Bacillus safensis FO-36b and *Bacillus pumilus* SAFR-032: A Whole Genome Comparison of Two Spacecraft Assembly Facility Isolates

Madhan R Tirumalai[1], Victor G. Stepanov[1], Andrea Wünsche[1], Saied Montazari[1], Racquel O. Gonzalez[1], Kasturi Venkateswaran[2], George. E. Fox[1§]

[1]Department of Biology and Biochemistry, University of Houston, Houston, TX, 77204-5001.

[2] Biotechnology & Planetary Protection Group, NASA Jet Propulsion Laboratories, California Institute of Technology, Pasadena, CA, 91109.

[§]Corresponding author:

Dr. George E. Fox

Dept. Biology & Biochemistry

University of Houston, Houston, TX 77204-5001

713-743-8363; 713-743-8351 (FAX); email: fox@uh.edu

Email addresses:

MRT: mrtirum2@central.uh.edu

VGS: vstepano@central.uh.edu

AW: andrea.wunsche@gmail.com

SM: montazaris@gmail.com

ROG: raquel.gonzalez100@yahoo.com

KV: kasthuri.j.venkateswaran@jpl.nasa.gov

GEF: fox@uh.edu

Keywords: Planetary protection, *Bacillus* endospores, extreme radiation resistance, peroxide resistance, genome comparison, phage insertions

Background

Microbial persistence in built environments such as spacecraft cleanroom facilities [1-3] is often characterized by their unusual resistances to different physical and chemical factors [1, 4-7]. Consistently stringent cleanroom protocols under planetary protection guidelines over several decades [1, 8-12], have created a special habitat for multi-resistant bacteria, many of which have been isolated and identified [13-19]. The potential of many of these isolates to possibly survive interplanetary transfer [2, 20-24] raises concern of potential forward and backward bacterial contamination. Understanding the survival mechanisms employed by these organisms is the key to controlling their impact on exobiology missions. In addition, their occurrence in the closed environments of the International Space Station, (ISS), could possibly impact the living conditions there as well [1-3, 25-27].

Two of the most studied organisms in the specialized econiches of spacecraft assembly facilities and the ISS are *B. safensis* FO-36b^T [28] (referred to as FO-36b henceforth) and *B. pumilus* SAFR-032 [16] (referred to as SAFR-032). These organisms are representative strains of the endospore producing *Bacillus* sp.[13, 16, 29-33]. Both strains produce spores that exhibit unusual levels of resistance to peroxide and UV radiation [24, 29, 34] that far exceed that of the dosimetric *B. subtilis* type strain (*B. subtilis subsp. subtilis str.* 168, referred to as BSU) [35]. A third strain, *B. safensis* MERTA-8-2 (referred to as MERTA), was initially isolated from the Mars Odyssey Spacecraft and associated facilities at the Jet Propulsion Laboratory and later also found on the Mars Explorer Rover (MER) before its launch in 2004. It has been reported that this

strain actually grows better on the ISS than on Earth [36]. However, the resistance properties of its spores have not been directly tested. A recent phylogenetic study of 24 *B. pumilus* and *B. safensis* strains, found FO-36b, and MERTA clustered together in a distinct group of *B. safensis* strains [37].

Previously a draft genome of FO-36b with as many as 408 contigs (https://www.hgsc.bcm.edu/microbiome/bacillus-pumilus-f036b) was compared to SAFR-032 and the type strain *B. pumilus* ATCC7061^T [38, 39] (referred to as ATCC7061). This comparison identified several genes and a mobile genetic element in SAFR-032 that may be associated with the elevated resistance [39]. Since this previous study was completed, minor corrections to the SAFR-032 gene order were made and the annotation was updated [40]. In addition, a draft genome of MERTA was reported [41]. Herein, we now report a complete genomic sequence for FO-36b and the results of a detailed comparison of these four genomes.

Methods

Sequencing of the *Bacillus safensis* FO-36b genome.

5µg of purified genomic DNA of FO-36b was digested with NEBNext dsDNA Fragmentase (New England Biolabs, Ipswich, MA) yielding dsDNA fragments in a size range of 50 bp up to 1000 bp. The fragments were fractionated on a 2% agarose gel, and those with the length from 300 bp to 350 bp were isolated as described [42]. The dsDNA fragments were converted to a shotgun DNA library using the TruSeq PCR-Free DNA Sample Preparation Kit LT (Illumina, San Diego, CA) according to the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq 2500 sequencer at the University of Arizona Genetic Core Facility (Tucson, AZ). A total of 10,812,117 pairs of 100 base-long reads with average Phred quality of 34.92/base were collected. The reads were processed with Sickle 1.33 [43] and Trimmomatic 0.32 [44] was used

to remove seven 3'-terminal low-quality bases, and to filter out the reads with average Phred quality below 16/base as well as reads containing unidentified nucleotides. Overall, 9,047,105 read pairs and 1,435,623 orphaned single reads with a total of 1,816,274,469 nucleotides were retained after the filtration step. The reads were assembled using the Abyss 1.5.2 *de novo* assembler [45] with the *kmer* parameter set at 64. The assembly consisted of 22 contigs with a total length of 3,753,329 bp. The average contig length was 170,605 bp (ranging from 352 to 991,464 bp), with an N50 contig length equal to 901,865 bp. Data from two previous FO-36b draft genomes (https://www.hgsc.bcm.edu/microbiome/bacillus-pumilus-f036b; and https://www.ncbi.nlm.nih.gov/biosample/SAMN02746691) did not provide the additional information needed to order the 22 remaining contigs.

Instead, connections between the contigs were obtained by systematic PCR screening using LongAmp *Taq* DNA polymerase (New England Biolabs, Ipswich, MA) and near-terminal outward-facing primers. The amplicons were gel purified and sequenced by the Sanger method at SeqWright, Inc (Houston, TX). This allowed closure of all the gaps between the contigs. The complete FO-36b genome sequence comprises 3.77 Mb and has G+C content of 41.74%.

B. safensis FO-36b genome annotation

The FO-36b genome was annotated using the NCBI's Prokaryotic Genome Annotation Pipeline [46]. 3850 ORFs and 40 non-coding RNAs and riboswitches were predicted and the results were deposited in Genbank under accession number CP010405.

92 **Genomes used in comparisons**

93 The recently updated complete sequence of the SAFR-032 genome was obtained from NCBI
94 (CP000813.4). The draft genomes of ATCC7061^T (Refseq accession no:
95 NZ_ABRX00000000.1), consisting of 16 contigs and MERTA consisting of 14 contigs (Refseq
96 accession no: GCF_000972825.1) were obtained from the public databases of the National
97 Center for Biotechnology Information (NCBI). Several additional *B. safensis* and *B. pumilus*
98 draft genomes from various sources have also been deposited in the NCBI database in recent
99 years. However, these genomes get excluded when performing a global Genbank Blast (NT)
100 analysis. To avoid this potential problem, these additional draft genomes were separately
101 retrieved from the Genbank repository (*B. pumilus* genomes,
102 https://www.ncbi.nlm.nih.gov/genome/genomes/440; *B. safensis* genomes,
103 https://www.ncbi.nlm.nih.gov/genome/genomes/13476) and locally integrated into the Genbank
104 NT database. The resulting local database allowed inclusion of these genomes in subsequent
105 Blast (NT) studies. Overall, the analysis involved 65 *B. pumilus* and *B. safensis* genomes
106 (including the FO-36b, MERTA, SAFR-032 and ATCC7061 genomes). The names of the
107 genomes used are given in Additional file 1: Table S1
108
109 **BLAST studies**
110 Individual gene and protein sequences from the FO-36b genome, were blasted against each other
111 as well as against the genomes of SAFR-032, MERTA and ATCC7061 using the standalone
112 version of NCBI's BLAST program [47]. The comprehensive search included blastN and blastX
113 for the nucleotide sequences and blastP for the protein sequences. Additionally, global blast was
114 performed on the sequences against the updated NR/NT databases downloaded from the NCBI

on the Opuntia Cluster at the Center of Advanced Computing and Data Systems at the University of Houston.

Genes with BLAST results in which the best hit had an e-value greater than (an arbitrary) 0.0001 were considered absent from the target genome, while those with BLAST e-values below e-10 were considered to be matches. Genes with e-values between e-20 and 0.0001 were further analyzed by aligning the sequence of the entire gene neighborhood with the corresponding region in the other genomes to ascertain/verify the BLAST results as well as to look for unusual features in the sequence. Gene/protein sequence alignments were performed using Bioedit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Phage analysis

The online tool PHAST [48, 49] was used to predict and annotate potential phage elements in the genomes. Comparative analysis of the respective homologs on the other genomes, were performed to map the respective corresponding phage regions on the other genomes.

Whole Genome Phylogenetic Analysis (WGPA) and Genome-Genome Distance Studies (GGDC)

In order to obtain an overall view of relationships among the various genomes, we used seven additional genomes thereby forming a complete set of 72 strains. Overall, the genomes included 65 *B. pumilus* and *B. safensis* genomes (including those of FO-36b, MERTA, SAFR-032 and ATCC7061), four representative strains from the *B. altitudinis* complex, *viz*., *B. aerophilus* C772, *B. altitudinis* 41KF2b, *B. cellulasensis* NIO-1130(T), and, *B. stratosphericus* LAMA 585.

The genomes of *Geobacillus kaustophilus*, and *B. subtilis* served as outliers in the *Firmicutes* group, while the genome of Gram-negative *E. coli* MG1655, served as a non-*Firmicutes* outlier.

A whole-genome-based phylogenetic analysis was conducted using the latest version of the Genome-BLAST Distance Phylogeny (GBDP) method [50] as previously described [51]. Briefly, BLAST+ [52] was used as a local alignment tool and distance calculations were done under recommended settings (greedy-with-trimming algorithm, formula D5, e-value filter 10e-8). 100 pseudo-bootstrap replicates were assessed under the same settings each. Finally, a balanced minimum evolution tree was inferred using FastME v2.1.4 with SPR post processing [53]. Replicate trees were reconstructed in the same way and branch support was subsequently mapped onto the tree. The final tree was rooted at the midpoint [54]. The genomes were also compared using the in-silico genome-to-genome comparison method, for genome-based species delineation and genome-based subspecies delineation based on intergenomic distance calculation [50, 55]**.**

In order to confirm the reasonableness of these results, a separate analysis was conducted using DNA gyrase A (*gyrA*), which has often been used for single gene phylogenetic studies [28, 56-60]. *gyrA* is preferable to 16S rRNA in this case, because many of the 16S rRNAs are too similar [61] .

The *gyrA* sequences were bioinformatically isolated from all 72 genomes and aligned using Bioedit ((http://www.mbio.ncsu.edu/BioEdit/bioedit.html), ClustalW, and MEGA [62, 63] with MUSCLE. Maximum Likelihood, Neighbor-Joining and Minimum Evolution trees were built using MEGA. The Maximum Likelihood tree was built using the Tamura-Nei model [64]. The tree with the highest log likelihood (-18473.7156) was used. Initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a

matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The topology with superior log likelihood value was selected.

