Genome sequence of the plant-growth-promoting bacterium Bacillus velezensis EU07

Preprint DOI: Posted on April 17, 2024
https://doi.org/10.1099/acmi.0.000762.v2

Data note: Genome sequence of the plant-growth-promoting bacterium *Bacillus velezensis* EU07

1.1 Author names

Ömür Baysal [1,2] (ORCID: 0000-0001-5104-0983); David J. Studholme [3] (ORCID: 0000-0002-3010-6637); Catherine Jimenez-Quiros [2] (ORCID: 0000-0002-8306-7109); Mahmut Tör [2] (ORCID: 0000-0002-4416-5048)

Corresponding author: DJS

1.2 Affiliation(s)

1 Faculty of Science Department of Molecular Biology and Genetics, Mugla Sitki Kocman University. Turkey.

2 Department of Biological Sciences, University of Worcester. United Kingdom.

3 Biosciences, University of Exeter. United Kingdom.

1.3 Corresponding author and email address

David Studholme: d.j.studholme@exeter.ac.uk

1.4 Keywords

Bacillus velezensis; Biological control; Plant growth promoting; Genome sequence

1.5 Repositories:

- BioProject PRJNA743875: https://www.ncbi.nlm.nih.gov/bioproject/743875
- Assembly GenBank accession number: GCA_019997305.2: https://www.ncbi.nlm.nih.gov/nuccore/JAIFZJ000000000
- NCBI RefSeq accession number: GCF_019997305.2
- Sequence Read Archive accession number: SRX22864526
- https://github.com/davidjstudholme/bacillus_EU07
- https://doi.org/10.5281/zenodo.10968102.

2. Abstract

Many Gram-positive spore-forming rhizobacteria of the genus *Bacillus* show potential as biocontrol biopesticides that promise improved sustainability and ecological safety in agriculture. Here we present a draft-quality genome sequence for *Bacillus velezensis* EU07, which shows growth-promotion in tomato plants and biocontrol against *Fusarium* head blight. We found that the genome of EU07 is almost identical to that of the commercially used strain QST713 but identified 46 single-nucleotide differences that distinguish these strains from each other. The availability of this genome sequence will facilitate future efforts to unravel the genetic and molecular basis for its beneficial properties.

3. Data summary

In this study, we generated genome sequence data, which has been deposited in public databases:

- BioProject PRJNA743875: https://www.ncbi.nlm.nih.gov/bioproject/743875
- Assembly GenBank accession number: GCA_019997305.2: https://www.ncbi.nlm.nih.gov/nuccore/JAIFZJ000000000
- NCBI RefSeq accession number: GCF_019997305.2
- Sequence Read Archive (SRA) accession number: SRR27184279

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

4. Introduction

Many Gram-positive spore-forming rhizobacteria of the genus *Bacillus* show potential as biocontrol biopesticides that promise improved sustainability and ecological safety in agriculture [1–3]. Here, we present genomic sequencing data for *Bacillus* strain Egem-Utku 07, hereafter known as EU07. This strain was previously isolated from the rhizosphere of diseased tomato plants [4] in an effort to collect strains that could inhibit the soilborne pathogen *Fusarium oxysporum* f. sp. *radicis-lycopersici* [4], which causes crown rot on tomato. We demonstrated that EU07 inhibits this pathogen *in vitro* [4]. Furthermore, EU07 promotes growth and inhibits fusarium head blight *in planta* [5]. We

previously established that EU07 is a member of the genus *Bacillus*, but its precise species identity was ambiguous. Furthermore, in the absence of sequence data, little was known about the potential molecular mechanisms for its beneficial properties. Here, we present a draft-quality genome sequence assembly and genomic sequence reads from strain EU07). This dataset will help to better understand EU07's phylogeny and taxonomy and provide a resource to assist elucidation of the molecular mechanisms of its beneficial traits.

5. Materials and methods

5.1 Bacterial strain and isolation of genomic DNA

We isolated genomic DNA from bacterial strain EU07 from fresh liquid culture grown for 24-hour in Nutrient Broth pH 7.2. We note that this medium provides a laboratory environment quite different from the bacterium's normal soil environment. The liquid culture was inoculated from a single colony and therefore was assumed to be clonal. We used the ISOLATE II Genomic DNA Kit (Bioline), following the manufacturer's instructions. The quality and concentration of the gDNA were assessed using the NanoDrop 2000c (ThermoFisher Scientific).

5.2 DNA sequencing

Genomic DNA was sent to the University of Exeter's Sequencing Facility (https://biosciences.exeter.ac.uk/sequencing/) for Illumina Nextera XT library preparation and sequencing on the Illumina MiSeq platform to generate 748,528 pairs of 300-bp reads with a mean insert size of approximately 400 bp.

5.3 Genome sequence assembly

We performed adapter trimming and quality filtering on the MiSeq reads using Trim Galore version 0.6.7 [6], which incorporates Cutadapt version 3.5 [7]. The `-q` parameter was set to 30 and we used the `--paired` option. The resulting cleaned read-pairs served as input for *de-novo* assembly using SPAdes version 3.13.1 [8] with the `--careful` option. The resulting scaffolds and contigs were re-ordered against the reference genome of strain FZB42 with the Mauve Contig Mover [9]. Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline version 6.6 [10] after submission of the genome assembly. The command lines are documented on GitHub at https://github.com/davidjstudholme/bacillus_EU07/tree/main/assembly and in the Zenodo repository [11].

5.4 Assessment of genome-assembly quality

We calculated assembly statistics using QUAST version 5.2.0 [12]. We checked read coverage of the genome assembly by aligning the EU07 reads against the EU07 assembly and calculating alignment statistics with Qualimap version 2.3 [13]. The alignment was performed using BWA-mem version 0.7.17 [14]; then we reformatted and sorted the output using SAMtools version 1.13 [15]. The full details of the command lines are documented at https://github.com/davidjstudholme/bacillus_EU07/blob/main/assemblyQC/README.md and in the Zenodo repository [11].

