# WHOLE GENOME SEQUENCING OF *Bacillus altitudinis* AGE-B8, A POTENTIAL CELLULOLYTIC BACTERIUM ISOLATED FROM ROTTED SUGAR CANE BAGASSE

**A. AGEEZ**[1,2,*]

1. *Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), 9 Gamaa St, Giza, 12619, Egypt*

2. *Faculty of Biotechnology, MSA University, 6 October City, Egypt*

* Corresponding author: aageez@msa.eun.eg. Tel: +20 12 2766 0233 Fax: +20 23 568 9519

**Keywords**: lignocellulosic degrading, *B. altitudinis*, Phylogenetic analysis, CMCase, FPase, Whole-genome sequencing, genome annotation,

List of **Abbreviations**

*B. altitudinis*: *Bacillus altitudinis*; CMC carboxymethyl cellulose; CMCase: carboxymethyl cellulase; FPase: filter-paper enzyme; *gry*B: gyrase B subunit; *rpo*B: RNA polymerase β subunit; SCB: Sugarcane Bagasse; *pyc*A: pyruvate carboxylase.

Biofuel produced from agricultural wastes can be an important renewable and sustainable source to replace the depleting fossil fuels (Luo *et al.*, 2019). Successful conversion of lignocellulose waste to fermentable sugars involves three successive steps: (1) size reduction, (2) pretreatment/fractionation, and (3) enzymatic hydrolysis (Kucharska *et al.*, 2018). Enzymatic hydrolysis of natural lignocellulosic materials is considered the most difficult technological challenges that needed to be overcome (Mosier *et al.*, 2005). Due to the limited success of protein engineering to enhance the performance of existing lignocellulose degrading enzymes, retrieving enzymes from naturally evolved biomass-degrading microbial communities offers a promising strategy for the identification of new lignocellulolytic enzymes with potentially improved activities (Rubin, 2008).

Cellulase enzyme cocktails from fungus species are favored over bacteria due to the ability of the former to produce abundant amounts of cellulolytic enzymes and often less complex than bacterial cellulase (Gusakov, 2011). Compared to its counterparts, bacteria have many favored characteristics such as i) it has a higher growth rate than fungi that led to higher

recombinant production of enzymes, ii) bacterial cellulases are more complex that provide increased function and synergy iii) bacteria inhabit a wide variety of environments, which results in cellulolytic strains that are extremely resistant to environmental stresses, and hence able to produce cellulolytic enzymes stable under extreme conditions (Maki *et al.,* 2009; Pham *et al.,* 2019).

Bagasse has various higher-value uses as a second-generation biofuel, fibers for paper, xylan-based products such as xylooligosaccharides, substrate for single-cell protein, and enzymes, or other high-value microbial products (Gebbie *et al.*, 2020; Pandey *et al.*, 2000). Culturing, and metagenomics analysis of bagasse samples has discovered a unique microbial community compared to other lignocellulosic environments, which lead to the identification of new biomass-degrading microbes and enzymes (García-Huante *et al.*, 2017; Mhuantong *et al.,* 2015; Rattanachomsri *et al.,* 2011).

Whole-genome sequencing is an innovative technology to analyze all DNA sequence of an organism efficiently. Whole-genome sequencing technology was used to investigate several potential lignocellulose degrading bacteria during recent years. Full genome sequencing was performed for various bacterial strains such as *Amycolatopsis* sp. strain ATCC 39116 (Davis *et al.,* 2012), *Novosphingobium* sp. strain MBES04 isolated from sunken wood (Ohta *et al.*, 2015), *Pseudomonas* sp. strain YS-1p (Prabhakaran *et al.,* 2016), the *β-proteobacterium Cupriavidus basilensis B-8* (Shi *et al.*, 2013), *Burkholderia* sp. Strain LIG30 isolated from wet tropical forest soil (Woo *et al.,* 2014b), and *Klebsiella* sp. strain BRL6-2 (Woo *et al.,* 2014a). These studies reveal a tremendous amount of high-quality information about the agricultural waste degrading mechanisms adopted by several microorganisms.

The current report describes the isolation of cellulose-degrading bacterial isolate, AGE-B8, isolated from rotted sugar cane bagasse. Various parameters such as pH, incubation temperature, incubation time, and different carbon sources were tested for identifying the best cellulases induction conditions for the isolate. A draft assembly of the bacterial isolate genome was performed. The isolate was molecularly identified using multilocus sequence analysis as *B. altitudinis*. A total of 3,540 genes were identified in the produced bacterial genome assembly.

## MATERIALS AND METHODS

### Bacterial isolation and qualitative screening of their cellulolytic activity.

In this study, cellulose-degrading bacteria were collected from rotted bagasse. Twenty grams of the rotted bagasse were added to 180 ml sterilized distilled water in a 250 ml Erlenmeyer flask and shooked for one hour at 30ºC. One ml of sample suspension was transferred into 9 ml of sterilized distilled water; serial di-

luted to $10^{-7}$, and one hundred µl were plated on nutrient agar plates then incubated at 30ºC for four days. Bacterial colonies were picked and sub-cultured to obtain a pure culture. Cellulase producing bacteria were screened on selective carboxymethyl cellulose agar containing (g/L): $NaNO_3$ 2.0, $KH_2$ $PO_4$ 1.0, $MgSO_4$ .$7H_2O$ 0.5, KCl 0.5, carboxymethyl cellulose sodium salt 10.0, peptone 0.2, agar 17.0. Plates were spot inoculated with a suspension of pure culture and incubated at 30ºC. The endoglucanase activity of the isolated bacterial isolates was tested through the detection of hydrolyzing clear zones on agar plates containing CMC-Na and Iodine staining (Kasana *et al.,* 2008). After 4 days, plates were flooded with gram's iodine stain (2.0 g KI and 1.0 g iodine in 300 ml distilled water) for 3 to 5 minutes. The diameter of the zone of decolorization around each well was recorded.

