

RESEARCH ARTICLE

Bacillus pumilus SAFR-032 Genome Revisited: Sequence Update and Re-Annotation

Victor G. Stepanov¹, Madhan R. Tirumalai¹, Saied Montazari¹, Aleksandra Checinska², Kasthuri Venkateswaran², George E. Fox^{1*}

1 Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America, **2** Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, United States of America

* fox@uh.edu



OPEN ACCESS

Citation: Stepanov VG, Tirumalai MR, Montazari S, Checinska A, Venkateswaran K, Fox GE (2016) *Bacillus pumilus* SAFR-032 Genome Revisited: Sequence Update and Re-Annotation. PLoS ONE 11 (6): e0157331. doi:10.1371/journal.pone.0157331

Editor: Peter Setlow, University of Connecticut, UNITED STATES

Received: May 5, 2016

Accepted: May 29, 2016

Published: June 28, 2016

Copyright: © 2016 Stepanov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The final revised version of the annotated *B. pumilus* SAFR-032 genome sequence was deposited in Genbank, accession number CP000813.4 (<http://www.ncbi.nlm.nih.gov/nucleotide/CP000813>).

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Bacillus pumilus strain SAFR-032 is a non-pathogenic spore-forming bacterium exhibiting an anomalously high persistence in bactericidal environments. In its dormant state, it is capable of withstanding doses of ultraviolet (UV) radiation or hydrogen peroxide, which are lethal for the vast majority of microorganisms. This unusual resistance profile has made SAFR-032 a reference strain for studies of bacterial spore resistance. The complete genome sequence of *B. pumilus* SAFR-032 was published in 2007 early in the genomics era. Since then, the SAFR-032 strain has frequently been used as a source of genetic/genomic information that was regarded as representative of the entire *B. pumilus* species group. Recently, our ongoing studies of conservation of gene distribution patterns in the complete genomes of various *B. pumilus* strains revealed indications of misassembly in the *B. pumilus* SAFR-032 genome. Synteny-driven local genome resequencing confirmed that the original SAFR-032 sequence contained assembly errors associated with long sequence repeats. The genome sequence was corrected according to the new findings. In addition, a significantly improved annotation is now available. Gene orders were compared and portions of the genome arrangement were found to be similar in a wide spectrum of *Bacillus* strains.

Introduction

The endospore forming species, *Bacillus pumilus*, naturally occurs in the plant root systems of tobacco, pepper, cucumber and tomato, as well as in the leaves of soybean crops [1, 2]. In this context, it can confer systemic protection against several plant diseases [3]. It was therefore unexpected when the strain *B. pumilus* SAFR-032 was isolated from a spacecraft assembly facility, and its spores were found to be extremely resistant to radiation, desiccation, and hydrogen peroxide treatment [4, 5].

Subsequently, SAFR-032 spores were found to exhibit high survival rates under exposure to outer space conditions in experiments on board the International Space Station (ISS) [6, 7]. The spaceflight data generated in these studies pointed to the risks of forward contamination of celestial bodies by microbial contaminants carried by landing probes. As a result, *B. pumilus*

SAFR-032 has been classified as an extreme microorganism according to the planetary protection standards [7], and established as an important reference organism in studies on bacterial spore viability [8–13]. The unique resistance of *B. pumilus* SAFR-032 attracted the attention of microbiologists. This interest culminated in the publication of a complete annotated genome sequence of the SAFR-032 strain in 2007 [14]. This preceded by several years the genome of the *B. pumilus* type strain, ATCC 7061^T. Because of its early availability and importance as a model organism, the SAFR-032 genome is often used as a source of genetic information that is considered likely to be representative of the entire *B. pumilus* species group [15–22]. The SAFR-032 genome has also been used as a reference for assembly, correction and annotation of closely related genomes [23–25].

In recent years, complete genome sequences were determined for several other *B. pumilus* strains [19, 23, 26, 27]. Their comparisons with the SAFR-032 genome revealed discrepancies in gene order, which were possibly due to errors in the original SAFR-032 genome assembly. New studies, described herein, have corrected what were in fact errors. In addition, the SAFR-032 genome annotation was updated to include recently identified genes and correct boundaries of known coding sequences.

Materials and Methods

Media, cultivation, and genomic DNA extraction

B. pumilus SAFR-032 culture growth was initiated from the original stock archived at Jet Propulsion Laboratory (JPL), Pasadena, CA. The cells from the stock were streaked on Tryptic Soy Agar and incubated at 30°C until colonies appeared. A single colony was inoculated into 5 mL of Tryptic Soy Broth, and the culture was grown overnight at 30°C and 200 rpm. The genomic DNA was extracted from 1 mL aliquot using Maxwell 16 DNA/RNA Extraction System (Promega, Madison, WI), and eluted with water [28].