A Minimum Evolution (ME) Tree was built using the method described by Rzhetsky and Nei (1992) [65]. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm [66] at a search level of 1. The Neighbor-Joining (NJ) Tree was built using the method described by Saitou and Nei (1987) [67].

For both the ME and NJ trees, the optimal tree(s) with the sum of branch length = 1.62873358 was derived. The evolutionary distances were computed using the Maximum Composite Likelihood method [68] and are in the units of the number of base substitutions per site.

The analysis involved 72 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 2424 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [69].

The Mauve alignment [70] program was used to align the previous draft FO-36b sequence (GCA_000691165.1 / ASJD00000000) with the current updated sequence (CP010405).

Screening Genomes for Antibiotic resistance genes.

A global analysis of each of the four genomes was performed to identify possible antibiotic resistance loci. This was done using the reference sequences of the Comprehensive Antibiotic Resistance Database ("CARD") [71], In addition a search for potential ‘resistome(s)’ was undertaken using the Resistance Gene Identifier feature of the CARD database for the four genomes.

Results

Unique and characteristic genes

Genes are considered to be characteristic if they are present in FO-36b, but absent in the other three organisms examined here. Unique genes are those that are not only absent in the other three genomes, but have not yet been found in any other genome**.** 307 ORFs found in FO-36b are not shared by SAFR-032. Sixty five of these ORFs did not have homologs in the genomes of ATCC7061 or MERTA and are therefore considered characteristic (Table 1). Although most are open reading frames that code for hypothetical proteins, six genes suggest that FO-36b has a CRISPR system. The likely presence of a CRISPR system is shared by 5 other *B. safensis* genomes and 8 other *B. pumilus* genomes (Additional file 2: Table S2). Among the 49 hypothetical protein coding ORFs, 26 are predicted to be part of phage element(s).

The analysis was extended to all available genomes of *B. safensis* (https://www.ncbi.nlm.nih.gov/genome/genomes/13476) and *B. pumilus* (https://www.ncbi.nlm.nih.gov/genome/genomes/440). Nine ORFs/genes classified as FO-36b characteristic are absent from all the *B. safensis* and *B. pumilus* genomes available in the NCBI database. These nine genes are totally unique to FO-36b with no homologs in the entire NR/NT databases (Table 2). Four of these are part of predicted phage elements. In addition, there are four genes with fewer than five homologs found in other *B. pumilus*/ *B. safensis* genomes (Table 3). Overall 217 SAFR-032 ORFs are not shared by *B. safensis* FO-36b. Sixty three of the 65 FO-36b characteristic ORFs are absent in 28 of the 61 total *B. safensis*, *B. pumilus*, and *Bacillus sp.* WP8 genomes. 18 are absent in all the *B. safensis* genomes, while 15 are not found in any of the *B. pumilus* genomes (Additional file 3: Table S3).

Phage insertions

The genome of FO-36b contains two phage insertions, namely the *Bacillus* bacteriophage SPP1 (NC_004166.2) insertion and the *Brevibacillus* phage Jimmer 1 (NC_029104.1) insertion. The SPP1 insertion, (Figure 1), consists of 62 genes (RS87_02955 to RS87_03255). Abbreviated versions are found in the MERTA strain (4 genes) and the ATCC7061 strain (3 genes), (Figure 1 and Figure 2). Portions of this element can also be detected in other *B. safensis/ B. pumilus* strains by sequence comparison.

The *Brevibacillus* phage Jimmer 1 (NC_029104.1) insertion is found to some extent in all 60 draft genomes belonging to the *B. safensis/ B. pumilus* family and the one *Bacillus sp* WP8. In the FO-36b genome, this phage element contains 94 genes (RS87_14155 to RS87_14625). The entire stretch of this insertion can be divided into three blocks, block A (30 genes, RS87_14155 to RS87_14305), block B (30 genes, RS87_14310 to RS87_14455) and block C (34 genes, RS87_14460 to RS87_14625). A major chunk of block C (26 genes RS87_14460 to RS87_14590) is a duplication of block A. The overall scheme of this unique duplication within the insertion is given in Figure 3.

A similar version of the Jimmer-1 phage region is found in the non-resistant ATCC7061 (Figure 4). In this case, the block A like region is comprised of 32 ORFs (30 genes and 2 pseudogenes, BAT_0021 to BAT_0052). The block C analog is formed from a cluster of 32 ORFs (29 genes and 3 pseudogenes, BAT_0175 to BAT_0206). Finally, a total of 42 ORFs (41 genes and 1 pseudogene, BAT_0053 to BAT_0094) comprise the equivalent of Block B from FO-36b (Figure 4).

The MERTA and SAFR-032 strains show equivalent regions of block A and block C from FO-36b. However, both block B and the duplication of the block A equivalent region are

missing in these strains (Figure 5 and Figure 6). The genome of the non-resistant spore producing BSU strain contains the block A and block C equivalents in stretches of 28 ORFs/genes (BSU12810 to BSU12580) and 30 ORFs/genes (BSU12810 to BSU12560) respectively, while block B is entirely missing. However, a major chunk of block A (RS87_14200 to RS87_14300) equivalent region in BSU is duplicated in a stretch of 20 ORFs/genes (BSU25980 to BSU26190) (Figure 7). In general, the occurrence of phage insertion regions and genes therein such as the dUTPase and RecT genes do not appear to be strongly correlated with resistance properties.

Genes Shared by FO-36b, SAFR-032, and MERTA but missing in ATCC7061

We had earlier reported that a total of 65 genes that were shared by SAFR-032 and FO-36b, were not found in the ATCC7061 strain [38]. Because they correlate with the presence or absence of resistance, these genes are of potential interest. A re-analysis of this list of genes extending to the MERTA strain showed that 59 of these genes are indeed shared by the MERTA strain as well (Additional file 4: Table S4). All of these genes are shared by at least several of the available 61 *B. pumilus*, *B. safensis* and *Bacillis sp.* WP8 draft genomes. However, since the resistance properties of these organisms have typically not been examined, it is not immediately possible to determine if the correlation can be extended to these strains.

Antibiotic Resistance loci in the genomes

The four genomes showed vast differences in the number of antibiotic resistance related mutations that were identified by the CARD [71] search. FO-36b, SAFR-032, MERTA and ATCC7061 had 670, 587, 317, and 495 mutations respectively. BSU comparatively had 861

such mutations. All the four genomes share "*cat86*", which is a chromosome-encoded variant of the *cat* gene found in *Bacillus pumilus* [72], belonging to the AMR (antimicrobial resistance) gene gamily of chloramphenicol acetyltransferase (CAT).

Phylogenetic analysis

Previous efforts to define the phylogenetic relationship between various *B. safensis* and *B. pumilus* strains relied on 24 genomes including the unpublished draft sequence (ASJD00000000) of *B. safensis*. Comparing this earlier version with our updated corrected sequence assembly using Mauve shows our version differs considerably (Additional file 5: Figure S1). Given this and the large number of additional draft genomes, it was concluded that a re-analysis would be appropriate. Whole Genome Phylogenetic Analysis and Genome-genome distance analysis were used to examine relationships among the strains. The results of the WGPA are shown in Figure 8, while the GGDC results are given in Additional file 1: Table S1. The phylogenetic trees are consistent with the earlier work (38). Two large clusters are seen. The first consists primarily of strains of *B. pumilus* with no *B. safensis* strains included. The first major cluster is itself broken into two large sub clusters, the first one of which includes both SAFR-032 and ATCC7061. The second sub cluster includes strains from the *B. altitudinis* complex (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1792192), as well as other strains recently reported to be *B. pumilus*. The second major cluster consists primarily of *B. safensis* isolates but does include several likely misnamed *B. pumilus* strains too. This latter cluster includes both the FO-36b and the MERTA8-2 strains.

To further ascertain this observation, a maximum likelihood tree was obtained for the gene *gyrA* (Figure 9), which further supports the WGPA and GGDC analysis. Alternative tree

constructions of *gyrA* are provided as Additional file 6: Figure S2 and Additional file 7: Figure S3.

Discussion

If there is a single group of genes accounting for the elevated spore resistances seen in various strains of *B. pumilus* and *B. safensis* then the relevant genes should be shared by all three strains examined here but absent in the type strain. The fact that the extent of resistance and type of resistance (radiation, desiccation etc.) varies suggests there may not be a single set of genes involved. In any event, the distinctions in resistance seen may occur due to regulatory differences resulting in key genes associated with resistance being expressed at different levels or for different times [73-75]. Although not correlated with resistance information, it is of interest that in FO-36b, there is a dUTPase and a DNA recombinase gene included in the *Bacillus* bacteriophage SPP1 (NC_004166.2) homologous region.

Phage insertions

Conjugative elements and phage-mediated insertions play major roles in the evolution of bacteria [76] by contributing to the genetic variability between closely related bacterial strains[77]. Such variability is often implicated in the phenotypical differences such as bacterial pathogenesis [77-80]. Bacteriophage-mediated horizontal gene transfer enhances bacterial adaptive responses to environmental changes such as the rapid spread of antibiotic resistance [81]. Furthermore, phages mediate inversions, deletions and chromosomal rearrangements, which help shunt genes that could directly impact the phenotype between related strains [77] or between phylogenetically distant strains via horizontal gene transfer (HGT)[82]. All of these evolutionary events have implications for selection and fitness.

297 The first phage insertion in FO-36b is homologous to the *Bacillus* bacteriophage SPP1.
298 The SPP1 is a 44-kb virulent *Bacillus subtilis* phage, well-known for its ability to mediate
299 generalized transduction, a widespread mechanism for the transfer of any gene from one
300 bacterium to another [83]. The second insertion is homologous to *Brevibacillus* phage Jimmer 1,
301 which is one of several myoviruses that specifically target *Paenibacillus larvae*, a *Firmicute*
302 bacterium, as a host [84].

303 The *B. safensis* strain lacks the ICEBs1-like element that was previously found in SAFR-
304 032 and as an incomplete analog in ATCC7061 [39]. As reported earlier [39], the ICEBs1-like
305 element does harbor some SAFR-032 unique genes and thus, their presence was suggested as
306 being possibly responsible for the resistance properties of SAFR-032. The absence of the
307 ICEBs1-like element in the FO-36b genome suggests that this may not be the case. FO-36b has
308 an established phenotype showing spore resistance to peroxide exceeding that of the other JPL-
309 CRF isolates [13]. SAFR-032 spores have been demonstrated to show resistance to UV radiation
310 exceeding that of the other JPL-CRF isolates [16]. Given that both FO-36b and SAFR-032
311 harbor genes unique to each of them, on their respective phage elements (the two insertion
312 elements in the case of FO-36b that are reported here and the ICEBs1-like element in the case of
313 SAFR-032), a role of these unique genes in their respective unique spore phenotypes cannot be
314 entirely ruled out.