**Quantitative screening of cellulolytic activity of isolated bacteria**

The pure cultures of bacterial isolates were inoculated with 3 ml (1 x $10^7$ CFU/ml) in modified basal mineral medium (g/L): $KH_2PO_4$, 2g; $(NH_4)_2SO_4$, 1.4g; Urea, 0.3g; $CaCl_2 \cdot 2H_2O$, 0.3 g; $MgSO_4 \cdot 7H_2O$, 0.3 g; peptone , 1.0 g; Tween 80, 0.2%; $CoCl_2 \cdot 6H_2O$, 2.0 mg; $FeSO_4 \cdot 7H_2O$, 5.0 mg; $MnSO_4 \cdot 2H_2O$, 1.6 mg; $ZnSO_4 \cdot 7H_2O$, 1.4 mg; Distilled water , 1000 ml. One gram of treated sugar cane bagasse (Gutierrez-Correa and Tengerdy, 1997) was added into 100 ml of the production medium (pH 5.5) in 250 ml Erlenmeyer flask then incubated at 30ºC with an agitation speed of 180 rpm in a rotary shaking incubator. The cultures filtrate were collected after four days by centrifugation at 6000 rpm for 15 min and the supernatant was used as a source of crude cellulase enzyme (Vega *et al.*, 2012). filter-paper enzyme (FPase) and carboxymethyl cellulase (CMCase) activities were determined using the method of Ghose, (1987). The activity of the enzymes was expressed in U/ml. One unit (IU) of enzyme activity is defined as the amount of enzyme required to release 1µmol of glucose in the reaction mixture per minute under specified conditions.

**Optimization of parameters for enzymatic hydrolysis**

The cellulolytic activities of the bacterial isolates were tested at varying pH ranged from 4-10, varying temperatures ranged from 30ºC - 37ºC, and varying incubation periods ranged from 4-10 days, while all other parameters were kept constant. For carbon source induction experiments; sugar cane bagasse (SCB), as natural cellulose and carbon methylcellulose (CMC), as artificial cellulose, were tested individually as a sole carbon source in a basal mineral media at 30ºC and 170 rpm. The production media was inoculated by 3ml (1 x $10^7$ CFU/ml) and all the flasks were incubated in a shaking incubator at 180 rpm (New Brunswick, USA). Enzyme assay of both FPase and CMCase was determined for each experiment as described before.

**Identification of bacterial isolate using 16S rRNA sequencing**

Genomic DNA of AGE-B8 isolate was purified using GeneJET Genomic DNA Purification Kit (ThermoFisher®, USA). The Primers 27F (AGAGTTTGATCMTGGCTCAG) and 1492R (TACGGYTACCTTGTTACGACTT) were used to amplify nearly full-length *16S rRNA* gene (Youseif *et al.*, 2021). The PCR product was purified, and the sequencing reaction of the PCR product was carried out using an ABI BigDye Terminator V3.1 cycle sequencing kit and ABI PRISM 3500 genetic analyzer (Applied Biosystem, USA).

**Library amplification and genome sequencing of *AGE-B8* isolate.**

Genomic DNA extraction was performed using the GeneJET Genomic DNA Purification Kit (ThermoFisher®, USA). DNA extraction quality was performed using DeNovix DS-11 (DENovix, USA). Sequencing library was prepared by random fragmentation of the DNA, followed by 5' and 3' adapter ligation using truseq nano DNA library kit (Illumina, USA), and the size of PCR enriched fragments was confirmed by electrophoresis on Agilent Technologies 2100 Bioanalyzer (Agilent, USA). For cluster generation, the library was loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. The sequencing was performed using NovaSeq 6000 (Illumina, USA). Library preparation and sequencing were performed at macrogen (Seoul, Korea).

**Genome assembly and genome annotation of *AGE-8B***

FastQC (V0.11.8) was used to check the high throughput sequencing data. Adapter sequences and low quality reads quality filtering of the fastaq raw files was performed using Trimmomatic (Bolger *et al.,* 2014). Map with BWA-MEM (Galaxy Version 0.7.17.1) was used to map the filtered reads to the *Bacillus altitudinis* reference genome (Li and Durbin, 2009). SAMTools was used to manipulate the SAM/BAM files produced from the BWA-MEM software (Li *et al.,* 2009). Variants (SNPs and short indels) were identified by SnpEff (v4.3t) (Cingolani *et al.*, 2012). Genome annotation was performed using Prokka Prokaryotic genome annotation (Seemann, 2014), and RAST (Rapid Annotation using Subsystem Technology (Aziz *et al.*, 2008).

**Phylogenetic analysis**

The *16S rRNA* gene sequencing reads were edited and assembled using Sequencher (Gene Code Corporation, MI, USA). The relevant sequences were obtained using the nucleotide BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and aligned using ClustalW version1.8 (Altschul *et al.*, 1997). The *gry*B, *rpo*B, *pyc*A housekeeping genes were identified from the draft genome assembly, BLASTed in GenBank

(www.ncbi.nlm.nih.gov) using the nucleotide BLAST program, and were aligned using ClustalW version1.8 (Altschul *et al.*, 1997). The maximum likelihood (ML) algorithm (Saitou and Nei, 1987) was selected for performing the phylogenetic analyses. Phylogenetic analyses were performed using MEGA 6.06 (Kumar *et al.*, 2016) using the Tamura-Nei model. Bootstrap support (BT) for each node was evaluated with1000 replicates.

## RESULTS AND DISCUSSION

### Isolation, screening, and identification of most efficient cellulase producing bacteria

Twenty bacterial isolates were acquired based on their ability to grow on basal mineral medium containing sugarcane bagasse pulp as the sole carbon source on solid media. The hydrolyzing zone diameter and colony diameter were recorded for all tested isolates, and only isolates hydrolyzing zone diameter larger than 10 mm, ten isolates, were selected for quantitative analysis. The ten bacterial isolates showing the highest clear zone were further quantitatively screened using DNS (3,5-dinitrosalicylic acid) method (Adney and Nrel, 1996). Bacterial isolates were inoculated in production media containing Sugarcane Bagasse (SCB) as the sole carbon source for seven days at 30ºC, and the induced exoglucanase (FPase) and endoglucanase (CMCase) were recorded (Table 1). The bacterial isolates showed FPase activity in a range from 0.022 to 0.092 U/ml, while the CMCase activity was ranged from 0.31 to 0.527 U/ml. The bacterial isolate AGE-B8 showed the highest CMCase activity with 0.527 U/ml in addition to 0. 092 U/ml for FPase, and was selected for further investigations. The results show a direct relationship between the enzyme concentration and the diameter of the cleared zone in the Congo-red staining method. The results are in accordance with other reports that confirmed this pattern (Bradner *et al.*, 1999; Florencio *et al.*, 2012; Teather and Wood, 1982).