Primer design

Since the putative sequence misconnections were associated with the repeating genome segments (ribosomal RNA operons or transposase genes), the amplification primers were designed to bind to the sites immediately adjacent to the repeat units. Primer picking was performed using Primer3 v. 2.3.6 [29], primer binding specificity was confirmed using Blast 2.2.29 + [30]. For each questionable junction, 2 primer pairs were selected according to the following criteria: melting temperature 54–56°C, amplicon size not exceeding 10,000 bp, no significant binding outside of the intended priming sites, no significant secondary structure or self-binding. Amplification primers were also used for sequencing the terminal parts of the corresponding amplicons. The internal regions of the rDNA amplicons were partially assessed using universal primers targeting intergenic 16S–23S and 23S–5S spacers. Sequences of all primers used in this study are presented in Table 1.

Amplicon synthesis, purification and sequencing

PCR amplification of the genomic DNA fragments covering questionable *B. pumilus* SAFR-032 genome regions was performed using Q5 High Fidelity PCR kit (New England Biolabs, Ipswich, MA). The reaction mixtures contained 0.16 µg of genomic DNA, 1 µM of each amplification primer and 40 µL of Q5 High Fidelity Master Mix in a final volume of 80 µL. After initial incubation at 94°C for 90 sec, the reactions were carried out for 37 cycles at 94°C for 30 sec, 52°C for 30 sec, and 72°C for 5 min. Final extension was performed at 72°C for 10 min. Then, the entire reaction mixtures were separated by electrophoresis on 1% agarose gel containing

Table 1. Amplification and sequencing primers.

Primer ID	Sequence
Amplification primers	
RRNAO_AC01_fwd	AATTCTTATCTCTTCTTTACAACAAG
RRNAO_AC01_rev	AAGATCATATTCCATTGCTTTATAAC
RRNAO_AC04_fwd	AATGAGAAAGTACTTGAAATCATTAG
RRNAO_AC04_rev	TTAAGAAGAATGGATTGTAAATAACG
RRNAO_BD02_fwd	GAATCATTAGAATTATGAAGGAAGAG
RRNAO_BD02_rev	AGCAATTGATACAAGATAATACTTTG
RRNAO_BD09_fwd	TCATTAGAATTATGAAGGAAGAGAAG
RRNAO_BD09_rev	TTAAATACGTAATACTTTACCAATC
RRNAO_CB00_fwd	CTGTAATATTGTTGTTTATGTCAGG
RRNAO_CB00_rev	CTTCTATATCTCTTCTAACACTTG
RRNAO_CB07_fwd	TTACATGTATCTATTAACCCTGTAAC
RRNAO_CB07_rev	TGATATATACATAATCACTACGAGAC
TRANS_L05_fwd	ATATTTATGGTGATACAATACAAGAG
TRANS_L05_rev	TCTTCAATACAGTTCAACATATAATG
TRANS_L09_fwd	GTTTCGACTACTTAACGATTAATAAC
TRANS_L10_rev	AAATAATCAAGAAACCTATATCGAAG
TRANS_R01_fwd	GAAATATTTATTCTCGCATTATGAAC
TRANS_R01_rev	CCTATTACATGCTTTGTTCTTC
TRANS_R03_fwd	AAATATTTATTCTCGCATTATGAACC
TRANS_R05_rev	CTATTACATGCTTTGTTCTTC
Sequencing primers	
16s_bpum_seq	GCTGGAATCGCTAGTAATCGCGGATCAGCATG
23s_bpum_seq1	CACTAGGGAGTATTTAGCCTTGGGAGATGGTC
23s_bpum_seq2	CGCAAGGAAGTAAGATCCCTGAAAGATGATC

doi:10.1371/journal.pone.0157331.t001

0.2 µg/mL ethidium bromide. The bands of relevant size were excised from the gel, and the DNA was isolated from the gel slices on silica spin columns [31]. The purified rDNA amplicons were partially sequenced by the Sanger method at SeqWright, Inc (Houston, TX) using 5 primers for each amplicon: 16s_bpum_seq, 23s_bpum_seq1, 23s_bpum_seq2, and 2 amplification primers. Transposase gene amplicons were sequenced using amplification primers only.

Genome re-assembly

The *B. pumilus* SAFR-032 genome sequence update was performed using Artemis genome viewer v. 16.0.0 [32]. Validation of the re-assembled sequence was done using the Seaview 4.5.3 sequence aligner [33] and Mauve multiple genome aligner, development snapshot 2015-02-13 [34]. Visualization of the aligned genomic fragments utilized Easyfig [35].

Annotation

The re-assembled new version of the *B. pumilus* SAFR-032 genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [36] and the RAST pipeline [37]. The results were compared with those of the existing annotated genome (NC_009848.1/CP000813.1). The comparison was conducted by manually aligning the predicted ORFs/genes (including the flanking 20–30 bases upstream) using the BioEdit sequence alignment software [38].