315 Furthermore, more than one-half of the *in silico* predicted phage gene products are
316 hypothetical proteins without any assigned functions [85-89]. Comparative genomic approaches
317 use closely related phages from different host organisms and exploit the modular organization of
318 phage genomes [90]. However, these methods are not adequate to address the hypothetical

319 protein coding ORFs that are unique to phage insertions found in a given microbial strain that
320 displays unique phenotypes as in the case of FO-36b and SAFR-032.

321 Hypothetical phage proteins are considered potential candidates for bacterial detection
322 and antimicrobial target selection. In recent times, efforts towards discovering phage-based
323 antimicrobials have led to the experimental characterization of specific phage proteins [91]. The
324 identification of hypothetical ORFs unique to FO-36b and SAFR-032 phage insertion elements
325 mark them out as potential biomarker candidates for the identification/detection of such strains.

326 The distribution of the phage elements is not consistently associated with resistance
327 properties. The Jimmer1 phage includes many genes found in all the strains whether resistant or
328 not. The previously highlighted ICEBs1 like element found in the resistant SAFR-032 is not
329 found in the resistant FO-36b strain. The SPP1 element found in the resistant ATCC7061 strain
330 is missing in SAFR-032. One might speculate that individual phage elements might have been
331 transferred to the main genome in the last two cases thereby maintaining consistency with
332 resistance properties. However, no examples of this were found.

333

334 **Non-phage associated genes**

335 Genes shared by the three resistant spore producing strains but not the non-resistant ATCC7061
336 strain are candidates for association with thee resistance properties. Of the 65 ORFs we had
337 reported earlier to be uniquely shared by SAFR-032 and FO-36b [38], 59 are shared by the
338 MERTA strain (Additional file 4: Table S4). When the analysis is extended to all 61 genomes it
339 was found that in each case at least one additional organism had a homolog to the candidate
340 gene. For example, one of these ORFs (FO-36b locus tag RS87_09285), is found to be shared by
341 *B. safensis* MROC1 (isolated from the feces of *Gallus gallus*) and *B. safensis* RP10 (isolated

342 from soils contaminated with heavy metals in Chile). Most of the strains containing these genes
343 are isolates from environments that have some extreme stress component. However, it is not
344 known if the stress component would include resistance to radiation or peroxide. Based on their
345 names alone, some of these strains, such as *B. altitudinis*, and *B. stratosphericus* may be of
346 special interest for further comparison and investigation of their spore resistance properties.
347
348 **Highly unique open reading frames**
349 The nine FO-36b ORFs (hypothetical proteins) that were found to be absent from all the *B.*
350 *safensis/ B. pumilus* (and the *Bacillus sp*. WP8) genomes available in the NCBI database (Table
351 2A) may be envisioned as possibly contributing to the FO-36b spore resistance. Four of these
352 highly unique ORFs are found on phage elements (one ORF, RS87_03140 on the *Bacillus*
353 bacteriophage SPP1 insertion and three ORFs, viz., RS87_14155, RS87_14285, and
354 RS87_14310 on the *Brevibacillus* phage Jimmer 1 insertion). This is similar to the situation with
355 the ICEBs-1 like element in SAFR-032 that harbors unique SAFR-032 ORFs [39]. Four other
356 ORFs had fewer than 5 homologs found in other *B. pumilus*/*B. safensis* genomes. Two of these
357 four ORFs, are also found on the phage elements and hence could be random remnants of lateral
358 transfer.
359
360 **Genes involved in peroxide resistance and DNA repair**
361 We had previously reported 15 peroxide resistance genes in SAFR-032, of which 2 were not
362 shared by either the earlier draft version of FO-36b, or the type strain ATCC7061 [38]. Five of
363 these peroxide genes were uniquely shared by SAFR-032 and the earlier draft version of the FO-
364 36b genome. Of the 8 SAFR-032 DNA repair genes reported then, 5 were not shared by FO-36b

or ATCC7061. We verified those results against the now complete FO-36b genome, and the status of the genes remains the same as before.

We also looked at the gene coding for 'Dps', which is a DNA-binding protein. Dps is very well-characterized for providing protection to cells during exposure to severe environmental conditions such as oxidative stress and nutritional deprivation in gram negative bacteria such as *E. coli* [92] as well as gram positive *Firmicutes* species such as *Staphylococcus aureus* [93], *B. subtilis* [94], *B. anthracis* [95, 96] and *B. cereus* [97, 98]. With its tripartite involvement in DNA binding, iron sequestration, and ferroxidase activity, Dps plays important roles in iron and hydrogen peroxide detoxification and acid resistance [99, 100]. The homolog for the *dps* gene in *Bacillus* strains is '*mrgA'* [101], which is highly conserved amongst the resistant spore-producing FO-36b and SAFR-032, as well as the non-resistant spore-producing ATCC7061 strain. Likewise, other peroxide resistance genes were checked for their presence/absence and were all found conserved in the four genomes. Thus it is unlikely that any of these genes play any role in the resistances seen in *B. safensis* FO-36b and *B. pumilus* SAFR-032.

Antibiotic resistance

There is increasing concern about bacterial pathogenicity under microgravity and/or in human spaceflight [102]. This is validated by reports that several microbial strains isolated from, or exposed to space environments, show resistance to desiccation, heat-shock, and/or applied antibiotics [103, 104]. A global analysis of the four genomes was undertaken to identify the presence of known antibiotic resistance related mutations. It was found that the FO-36b and SAFR genomes had significantly larger numbers (approximately 100-200 more) of the mutations as compared with the MERTA and ATCC7061 genomes. On a comparative scale, the genome of

BSU had almost 200 more AMR related mutations. The mere presence or the number of these mutations as such cannot be linked with the respective antibiotic resistance properties of these strains. However, further analysis of antibiotic susceptibility of these strains is warranted to establish how they differ from other strains.

Phylogenetic analysis

The current study used Whole Genome Phylogenetic Analysis methodology to delineate phylogenetic distances based on whole genomes of organisms. This and the separate genome-genome distance analysis are consistent with, but more detailed than the earlier study [38]. Additionally, the "*gyrA*" tree analysis was found to support the WGPA and GGDC results. In agreement with the earlier studies, the *B. safensis*/ *B. pumilus* strains form a coherent cluster with three large sub clusters (Figure 8, 9). One of the large sub clusters includes the FO-36b, and MERTA strains as well as all other *B. safensis* strains. *B. pumilus* strains in this grouping may be usefully renamed as *B. safensis*. SAFR-032 and ATCC7061 are in a second sub cluster that is exclusively populated with *B. pumilus* strains. The third sub cluster includes all members of the *B. altitudinis* group and many *B. pumilus* strains.

Conclusions

A recent report [105] has implicated that the opposing effects of environmental DNA damage and DNA repair result in elevated rates of genome rearrangements in radiation-resistant bacteria that belong to multiple, phylogenetically independent groups including *Deinococcus*. This view is not consistent with the four genomes examined in detail here as few arrangements are observed. Comparison with earlier results [38, 39] did not yield anything new and thus although candidates continue to exist, no specific gene has been identified as likely being responsible for

410 the resistances exhibited by these organisms. The differences in resistance properties can easily
411 be attributed to changes in expression level but of what gene or genes? With a larger
412 phylogenetic tree now available, it should be possible to select a representative subset of strains
413 for further resistance studies as well as sequencing.

414

415 ## List of abbreviations used

416 FO-36b - *B. safensis* FO-36b ^{T} (Genbank Accession no: - CP010405).

417 SAFR-032 – *B. pumilus* SAFR-032.

418 ATCC7061 – *B. pumilus* ATCC7061^{T}.

419 MERTA - *B. safensis* JPL-MERTA-8-2.

420 BSU - *B. subtilis subsp. subtilis str*. 168.

421

422 ## Declarations

423 ### **1.** Ethics approval and consent to participate

424 Not Applicable

425 ### **2.** Consent for publication

426 Not applicable

427 ### **3.** Availability of data and material

428 The datasets used and analyzed within the current study are available from the NCBI
429 Website as referenced in the paper. The sequence of the *B. safensis* FO-36b strain is being
430 deposited with the NCBI/Genbank with the accession number CP010405. Until the deposit is
431 complete, the data will be available from the corresponding author.

4. Competing Interests

All authors declare they have no competing interests.

5. Funding

This work was funded in part by NASA Grant NNX14AK36G to GEF. Discussions relating to the NASA project led to this work and provided resources for the design of the study, and writing of the manuscript.

6. Author's contributions

KJV, VGS, and GEF conceived and designed the study. MRT annotated and curated the annotated genome, analyzed the data, performed the comparative genome analysis. SM and MRT prepared the tables and the figures. VGS prepared the library for NextGen sequencing and processed the resulting data. AW performed the sequencing. VGS and ROG conducted local sequencing studies to order contigs and close the genome. KRG provided genomic DNA. MRT, VGS and GEF prepared a draft paper which was finalized with help from all the authors. All authors read and approved the final manuscript.

7. Acknowledgements

The authors acknowledge the use of the Maxwell/Opuntia Cluster and the advanced support from the Center of Advanced Computing and Data Systems at the University of Houston to carry out the research presented here. The authors would also like to thank Dr. Jan Meier-Kolthoff, Department of Bioinformatics, Leibniz Institute DSMZ (German Collection of Microorganisms and Cell Cultures), Inhoffenstraße 7 B, 38124 Braunschweig, Germany, for help with the Whole Genome Phylogenetic Analysis. This work was supported in part by NASA Grant NNX14AK36G to GEF.

8. Authors' information

Madhan R Tirumalai[1], Victor G. Stepanov[1], Saied Montazari[1], Racquel O. Gonzalez[1], Andrea Wünsche[1], Kasturi Venkateswaran[2], George. E. Fox[1§]

[1]Department of Biology and Biochemistry, University of Houston, Houston, TX, 77204-5001.

[2] Biotechnology & Planetary Protection Group, NASA Jet Propulsion Laboratories, California Institute of Technology, Pasadena, CA, 91109.