### Optimization of cellulase production

Cellulase production from AGE-B8 isolate was assessed after fermentation at the different parameters. The effect of pH on the induction of exoglucanase (FPase) and endoglucanase (CMCase) was assessed for the AGE-B8 bacterial isolate at 30ºC (Fig. 1 A). The results illustrated by Fig. (1. A) show that both Fbase and CMCase gradually increases as the pH values increased and reached maximum values at pH from 5.5 to 6.0 . At a pH of 5.5, the values were 0.532 and 0.094 U/ml for CMCase and FBase, respectively.

The results illustrated by Fig. (1) clearly show that CMCase production, expressed as enzyme activity, gradually increased as the incubation period increased reaching the maximum (0.72 U/ml ±0.004) at the tenth day, when incubated at 30ºC. The maximum activity for CMCase was 0.713 U/ml ± 0.02 on day 7 when the bacterial isolate incubated at 34ºC, and 0.80 U/ml ± 0.05 when incubated at 37ºC. The FPase production showed a similar pattern. It gradually increased as

the incubation period increased reaching the maximum (0.082 U/ml ±0.004) on the tenth day, when incubated at 30ºC. The maximum activity for FPase was 0.274 U/ml ± 0.04 on day 7 when the isolate incubated at 34ºC, and 0.191 U/ml ± 0.03 when incubated at 37ºC.

Sugarcane bagasse and carboxymethyl cellulose were tested for their effects on cellulase production. Data presented in Fig (2) showed that Bagasse is a better sole source of carbon for cellulase production as cellulase activity was 0.072U/ml for FPase and 0.54 U/ml for CMCase (Fig. 2).

**Identification of *AGE-B8* using *16S rRNA***

Nearly full-length *16S rRNA* gene (1500 bp) was successfully amplified and sequenced. The *16S rRNA* sequence of AGE-B8 was submitted to the GenBank under accession numbers LC593153. Nucleotide BLAST of the *16S rRNA* showed the highest similarity to *B. altitudinis* strain GQYP101 (Accession number CP040514.1), *B. altitudinis* strain BIM (Accession number CP063360.1), *B. altitudinis* strain CHB19 (Accession number CP043559.1), *B. pumilus* strain TUAT1 (Accession number AP044928.1), *B. pumilus* strain SH B11 (Accession number C010997.1P), *B. cellulasensis* strain NJ-M2 (Accession number CP012329.1), *B. cellulasensis* strain GQYP101 (Accession number CP0405141.1), *B. cellulasensis* strain NJ-V2 (Accession number CP012482.1). Maximum likelihood (ML) phylogenetic tree based on *16S rRNA* was generated to determine the evolutionary relatedness of the bacterial isolates AGE-B8 (Fig. 3). According to the sequence similarity, the AGE-B8 isolate was grouped and could not be differentiated from the aforementioned strains (Fig. 3). *B. pumilus*, *B. cellulasensis* and *B. altitudinis*, form one group that can be differentiated from other closely related group. Some *Bacillus* isolates closely related to the *B. pumilus*, are not easily distinguished from *B. cellulasensis* and *B. altitudinis* by *16S rRNA* gene sequence alone, which are nearly identical in *16S rRNA* gene sequence, sharing similarity over 99.5% (Liu *et al.*, 2013).

**Genome assembly and genome annotation of AGE-B8 bacterial isolate**

For sequencing, genomic DNA was extracted, and one paired-ends library was performed. The genome was sequenced using the Novaseq 6000 (Illumina). In total, 15,882,070 reads with 2,398,192,570 bases were generated, achieving about 600X-fold coverage of the *B. pumilus*, *B. cellulasensis* and *B. altitudinis* reference genomes. Trimmomatic was used to remove adapter sequences and low quality reads to reduce biases in further analysis. After quality trimming at Q30, (~93.5%) of total length passed the filtering. Table (2) summarizes the resulting sequencing data and related quality metrics, before and after the quali-

ty trimming. The GC% of the reads was 41.57%.

The reads were submitted to NCBI under project number PRJNA675729, BioSample number SAMN16712403, and SRA accession number SRR13084424. To my knowledge, this is the first draft genome for *Bacillus altitudinis* with ability for lignocellulosic degrading. The reference genome strategy was applied for building the draft genome assembly. The NCBI reference genome of the *B. altitudinis* (accession number GCF_004563755.2) was selected as reference for aligning the reads sequences. More than 94% of the filtered reads could be aligned to it, compared to 85% match in case of *B. pumilus, B. cellulasensis.* The BWA-MEM (Galaxy Version 0.7.17.1) was used to map the filtered reads to the *Bacillus altitudinis* reference genome. The reads bases can be aligned to ninety-two percent of the 3,867,833 bases of the *B. altitudinis* reference genome (Table 3). The average alignment depth was 504.14.

Genome annotation was performed using Prokka Prokaryotic genome annotation (Seemann, 2014). The resulting assembly generated 208 contigs, with 3,575,132 bases, with N50 of 51,914. The assembly identified twenty-two rRNA, 74 tRNA, one tmRNA, and 3,540 coding regions. The functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology (Aziz *et al.*, 2008). The annotated genome revealed 232 genes potentially involved in carbohydrates including 76 Central carbohydrate metabolism genes (Fig. 4). These genes, with the mutations identified within it, can be characterized for the identification of new biomass-degrading enzymes, and testing its efficient conversion processes on bagasse that is characterized by the recalcitrant nature of its polymers.

**Identification of *AGE-B8* using Multilocus Sequence Analysis**

To confirm the taxonomy of the AGE-B8 isolate, phylogenetic analysis based upon the Multilocus Sequence Analysis (MLSA) of *gyr*B, *rpo*B, *pyc*A housekeeping genes was constructed. The sequence of *gyr*B, *rpo*B, *pyc*A housekeeping genes were extracted from the AGE-B8 assembly and aligned to those of bacterial strains that showed the highest similarity of 16S rRNA to AGE-B8 strain. The maximum likelihood (ML) phylogenetic tree based on *gyr*B, *rpo*B, *pyc*A housekeeping genes were generated and confirmed the identification of the AGE-B8 isolated as *B. altitudinis* isolate. The AGE-B8 can be grouped in a separate clade and could be differentiated from the *B. pumilus*, *B. cellulasensis* strains (Fig. 5). The bacteria of the *B. pumilus*, groups are repeatedly misnamed at the species level due to the high similarity in their *16S rRNA* gene sequence. The use of housekeeping genes can be used as a molecular marker to distinguish these closely related strains (Liu *et al.*, 2013). Inside the *B.* subtilis group, the gyrB gene has a better discrimination power than *16S rRNA* (Wang *et al.*, 2007). Also three concatenated housekeeping genes was used to

distinguish closely related rhizobial isolates (Youseif *et al.*, 2014). In this study, the *gyr*B, *rpo*B, *pyc*A housekeeping gene could successfully identify the *B. altitudinis* isolates from the *B. pumilus*, *B. cellulasensis* strains.