5.5 Average nucleotide identity (ANI)

We used fastANI [16] to calculate average nucleotide identity (ANI) between the genome of EU07 and each of the *B. amyloliquefaciens* group (taxonomy ID: 1938374) genome assemblies retrieved from GenBank [17, 18]. The exact command lines are documented on GitHub at https://github.com/davidjstudholme/bacillus_EU07/ and the Zenodo repository [11]

5.6 Phylogenomics

To generate a maximum-likelihood phylogenetic tree based on genome-wide single-nucleotide polymorphisms (SNPs), we used PhaME [19] with FastTree [20]. The exact command lines used are documented at https://github.com/davidjstudholme/bacillus_EU07/ and the Zenodo repository [11]. The resulting tree was rendered using the Interactive Tree of Life (IToL) 6.8.1 [21]

5.7 Whole-genome alignment

Genome sequences were aligned using progressiveMauve version 2.4.0 [22] after first re-ordering the contigs against the reference genome of strain KNU-28 [23] with the Mauve Contig Mover [9]. The resulting alignment was visualised using Mauve snapshot_2015-02-25 [24]. The exact command lines used are documented at https://github.com/davidjstudholme/bacillus_EU07/ and the Zenodo repository [11].

5.8 Further whole-genome analyses

We used the Proksee web server [25] to perform several analyses of the assembled EU07 genome. This included BLASTN searches against 888 related genomes, annotation of horizontally acquired genomic regions with Alien Hunter [26] and identification of bacteriophage sequences using

VirSorter [27, 28] and Phigaro [29]. Variant-calling was performed using the Parsnp tool in Harvest [30].

6. Results and Discussion

6.1 Genome sequencing and assembly

We generated 748,528 pairs of 300-bp Illumina MiSeq sequencing reads from EU07 genomic DNA. This represents approximately 100X coverage of the 4.2-Mb genome. Trimming and filtering with Trim Galore left 715,442 pairs of reads, with lengths ranging from 20 to 300 bp. *De-novo* assembly with SPAdes yielded 266 contigs with a total length of 4.2 Mbp and N_{50} length of 52.8 kb. This was deposited in GenBank via the NCBI Submission Portal under accession number GCA_019997305.2. The NCBI's contamination filtering removed five contigs, leaving 261. The NCBI PGAP annotation system predicted 4273 genes, of which 4081 encode putative proteins. The results of NCBI's quality check with CheckM v1.2.2 [31, 32] revealed a completeness of 98.16 % (85^{th} percentile) and 0.47 % contamination.

Alignment of sequencing reads against the genome assembly and analysis with Qualimap revealed a mean coverage of 93.25 X and standard deviation of 89.87. Almost all of the genome assembly (99.96% had at least 1 X coverage and 97.59% of the assembly has at least 10 X coverage. The full Qualimap report and output files are available in the Zenodo repository [11], allowing users of this data to take coverage into account when performing analyses. We note that the contig with least coverage is JAIFZJ020000237.1, having only 1.04 X coverage. Nevertheless, BLAST searches reveal that this contig shows very high levels of sequence similarity to genomes of other *Bacillus velezensis* strains, increasing confidence in its validity.

6.2 EU07 belongs to the species *Bacillus velezensis*

Previously, the phylogenetic and taxonomic position of strain EU07 had been ambiguous and we previously referred to it a '*B.* sp.' and '*B. subtilis*' [4, 5]. To identify the species to which strain EU07 belongs, we uploaded the genome assembly to the Type Strain Genome Server (TYGS) [33]. This classified EU07 to the species with *Bacillus amyloliquefaciens*. Among the sequenced type strains in TYGS, the most similar to EU07 was FZB42 [34], which is the type strain of *B. amyloliquefaciens*

subsp. *plantarum* [35]. However, this taxon is now considered to be synonymous with *B. velezensis* and distinct from *B. amyloliquefaciens* [36]. Hereafter, we refer to our strain as *B. velezensis* EU07.

6.3 EU07 belongs to a clade of plant-associated strains of *B. velezensis*

To identify previously sequenced similar genomes, we calculated average nucleotide identity (ANI) between *B. velezensis* EU07 and all 888 genome assemblies available in GenBank for *B. amyloliquefaciens* group. This revealed that EU07 shares more than 99.9 % ANI with 13 previously sequenced genomes. Table 1 lists the genomes showing the highest levels of ANI to that of *B. velezensis* EU07. This includes strains that previously have been classified variously as *B. amyloliquefaciens* or *B. velezensis*. However, they all fall within the *B. velezensis* clade [36–38] and should be considered to belong to that species. To further elucidate the evolutionary relationships of EU07, we generated a phylogenomic tree including these closely related strains and the relevant type strains; this is presented in Figure 1. Consistent with the ANI results, strain EU07 falls within a clade that includes the same 13 strains that showed greatest ANI with EU07. Alignment of these genomes with Mauve (Figure 2) reveals extensive conservation and co-linearity of the chromosome sequence among these strains. Comparison of the EU07 chromosome versus the genome sequences of related strains, as shown in Figure 3, revealed that most of the presence—absence polymorphism was associated with loci predicted to originate from bacteriophage genomes.

Among the strains closely related to EU07 are several that have previously been described as having growth-promoting and/or pathogen-inhibitory properties. For example, strain BS006 was isolated from roots of *Physalis peruviana* in Colombia and promotes growth in banana [39]. Strain KNU-28 was isolated from peach leaves in Korea [23]. Strain ALB79 was isolated from grapes in northern California and shown to inhibit the growth of *Listeria monocytogenes* in vitro [40], while strain QST713 is used commercially (Serenade, Bayer) to protect mushroom crops against green mould disease and promotes growth in banana [37, 41], among other applications. The endophytic *Bacillus* strain DMW1 was isolated from the inner tissues of potato tubers and exhibited strong biocontrol activity [42]. The near-identity of these genome sequences, independently isolated from plants in diverse geographical locations, suggests that EU07 is a member of a widely disseminated lineage of *B. velezensis* with biocontrol and growth-promoting properties. The molecular mechanisms and genetic determinants of these properties have been extensively reviewed elsewhere [43–45] and include gene-clusters for secondary metabolites such as bacilysin, fengycin and macrolactin, which are conserved in *B. velezensis* lineage that includes BS006 and EU07 [38].