Genome alignments

Alignments of *B. pumilus* genomes were constructed using the Progressive Mauve Aligner [34]. A core alignment included the original *B. pumilus* SAFR-032 genome version, CP000813.1, its updated version, CP000813.4, and the following complete genomes: *B. pumilus* WP8 (CP010075.1, *de novo* assembly of Illumina MiSeq and PacBio reads), *B. pumilus* MTCC B6033 (CP007436.1, *de novo* assembly of PacBio reads), *B. pumilus* NJ-M2 (CP012329.1, *de novo* assembly of Illumina and Sanger reads), *B. pumilus* NJ-V2 (CP012482.1, *de novo* assembly of Illumina and Sanger reads), *B. pumilus* TUAT1 (AP014928.1, *de novo* assembly of 454 and Sanger reads). An extended alignment additionally included the complete genomes of *B. pumilus* GR-8 (CP009108.1, a reference-driven assembly of Illumina reads with the *B. pumilus* MTCC B6033 genome as a reference), *B. pumilus* W3 (CP011150.1, a reference-driven assembly of Illumina reads with the *B. pumilus* SAFR-032 genome as a reference), *B. pumilus* ku-bf1 (CP014165.1, a reference-driven assembly of Illumina reads with the *B. pumilus* W3 genome as a reference). Since some of the genomes were deposited in reverse-complement orientation to a standard layout, they were inverted to bring their orientation to normal. The first position of each genome was set at the first position of *dnaA* gene.

Results

Identifying and Correcting Sequencing Problems

Six *de novo* assembled complete genomes of *B. pumilus* strains SAFR-032 (CP000813.1), WP8 (CP010075.1), MTCC B6033 (CP007436.1), NJ-M2 (CP012329.1), NJ-V2 (CP012482.1), and TUAT1 (AP014928.1) were aligned using the Mauve aligner to investigate conservation of *B. pumilus* genome layout. The original assembly of the SAFR-032 genome deviates from the consensus gene order near the origin of bacterial chromosome replication (Fig 1A). A conserved genome fragment encompassing 19 protein-encoding genes, 1 stand-alone tRNA gene and 1 rRNA operon was found displaced. This fragment is 522 kbp downstream of its location in the other genomes as related to the position of chromosomal replication initiation protein *dnaA*. Another SAFR-032 genome segment containing 13 protein-encoding genes and 1 tRNA gene, and flanked by transposase coding sequences was found to be inverted as compared to the genomes of other strains (Fig 1B).

Both problematic fragments are flanked with long sequence repeats, rRNA operons in one case and transposase genes in the other. This makes it look like genome assembly errors were caused by incorrect contig joining over repeat regions. Better agreement with other *B. pumilus* strains can be achieved by moving a stretch of genes BPUM_0503-BPUM_0520 and the ribosomal operon 5 (BPUM_r0013, BPUM_r0013, BPUM_r0015) immediately downstream of the ribosomal operon 1 (BPUM_r0001, BPUM_r0002, BPUM_r0003), and by inverting a genome segment harboring genes from BPUM_0842 to BPUM_054 (Fig 2A). Validity of this rearrangement has been confirmed by selective PCR amplification of the affected SAFR-032 genome regions followed by dye dideoxy terminator sequencing of the amplicons (Fig 2B).

Comparison of New PGAP Annotation with Original Annotation and Resolution of Differences

Although the SAFR-032 genome rearrangement did not affect the existing gene boundaries, the genome sequence was re-annotated to bring it to the most up-to-date state. The primary annotation was performed with NCBI PGAP [36] and the results were compared with the original (CP000813.1) annotation. A surprisingly large number of differences were found and significant effort was required to resolve them.

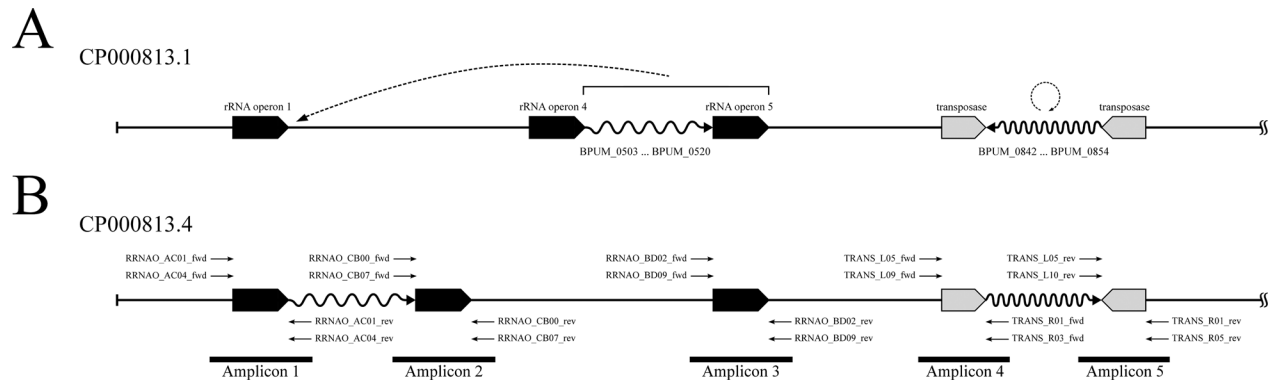


Fig 2. Proposed *B. pumilus* SAFR-032 genome rearrangements (A) and scheme of their validation (B). Not to scale. rRNA operons and transposase genes are shown as black- and grey-filled rectangular arrows, respectively. The problematic fragments between sequence repeats are shown as waved arrows. The proposed sequence rearrangements are shown as dashed arrows. Relative positions of amplification primers are shown by thin straight arrows. The span of PCR amplicons is marked by thick black lines. Sequence start corresponds to the first codon of *dnaA* gene.