**Genetic variation analysis of *AGE-8B* bacterial isolate**

Comparing with the reference genome, *B. altitudinis* (accession number GCF_004563755.2), the produced genome assembly was tested for genetic variation using the SNPeff software. After removing duplicates and identifying variants, information on each variant was gathered and classified. Compared with the reference genome, 375 deletions, 398 insertions, and 60,009 SNP were identified. Supplementary file 1 lists all deletions, insertions, and SNPs along with their annotations (Supp. File 1).

Base changes (DNA substitution) are of two types. Interchanges of purines (A <-> G), or of pyrimidines (C <-> T) are transitions, while interchanges of purine for pyrimidine bases, and vice versa, are transversions. Generally, transversions are more likely to cause amino acid sequence changes. The number of transition (Ts) and transversion (Tv), and the Ts/Tv ratio were calculated using the base change count. Base changes (DNA substitution) are of two types. Interchanges of purines (A <-> G), or of pyrimidines (C <-> T) are transitions, while interchanges of purine for pyrimidine bases, and vice versa, are transversions. A total of 40,521 transitions and 19,488 transversions were identified (Table 4). Generally, transversions are more likely to cause amino acid sequence changes.

To find out the annotation information such as amino acid changes by variants, SnpEff was used. Table (5) shows the number of variants per type (based on the representative transcript). Seventy-eight percent of the identified 53,790 variations were synonymous. Twenty-one percent were missense_variant (Table 5). A list of 60,765 identified variants is listed in supp. file (2).

## SUMMARY

Bagasse is a major source of lignocellulose that is produced in a huge amount in Egypt. It can be used as a second-generation biofuel starting material, fibers for paper, and substrate for high-value microbial products. The present study describes the isolation of cellulose-degrading bacterial strain, AGE-B8, isolated from rotted sugar cane bagasse. AGE-B8 bacterium was tested for the cellulases induction under various parameters such as pH, incubation temperature, incubation time, carbon sources. The best filter-paper enzyme (FPase) activity was obtained after incubation for 7 days at 34ºC, while best carboxymethyl cellulase (CMCase) activity was obtained after incubation for 7 days at 37ºC. The strain was identified using multilocus sequence analysis as *Bacillus altitudinis*. A draft assembly of the AGE-B8 genome was performed. The assembly was submitted to NCBI under project number

PRJNA675729, BioSample number SAMN16712403, and SRA accession number SRR13084424. A total of 3,540 genes were identified in the produced bacterial genome assembly. Gene annotation and genetic variation analysis was performed for the identified isolate. A detailed list of all mutations, in reference to the NCBI *Bacillus altitudinis* reference genome, was identified. The results will significantly contribute to a better understanding of the genetics of the lignocellulosic degrading bacteria.

## REFERENCES

Adney B., and J. Nrel B., (1996). Measurement of cellulase activities. In: Labora-tory Analytical Procedure No. 006. Golden, CO: National RenewableEnergy Laboratory.

Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhanger W., and Lipman D. J., (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. 25:.3389-3402. https: //doi.org/10.1093/nar/25.17.3389

Aziz R. K., Bartels D., Best A., DeJongh M., Disz T., Edwards R. A., and Zagnitko O., (2008). The RAST Server: Rapid annotations using subsystems technology. BMC Genomics, 9:1-15. https://doi.org/ 10.1186/1471-2164-9-75

Bolger A. M., Lohse M., and Usadel B., (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30: 2114-2120.https://doi.org/10.1093/bioinformatics/btu170

Bradner J. R., Gillings M., and Nevalainen K. M. H., (1999). Qualitative assessment of hydrolytic activities in antarctic microfungi grown at different temperatures on solid media. World Journal of Microbiology and Biotechnology, 15: 143-145. https://doi.org/ 10.1023/A:1008855406319

Cingolani P., Platts A., Wang L. L., Coon M., Nguyen T., Wang L., and Ruden D. M., (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6: 80-92. https://doi.org/10.4161 /fly.19695

Davis J. R., Goodwin L. A., Woyke T., Teshima H., Bruce D., Detter C., and Sello J. K., (2012). Genome sequence of Amycolatopsis sp. strain ATCC 39116, a plant biomass-degrading actinomycete. Journal of Bacteriology, (2012): 2396-2397. https://doi.org/10.1128 /JB.00186-12

Florencio C., Couri S., and Farinas C. S., (2012). Correlation between agar plate screening and solid-state fermentation for the prediction of

cellulase production by trichoderma strains. Enzyme Research, 2012. https://doi.org/10.1155/2012/793708

García-Huante Y., Cayetano-Cruz M., Santiago-Hernández A., Cano-Ramírez C., Marsch-Moreno R., Campos J. E., and Hidalgo-Lara M. E., (2017). The thermophilic biomass-degrading fungus Thielavia terrestris Co3Bag1 produces a hyperthermophilic and thermostable β-1,4-xylanase with exo- and endo-activity. Extremophiles, 21: 175-186.https://doi.org/10.1007/s00792-016-0893-z

Gebbie L., Dam T. T., Ainscough R., Palfreyman R., Cao L., Harrison M., and Speight R., (2020). A snapshot of microbial diversity and function in an undisturbed sugarcane bagasse pile. BMC Biotechnology, 20: 1-16. https://doi.org/10.1186 /s12896-020-00609-y

Ghose T. K., (1987). Measurement of cellulase activities. Pure and Applied Chemistry, 59: 257-268. https:// doi.org/10.1351/pac198759020257

Gusakov A. V., (2011). Alternatives to Trichoderma reesei in biofuel production. Trends in Biotechnol-ogy,. (9):.419-425. https://doi.org/ 10.1016/ j.tibtech.2011.04.004

Gutierrez-Correa M., and Tengerdy R. P., (1997). Production of cellulase on sugar cane bagasse by fungal mixed culture solid substrate fermentation. Biotechnology Letters, 19: 665-667. https://doi.org/10.1023/A:1018342916095

Kasana R. C., Salwan R., Dhar H., Dutt S., and Gulati A., (2008). A rapid and easy method for the detection of microbial cellulases on agar plates using Gram's iodine. Current Microbiology, 57: 503-507. https://doi.org/10.1007/s00284-008-9276-8

Kucharska K., Rybarczyk P., Hołowacz I., Łukajtis R., Glinka M. and Kamiński M., (2018). Pretreatment of lignocellulosic materials as substrates for fermentation processes. Molecules, 23: 2937.