Since our previous phenotypic comparisons between strains EU07 and QST713 revealed differences in their abilities to suppress fungal growth, we compared their genome sequences to identify possible genetic determinants of the observed differences. Their genomes are almost identical, with no detectable differences in their gene contents. However, we identified 46 single-nucleotide differences that are listed in Table 3. These differences appear to be non-uniformly distributed across the genome. For example, 20 of the 46 SNPs occur within a single gene that encodes the beta subunit of a class-1b ribonucleoside-diphosphate reductase [46] (RefSeq: WP_108702400.1; locus tag: BVQ_RS09140). This suggests that these differences might be explained by a recombination events associated with horizontal genetic transfer rather than point mutations. We also identified some sequence differences between EU07 and QST713 in the intergenic regions between several tRNA genes (GenBank: JAIFZJ010000168.1). These genetic differences may explain the previously observed differences observed between the DNA fingerprints of these two strains when previously assayed using RAPDs [4].

6.4 Conclusion

Genome sequencing of potential biocontrol strain EU07 revealed that it belongs to the species *B. velezensis*, a species often closely associated with plant roots and well known for promoting plant growth and biocontrol. The EU07 strain is genetically almost identical to the commercially used strain QST713 (Serenade®) and several other previously sequenced and characterized strains; however, we identified several genes containing single-nucleotide differences that can distinguish between EU07 and QST713. Strain EU07 is more distantly related to the commercially used *B. velezensis* strain FZB24 (TAEGRO®), previously known as the type-strain of *B. amyloliquefaciens* subsp. *plantarum*. The availability of this genome sequence will facilitate future efforts to unravel the genetic and molecular basis for its beneficial properties.

7. Figures and tables

GenBank accession number	Reference	Strain	ANI (%)
GCA_004421045.1	[47]	**"*B. amyloliquefaciens*" FS1092**	99.99
GCA_021228895.1	[48]	***B. velezensis* A4P130**	99.99
GCA_003986895.1		***B. velezensis* BE2**	99.99
GCA_007678125.1	[49]	***B. velezensis* DE0189**	99.99
GCA_003073255.1	[37]	***B. velezensis* QST713**	99.99
GCA_026156445.1	[50]	***B. velezensis* CHBv2**	99.98
GCA_001709055.1		***B. velezensis* CFSAN034339**	99.98
GCA_019093835.1		**"*B. amyloliquefaciens*" BK**	99.98
GCA_014791945.1		**"*B. amyloliquefaciens*" INH2-4b**	99.98
GCA_028609625.1	[42]	***B. velezensis* DMW1**	99.98
GCA_003149795.1	[40]	**"*B. amyloliquefaciens*" ALB79**	99.95
GCA_024300805.1	[23]	**"*B. amyloliquefaciens*" KNU-28**	99.95
GCA_001278635.1	[39]	**"*B. amyloliquefaciens*" BS006**	99.94
GCA_024134605.1		*B. velezensis* 2987tsa1	99.12
GCA_000817575.1	[51]	"*B. amyloliquefaciens*" TF28	99.10
GCA_034060585.1		*B. velezensis* Y-4	99.07
GCA_010671715.1	[52]	*B. velezensis* HU-91	99.07
GCA_009193045.1	[53]	*B. velezensis* BPC6	99.07
GCA_034061945.1		*B. velezensis* YN-2A	99.05
GCA_026786545.1		*B.* velezensis NRRL B-59289	99.04
GCA_024138555.1	[54]	"*B. amyloliquefaciens*" TPS17	99.04
GCA_029866505.1	[55]	"*B. amyloliquefaciens*" MN-13	99.03
GCA_000341875.1	[56]	*B. velezensis* UCMB5036	99.02
GCA_009789615.1	[57]	*B. velezensis* BA-26	99.02
GCA_029910295.1		*B. velezensis* PT4	99.01
GCA_009738165.1	[58]	*B. velezensis* HN-Q-8	99.01
GCA_021559715.1	[59]	*B. velezensis* CF57	99.01
GCA_012647845.1	[60]	*B. velezensis* UCMB5140	99.01

198 **Table 1. Genomes that share more than 99 % average nucleotide identity (ANI) with *B. velezensis***
199 **EU07.**

200

201

GenBank accession	**Taxon**	Reference
GCA_003149795.1	""*B. amyloliquefaciens*"" ALB79	[40]
GCA_019093835.1	"*B. amyloliquefaciens*" BK	
GCA_001278635.1	"*B. amyloliquefaciens*" BS006	[39]
GCA_000196735.1	*B. amyloliquefaciens* DSM7 ^{T}	[34]
GCA_004421045.1	"*B. amyloliquefaciens*" FS1092	[47]
GCA_014791945.1	"*B. amyloliquefaciens*" INH2-4b	
GCA_024300805.1	"*B. amyloliquefaciens*" KNU-28	[23]
GCA_029866505.1	"*B. amyloliquefaciens*" MN-13	[55]
GCA_000817575.1	"*B. amyloliquefaciens*" TF28	[51]
GCA_024138555.1	"*B. amyloliquefaciens*" TPS17	[54]
GCA_000262045.1	*B. siamensis* KCTC 13613 ^{T}	[61]
GCA_024134605.1	*B. velezensis* 2987tsa1	
GCA_021228895.1	*B. velezensis* A4P130	[48]
GCA_001647965.1	*B. velezensis* AP194	[62]
GCA_009789615.1	*B. velezensis* BA-26	[57]
GCA_003986895.1	*B. velezensis* BE2	
GCA_009193045.1	*B. velezensis* BPC6	[53]
GCA_003431885.1	*B. velezensis* (*B. methylotrophicus*) CBMB205 ^{T}	[63]
GCA_021559715.1	*B. velezensis* CF57	[59]
GCA_001709055.1	*B. velezensis* CFSAN034339	
GCA_026156445.1	*B. velezensis* CHBv2	[50]
GCA_007678125.1	*B. velezensis* DE0189	[49]