doi:10.1371/journal.pone.0157331.g002

(S2 Table). These questionable gene boundaries were examined using the RAST pipeline [37]. In 197 cases RAST supported the new PGAP ORF coordinates. In 118 cases RAST placed the start codon position as it was in CP000813.1, and thus the earlier annotation was retained. Finally, in 70 cases RAST did not support either annotation. Hands on visual examination made it clear that the problematic ORFs do not have recognizable ribosome-binding sites in front of the candidate start codons. In the absence of experimental data, this makes the prediction of ORF start positions in these 70 cases unreliable, and does not allow a clear choice between several available alternatives. Therefore in the final updated annotation (CP000813.4), for consistency reasons, the ORFs with ambiguous coordinates were left as annotated by PGAP.

In the PGAP annotation, 58 entries were marked as pseudogenes (S3 Table). The two annotations agreed on 35 pseudogenes. The new PGAP version had 23 extra pseudogenes, 22 of which were previously regarded as valid genes, and 1 was newly identified. After examination of sequence alignments with close homologs, it was concluded that nine of these are indeed genuine pseudogenes with either in-frame stop codons or deletion of a substantial part of the coding sequence. The remaining fourteen examples rather correspond to properly expressed genes as it follows from homology analysis. Thus, only 44 pseudogenes were retained in the final CP000813.4 annotation.

BPUM_2060: A Special Case

BPUM_2060 was annotated by PGAP as a pseudogene of a "bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase" (*ribA*). This gene is a part of the four gene operon *ribTHAE* that codes for enzymes involved in the riboflavin biosynthetic pathway [39]. The pseudogene status was assigned to BPUM_2060 because it maps to only the N-terminal half of the reference protein (WP_006636466.1, product of *ribA* gene from *Bacillus sonorensis*), which makes BPUM_2060 look like a truncated version of the reference. However, it appears that in some *Bacillus* species, GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone 4-phosphate synthase are two distinct polypeptides encoded by separate genes. In *B. pumilus* SAFR-032, BPUM_2060 gene encodes 3,4-dihydroxy-2-butanone-4-phosphate synthase while two other genes, BPUM_0700 and BPUM_3531, each encodes GTP cyclohydrolase II. Besides *B. pumilus* strains, the same situation exists in *B. safensis*, *B. altitudinis*, *B. xiamenensis*, *B. stratosphericus*, *B. isronensis*, *B. simplex*, and *B. butanolivorans*. It can also be seen outside of the

Bacillus genus, for example, in *Planococcus donghaensis*, *P. halocryophilus*, *P. antarcticus*, *Solibacillus silvestris*, *Lysinibacillus fusiformis*, *L. sphaericus*, *L. varians*, *Paenibacillus polymyxa*, *Kurthia massiliensis*, *K. huakuii*, and *Oceanobacillus manasiensis*. Therefore, BPUM_2060 was re-annotated as a valid gene. This example demonstrates that an improperly selected reference may cause an annotation artifact. Clearly, in the present day scenario of rapidly evolving sequencing technologies, automated annotation of genomes is likely to be error-prone, and manual curation is absolutely necessary to maintain better standards of quality of a genome and its various components.

Final Updated Annotation

Overall, the updated *B. pumilus* SAFR-032 genome annotation contains 3710 valid protein-encoding genes, 44 likely pseudogenes, 21 rRNA genes grouped in 7 rRNA operons, 72 tRNA genes, one tmRNA, 4 other non-coding RNA genes, and 16 riboswitches. The updated annotation includes 38 genes and 2 riboswitches that were not present in the original CP000813.1 genome annotation (S4 Table). Three of the newly identified genes, BPUM_03795, BPUM_04085, and BPUM_04135, are transposases, which brings a total number of mobile genetic elements in *B. pumilus* SAFR-032 genome to 20. Another two genes, BPUM_04115 (coding for the “capsular biosynthesis protein CpsH”) and BPUM_04155 (coding for a “spore coat protein”) with probable roles in cellular structural integrity and sporulation, respectively, may be an addition to the candidate genes involved in the unusual resistances exhibited by *B. pumilus* SAFR-032 [10, 11]. Among the rest, seven genes are predicted to be involved in transport and secretion, two genes encode regulatory proteins, three genes encode enzymes, and five genes are predicted to encode stable cellular RNAs (three tRNAs, tmRNA, and 6S RNA). The final revised version of the annotated *B. pumilus* SAFR-032 genome sequence was deposited in Genbank under accession number CP000813.4.

