DSM 2266</td>
</tr>
<tr>
<td>betalactone</td>
<td>From 1822796</td>
<td>To 1850340</td>
<td>fengycin</td>
<td>NRP</td>
<td>53%</td>
<td>Bacillus velezensis FZB42</td>
</tr>
<tr>
<td>terpene</td>
<td>From 1919077</td>
<td>To 1940954</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>T3PKS</td>
<td>From 1980564</td>
<td>To 2021661</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>betalactone</td>
<td>From 2479683</td>
<td>To 2512087</td>
<td>Bottromycin A2</td>
<td>RiPP:Bottromycin</td>
<td>6%</td>
<td>Streptomyces sp. BC16019</td>
</tr>
<tr>
<td>other</td>
<td>From 3419395</td>
<td>To 3460816</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>RiPP-like</td>
<td>From 3631941</td>
<td>To 3642285</td>
<td>bacillysin</td>
<td>other</td>
<td>85%</td>
<td>Bacillus velezensis FZB42</td>
</tr>
<tr>
<td>NRPS</td>
<td>From 3710873</td>
<td>To 3760583</td>
<td>bacillibactin</td>
<td>NRP</td>
<td>53%</td>
<td>Bacillus subtilis subsp. subtilis str. 168</td>
</tr>
</tbody>
</table>

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 structure. Biological herbicides are presently being used as a new weed control method because of their wide herbicide spectrum, environmental protection and long persistence. 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 herbicides. Rust fungus (*Puccinia specazzinii*) can control *Mikania micrantha* invading Papua New Guinea and can be used as a traditional biological control agent. 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 A2\(^{50}\). 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 environment\(^{45-47}\). 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](http://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/](http://creativecommons.org/licenses/by/4.0/).

**References**


Figures
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