Kumar S., Stecher G., and Tamura K., (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution, 33: 1870-1874. https://doi.org/ 10.1093/molbev/msw054

Li H., and Durbin R., (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25: 1754-1760. https: //doi.org/10.1093 /bioinformatics/btp324

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., and Durbin R., (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25: 2078-2079. https://doi.org/ 10.1093 /bioinformatics/btp352

Liu Y., Lai Q., Dong C., Sun F., Wang L., Li G., and Shao Z., (2013). Phylogenetic diversity of the Bacillus pumilus group and the marine ecotype revealed by multilocus sequence analysis. PLoS ONE, 8: 1-11. https://doi.org/ 10.1371/ journal.pone.0080097

Luo X., Liu J., Zheng P., Li M., Zhou Y., Huang L., Chen L. and Shuai L., (2019). Promoting enzymatic hydrolysis of lignocellulosic biomass by inexpensive soy protein. Biotechnology for biofuels, 12: 1-13.

Maki M., Leung K. T., and Qin W., (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. International Journal of Biological Sciences, 5: 500-516. https://doi. org/10.7150 /ijbs.5.500

Mhuantong W., Charoensawan V., Kanokratana P., Tangphatsornruang S., and Champreda V., (2015). Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. Biotechnology for Biofuels, 8(1). https://doi .org/10.1186/s13068-015-0200-8

Mosier N., Wyman C., Dale B., Elander R., Lee Y. Y., Holtzapple M., and Ladisc M., (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresource Technology, 96: 673-686. https://doi. org /10.1016/j.biortech.2004.06.025

Ohta Y., Nishi S., Kobayashi K., Tsubouchi T., Iida K., Tanizaki A., and Hatada Y., (2015). Draft genome sequence of Novosphingobium sp. strain MBES04, isolated from sunken wood from Suruga Bay, Japan. Genome Announcements, 3(1). https://doi.org/10.1128 / genomeA. 01373-14

Pandey A., Soccol C. R., Nigam P., and Soccol V. T., (2000). Biotechnological potential of agro-industrial residues. I: Sugarcane bagasse. Bioresource Technology, 74: 69-80. https://doi. org /10.1016/S0960-8524(99)00142-X

Pham J. V., Yilma M. A., Feliz A., Majid M.T., Maffetone N., Walker J. R., Kim E., Cho H. J., Reynolds J. M., Song M. C. and Park S. R., (2019). A review of the microbial production of bioactive natural products and biologics. Frontiers in microbiology, 10: 1404.

Prabhakaran M., Couger M. B., Jackson C. A., Weirick T., and Fathepure B. Z., (2016). Genome sequences of the lignin-degrading Pseudomonas sp. strain YS-1p and *Rhizobium* sp. strain YS-1r isolated from decaying wood. Genome Announcements, 3(2). https://doi.org/10.1128/genomeA.00019-15

Rattanachomsri U., Kanokratana P., Eurwilaichitr L., Igarashi Y., and Champreda V., (2011). Culture-independent phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles. Bioscience, Biotechnology and Biochemistry, 75: 232-239. https: //doi.org/10.1271/bbb.100429

Rubin E. M., (2008). Genomics of cellulosic biofuels. Nature, 454(7206): 841-845. https://doi .org/10.1038 /nature07190

Saitou N., and Nei M., (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4: 406-425. https:// doi.org/10.1093/oxfordjournals.molbev.a040454

Seemann T., (2014). Prokka: Rapid prokaryotic genome annotation. Bioinformatics, 30: 2068-2069. https://doi.org/10.1093/bioinformatics/btu153

Shi Y., Chai L., Tang C., Yang Z., Zhang H., Chen R., and Zheng Y., (2013). Characterization and genomic analysis of kraft lignin biodegradation by the beta-proteobacterium Cupriavidus basilensis B-8. Biotechnology for Biofuels, 6(1). https:// doi.org/ 10.1186/1754-6834-6-1

Teather R. M. and P. Wood J., (1982). Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Applied and Environmental Microbiology, 43: 777-780. https:// doi.org/10.1128/aem.43.4.777-780.1982

Vega K., Villena G. K., Sarmiento V. H., Ludeña Y., Vera N., and Gutiérrez-Correa M., (2012). Production of Alkaline Cellulase by Fungi Isolated from an Undisturbed Rain Forest of Peru. Biotechnology Research International, 2012: 1-7. https://doi.org/ 10.1155/2012/934325

Wang L. T., Lee F. L., Tai C. J., and Kasai H., (2007). Comparison of gyrB gene sequences, *16S rRNA* gene sequences and DNA-DNA hybridization in the *Bacillus subtilis* group. International Journal of Systematic and Evolutionary Microbiology, 57: 1846-1850.https: //doi.org/10.1099 /ijs.0.64685-0

Woo H. L., Ballor N. R., Hazen T. C., Fortney J. L., Simmons B., Davenport K. W., and DeAngelis K. M., (2014a). Complete genome sequence of the lignin-degrading bacterium *Klebsiella* sp. strain BRL6-2. Standards in Genomic Sciences, 9(1): 19. https://doi .org/10.1186 /1944-3277-9-19

Woo H. L., Utturkar S., Klingeman D., Simmons B. A., DeAngelis K. M., Brown S. D., and Hazen T. C., (2014b). Draft genome sequence of the lignin-degrading Burkholderia sp. strain LIG30, isolated from wet tropical forest soil. Genome Announcements, 2(3). https://doi. org/ 10.1128/ genomeA.00637-14

Youseif S. H., Abd El-Megeed F. H., Ageez A., Cocking E. C., and Sale S. A., (2014). Phylogenetic multilocus sequence analysis of native rhizobia nodulating faba bean (*Vicia faba* L.) in Egypt. Systematic and Applied Microbiology, 37: 560-569. https:// doi.org/10.1016/j.syapm.2014.10.0 01

Youseif Sameh H, El-megeed F. H. A., Mohamed A. H., Ageez A., Veliz E., and Martínez-romero E., (2021). Diverse Rhizobium strains isolated from root nodules of Trifolium alexandrinum in Egypt and symbiovars. Systematic and Applied Microbiology, 44(1): 126156. https: //doi.org/10.1016/j.syapm.2020.12 6156.