GCA_028609625.1	*B. velezensis* DMW1	[42]
GCA_000015785.2	*B. velezensis* (*B. amyloliquefaciens* subsp. *plantarum*) FZB42 ^{T}	[34]
GCA_009738165.1	*B. velezensis* HN-Q-8	[58]
GCA_010671715.1	*B. velezensis* HU-91	[52]
GCA_001461835.1	*B. velezensis* (= *B. oryzicola*) KACC 18228 ^{T}	[64]
GCA_001267695.1	*B. velezensis* KCTC 13012	[65]
GCA_001461825.1	*B. velezensis* NRRL B-41580 ^{T}	[36]
GCA_026786545.1	*B. velezensis* NRRL B-59289	
GCA_026787705.1	*B. velezensis* NRRL BD-154	
GCA_029910295.1	*B. velezensis* PT4	
GCA_003073255.1	*B. velezensis* QST713	[37]
GCA_000341875.1	*B. velezensis* UCMB5036	[56]
GCA_012647845.1	*B. velezensis* UCMB5140	[60]
GCA_034060585.1	*B. velezensis* Y-4	
GCA_034061945.1	*B. velezensis* YN-2A	
GCA_019997305.1	*B. velezensis* EU07	This study

202

203 Table 2. Genome sequences included in the phylogenomic analysis.

204

205

Position in CP025079.1	Nucleotide in QST713	Nucleotide in EU07	Amino-acid change	Predicted gene product
21222	A	G	K -> E	BVQ_RS00080: serine--tRNA ligase
230096	A	C	E -> A	BVQ_RS21890: non-ribosomal peptide synthetase
230098	A	C	K -> Q	
230111	C	A	A -> E	
530737	T	G	Y -> *	BVQ_RS02595: hypothetical protein
530789	T	G	L -> V	
530811	T	G	I -> S	
531288	T	G	I -> S	
705298	A	C	F -> V	BVQ_RS03655: GNAT family N-acetyltransferase
855165	A	C	Non-coding	
1168486	A	C	Non-coding	
1215136	A	C	F -> C	BVQ_RS06330: contact-dependent growth inhibition system immunity protein
1851920	T	G	F -> L	BVQ_RS09140: class 1b ribonucleoside-diphosphate reductase subunit beta
1851923	A	T	G -> G (synonymous)	
1851925	C	A	T -> K	
1851929	G	T	K -> N	
1851932	A	G	E -> E (synonymous)	
1851935	A	G	Q -> Q (synonymous)	
1851938	C	T	D -> D (synonymous)	
1851941	T	G	T -> T (synonymous)	
1851944	T	C	Y -> Y (synonymous)	
1851950	A	G	K -> K (synonymous)	

1851953	T	G	V -> V (synonymous)	
1851954	T	C	L -> L (synonymous)	
1851956	A	C	L -> F	
1851959	T	C	A -> A (synonymous)	
1851962	A	C	G -> G (synonymous)	
1851965	T	G	L -> L (synonymous)	
1851969	T	C	L -> L (synonymous)	
1851971	A	G	L -> L (synonymous)	
1851972	T	C	L -> L (synonymous)	
1851974	G	T	L -> F	
1878004	T	G	Non-coding	
2191740	T	C	D -> G	BVQ_RS10680: cysteine hydrolase family protein
2415378	C	A	Non-coding	
2415381	C	A	Non-coding	
2415440	C	A	Non-coding	
2722225	G	T	Non-coding	
2722243	T	G	Non-coding	
3268938	G	T	A -> E	BVQ_RS16510: class 1 isoprenoid biosynthesis enzyme
3269022	T	G	N -> T	
3467035	A	C	Non-coding	
3489562	A	G	F -> F (synonymous)	BVQ_RS17685: lantibiotic immunity ABC transporter MutG family permease subunit
3490697	T	A	I -> I (synonymous)	BVQ_RS17690: lantibiotic immunity ABC transporter MutE/EpiE family permease subunit
3573178	T	A	Non-coding	
4000822	T	G	Non-coding	

207 **Table 3. Forty-six single-nucleotide polymorphisms between *B. velezensis* strains EU07 and**
208 **QST713.**