[§]Corresponding author:

Dr. George E. Fox

Dept. Biology & Biochemistry

University of Houston

Houston, TX 77204-5001

713-743-8363; 713-743-8351 (FAX)

fox@uh.edu

Email addresses:

MRT: mrtirum2@mail.uh.edu

VGS: vstepano@central.uh.edu

AW: andrea.wunsche@gmail.com

SM: SM: montazaris@gmail.com

ROG: raquel.gonzalez100@yahoo.com

KV: kasthuri.j.venkateswaran@jpl.nasa.gov

GEF: fox@uh.edu

Endnotes

None

References

1. Moissl-Eichinger C, Auerbach AK, Probst AJ, Mahnert A, Tom L, Piceno Y, Andersen GL, Venkateswaran K, Rettberg P, Barczyk S *et al*: **Quo vadis? Microbial profiling revealed strong effects of cleanroom maintenance and routes of contamination in indoor environments**. *Sci Rep* 2015, **5**:9156.
2. Venkateswaran K, La Duc MT, Horneck G: **Microbial existence in controlled habitats and their resistance to space conditions**. *Microbes Environ* 2014, **29**(3):243-249.
3. Venkateswaran K, Vaishampayan P, Cisneros J, Pierson DL, Rogers SO, Perry J: **International Space Station environmental microbiome - microbial inventories of ISS filter debris**. *Appl Microbiol Biotechnol* 2014, **98**(14):6453-6466.
4. Rummel JD, Race MS, Horneck G: **Ethical considerations for planetary protection in space exploration: a workshop**. *Astrobiology* 2012, **12**(11):1017-1023.
5. Debus A: **The European standard on planetary protection requirements**. *Research in microbiology* 2006, **157**(1):13-18.
6. Mora M, Mahnert A, Koskinen K, Pausan MR, Oberauner-Wappis L, Krause R, Perras AK, Gorkiewicz G, Berg G, Moissl-Eichinger C: **Microorganisms in Confined Habitats: Microbial Monitoring and Control of Intensive Care Units, Operating Rooms, Cleanrooms and the International Space Station**. *Front Microbiol* 2016, **7**:1573.
7. Mahnert A, Vaishampayan P, Probst AJ, Auerbach A, Moissl-Eichinger C, Venkateswaran K, Berg G: **Cleanroom Maintenance Significantly Reduces Abundance but Not Diversity of Indoor Microbiomes**. *PLoS One* 2015, **10**(8):e0134848.
8. Rummel JD, Meyer MA: **A consensus approach to planetary protection requirements: recommendations for Mars lander missions**. *Adv Space Res* 1996, **18**(1-2):317-321.
9. Venkateswaran K, Chung S, Allton J, Kern R: **Evaluation of various cleaning methods to remove bacillus spores from spacecraft hardware materials**. *Astrobiology* 2004, **4**(3):377-390.
10. DeVincenzi DL, Stabekis PD: **Revised planetary protection policy for solar system exploration**. *Adv Space Res* 1984, **4**(12):291-295.
11. Pillinger JM, Pillinger CT, Sancisi-Frey S, Spry JA: **The microbiology of spacecraft hardware: lessons learned from the planetary protection activities on the Beagle 2 spacecraft**. *Res Microbiol* 2006, **157**(1):19-24.
12. Race MS: **Anticipating the reaction: public concern about sample return missions**. *Planet Rep* 1994, **14**(4):20-22.
13. 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.

14. La Duc MT, Satomi M, Venkateswaran K: **Bacillus odysseyi sp. nov., a round-spore-forming bacillus isolated from the Mars Odyssey spacecraft**. *Int J Syst Evol Microbiol* 2004, **54**(Pt 1):195-201.
15. Schuerger AC, Mancinelli RL, Kern RG, Rothschild LJ, McKay CP: **Survival of endospores of Bacillus subtilis on spacecraft surfaces under simulated martian environments: implications for the forward contamination of Mars**. *Icarus* 2003, **165**(2):253-276.
16. 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):159-163.
17. Vaishampayan P, Miyashita M, Ohnishi A, Satomi M, Rooney A, La Duc MT, Venkateswaran K: **Description of Rummeliibacillus stabekisii gen. nov., sp. nov. and reclassification of Bacillus pycnus Nakamura et al. 2002 as Rummeliibacillus pycnus comb. nov**. *Int J Syst Evol Microbiol* 2009, **59**(Pt 5):1094-1099.
18. Vaishampayan P, Probst A, Krishnamurthi S, Ghosh S, Osman S, McDowall A, Ruckmani A, Mayilraj S, Venkateswaran K: **Bacillus horneckiae sp. nov., isolated from a spacecraft-assembly clean room**. *Int J Syst Evol Microbiol* 2010, **60**(Pt 5):1031-1037.
19. Vaishampayan P, Roberts AH, Augustus A, Pukall R, Schumann P, Schwendner P, Mayilraj S, Salmassi T, Venkateswaran K: **Deinococcus phoenicis sp. nov., an extreme ionizing-radiation-resistant bacterium isolated from the Phoenix Lander assembly facility**. *Int J Syst Evol Micr* 2014, **64**(Pt 10):3441-3446.
20. Msadek T, Dartois V, Kunst F, Herbaud ML, Denizot F, Rapoport G: **ClpP of Bacillus subtilis is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation**. *Mol Microbiol* 1998, **27**(5):899-914.
21. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Medigue C *et al*: **From a consortium sequence to a unified sequence: the Bacillus subtilis 168 reference genome a decade later**. *Microbiology* 2009, **155**(Pt 6):1758-1775.
22. Benardini JN, Sawyer J, Venkateswaran K, Nicholson WL: **Spore UV and acceleration resistance of endolithic Bacillus pumilus and Bacillus subtilis isolates obtained from Sonoran desert basalt: implications for lithopanspermia**. *Astrobiology* 2003, **3**(4):709-717.
23. Horneck G, Moeller R, Cadet J, Douki T, Mancinelli RL, Nicholson WL, Panitz C, Rabbow E, Rettberg P, Spry A *et al*: **Resistance of Bacterial Endospores to Outer Space for Planetary Protection Purposes-Experiment PROTECT of the EXPOSE-E Mission**. *Astrobiology* 2012, **12**(5):445-456.
24. 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):487-497.
25. Checinska A, Probst AJ, Vaishampayan P, White JR, Kumar D, Stepanov VG, Fox GE, Nilsson HR, Pierson DL, Perry J *et al*: **Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities**. *Microbiome* 2015, **3**:50.

26. Mora M, Perras A, Alekhova TA, Wink L, Krause R, Aleksandrova A, Novozhilova T, Moissl-Eichinger C: **Resilient microorganisms in dust samples of the International Space Station-survival of the adaptation specialists**. *Microbiome* 2016, **4**(1):65.
27. Moissl-Eichinger C, Cockell C, Rettberg P: **Venturing into new realms? Microorganisms in space**. *FEMS microbiology reviews* 2016, **40**(5):722-737.
28. Satomi M, La Duc MT, Venkateswaran K: **Bacillus safensis sp. nov., isolated from spacecraft and assembly-facility surfaces**. *Int J Syst Evol Microbiol* 2006, **56**(Pt 8):1735-1740.
29. Ghosh S, Osman S, Vaishampayan P, Venkateswaran K: **Recurrent isolation of extremotolerant bacteria from the clean room where Phoenix spacecraft components were assembled**. *Astrobiology* 2010, **10**(3):325-335.
30. La Duc MT, Dekas A, Osman S, Moissl C, Newcombe D, Venkateswaran K: **Isolation and characterization of bacteria capable of tolerating the extreme conditions of clean room environments**. *Appl Environ Microbiol* 2007, **73**(8):2600-2611.
31. La Duc MT, Kern R, Venkateswaran K: **Microbial monitoring of spacecraft and associated environments**. *Microb Ecol* 2004, **47**(2):150-158.
32. Opfell JB, Bandaruk W: **Microbial contaminants in the interiors of spacecraft components**. *Life Sci Space Res* 1966, **4**:133-165.
33. Puleo JR, Fields ND, Bergstrom SL, Oxborrow GS, Stabekis PD, Koukol R: **Microbiological profiles of the Viking spacecraft**. *Appl Environ Microbiol* 1977, **33**(2):379-384.
34. Newcombe DA, Schuerger AC, Benardini JN, Dickinson D, Tanner R, Venkateswaran K: **Survival of spacecraft-associated microorganisms under simulated martian UV irradiation**. *Appl Environ Microbiol* 2005, **71**(12):8147-8156.
35. Nicholson WL, Setlow B, Setlow P: **UV photochemistry of DNA in vitro and in Bacillus subtilis spores at earth-ambient and low atmospheric pressure: implications for spore survival on other planets or moons in the solar system**. *Astrobiology* 2002, **2**(4):417-425.
36. Coil DA, Neches RY, Lang JM, Brown WE, Severance M, Cavalier D, Eisen JA: **Growth of 48 built environment bacterial isolates on board the International Space Station (ISS)**. *PeerJ* 2016, **4**:e1842.
37. Espariz M, Zuljan FA, Esteban L, Magni C: **Taxonomic Identity Resolution of Highly Phylogenetically Related Strains and Selection of Phylogenetic Markers by Using Genome-Scale Methods: The Bacillus pumilus Group Case**. *PLoS One* 2016, **11**(9):e0163098.
38. Tirumalai MR, Rastogi R, Zamani N, O'Bryant Williams E, Allen S, Diouf F, Kwende S, Weinstock GM, Venkateswaran KJ, Fox GE: **Candidate genes that may be responsible for the unusual resistances exhibited by Bacillus pumilus SAFR-032 spores**. *PLoS One* 2013, **8**(6):e66012.
39. 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):767-774.
40. Stepanov VG, Tirumalai MR, Montazari S, Checinska A, Venkateswaran K, Fox GE: **Bacillus pumilus SAFR-032 Genome Revisited: Sequence Update and Re-Annotation**. *PLoS One* 2016, **11**(6):e0157331.

41. Coil DA, Benardini JN, Eisen JA: **Draft Genome Sequence of Bacillus safensis JPL-MERTA-8-2, Isolated from a Mars-Bound Spacecraft**. *Genome Announc* 2015, **3**(6).
42. Borodina TA, Lehrach H, Soldatov AV: **DNA purification on homemade silica spin-columns**. *Anal Biochem* 2003, **321**(1):135-137.
43. **Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software].** [Available at https://github.com/najoshi/sickle.]
44. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
45. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I: **ABySS: a parallel assembler for short read sequence data**. *Genome Res* 2009, **19**(6):1117-1123.
46. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V *et al*: **Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation**. *Omics* 2008, **12**(2):137-141.
47. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL: **NCBI BLAST: a better web interface**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W5-9.
48. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS: **PHAST: a fast phage search tool**. *Nucleic acids research* 2011, **39**(Web Server issue):W347-352.
49. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS: **PHASTER: a better, faster version of the PHAST phage search tool**. *Nucleic acids research* 2016, **44**(W1):W16-21.
50. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M: **Genome sequence-based species delimitation with confidence intervals and improved distance functions**. *BMC Bioinformatics* 2013, **14**:60.
51. Meier-Kolthoff JP AA, Klenk H-P, Göker M: **Highly parallelized inference of large genome-based phylogenies**. *Concurrency and Computation: Practice and Experience* 2014, **26**(10):1715–1729
52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC Bioinformatics* 2009, **10**:421.
53. Lefort V, Desper R, Gascuel O: **FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program**. *Molecular Biology and Evolution* 2015, **32**(10):2798-2800.
54. Farris JS: **Estimating Phylogenetic Trees from Distance Matrices**. *American Naturalist* 1972, **106**(951):645-&.
55. Meier-Kolthoff JP, Klenk HP, Goker M: **Taxonomic use of DNA G plus C content and DNA-DNA hybridization in the genomic age**. *Int J Syst Evol Micr* 2014, **64**:352-356.
56. Levican A, Collado L, Figueras MJ: **Arcobacter cloacae sp nov and Arcobacter suis sp nov., two new species isolated from food and sewage**. *Systematic and Applied Microbiology* 2013, **36**(1):22-27.
57. Shamsi H, Mardani K, Ownagh A: **Phylogenetic analysis of Escherichia coli isolated from broilers with colibacillosis based on gyrA gene sequences**. *Can J Vet Res* 2017, **81**(1):28-32.
58. Abdelbaqi K, Menard A, Prouzet-Mauleon V, Bringaud F, Lehours P, Megraud F: **Nucleotide sequence of the gyrA gene of Arcobacter species and characterization of human ciprofloxacin-resistant clinical isolates**. *Fems Immunol Med Mic* 2007, **49**(3):337-345.