Discussion

High-throughput shotgun sequencing methods generate a steady stream of genomic sequences filling public databases. Currently, the majority of deposited prokaryotic genomes have draft status. Thus, they represent a collection of contigs or scaffolds rather than a single circularly closed chromosome. While being acceptable for mutation mapping, strain typing, and metabolic pathways identification, the draft genome assemblies are not sufficient for studying large-scale genome architecture, which is needed for better understanding of genome function and evolution. This creates a demand for completed genome sequences, even though the assembly finishing is costly and time-consuming. The major obstacle on the way towards a complete genome is posed by long sequence repeats, which may cause assembly errors because of ambiguities in contig order. The present study demonstrates that such errors can be revealed using multiple alignments of closely related genomes.

All *de novo* assembled complete *B. pumilus* genomes shows absolute conservation of gene order in the region immediately downstream of the replication origin, between *dnaA* and *metS* (*metG*) genes. This observed genetic layout is well preserved far beyond the *B. pumilus* strains. With minor modifications it can be seen in *B. subtilis*, *B. mojavensis*, *B. atrophaeus*, *B. amyloliquefaciens*, *B. licheniformis*, *B. megaterium*, *B. endophyticus*, *B. smithii*, *B. cytotoxicus*, in the entire *B. cereus* group of species, and throughout *Geobacillus* and *Anoxybacillus* genera (S1 Fig). Thus, the deviation of *B. pumilus* SAFR-032 genome sequence from such a pronounced trend strongly indicated a high possibility of an assembly error. Direct sequencing of the problematic locus confirmed the presence of an error. The corrected SAFR-032 sequence exhibits perfect agreement with the consensus gene distribution pattern near the replication origin.

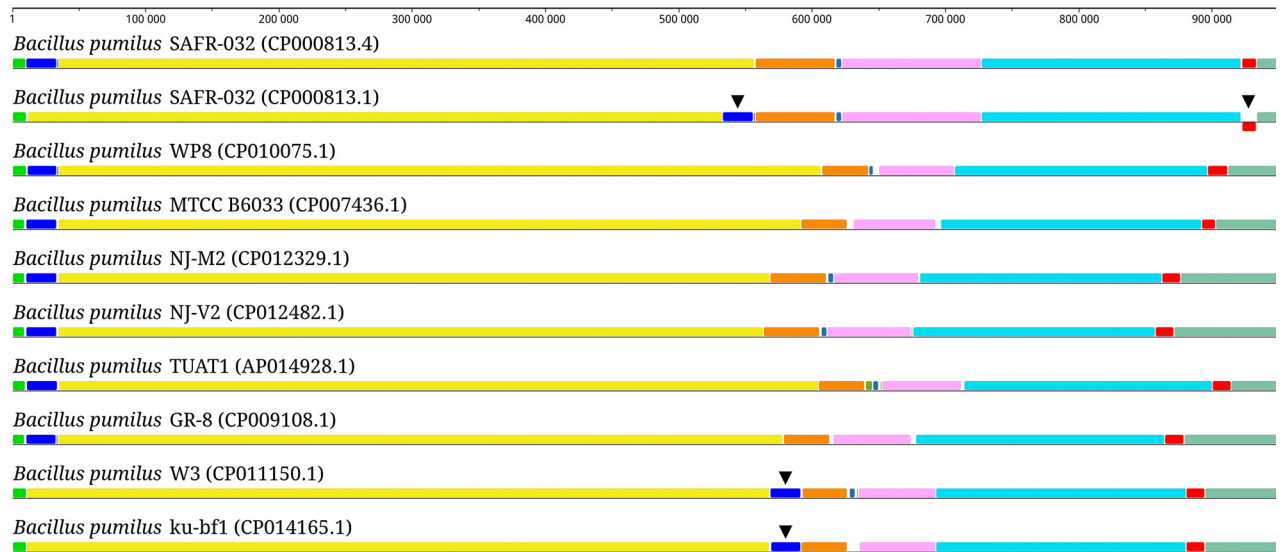


Fig 3. Comparative genome analysis of complete *B. pumilus* genomes. Multiple genome alignments were performed with Progressive Mauve Aligner [34]. Related segments have the same color in all aligned genomes. Inverted segments are shown below a genome's center line. Only the first 950,000 bp of each genome are shown for the entire alignment (S2 Fig). Both *de novo* and reference-driven genome assemblies are presented. The problematic genome fragments are marked with black triangles.

doi:10.1371/journal.pone.0157331.g003

The second problematic segment in the SAFR-032 genome covering genes from *gpmB* (BPUM_0834) to *cspB* (BPUM_0862) is significantly less conserved. Among six *de novo* assembled *B. pumilus* genomes, only strains WP8, NJ-M2, NJ-V2, and TUAT1 contain *ssuBACD* operon in this location while in SAFR-032 and MTCC B6033 it is missing. The SAFR-032 gene BPUM_0842 encoding an YncM-like protein maps to the QR42_04470 gene in the aligned fragment of the WP8 genome but is absent in the corresponding genomic regions of other strains. Two identical transposase genes are present only in the SAFR-032 genome fragment and flank the gene stretch from BPUM_0842 to BPUM_0854, which was inverted in the earlier *B. pumilus* SAFR-032 genome version, CP000813.1, as related to other *B. pumilus* genomes. The relatively low local conservation of the gene order has made the indications of sequence misassembly look weaker than in the previous case. Yet, the reversal of a 11,000 bp-long genome fragment was regarded as a sufficiently substantial deviation from the consensus gene pattern to justify its re-evaluation.