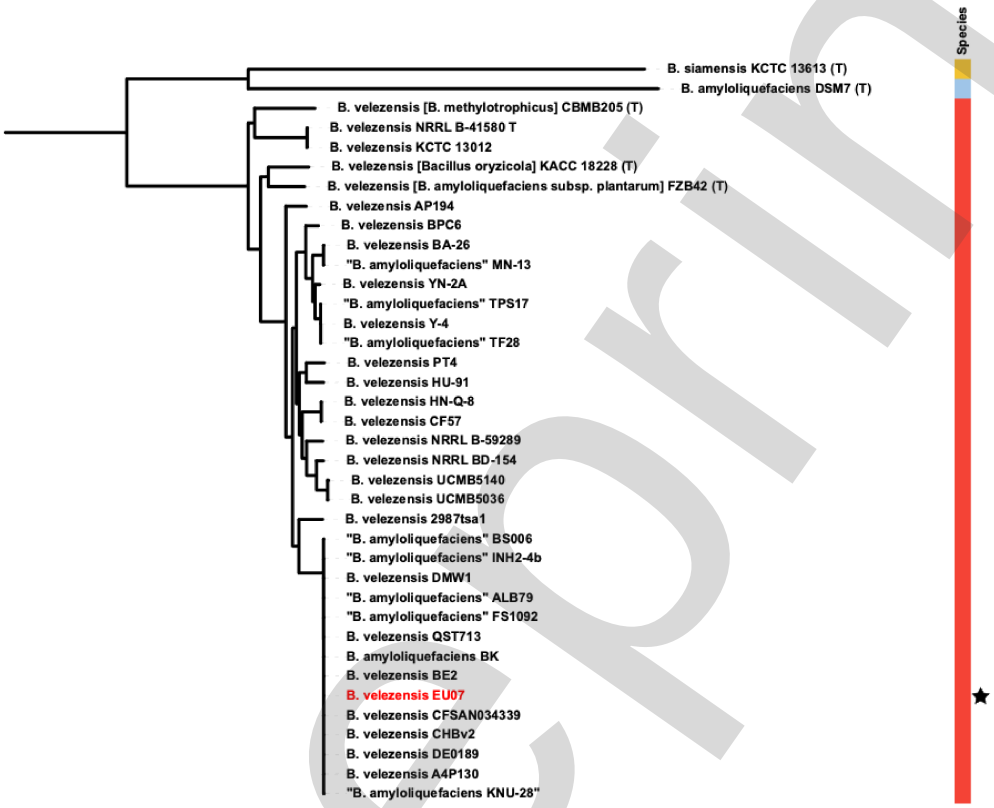

Figure 1. Phylogenetic position of *B. velezensis* EU07 within the *B. amyloliquefaciens* group. The phylogenomic maximum-likelihood tree was generated using PhaME and FastTree. The black star highlights the position of strain EU07, whose genome sequence is presented in the present study. The configuration file and the treefiles are deposited on GitHub at https://github.com/davidjstudholme/bacillus_EU07. Accession numbers for the genome assemblies can be found in Table 2. The tree can be viewed interactively at https://itol.embl.de/tree/144173231522426917024746 08.

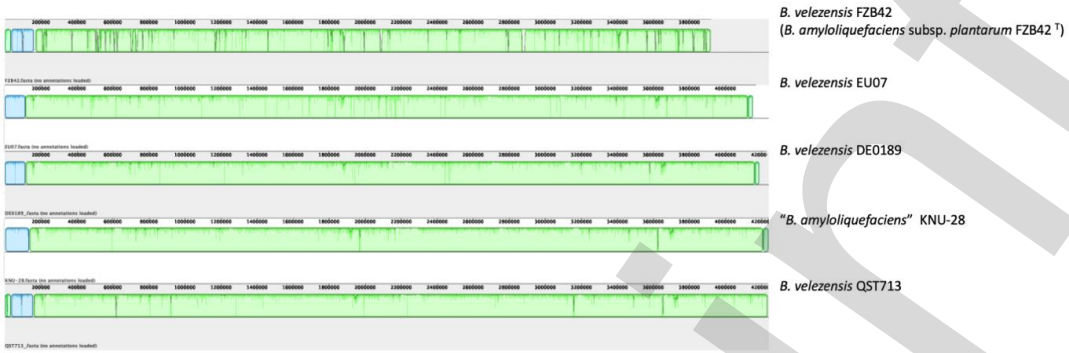

Figure 2. Whole-genome sequence alignment between *B. velezensis* EU07 and closely related strains. Genome sequences were re-ordered, aligned and visualised using Mauve. Accession numbers for the genome assemblies can be found in Table 2.

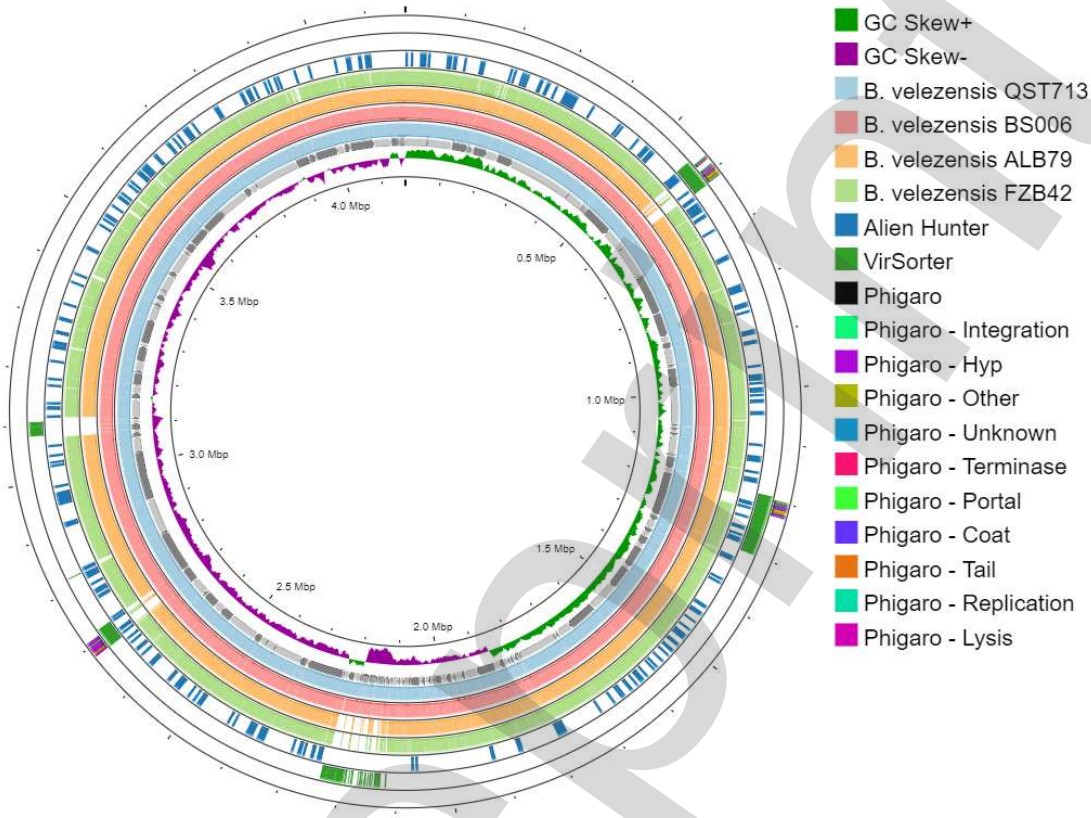

Figure 3. Overview of the genome of *B. velezensis* EU07 and comparison with closely related genomes. The circular plot of the EU07 chromosome was generated using Proksee. Data are arranged in nine concentric circular tracks as follows: (1) GC skew, (2) EU07 contigs, (3) BLASTN hits against QST713 genome, (4) BLASTN hits against BS006 genome, (5) BLASTN hits against ALB79 genome, (6) BLASTN hits against FZB542 genome, (7) predicted horizontally-acquired regions predicted by Alien Hunter, (8) phage loci predicted by VirSorter and (9) phage loci predicted by Phigaro.