59. La Duc MT, Satomi M, Agata N, Venkateswaran K: **gyrB as a phylogenetic discriminator for members of the Bacillus anthracis-cereus-thuringiensis group**. *J Microbiol Meth* 2004, **56**(3):383-394.
60. Dickinson DN, La Duc MT, Satomi M, Winefordner JD, Powell DH, Venkateswaran K: **MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of Bacillus pumilus spores**. *J Microbiol Methods* 2004, **58**(1):1-12.
61. Fox GE, Wisotzkey JD, Jurtshuk P, Jr.: **How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity**. *Int J Syst Bacteriol* 1992, **42**(1):166-170.
62. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Molecular Biology and Evolution* 2016, **33**(7):1870-1874.
63. Kumar S, Stecher G, Peterson D, Tamura K: **MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis**. *Bioinformatics* 2012, **28**(20):2685-2686.
64. Tamura K, Nei M: **Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees**. *Molecular Biology and Evolution* 1993, **10**(3):512-526.
65. Rzhetsky A, Nei M: **A Simple Method for Estimating and Testing Minimum-Evolution Trees**. *Molecular Biology and Evolution* 1992, **9**(5):945-967.
66. Kumar MNaS: **Molecular Evolution and Phylogenetics**: Oxford University Press, New York.; 2000.
67. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Molecular Biology and Evolution* 1987, **4**(4):406-425.
68. Tamura K, Nei M, Kumar S: **Prospects for inferring very large phylogenies by using the neighbor-joining method**. *P Natl Acad Sci USA* 2004, **101**(30):11030-11035.
69. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0**. *Molecular Biology and Evolution* 2013, **30**(12):2725-2729.
70. Darling AC, Mau B, Blattner FR, Perna NT: **Mauve: multiple alignment of conserved genomic sequence with rearrangements**. *Genome Res* 2004, **14**(7):1394-1403.
71. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN *et al*: **CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database**. *Nucleic Acids Research* 2017, **45**(D1):D566-D573.
72. Harwood CR, Williams DM, Lovett PS: **Nucleotide sequence of a Bacillus pumilus gene specifying chloramphenicol acetyltransferase**. *Gene* 1983, **24**(2-3):163-169.
73. Vertessy BG, Toth J: **Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases**. *Acc Chem Res* 2009, **42**(1):97-106.
74. Nagy GN, Leveles I, Vertessy BG: **Preventive DNA repair by sanitizing the cellular (deoxy)nucleoside triphosphate pool**. *Febs J* 2014, **281**(18):4207-4223.
75. Szabo JE, Nemeth V, Papp-Kadar V, Nyiri K, Leveles I, Bendes AA, Zagyva I, Rona G, Palinkas HL, Besztercei B *et al*: **Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control**. *Nucleic acids research* 2014, **42**(19):11912-11920.

76. Salmond GP, Fineran PC: **A century of the phage: past, present and future**. *Nature reviews Microbiology* 2015, **13**(12):777-786.
77. Brussow H, Canchaya C, Hardt WD: **Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion**. *Microbiology and molecular biology reviews : MMBR* 2004, **68**(3):560-602, table of contents.
78. Neely MN, Friedman DI: **Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release**. *Molecular microbiology* 1998, **28**(6):1255-1267.
79. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H: **Phage as agents of lateral gene transfer**. *Current Opinion in Microbiology* 2003, **6**(4):417-424.
80. Touchon M, de Sousa JAM, Rocha EPC: **Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer**. *Current Opinion in Microbiology* 2017, **38**:66-73.
81. Brown-Jaque M, Calero-Caceres W, Muniesa M: **Transfer of antibiotic-resistance genes via phage-related mobile elements**. *Plasmid* 2015, **79**:1-7.
82. Zinder ND, Lederberg J: **Genetic exchange in Salmonella**. *Journal of bacteriology* 1952, **64**(5):679-699.
83. Yasbin RE, Young FE: **Transduction in Bacillus-Subtilis by Bacteriophage Spp1**. *J Virol* 1974, **14**(6):1343-1348.
84. Merrill BD, Grose JH, Breakwell DP, Burnett SH: **Characterization of Paenibacillus larvae bacteriophages and their genomic relationships to firmicute bacteriophages**. *BMC Genomics* 2014, **15**.
85. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W: **Bacteriophage T4 genome**. *Microbiol Mol Biol R* 2003, **67**(1):86-+.
86. Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, Rasko DA: **Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing Escherichia coli**. *Front Cell Infect Microbiol* 2012, **2**.
87. Cui XL, You JJ, Sun L, Yang XJ, Zhang T, Huang KC, Pan XW, Zhang FJ, He Y, Yang HJ: **Characterization of Pseudomonas aeruginosa Phage C11 and Identification of Host Genes Required for Virion Maturation**. *Sci Rep* 2016, **6**.
88. O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF, Ross RP: **Genome of Staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content**. *Journal of Bacteriology* 2004, **186**(9):2862-2871.
89. Varani AD, Souza RC, Nakaya HI, de Lima WC, de Almeida LGP, Kitajima EW, Chen J, Civerolo E, Vasconcelos ATR, Van Sluys MA: **Origins of the Xylella fastidiosa Prophage-Like Regions and Their Impact in Genome Differentiation**. *PLoS One* 2008, **3**(12).
90. Hendrix RW, Hatfull GF, Smith MCM: **Bacteriophages with tails: chasing their origins and evolution**. *Research in Microbiology* 2003, **154**(4):253-257.
91. Liu J, Dehbi M, Moeck G, Arhin F, Bauda P, Bergeron D, Callejo M, Ferretti V, Ha NH, Kwan T *et al*: **Antimicrobial drug discovery through bacteriophage genomics**. *Nat Biotechnol* 2004, **22**(2):185-191.
92. Almiron M, Link AJ, Furlong D, Kolter R: **A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli**. *Genes & development* 1992, **6**(12B):2646-2654.

93. Ushijima Y, Yoshida O, Villanueva MJ, Ohniwa RL, Morikawa K: **Nucleoid clumping is dispensable for the Dps-dependent hydrogen peroxide resistance in Staphylococcus aureus**. *Microbiology* 2016, **162**(10):1822-1828.

94. Antelmann H, Engelmann S, Schmid R, Sorokin A, Lapidus A, Hecker M: **Expression of a stress- and starvation-induced dps/pexB-homologous gene is controlled by the alternative sigma factor sigmaB in Bacillus subtilis**. *Journal of Bacteriology* 1997, **179**(23):7251-7256.

95. Tu WY, Pohl S, Gizynski K, Harwood CR: **The iron-binding protein Dps2 confers peroxide stress resistance on Bacillus anthracis**. *Journal of Bacteriology* 2012, **194**(5):925-931.

96. Schwartz JK, Liu XS, Tosha T, Diebold A, Theil EC, Solomon EI: **CD and MCD spectroscopic studies of the two Dps miniferritin proteins from Bacillus anthracis: role of O2 and H2O2 substrates in reactivity of the diiron catalytic centers**. *Biochemistry* 2010, **49**(49):10516-10525.

97. Shu JC, Soo PC, Chen JC, Hsu SH, Chen LC, Chen CY, Liang SH, Buu LM, Chen CC: **Differential regulation and activity against oxidative stress of Dps proteins in Bacillus cereus**. *Int J Med Microbiol* 2013, **303**(8):662-673.

98. Wang SW, Chen CY, Tseng JT, Liang SH, Chen SC, Hsieh C, Chen YH, Chen CC: **orf4 of the Bacillus cereus sigB gene cluster encodes a general stress-inducible Dps-like bacterioferritin**. *Journal of Bacteriology* 2009, **191**(14):4522-4533.

99. Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, Chiancone E, Chasteen ND: **Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli**. *The Journal of biological chemistry* 2002, **277**(31):27689-27696.

100. Chiancone E: **Dps proteins, an efficient detoxification and DNA protection machinery in the bacterial response to oxidative stress**. *RENDICONTI LINCEI* 2008, **19**(3):261.

101. Chen L, Helmann JD: **Bacillus subtilis MrgA is a Dps(PexB) homologue: evidence for metalloregulation of an oxidative-stress gene**. *Molecular microbiology* 1995, **18**(2):295-300.

102. Taylor PW: **Impact of space flight on bacterial virulence and antibiotic susceptibility**. *Infect Drug Resist* 2015, **8**:249-262.

103. Urbaniak C, Sielaff AC, Frey KG, Allen JE, Singh N, Jaing C, Wheeler K, Venkateswaran K: **Detection of antimicrobial resistance genes associated with the International Space Station environmental surfaces**. *Sci Rep* 2018, **8**(1):814.

104. Zhang D, Chang D, Zhang X, Yu Y, Guo Y, Wang J, Li T, Xu G, Dai W, Liu C: **Genome Sequence of Escherichia coli Strain LCT-EC52, Which Acquired Changes in Antibiotic Resistance Properties after the Shenzhou-VIII Mission**. *Genome Announc* 2014, **2**(2).

105. Repar J, Supek F, Klanjscek T, Warnecke T, Zahradka K, Zahradka D: **Elevated Rate of Genome Rearrangements in Radiation-Resistant Bacteria**. *Genetics* 2017, **205**(4):1677-1689.

Figures

Figure 1. The *Bacillus* bacteriophage SPP1 (NC_004166) homologous region in the *B. safensis* FO-36b genome, as compared with the equivalent genomic regions of *B. pumilus* ATCC7061^{T}, *B. pumilus* SAFR-032 and *B .subtilis subsp. subtilis str.* 168. The locus tag numbers are given inside the boxes/rectangles. Red diamonds denote absence of a single gene/homolog. Red rectangle denotes absence of a series/cluster of ORFs/genes. Green box encloses the phage insertion region. Green diamond denotes absence of a single gene/homolog within the phage. "hyd" = hydrolase, "chp" = conserved hypothetical protein, "pept" = peptidase, "hp" = hypothetical protein, "TR" = transcriptional regulator, "Ps"= pseudogene, "lp" = lipoprotein, “gsp” = group specific protein, “oxi” = oxidase.