Conclusions

Possible assembly errors in the previously published *B. pumilus* SAFR-032 genome were evaluated and corrected. To accommodate these changes, re-annotation was required. Despite the fact no coding sequences were changed this led to significant changes in the annotation that required considerable effort to evaluate. This points to continuing difficulties in automated annotation and highlights the importance of being cautious when relying on it. The immediate consequences of the SAFR-032 genome sequence update reflect the organism's long-standing role as a reference for closely related species. The SAFR-032 genome was previously used as an assembly template in reference-driven assembly of *B. pumilus* W3 genome, which in its own turn was used as a reference for *B. pumilus* ku-bf1 genome assembly. Therefore, it is not surprising that both W3 and ku-bf1 genomes exhibit the same deviant layout that was seen in the CP000813.1 version of the SAFR-032 genome (Fig 3 and S2 Fig). As in the case of *B. pumilus*

SAFR-032, these two genomes may need re-evaluation. Furthermore, all intergenomic synteny studies that have involved *B. pumilus* SAFR-032 strain may need to be critically revised.

Supporting Information

S1 Fig. Synteny analysis of gene cluster *dnaA*—*metS* (*metG*) in family *Bacillaceae*. Red dashed outline shows the limits within which the synteny is generally preserved. (A) BLASTN alignments of *dnaA*—*metS* (*metG*) segments visualized with Easyfig. Protein-coding sequences, rRNA and tRNA genes are colored in cyan, magenta and red, respectively. (B) Fragment of PATRIC phylogenetic tree of order *Bacillales*. The tree is based on genome-wide analysis of homologous protein groups.

(TIF)

S2 Fig. Comparative genome analysis of complete *B. pumilus* genomes. Multiple genome alignments were performed with Progressive Mauve Aligner. Related segments are identically colored in all aligned genomes, and are connected with a line of the same color through the entire alignment. Inverted segments are shown below a genome's center line.

(TIF)

S1 Table. List of genes in the original *B. pumilus* SAFR-032 genome annotation (CP000813.1), which were not included by NCBI PGAP into the new annotation (CP000813.4).

(XLSX)

S2 Table. List of genes with ambiguous start codons in the *B. pumilus* SAFR-032 genome annotation.

(XLSX)

S3 Table. List of pseudogenes in the new *B. pumilus* SAFR-032 genome annotation (CP000813.4).

(XLSX)

S4 Table. List of new genes and pseudogenes identified by NCBI PGAP in the new *B. pumilus* SAFR-032 genome annotation (CP000813.4).

(XLSX)

Author Contributions

Conceived and designed the experiments: VS GF KV. Performed the experiments: VS AC. Analyzed the data: MT VS SM AC. Contributed reagents/materials/analysis tools: AC KV. Wrote the paper: VS GF KV MT AC SM.

References

1. Choudhary DK, Johri BN. Interactions of *Bacillus* spp. and plants—with special reference to induced systemic resistance (ISR). *Microbiol Res*. 2009; 164(5):493–513. doi: [10.1016/j.micres.2008.08.007](https://doi.org/10.1016/j.micres.2008.08.007) PMID: [18845426](https://pubmed.ncbi.nlm.nih.gov/18845426/)
2. Arias RS, Sagardoy MA, van Vuurde JW. Spatio-temporal distribution of naturally occurring *Bacillus* spp. and other bacteria on the phylloplane of soybean under field conditions. *J Basic Microbiol*. 1999; 39(5–6):283–92. PMID: [10629969](https://pubmed.ncbi.nlm.nih.gov/10629969/)
3. Kloepper JW, Ryu CM, Zhang S. Induced Systemic Resistance and Promotion of Plant Growth by *Bacillus* spp. *Phytopathology*. 2004; 94(11):1259–66. doi: [10.1094/PHYTO.2004.94.11.1259](https://doi.org/10.1094/PHYTO.2004.94.11.1259) PMID: [18944464](https://pubmed.ncbi.nlm.nih.gov/18944464/)