8. Author statements

8.1 Author contributions

Conceptualization: OB and MT

Data Curation: All authors

Formal Analysis: All authors

Investigation: All authors

Writing – Original Draft: OB, DJS and MT

Writing – Review & Editing: All authors

8.2 Conflicts of interest

The author(s) declare that there are no conflicts of interest.

8.3 Funding information

Support for C. Jimenez-Quiros from the University of Worcester is gratefully acknowledged.

This project utilised equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1)"

8.4 Acknowledgements

We acknowledge the University of Exeter's DNA Sequencing Facility for performing library preparation and sequencing and the University of Exeter's Research IT team for provision of Advanced Research Computing infrastructure used for sequence analysis.

9. References

1. Saxena AK, Kumar M, Chakdar H, Anuroopa N, Bagyaraj DJ. *Bacillus* species in soil as a natural resource for plant health and nutrition. Journal of Applied Microbiology. 2020;128:1583–94.

2. Aloo BN, Makumba BA, Mbega ER. The potential of Bacilli rhizobacteria for sustainable crop production and environmental sustainability. Microbiological Research. 2019;219:26–39.

3. Hashem A, Tabassum B, Fathi Abd_Allah E. *Bacillus subtilis*: A plant-growth promoting rhizobacterium that also impacts biotic stress. Saudi Journal of Biological Sciences. 2019;26:1291–7.

4. Baysal Ö, Çalişkan M, Yeşilova Ö. An inhibitory effect of a new *Bacillus subtilis* strain (EU07) against *Fusarium oxysporum* f. sp. *radicis-lycopersici*. Physiological and Molecular Plant Pathology. 2008;73:25–32.

5. Jimenez-Quiros C, Okechukwu EC, Hong Y, Baysal O, Tör M. Comparison of Antifungal Activity of *Bacillus* Strains against *Fusarium graminearum In Vitro* and *In Planta*. Plants. 2022;11.

6. Krueger F. Babraham Bioinformatics - Trim Galore! http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed 14 Jul 2020.

7. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10–2.

8. Bankevich A, Nurk S, Antipov D, Gurevich A a, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology. 2012;19:455–77.

9. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics (Oxford, England). 2009;25:2071–3.

10. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic acids research. 2016;44:6614–24.

11. Studholme D. davidjstudholme/bacillus_EU07: v1.0. 2024.

12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics (Oxford, England). 2013;29:1072–5.

13. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32:292–4.

14. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;:3.

15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

16. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature communications. 2018;9:5114.

17. Benson DA. GenBank. Nucleic Acids Research. 2004;33 Database issue:D34–8.

18. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. Nucleic acids research. 2019;47:D94–9.

19. Shakya M, Ahmed SA, Davenport KW, Flynn MC, Lo C-C, Chain PSG. Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. Scientific reports. 2020;10:1723.

20. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5.

21. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 2021;49:W293–6.

22. Darling ACE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE. 2010;5:e11147.

23. Kim M-J, Park T, Jeong M, Lee G, Jung D-R, Shin J-H. Complete Genome Sequence of *Bacillus amyloliquefaciens* KNU-28 Isolated from Peach Leaves (*Prunus persica* [L.] Batsch). Microbiology resource announcements. 2022;11.

24. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research. 2004;14:1394–403.

25. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C-Y, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023;51:W484–92.

26. Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics. 2006;22:2196–203.

27. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985.

28. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9:37.

29. Starikova EV, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilina EN, et al. Phigaro: high-throughput prophage sequence annotation. Bioinformatics. 2020;36:3882–4.

30. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15:524.

31. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 2015;25:1043–55.

32. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2023;20:1203–12.

33. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nature Communications. 2019;10.

34. Rückert C, Blom J, Chen X, Reva O, Borriss R. Genome sequence of *B. amyloliquefaciens* type strain DSM7T reveals differences to plant-associated *B. amyloliquefaciens* FZB42. Journal of Biotechnology. 2011;155:78–85.

35. Borriss R, Chen X-H, Rueckert C, Blom J, Becker A, Baumgarth B, et al. Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7T and FZB42T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. Int J Syst Evol Microbiol. 2011;61 Pt 8:1786–801.

36. Dunlap CA, Kim S-J, Kwon S-W, Rooney AP. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. International Journal of Systematic and Evolutionary Microbiology. 2016;66:1212–7.

37. Pandin C, Le Coq D, Deschamps J, Védie R, Rousseau T, Aymerich S, et al. Complete genome sequence of *Bacillus velezensis* QST713: A biocontrol agent that protects *Agaricus bisporus* crops against the green mould disease. J Biotechnol. 2018;278:10–9.

38. Palazzini JM, Dunlap CA, Bowman MJ, Chulze SN. *Bacillus velezensis* RC 218 as a biocontrol agent to reduce *Fusarium* head blight and deoxynivalenol accumulation: Genome sequencing and secondary metabolite cluster profiles. Microbiological Research. 2016;192:30–6.

39. Gamez RM, Rodríguez F, Bernal JF, Agarwala R, Landsman D, Mariño-Ramírez L. Genome Sequence of the Banana Plant Growth-Promoting Rhizobacterium *Bacillus amyloliquefaciens* BS006. Genome announcements. 2015;3.

40. Tran TD, Huynh S, Parker CT, Hnasko R, Gorski L, McGarvey JA. Complete Genome Sequences of Three *Bacillus amyloliquefaciens* Strains That Inhibit the Growth of *Listeria monocytogenes In Vitro*. Genome announcements. 2018;6.