Figure 2. The *Bacillus* bacteriophage SPP1 (NC_004166) homologous region in the *B. safensis* FO-36b genome, as compared with the equivalent genomic region of *B. safensis* JPL_MERTA8-2. Red diamonds denote absence of a single gene/homolog. Red rectangle denotes absence of a series/cluster of ORFs/genes. Green box encloses the phage insertion region.

Figure 3. Overall scheme of the *Brevibacillus* phage Jimmer1 (NC_029104) phage insertion in the *B. safensis* FO-36b genome. The three blocks A, B and C and the genes they encompass are shown. The first part of Block C is a duplication of Block A.

Figure 4. The *Brevibacillus* phage Jimmer1 (NC_029104) phage insertion in the *B. safensis* FO-36b genome as compared with the equivalent region in the genome of *B. pumilus* ATCC7061^{T}. Black box encloses the phage insertion region(s). Green (dashed line) box corresponds to block

A. Green (dotted line) box corresponds to block B. Blue (dashed line) box corresponds to block C. Red (dashed line) box encloses 'terminase' genes. A diamond denotes absence of a single gene/homolog within the phage, while rectangle denotes absence of a cluster of genes/homologs. "hp" = hypothetical protein, "chp" = conserved hypothetical protein, "pp" = phage portal protein, "sp" = structural protein, "sgp" = spore germination protein, "int" = integrase.

Figure 5. The *Brevibacillus* phage Jimmer1 (NC_029104) phage insertion in the *B. safensis* FO-36b genome as compared with the equivalent region in the genome of *B. safensis* JPL_MERTA8-2. "hp" = hypothetical protein, "chp" = conserved hypothetical protein, "pp" = phage portal protein, "sp" = structural protein, "sgp" = spore germination protein, "int" = integrase.

Figure 6. The *Brevibacillus* phage Jimmer1 (NC_029104) phage insertion in the *B. safensis* FO-36b genome as compared with the equivalent region in the genome of *B. pumilus* SAFR-032. "hp" = hypothetical protein, "chp" = conserved hypothetical protein, "pp" = phage portal protein, "sp" = structural protein, "sgp" = spore germination protein, "int" = integrase.

Figure 7. The *Brevibacillus* phage Jimmer1 (NC_029104) phage insertion in the *B. safensis* FO-36b genome as compared with the equivalent region in the genome of *B. subtilis*. "hp" = hypothetical protein, "chp" = conserved hypothetical protein, "pp" = phage portal protein, "sp" = structural protein,"sgp" = spore germination protein, "int" = integrase.

Figure 8. Whole genome Phylogenetic Analysis (WGPA) using the latest version of the Genome-BLAST Distance Phylogeny (GBDP).

B. safensis FO-36b, *B. safensis* JPL_MERTA8-2B, *B. pumilus* SAFR-032, and *B. pumilus* ATCC7061^T are highlighted in red dash-lined rectangles

Figure 9. Molecular Phylogenetic analysis by the Maximum Likelihood method

B. safensis FO-36b, *B. safensis* JPL_MERTA8-2B, *B. pumilus* SAFR-032, and *B. pumilus* ATCC7061^T are highlighted in red dash-lined rectangles.

Tables and captions

Table 1: List of *B. safensis* FO-36b genes not shared by *B. pumilus* SAFR-032, *B. pumilus* ATCC7061^T and *B. safensis* JPL_MERTA8-2.

^ - Genes/ORFs not found in either *B. pumilus* SAFR-032, or, *B. pumilus* ATCC7061^T, or, *B. safensis* JPL_MERTA8-2 or any other *B. pumilus* and *B. safensis* genomes. HP – Hypothetical protein(s)

Table 2: *B. safensis* F0-36b unique genes.

* Genes that are part of phage elements

Table 3: *B. safensis* FO-36b genes (hypothetical proteins) with fewer than 5 homologs

* Genes that are part of phage elements

Additional files

Additional file 1: Table S1. In silico DNA-DNA hybridization (DDH) values showing Genome-genome distance [50] relationship values for the genomes of various *B. pumilus, B. safensis, B. altitudinis* strains. The genomes of *Geobacillus kaustophilus, and B. subtilis subsp. subtilis str.* 168 serving as outliers in the *Firmicutes* group and that of gram-negative *E.coli* MG1655, as a non-*Firmicutes* outlier.

Additional file 2: Table S2. Presence and absence of the *B. safensis* FO-36b CRISPR module element protein(s) in the other *B. pumilus* / *B. safensis* genomes.

Additional file 3: Table S3. *B. safensis* FO-36b characteristic genes (ORFs/genes that are absent from *B. pumilus* SAFR-032, *B. pumilus* ATCC7061^{T}, and, *B. safensis* JPL-MERTA-8-2) and their occurrence (presence/absence) in the *B. pumilus*/*B. safensis* genomes available in the NCBI database. P: Present, A: Absent, *found on phage insertions.

Additional file 4: Table S4. *B. safensis* F0-36b genes reported earlier as shared by *B. pumilus* SAFR-032 and not found in the *B. pumilus* ATCC7061^{T} strain [38], compared with the *B. safensis* JPL-MERTA-8-2 strain, and the other *B. pumilus / B. safensis* genomes.

Additional file 5: Figure S1. Whole genome alignment of the previously existing *B. safensis* FO-36b sequence (GCA_000691165.1 / ASJD00000000) with our current updated sequence (CP010405) using Mauve [70].

881 **Additional file 6: Figure S2.** Molecular Phylogenetic analysis by the Neighbor-Joining method.
882 *B. safensis* FO-36b, *B. safensis* JPL_MERTA8-2B, *B. pumilus* SAFR-032, and *B. pumilus*
883 ATCC7061^{T} are highlighted in red dash-lined rectangles.
884
885 **Additional file 7: Figure S3.** Molecular Phylogenetic analysis using the Minimum Evolution
886 method. *B. safensis* FO-36b, *B. safensis* JPL_MERTA8-2B, *B. pumilus* SAFR-032, and *B.*
887 *pumilus* ATCC7061^{T} are highlighted in red dash-lined rectangles.
888
889

Table 1: List of *B. safensis* FO-36b (characteristic) genes not shared by *B. pumilus* SAFR-032, *B. pumilus* ATCC7061^T and *B. safensis* JPL_MERTA8-2.

Hypothetical protein coding ORFs (unknown function)	
Locus tag RS87_#	
01590 **02635-40** **02695-700** **02960** **03370** **03615-20** **04125** **04345** **06055^** **09165** **09820^** **12770^** **14125-30** **14140** **14150^** **15275^** **17540** **18710** **18745** **18755-60**	**HP (25)**
02980, **02995,** **03000,** **03010,** **03030-35,** **03050^,** **03065,** **03075-095,** **03110,** **03125,** **03195-200,** **03220,** **14285,** **14395,** **14400,** **14310^,** **14320^,** **14410.**	**HP on phages (24)**
03215 (ps)	**Pseudogene (on phage) (1)**

Locus tag RS87_#	Gene	Function
03015	recombinase RecT	DNA repair/recombination (2)
03060	dUTPase	
03190	alkaline phosphatase	Phage element components (2)
03210	phage tail protein	
03225	protein XhlA	
04350	CRISPR module RAMP protein Cmr1	Type III-B CRISPR element components (6)
04355	CRISPR-associated protein Cas10/Cmr2	
04360	CRISPR module-associated protein Cmr3	
04365	CRISPR module RAMP protein Cmr4	
04370	CRISPR module-associated protein Cmr5	
04375	CRISPR module RAMP protein Cmr6	
09105	protein IolH	Metabolism (4)
09130	5-deoxy-glucuronate isomerase	
09150	isomerase	
09160	myo-inosose-2 dehydratase	

^ - Genes/ORFs not found in either *B .pumilus* SAFR-032, or, *B. pumilus* ATCC7061^T, or, *B. safensis* JPL_MERTA8-2 or any other *B. pumilus* and *B. safensis* genomes.

HP – Hypothetical protein(s).

Table 2: FO-36b unique genes (hypothetical proteins)

Locus tag RS87_#
03140*
09820
12770
14110
14145
14150
14155*
14285*
14310*

*** Genes that are part of phage elements**

Table 3: FO-36b genes (hypothetical proteins) with fewer than 5 homologs

Locus tag RS87_#	
03030*	**only four homologs in** ***B.safensis* U41 (GCA_001938685.1),** ***B.safensis* U17-1 (GCA_001938705.1),** ***B.pumilus* CCMA-560 (GCA_000444805.1), and,** ***B.pumilus* strain 36R_ATNSAL (GCA_002744245.1).**
03050*	**only one homolog in *B.pumilus* strain 36R_ATNSAL (GCA_002744245.1).**
03110	**only two homologs in** ***B.safensis* 7783 (GCA_002276315.1), and,** ***B safensis* Bcs96 (GCA_002155005.1).**
04125	**only one homolog in *B.pumilus* PE09-72 (GCA_002174275.1).**