4. Link L, Sawyer J, Venkateswaran K, Nicholson W. Extreme spore UV resistance of *Bacillus pumilus* isolates obtained from an ultraclean Spacecraft Assembly Facility. *Microb Ecol.* 2004; 47(2):159–63. PMID: [14502417](#)
5. Kempf MJ, Chen F, Kern R, Venkateswaran K. Recurrent isolation of hydrogen peroxide-resistant spores of *Bacillus pumilus* from a spacecraft assembly facility. *Astrobiology.* 2005; 5(3):391–405. PMID: [15941382](#)
6. Horneck G, Moeller R, Cadet J, Douki T, Mancinelli RL, Nicholson WL, et al. Resistance of bacterial endospores to outer space for planetary protection purposes—experiment PROTECT of the EXPOSE-E mission. *Astrobiology.* 2012; 12(5):445–56. doi: [10.1089/ast.2011.0737](#) PMID: [22680691](#)
7. Vaishampayan PA, Rabbow E, Horneck G, Venkateswaran KJ. Survival of *Bacillus pumilus* spores for a prolonged period of time in real space conditions. *Astrobiology.* 2012; 12(5):487–97. doi: [10.1089/ast.2011.0738](#) PMID: [22680694](#)
8. Friedline AW, Zachariah MM, Middaugh AN, Garimella R, Vaishampayan PA, Rice CV. Sterilization Resistance of Bacterial Spores Explained with Water Chemistry. *J Phys Chem B.* 2015; 119(44):14033–44. doi: [10.1021/acs.jpcc.5b07437](#) PMID: [26435315](#)
9. Friedline A, Zachariah M, Middaugh A, Heiser M, Khanna N, Vaishampayan P, et al. Sterilization of hydrogen peroxide resistant bacterial spores with stabilized chlorine dioxide. *AMB Express.* 2015; 5:24. doi: [10.1186/s13568-015-0109-4](#) PMID: [25897406](#)
10. Tirumalai MR, Rastogi R, Zamani N, O'Bryant Williams E, Allen S, Diouf F, et al. Candidate genes that may be responsible for the unusual resistances exhibited by *Bacillus pumilus* SAFR-032 spores. *PLoS One.* 2013; 8(6):e66012. doi: [10.1371/journal.pone.0066012](#) PMID: [23799069](#)
11. Tirumalai MR, Fox GE. An ICEBs1-like element may be associated with the extreme radiation and desiccation resistance of *Bacillus pumilus* SAFR-032 spores. *Extremophiles.* 2013; 17(5):767–74. doi: [10.1007/s00792-013-0559-z](#) PMID: [23812891](#)
12. Checinska A, Fruth IA, Green TL, Crawford RL, Paszczynski AJ. Sterilization of biological pathogens using supercritical fluid carbon dioxide containing water and hydrogen peroxide. *Journal of microbiological methods.* 2011; 87(1):70–5. doi: [10.1016/j.mimet.2011.07.008](#) PMID: [21787810](#)
13. Checinska A, Paszczynski A, Burbank M. *Bacillus* and other spore-forming genera: variations in responses and mechanisms for survival. *Annu Rev Food Sci Technol.* 2015; 6:351–69. doi: [10.1146/annurev-food-030713-092332](#) PMID: [25705935](#)
14. Gioia J, Yerrapragada S, Qin X, Jiang H, Igboeli OC, Muzny D, et al. Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus* SAFR-032. *PLoS One.* 2007; 2(9):e928. PMID: [17895969](#)
15. Schmidt TR, Scott EJ 2nd, Dyer DW. Whole-genome phylogenies of the family *Bacillaceae* and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. *BMC Genomics.* 2011; 12:430. doi: [10.1186/1471-2164-12-430](#) PMID: [21864360](#)
16. Alcaraz LD, Moreno-Hagelsieb G, Eguiarte LE, Souza V, Herrera-Estrella L, Olmedo G. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics.* 2010; 11:332. doi: [10.1186/1471-2164-11-332](#) PMID: [20504335](#)
17. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, et al. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology.* 2009; 155(Pt 6):1758–75. PMID: [19383706](#) doi: [10.1099/mic.0.027839-0](#)
18. Ruckert C, Blom J, Chen X, Reva O, Borriss R. Genome sequence of *B. amyloliquefaciens* type strain DSM7(T) reveals differences to plant-associated *B. amyloliquefaciens* FZB42. *J Biotechnol.* 2011; 155(1):78–85. doi: [10.1016/j.jbiotec.2011.01.006](#) PMID: [21262282](#)
19. Villanueva J, Switala J, Ivancich A, Loewen PC. Genome Sequence of *Bacillus pumilus* MTCC B6033. *Genome Announc.* 2014; 2(2). doi: [10.1128/genomeA.00327-14](#) PMID: [24744340](#)
20. Niazi A, Manzoor S, Asari S, Bejai S, Meijer J, Bongcam-Rudloff E. Genome analysis of *Bacillus amyloliquefaciens* subsp. *plantarum* UCMB5113: a rhizobacterium that improves plant growth and stress management. *PLoS One.* 2014; 9(8):e104651. doi: [10.1371/journal.pone.0104651](#) PMID: [25119988](#)
21. Zeigler DR. The genome sequence of *Bacillus subtilis* subsp. *spizizenii* W23: insights into speciation within the *B. subtilis* complex and into the history of *B. subtilis* genetics. *Microbiology.* 2011; 157(Pt 7):2033–41. doi: [10.1099/mic.0.048520-0](#) PMID: [21527469](#)
22. Misirli G, Hallinan J, Rottger R, Baumbach J, Wipat A. BacillusRegNet: a transcriptional regulation database and analysis platform for *Bacillus* species. *J Integr Bioinform.* 2014; 11(2):244. doi: [10.2390/biecoll-jib-2014-244](#) PMID: [25001169](#)
23. Guan ZB, Cai YJ, Zhang YZ, Zhao H, Liao XR. Complete genome sequence of *Bacillus pumilus* W3: A strain exhibiting high laccase activity. *J Biotechnol.* 2015; 207:8–9. doi: [10.1016/j.jbiotec.2015.04.019](#) PMID: [25957807](#)