41. Tian L, Zhang W, Zhou G-D, Li S, Wang Y, Yang B, et al. A biological product of *Bacillus amyloliquefaciens* QST713 strain for promoting banana plant growth and modifying rhizosphere soil microbial diversity and community composition. Front Microbiol. 2023;14:1216018.

42. Yu C, Chen H, Zhu L, Song Y, Jiang Q, Zhang Y, et al. Profiling of Antimicrobial Metabolites Synthesized by the Endophytic and Genetically Amenable Biocontrol Strain *Bacillus velezensis* DMW1. Microbiol Spectr. 2023;11:e0003823.

366 43. Chowdhury SP, Hartmann A, Gao X, Borriss R. Biocontrol mechanism by root-associated *Bacillus*
367 *amyloliquefaciens* FZB42 – a review. Frontiers in Microbiology. 2015;6.

368 44. Chen XH, Koumoutsi A, Scholz R, Schneider K, Vater J, Süssmuth R, et al. Genome analysis of
369 *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. J Biotechnol.
370 2009;140:27–37.

371 45. Magno-Perez-Bryan MC, Martinez-Garcia PM, Hierrezuelo J, Rodriguez-Palenzuela P, Arrebola E,
372 Ramos C, et al. Comparative Genomics Within the *Bacillus* Genus Reveal the Singularities of Two
373 Robust *Bacillus amyloliquefaciens* Biocontrol Strains. Molecular plant-microbe interactions : MPMI.
374 2015;28:1102–16.

375 46. Zhang Y, Stubbe J. *Bacillus subtilis* class Ib ribonucleotide reductase is a dimanganese(III)-tyrosyl
376 radical enzyme. Biochemistry. 2011;50:5615–23.

377 47. Gu G, Gonzalez-Escalona N, Zheng J, Bolten S, Luo Y, Mafiz AI, et al. Genome Sequences of
378 *Brevundimonas naejangsanensis* Strain FS1091 and *Bacillus amyloliquefaciens* Strain FS1092,
379 Isolated from a Fresh-Cut-Produce-Processing Plant. Microbiol Resour Announc. 2020;9:e01448-19.

380 48. Becker R, Ulrich K, Behrendt U, Schneck V, Ulrich A. Genomic Characterization of *Aureimonas*
381 *altamirensis* C2P003—A Specific Member of the Microbiome of *Fraxinus excelsior* Trees Tolerant to
382 Ash Dieback. Plants (Basel). 2022;11:3487.

383 49. Zhang C, Song W, Ma HR, Peng X, Anderson DJ, Fowler VG, et al. Temporal encoding of bacterial
384 identity and traits in growth dynamics. Proc Natl Acad Sci U S A. 2020;117:20202–10.

385 50. Nøhr-Meldgaard K, Struve C, Ingmer H, Agersø Y. Intrinsic tet(L) sub-class in *Bacillus velezensis*
386 and *Bacillus amyloliquefaciens* is associated with a reduced susceptibility toward tetracycline. Front
387 Microbiol. 2022;13:966016.

388 51. Zhang S, Jiang W, Li J, Meng L, Cao X, Hu J, et al. Whole genome shotgun sequence of *Bacillus*
389 *amyloliquefaciens* TF28, a biocontrol entophytic bacterium. Stand Genomic Sci. 2016;11:73.

390 52. Wash P, Yasmin H, Ullah H, Haider W, Khan N, Ahmad A, et al. Deciphering the genetics of
391 antagonism and antimicrobial resistance in Bacillus velezensis HU-91 by whole genome analysis.
392 Journal of King Saud University - Science. 2023;35:102954.

393 53. Wangwang SUN, Li Y a. N, Changlong C, Yu T, Xiaoying LI, Jie C, et al. Identification and Biocontrol
394 Effect of Antagonistic Bacterium *Bacillus velezensis* BPC6 against Soft Rot and *Sclerotinia* Rot
395 Diseases on Lettuce. Chinese Journal of Biological Control. 2020;36:231.

396 54. Kuebutornye FKA, Lu Y, Abarike ED, Wang Z, Li Y, Sakyi ME. In vitro Assessment of the Probiotic
397 Characteristics of Three *Bacillus* Species from the Gut of Nile Tilapia, *Oreochromis niloticus*.
398 Probiotics Antimicrob Proteins. 2020;12:412–24.

399 55. Yang J, Gao MY, Li M, Li ZZ, Li H, Li HY. *Bacillus amyloliquefaciens* CotA degradation of the lignin
400 model compound guaiacylglycerol-β-guaiacyl ether. Letters in Applied Microbiology. 2018;67:491–6.

401 56. Manzoor S, Niazi A, Bejai S, Meijer J, Bongcam-Rudloff E. Genome Sequence of a Plant-
402 Associated Bacterium, *Bacillus amyloliquefaciens* Strain UCMB5036. Genome Announc.
403 2013;1:e0011113.

404 57. Wang B, Liu C, Yang X, Wang Y, Zhang F, Cheng H, et al. Genomics-guided isolation and
405 identification of active secondary metabolites of *Bacillus velezensis* BA-26. Biotechnology &
406 Biotechnological Equipment. 2021;35:895–904.

407 58. Zhao J, Zhou Z, Bai X, Zhang D, Zhang L, Wang J, et al. A novel of new class II bacteriocin from
408 *Bacillus velezensis* HN-Q-8 and its antibacterial activity on Streptomyces scabies. Front Microbiol.
409 2022;13:943232.

410 59. Zhang C, Chen L, Si H, Gao W, Liu P, Zhang J. Study on the characteristics and mechanisms of
411 nicosulfuron biodegradation by *Bacillus velezensis* CF57. J Basic Microbiol. 2020;60:649–58.

412 60. Reva ON, Larisa SA, Mwakilili AD, Tibuhwa D, Lyantagaye S, Chan WY, et al. Complete genome
413 sequence and epigenetic profile of *Bacillus velezensis* UCMB5140 used for plant and crop protection
414 in comparison with other plant-associated *Bacillus* strains. Appl Microbiol Biotechnol.
415 2020;104:7643–56.