**Supporting Files**

- **S1 File**: list of all deletions, insertions, and SNPs along with their annotations in the *Bacillus altitudinis* AGE-B8 genome. https://drive.google.com/file/d/1CfiYq1rrvax0AErYguNcy3MWqbqoJUjd/view?usp=sharing

- **S2 File:** A list of 60,765 identified variants in the *B. altitudinis* AGE-B8 genome. https://drive.google.com/file/d/1BfZJIrZ5g7kHbule5Dg044NgtAn-5aG_/view?usp=sharing

Table (1): Evaluation of Exoglucanase (FPase) activity and Endoglucanase (CMCase) activity of the cellulose-degrading bacterial isolates.

| Isolate code | D/d | FPase (IU/ml) | CMCase (IU/ml) |
|---|---|---|---|
| AGE-B1 | 27/2.1 | 0.037 ± 0.002 | 0.352 ± 0.002 |
| AGE-B2 | 28/2.2 | 0.032 ± 0.001 | 0.425 ± 0.004 |
| AGE-B3 | 18/2.0 | 0.022 ± 0.001 | 0.310 ± 0.003 |
| AGE-B4 | 29/2.1 | 0.075 ± 0.01 | 0.499 ± 0.002 |
| AGE-B5 | 24/2.3 | 0.042 ± 0.004 | 0.400 ± 0.003 |
| AGE-B6 | 23/1.8 | 0.056 ± 0.002 | 0.325 ± 0.001 |
| AGE-B7 | 24/2.2 | 0.053 ± 0.003 | 0.325 ± 0.004 |
| AGE-B8 | 31/1.9 | 0.092 ± 0.001 | 0.527 ± 0.008 |
| AGE-B9 | 17/2.0 | 0.025 ± 0.002 | 0.410 ± 0.005 |
| AGE-B10 | 29/2.1 | 0.086 ± 0.001 | 0.458 ± 0.002 |

"*D*/*d*": hydrolyzed zone diameter/colony diameter, in millimeter, on agar media containing CMC as a sole carbon source.

Table (2): Raw data Stats before and after trimming using the trimmomatic software.

| | Total read bases (bp) | Total number of reads | Total reads GC(%) | Q20(%) | Q30(%) |
|---|---|---|---|---|---|
| Before trimming | 2,398,192,570 | 15,882,070 | 41.58 | 96.20 | 90.73 |
| After trimming | 2,170,161,529 | 14,675,394 | 41.57 | 98.29 | 93.54 |

Table (3): Mapped data Stats.

| Ref. Length (GCF_004563755.2) | Mapped Sites (>=1x) | Total Reads | Mapped Reads | Mapped Bases Mean Depth |
|---|---|---|---|---|
| 3,867,833 | 3,586,315 (92.73%) | 14,675,394 | 13,813,626 (94.13%) | 504.14 |

Table (4): Transition, Transversion information table.

| Ref. Length | Number of SNPs | Transition | Transversion | Ts/Tv |
|---|---|---|---|---|
| 3,867,833 | 60,009 | 40,521 | 19,488 | 2.08 |

Table (5): Annotation type count.

| Type of annotation | Count | Ratio |
|---|---|---|
| synonymous_variant | 42,117 | 78.3% |
| missense_variant | 11,398 | 21.19% |
| stop_gained | 53 | 0.1% |
| frameshift_variant | 50 | 0.09% |
| downstream_gene_variant splice_region_variant | 40 | 0.07% |
| and stop_retained_variant | 31 | 0.06% |
| conservative_inframe_insertion | 18 | 0.03% |
| start_lost | 14 | 0.03% |
| conservative_inframe_deletion | 14 | 0.03% |
| upstream_gene_variant | 10 | 0.02% |

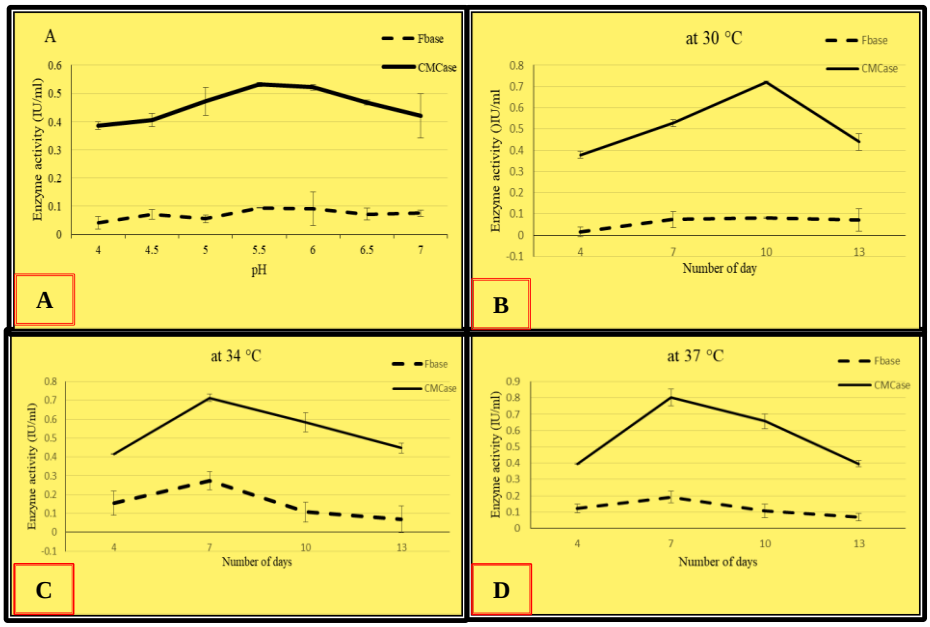


Fig. (1): Effect of pH and incubation time at different incubation temperatures on the activity of FPase and CMCase for *B. altitudinis* AGE-B8. (A): Effect of different pH, at 30ºC, on the production of cellulases by *B. altitudinis* AGE-B8. Effect of incubation period on the production of cellulases at incubation temperature 30ºC (B), 34ºC (C) and 37ºC (D).