24. Handtke S, Volland S, Methling K, Albrecht D, Becher D, Nehls J, et al. Cell physiology of the biotechnological relevant bacterium *Bacillus pumilus*-an omics-based approach. *J Biotechnol*. 2014; 192 Pt A:204–14. doi: [10.1016/j.jbiotec.2014.08.028](https://doi.org/10.1016/j.jbiotec.2014.08.028) PMID: [25281541](https://pubmed.ncbi.nlm.nih.gov/25281541/)
25. Wu X, Xu L, Gu W, Xu Q, He QY, Sun X, et al. Iterative genome correction largely improves proteomic analysis of nonmodel organisms. *J Proteome Res*. 2014; 13(6):2724–34. Epub 2014/05/09. doi: [10.1021/pr500369b](https://doi.org/10.1021/pr500369b) PMID: [24809469](https://pubmed.ncbi.nlm.nih.gov/24809469/)
26. Kang Y, Shen M, Wang H, Zhao Q. Complete Genome Sequence of *Bacillus pumilus* Strain WP8, an Efficient Plant Growth-Promoting Rhizobacterium. *Genome Announc*. 2015; 3(1). doi: [10.1128/genomeA.01452-14](https://doi.org/10.1128/genomeA.01452-14) PMID: [25614565](https://pubmed.ncbi.nlm.nih.gov/25614565/)
27. Yuan Y, Gao M. Genomic analysis of a ginger pathogen *Bacillus pumilus* providing the understanding to the pathogenesis and the novel control strategy. *Sci Rep*. 2015; 5:10259. doi: [10.1038/srep10259](https://doi.org/10.1038/srep10259) PMID: [25989507](https://pubmed.ncbi.nlm.nih.gov/25989507/)
28. Kwan K, Cooper M, La Duc MT, Vaishampayan P, Stam C, Benardini JN, et al. Evaluation of procedures for the collection, processing, and analysis of biomolecules from low-biomass surfaces. *Applied and environmental microbiology*. 2011; 77(9):2943–53. doi: [10.1128/AEM.02978-10](https://doi.org/10.1128/AEM.02978-10) PMID: [21398492](https://pubmed.ncbi.nlm.nih.gov/21398492/)
29. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic acids research*. 2012; 40(15):e115. doi: [10.1093/nar/gks596](https://doi.org/10.1093/nar/gks596) PMID: [22730293](https://pubmed.ncbi.nlm.nih.gov/22730293/)
30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. doi: [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421) PMID: [20003500](https://pubmed.ncbi.nlm.nih.gov/20003500/)
31. Borodina TA, Lehrach H, Soldatov AV. DNA purification on homemade silica spin-columns. *Anal Biochem*. 2003; 321(1):135–7. PMID: [12963065](https://pubmed.ncbi.nlm.nih.gov/12963065/)
32. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000; 16(10):944–5. PMID: [11120685](https://pubmed.ncbi.nlm.nih.gov/11120685/)
33. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*. 2010; 27(2):221–4. doi: [10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259) PMID: [19854763](https://pubmed.ncbi.nlm.nih.gov/19854763/)
34. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004; 14(7):1394–403. doi: [10.1101/gr.2289704](https://doi.org/10.1101/gr.2289704) PMID: [15231754](https://pubmed.ncbi.nlm.nih.gov/15231754/)
35. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011; 27(7):1009–10. doi: [10.1093/bioinformatics/btr039](https://doi.org/10.1093/bioinformatics/btr039) PMID: [21278367](https://pubmed.ncbi.nlm.nih.gov/21278367/)
36. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omics*. 2008; 12(2):137–41. doi: [10.1089/omi.2008.0017](https://doi.org/10.1089/omi.2008.0017) PMID: [18416670](https://pubmed.ncbi.nlm.nih.gov/18416670/)
37. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*. 2014; 42 (Database issue):D206–14. doi: [10.1093/nar/gkt1226](https://doi.org/10.1093/nar/gkt1226) PMID: [24293654](https://pubmed.ncbi.nlm.nih.gov/24293654/)
38. Hall T. BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*. 2011; 2:60–1.
39. Mironov VN, Kraev AS, Chikindas ML, Chernov BK, Stepanov AI, Skryabin KG. Functional organization of the riboflavin biosynthesis operon from *Bacillus subtilis* SHgw. *Molecular & general genetics: MGG*. 1994; 242(2):201–8. PMID: [8159171](https://pubmed.ncbi.nlm.nih.gov/8159171/)