416 61. Jeong H, Jeong D-E, Kim SH, Song GC, Park S-Y, Ryu C-M, et al. Draft genome sequence of the
417 plant growth-promoting bacterium *Bacillus siamensis* KCTC 13613T. J Bacteriol. 2012;194:4148–9.

418 62. Hassan MK. In vitro pectate lyase activity and carbon uptake assays and whole genome
419 sequencing of *Bacillus amyloliquefaciens* subsp. *plantarum* strains for a pectin defective pathway.
420 2021;:2021.01.03.425148.

421 63. Hwangbo K, Um Y, Kim KY, Madhaiyan M, Sa TM, Lee Y. Complete Genome Sequence of *Bacillus*
422 *velezensis* CBMB205, a Phosphate-Solubilizing Bacterium Isolated from the Rhizoplane of Rice in the
423 Republic of Korea. Genome Announc. 2016;4:e00654-16.

424 64. Chung EJ, Hossain MT, Khan A, Kim KH, Jeon CO, Chung YR. *Bacillus oryzicola* sp. nov., an
425 Endophytic Bacterium Isolated from the Roots of Rice with Antimicrobial, Plant Growth Promoting,
426 and Systemic Resistance Inducing Activities in Rice. Plant Pathol J. 2015;31:152–64.

427 65. Jeong H, Park S-H, Choi S-K. Genome Sequence of Antibiotic-Producing *Bacillus amyloliquefaciens*
428 Strain KCTC 13012. Genome Announc. 2015;3:e01121-15.

429

Line 28 - "sGram" change to "Gram".
We have now resolved this.

Line 28/29 - "biocontrol biopesticides agricultural" change to just "biocontrol pesticides" as agriculture is mentioned later in the sentence.
We have now done this.

46 - INTRODUCTION. For me this is far too short. It did not give me that much context about why this dataset was a vital part of ongoing research, or why it is important.

Thank you for this feedback. We have now substantially revised the Introduction section to improve clarity about the background of this strain. However, it remains fairly short. This is a Data Note paper. On the Journal's website at https://www.microbiologyresearch.org/article-types, they suggest the following article as an example of a Data Note: https://doi.org/10.1099/acmi.0.000655.v3. We note that its Introduction is of similar length to ours and understand that for this kind of article the Introduction should indeed be concise.

There is no reference associated to previous work isolating this strain,
Sorry for the lack of clarity. The isolation was previously described in reference number 4 (Baysal Ö, Çalişkan M, Yeşilova Ö. An inhibitory effect of a new *Bacillus subtilis* strain (EU07) against *Fusarium oxysporum* f. sp. *radicis-lycopersici*. Physiological and Molecular Plant Pathology. 2008;73:25–32.). We have now added another citation to this reference in the sentence about isolating the strain.

and the reader is not given much information as to the relevance of this EU07 strain. Why was it isolated in the first place?
The story of EU07's isolation is described in reference number 4. It was isolated in an effort to isolate strains that inhibit the soilborne pathogen *Fusarium oxysporum* f. sp. *radices-lycopersici*.

Is it a well studied strain?
It has been studied only in those studies that we already cited.

Is it part of a soil ecosystems natural defences against pathogens?
We do not know the answer to that question and therefore cannot provide an answer. In fact, it is not clear to the authors how one could go about empirically testing that hypothesis. It would require an experiment in which we remove the strain from the soil ecosystem and determine whether that has any impact on its "natural defences".

Is it an engineered strain used in agriculture?
No.

Also, authors write "it inhibits / it promotes" at a couple of points; I would put the name of the strain here instead to be specific.
Thank you for this excellent idea. We have made the suggested changes in the Introduction section.

Line 57 - Bacterial strain DNA isolation. More information needed here too. Was the DNA isolated from a clonal sample? Which broth was used? Was the broth similar to environmental conditions, or more similar to lab conditions?
For DNA extraction, the bacteria were grown in Nutrient Broth pH 7.2, which provides a laboratory environment quite different from the bacterium's normal soil environment. The culture was grown from a single colony and therefore clonal. We have now added this information to the Methods.

Line 91 - I would say how many related genomes here. I know it's mentioned later on (888?) but it should be put in the methods section.
We have added this now.

Line 98 - How consistent was the coverage? Did you have areas of high / low coverage that may affect the analysis of the data?
We checked this using Qualimap. We added a new section to the Methods: "5.4 Assessment of genome-assembly quality". Furthemore, we added the following text to the Results:

"Alignment of sequencing reads against the genome assembly and analysis with Qualimap revealed a mean coverage of 93.25 X and standard deviation of 89.87. Almost all of the genome assembly (99.96% had at least 1 X coverage and 97.59% of the assembly has at least 10 X coverage. The full Qualimap report and output files are available in the Zenodo repository [32], allowing users of this data to take coverage into account when performing analyses. We note that the contig with least coverage is JAIFZJ020000237.1, having only 1.04 X coverage. Nevertheless, BLAST searches reveal that this contig shows very high levels of sequence similarity to genomes of other Bacillus velezensis strains, increasing confidence in its validity."

Line 127 - Do you have any hypothesis as to why this strain would have more bacteriophage genomes within it? Is it more susceptible in some way?
No. There is no evidence that this genome has more bacteriophage genomes in it. We are not making that claim.

I suggest the authors afford more details to simply describe the processes about how to use the web servers that the authors mentioned.

The authors consider that providing tutorial material about the Proksee, TYGS and iTOL webservers is outside the scope of this data article, whose purpose is to describe the genomic sequencing dataset. These webservers provide their own documentation on how to use them.

1. L160, charcaterised, please correct to characterized.
Thank you for spotting that error. This is now corrected.

2. Figure 2, I suggest the authors add the species names of the genomes in the picture of Figure 2.
We have now added these labels to Figure 2.

3. The names of genus and species in the Reference section should be corrected to be italic.
We have fixed this now.