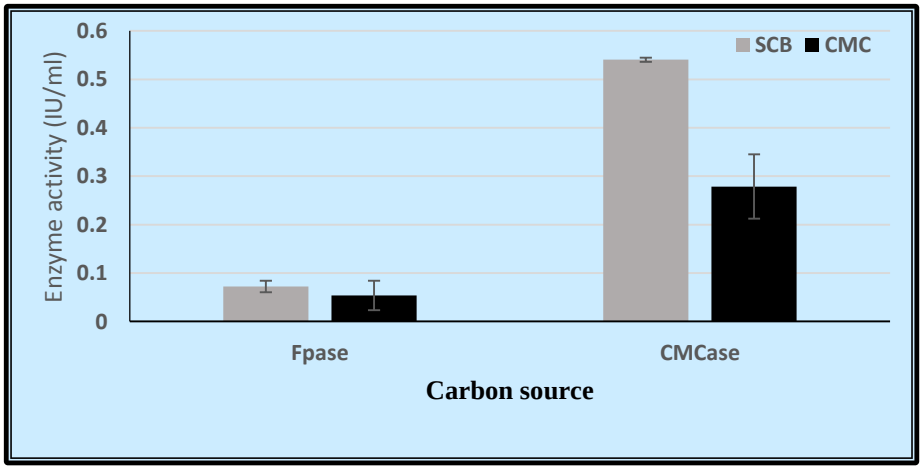


Fig. (2): Effect of SCB and CMC as sole carbon sources for the induction of FPase and CMCase.

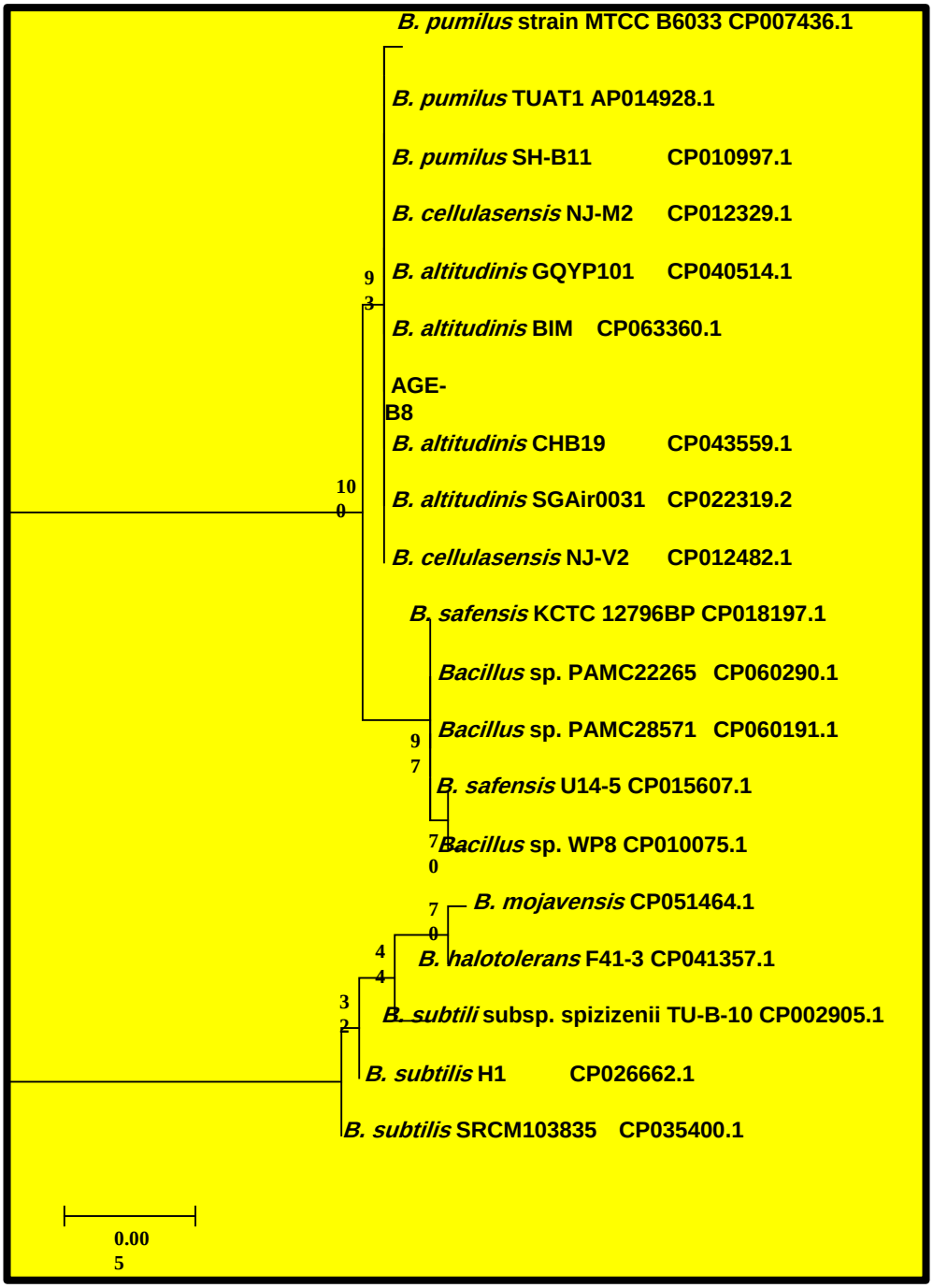


Fig. (3): Phylogenetic tree based upon the *16S rRNA* sequences obtained by maximum likelihood (ML) method. GenBank accession numbers are indicated next to the strain name. Bootstrap values were calculated from 1000 trees compared with the sequence of standard strains. Bootstraps are shown at the branching points.