*** Genes that are part of phage elements**

https://doi.org/10.1101/283937

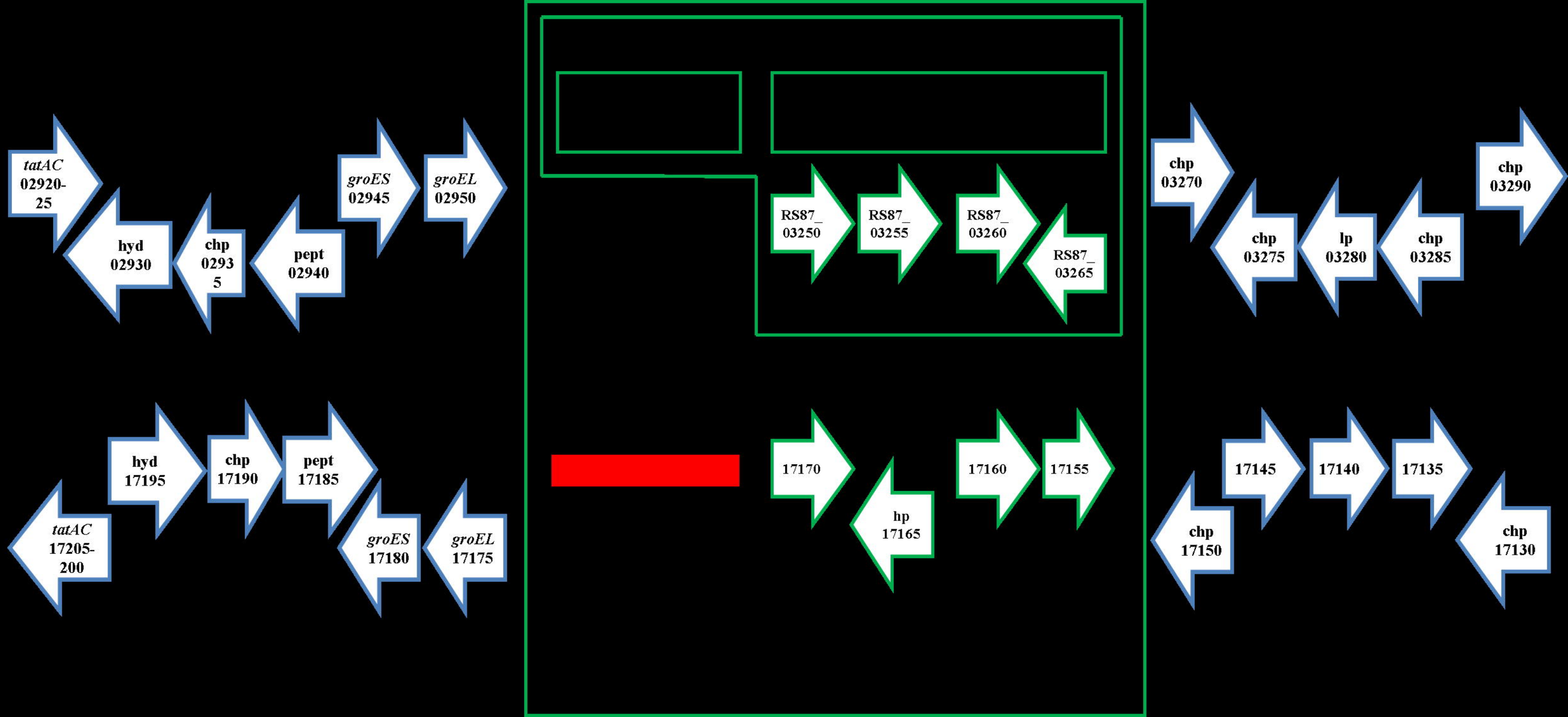
Phage 1 insertion

tatAC 02920-25
hyd 02930
chp 02935
pept 02940
groES 02945
groEL 02950
RS87_ 03260
RS87_ 03265
chp 03270
chp 03275
lp 03280
chp 03285
chp 03290

tatAC 3549-48
hyd 3547
3546
3545
groES 3544
groEL 3543
hp 3542
TR 3541
chp 3540
hp 3539
gsp 3538
Ps 3537-3536
oxi 3535
Ps 3534-3533
chp 3532
Ps 3531
lp 3530
chp 3529
chp 3528

tatAC 0529-30
hyde 0531
chp 0532
pept 0533
groES 0534
groEL 0535
chp 0536
chp 0537
lp 0538
chp 0539
chp 0540

tatAC 05980-90
YdiK 06000
YdiL 06010
groES 06020
groEL 06030
chp 06250


tatAC 02920-25
hyd 02930
chp 02935
pept 02940
groES 02945
groEL 02950
RS87_03250
RS87_03255
RS87_03260
RS87_03265
chp 03270
chp 03275
lp 03280
chp 03285
chp 03290
hyd 17195
chp 17190
pept 17185
tatAC 17205-200
groES 17180
groEL 17175
17170
hp 17165
17160
17155
17145
17140
17135
chp 17150
chp 17130

Phage 2 insertion

rapB 14115 — *chp* 14120 — *int* 14450 — *tRNA-Ala* 14455 — *chp* 14155 — *chp* 14175 — *pp* 14230 — *chp* 14270 — *pp* 14275 — 14280 — *chp* 14285 — *pp* 14290 — *pp* 14295 — 14300 — 14305 — *hp* 14310 — *chp* 14315 — *chp* 14320-325

chp 14460 — *pp* 14535 — *pp* 14575 — 14580 — 14590

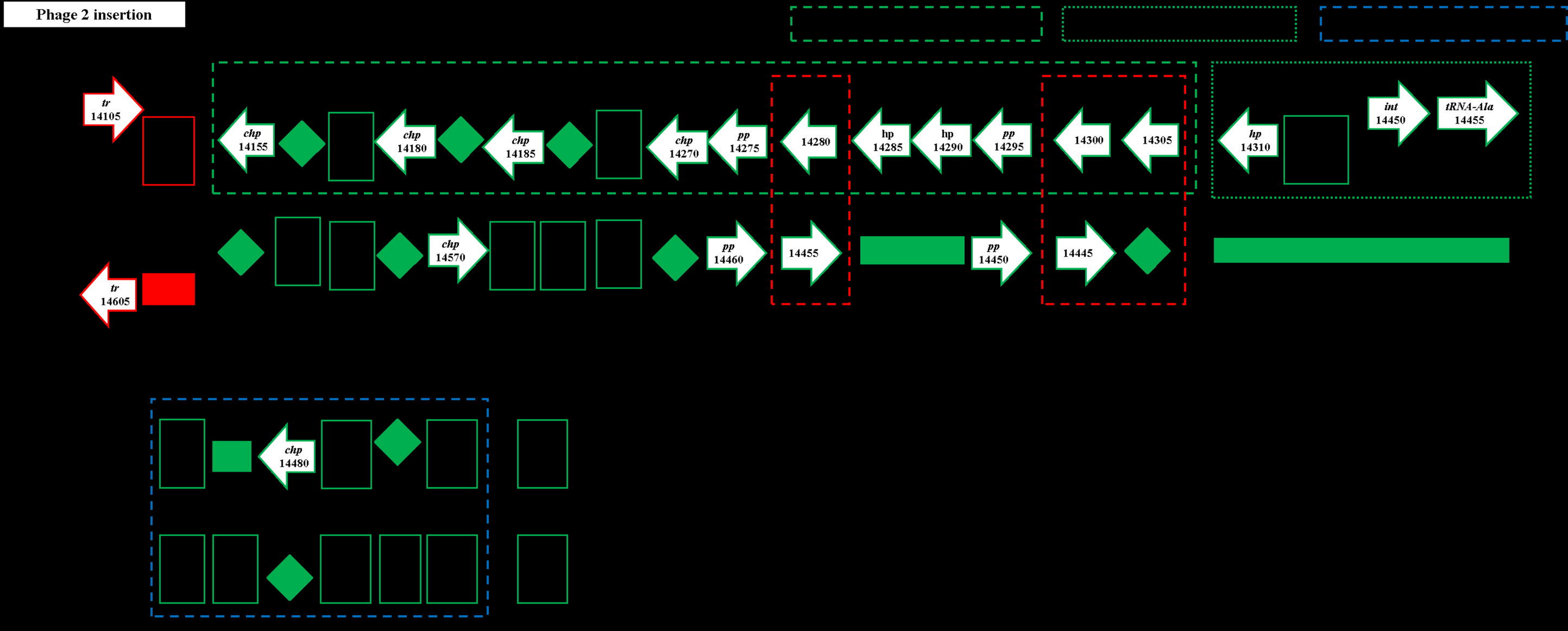
rapB 0014 — *chp* 0015 — *chp* 0024 — *ps* 0035 — *pp* 0036 — *pp* 0037 — *chp* 0045 — *pp* 0046 — 0047 — *hp* 0048 — *pp* 0049 — *Ps-pp* 0050 — 0051 — 0052 — *chp* 0053

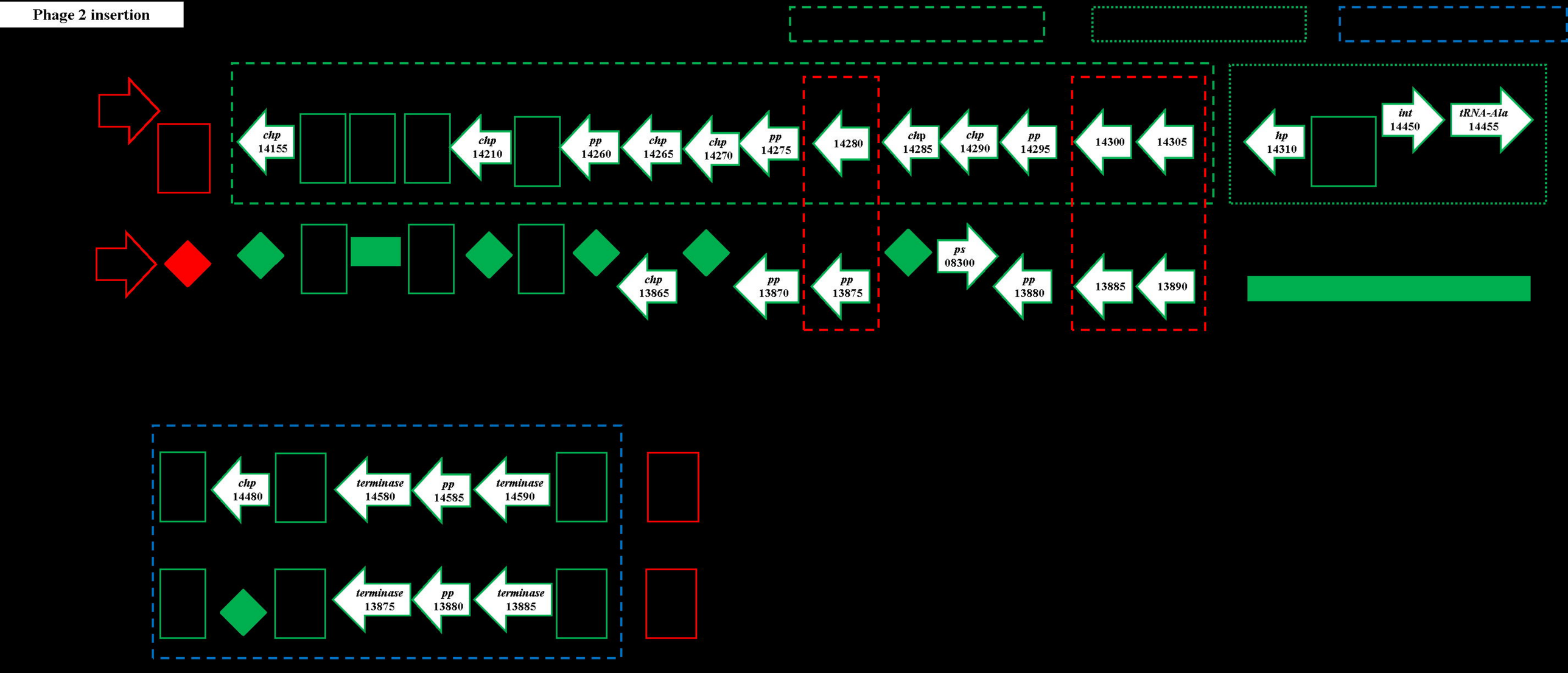
tr 14330 — *sp* 14335 — *tr* 14340 — *sgp* 14345 — *chp* 14365 — *chp* 14370 — *chp* 14375 — *pp* 14380 — *int* 14450 — *tRNA-A1a* 14455 — *hp* 14460 — 14465

tr 0054 — *hp* 0055 — *sp* 0056 — 0057-59 — *tr* 0060 — *hp* 0071 — *chp* 0072 — *chp* 0073-74 — *chp* 0075 — *helicase* 0076 — *chp* 0077 — *pp* 0078 — *int* 0093 — *tRNA-A1a* 0094 — *int* 0095 — 0099 — 0021

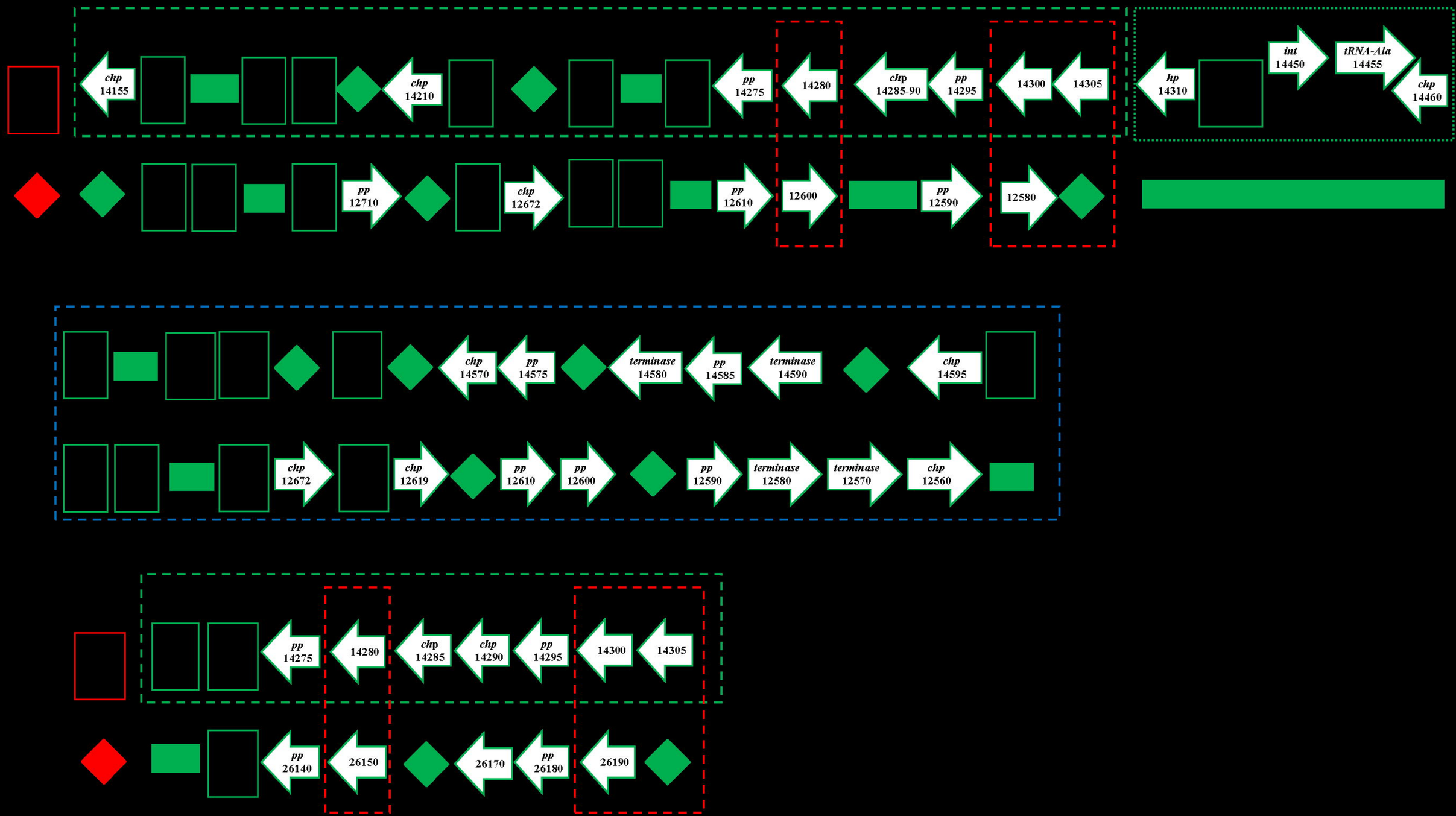
pp 14535 — *pp* 14575 — *terminase* 14580 — *pp* 14585 — 14590

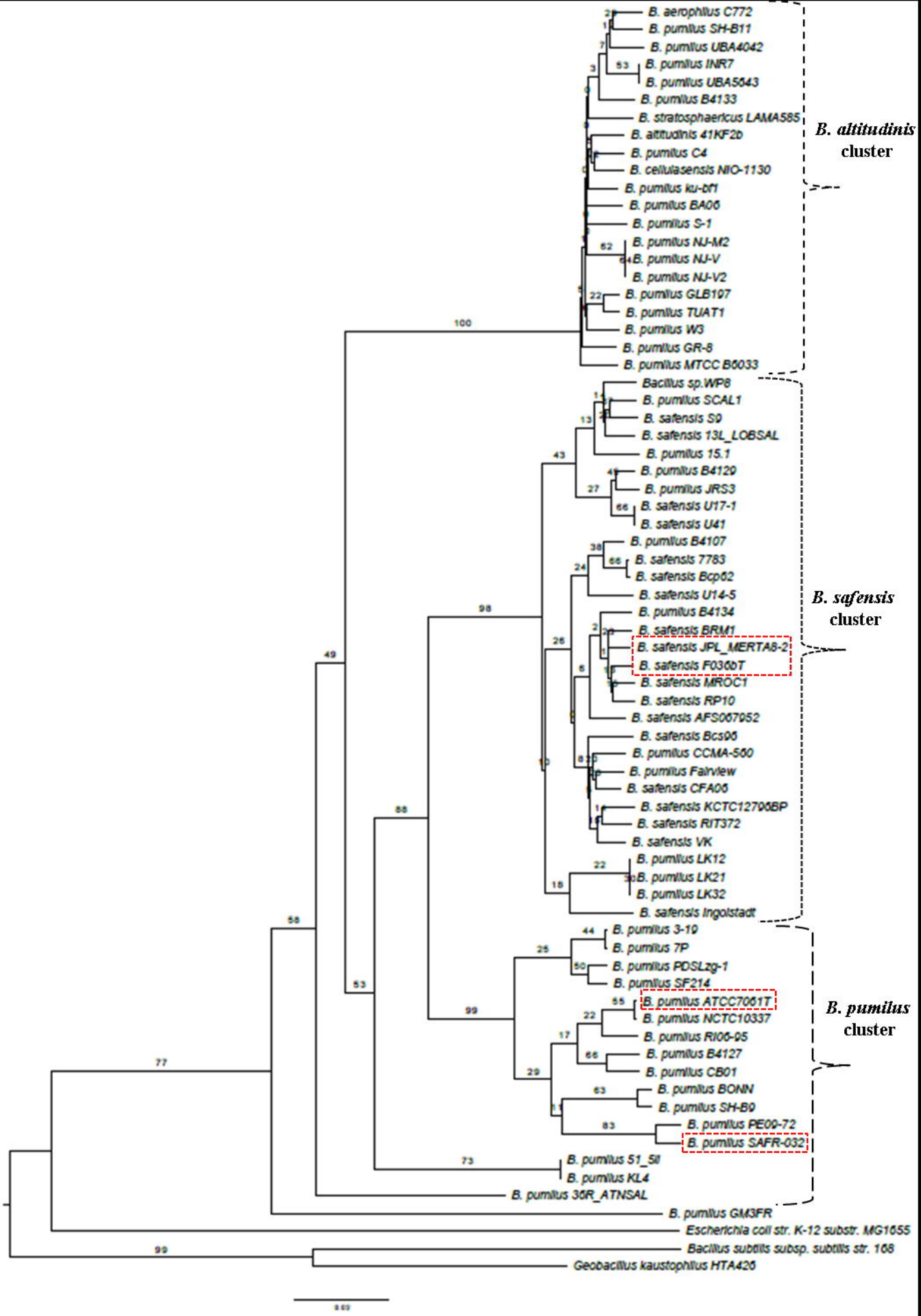
chp 0024 — *ps* 0035 — *pp* 0036 — *pp* 0037 — *chp* 0045 — *pp* 0046 — *terminase* 0047 — 0051 — 0052


Phage 2 insertion
tr
14105
chp
14155
chp
14180
chp
14185
chp
14270
pp
14275
14280
hp
14285
hp
14290
pp
14295
14300
14305
hp
14310
int
14450
tRNA-Ala
14455
chp
14570
pp
14460
14455
pp
14450
14445
tr
14605
chp
14480


Phage 2 insertion
chp 14155
chp 14210
pp 14260
chp 14265
chp 14270
pp 14275
14280
chp 14285
chp 14290
pp 14295
14300
14305
hp 14310
int 14450
tRNA-Ala 14455
chp 13865
pp 13870
pp 13875
ps 08300
pp 13880
13885
13890
chp 14480
terminase 14580
pp 14585
terminase 14590
terminase 13875
pp 13880
terminase 13885

Phage 2 insertion




B. aerophilus C772
B. pumilus SH-B11
B. pumilus UBA4042
B. pumilus INR7
B. pumilus UBA5043
B. pumilus B4133
B. stratosphaericus LAMA585
B. altitudinis 41KF2b
B. pumilus C4
B. cellulasensis NIO-1130
B. pumilus ku-bf1
B. pumilus BA06
B. pumilus S-1
B. pumilus NJ-M2
B. pumilus NJ-V
B. pumilus NJ-V2
B. pumilus GLB197
B. pumilus TUAT1
B. pumilus W3
B. pumilus GR-8
B. pumilus MTCC B6033
B. altitudinis cluster
Bacillus sp.WP8
B. pumilus SCAL1
B. safensis S9
B. safensis 13L_LOBSAL
B. pumilus 15.1
B. pumilus B4129
B. pumilus JRS3
B. safensis U17-1
B. safensis U41
B. pumilus B4107
B. safensis 7783
B. safensis Bcp62
B. safensis U14-5
B. pumilus B4134
B. safensis BRM1
B. safensis JPL_MERTA8-2
B. safensis F0306T
B. safensis MROC1
B. safensis RP10
B. safensis AFS067952
B. safensis Bcs96
B. pumilus CCMA-560
B. pumilus Fairview
B. safensis CFA06
B. safensis KCTC12796BP
B. safensis RIT372
B. safensis VK
B. pumilus LK12
B. pumilus LK21
B. pumilus LK32
B. safensis Ingolstadt
B. safensis cluster
B. pumilus 3-19
B. pumilus 7P
B. pumilus PDSLzg-1
B. pumilus SF214
B. pumilus ATCC7061T
B. pumilus NCTC10337
B. pumilus RI06-05
B. pumilus B4127
B. pumilus CB01
B. pumilus BONN
B. pumilus SH-B9
B. pumilus PE00-72
B. pumilus SAFR-032
B. pumilus 51_SII
B. pumilus KL4
B. pumilus 36R_ATNSAL
B. pumilus GM3FR
B. pumilus cluster
Escherichia coli str. K-12 substr. MG1655
Bacillus subtilis subsp. subtilis str. 168
Geobacillus kaustophilus HTA426