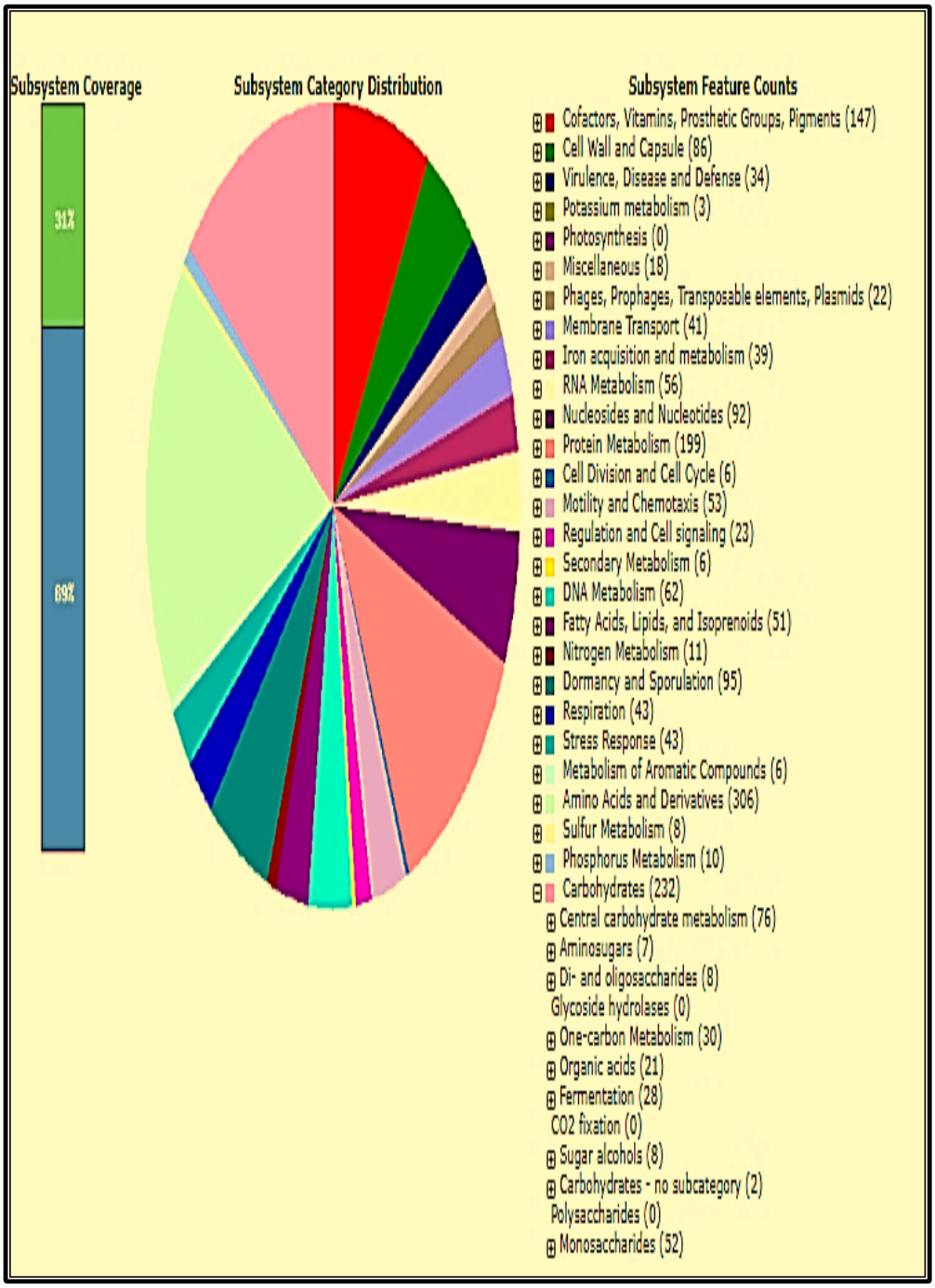


Fig. (4): Subsystem distribution of AGE-8B genes based on RAST annotation server.

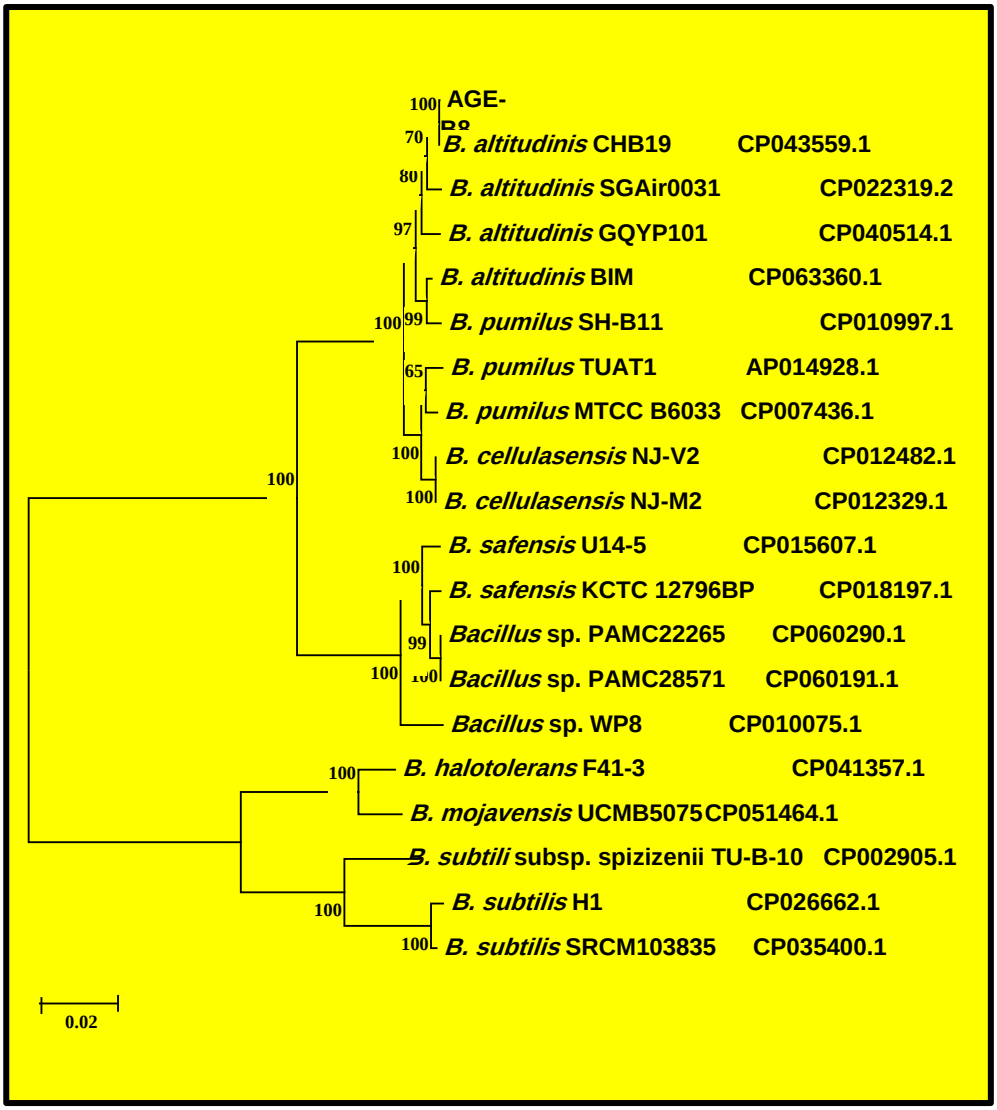


Fig. (5): ML tree based on three concatenated *gyr*B, *rpo*B, *pyc*A housekeeping genes. GenBank accession numbers are indicated next to the strain name. Bootstrap values were calculated from 1000 trees compared with the sequence of standard strains. Bootstraps are shown at the branching points.