bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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)
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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Background

Bacillus strains producing highly resistant spores have been isolated from cleanrooms and space craft assembly facilities. Organisms that can survive such conditions merit planetary protection concern and if that resistance can be transferred to other organisms, a health concern too. To further efforts to understand these resistances, the complete genome of *Bacillus safensis* strain FO-36b, which produces spore resistant to peroxide and radiation was determined. The genome was compared to the complete genome of *B. pumilus S*AFR-032, as well as draft genomes of *B. safensis* JPL-MERTA-8-2 and the type strain *B. pumilus* ATCC7061^{T}. In addition, comparisons were made to 61 draft genomes that have been mostly identified as strains of *B. pumilus* or *B. safensis*.

Results

The FO-36b gene order is essentially the same as that in SAFR-032 and other *B. pumilus* strains [1]. The annotated genome has 3850 open reading frames and 40 noncoding RNAs and riboswitches. Of these, 307 are not shared by SAFR-032, and 65 are also not shared by either MERTA or ATCC7061^{T}. The FO-36b genome was found to have ten unique reading frames and two phage-like regions, which have homology with the *Bacillus* bacteriophage SPP1 (NC_004166) and *Brevibacillus* phage Jimmer1 (NC_029104). Differing remnants of the Jimmer1 phage are found in essentially all *safensis/pumilus* strains. Seven unique genes are part of these phage elements. Comparison of *gyrA* sequences from FO-36b, SAFR-032, ATCC7061^{T}, and 61 other draft genomes separate the various strains into three distinct clusters. Two of these are subgroups of *B. pumilus* while the other houses all the *B. safensis* strains.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conclusions

It is not immediately obvious that the presence or absence of any specific gene or combination of genes is responsible for the variations in resistance seen. It is quite possible that distinctions in gene regulation can change the level of expression of key proteins thereby changing the organism's resistance properties without gain or loss of a particular gene. What is clear is that phage elements contribute significantly to genome variability. The larger comparison of multiple strains indicates that many strains named as *B. pumilus* actually belong to the *B. safensis* group.

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 [2-4] is often characterized by their unusual resistances to different physical and chemical factors [2, 5-8]. Consistently stringent cleanroom protocols under planetary protection guidelines over several decades [2, 9-13], have created a special habitat for multi-resistant bacteria, many of which have been isolated and identified [14-20]. The potential of many of these isolates to possibly survive interplanetary transfer [3, 21-25] 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 [2-4, 26-28]

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Two of the most studied organisms in the specialized econiches of spacecraft assembly facilities and the ISS are *B. safensis* FO-36b^T [29] (referred to as FO-36b henceforth) and *B. pumilus* SAFR-032 [17] (referred to as SAFR-032). These organisms are representative strains of the endospore producing *Bacillus* sp.[14, 17, 30-34]. Both strains produce spores that exhibit unusual levels of resistance to peroxide and UV radiation [25, 30, 35] that far exceed that of the dosimetric *B. subtilis* type strain (*B. subtilis subsp. subtilis str.* 168, referred to as BSU) [36]. A third strain, *B. safensis* MERTA-8-2 (referred to as MERTA), was initially isolated from the Mars Odyssey Spacecraft and associated facilities at Jet Propulson 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 [37]. 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 [38].

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 [39, 40] (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 [40]. Since this previous study was completed, minor corrections to the SAFR-032 gene order were made and the annotation was updated [1]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 2% agarose gel, and those with the length from 300 bp to 350 bp were isolated from the gel as described [42]. The dsDNA fragments were converted to a shotgun DNA library using 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 [Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) available at https://github.com/najoshi/sickle and Trimmomatic 0.32 [43] 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 Abyss 1.5.2 *de novo* assembler [44] 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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 [45]. 3850 ORFs and 40 non-coding RNAs and riboswitches were predicted and the results were deposited in Genbank under accession number CP010405.

Genomes used in comparisons

The recently updated complete sequence of the SAFR-032 genome was obtained from NCBI (CP000813.4). The draft genomes of ATCC7061^{T} (Refseq accession no: NZ_ABRX00000000.1), consisting of 16 contigs and MERTA consisting of 14 contigs (Refseq accession no: GCF_000972825.1) were obtained from the public databases of the National Center for Biotechnology Information (NCBI). Several additional *B. safensis* and *B. pumilus* draft genomes from various sources have also been deposited in the NCBI database in recent years. However, these genomes get excluded when performing a global Genbank Blast (NT) analysis. To avoid this potential problem, these additional draft genomes were separately retrieved from the Genbank repository (*B. pumilus* genomes, https://www.ncbi.nlm.nih.gov/genome/genomes/440; *B. safensis* genomes,

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

https://www.ncbi.nlm.nih.gov/genome/genomes/13476) and locally integrated into the Genbank NT database. The resulting local database allowed inclusion of these genomes in subsequent Blast (NT) studies.

BLAST studies

Individual gene and protein sequences from the FO-36b genome, were blasted against each other as well as against the genomes of SAFR-032, MERTA and ATCC7061 using the standalone version of NCBI's BLAST program [46]. The comprehensive search included blastN and blastX for the nucleotide sequences and blastP for the protein sequences. Additionally, global blast was 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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The online tool PHAST [47, 48] 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.

Genome-genome distance studies and phylogenetic tree comparison

The genomes of *B. pumilus* and *B. safensis* were 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 [49, 50]**.** Many of the genomes are in draft format, and therefore, do not have complete and reliable ribosomal RNA sequences. Under such circumstances, other housekeeping genes, such as the DNA gyrase genes (*gyrA* and *gyrB*) have often been used as alternate targets for phylogenetic studies, given that both are reported to be sufficiently conserved over a long period of evolution [29, 51-55]. Therefore, we used the sequence of the DNA gyrase gene (*gyrA*). The *gyrA* sequences were bioinformatically isolated from the *pumilus/safensis* genomes and aligned using Bioedit ((http://www.mbio.ncsu.edu/BioEdit/bioedit.html) using ClustalW and MEGA [56, 57] with MUSCLE. The analysis involved 72 nucleotide sequences of *gyrA*, that included the *gyrA* homologs from the 65 *B. pumilus* and *B. safensis* genomes (including the FO-36b, MERTA, SAFR-032 and ATCC7061 genomes), and *gyrA* homologs of the representative strains from the *B. altitudinis* complex, *viz*., *B. aerophilus* C772, *B. altitudinis* 41KF2b, *B. cellulasensis* NIO-1130(T), *B. stratosphericus* LAMA 585. The *gyrA* homologs of *Geobacillus kaustophilus*, and *B.subtilis* served as outliers in the *Firmicutes* group and that of gram-negative *E. coli* MG1655, as a non-*Firmicutes* outlier. Maximum Likelihood, Neighbor-Joining and Minimum Evolution trees were built using MEGA. Codon positions included were 1st+2nd+3rd+Noncoding. All

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

positions containing gaps and missing data were eliminated. There were a total of 2436 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [58]. The Mauve alignment [59] program was used to align the previous draft FO-36b sequence (GCA_000691165.1 / ASJD00000000) with our current updated sequence (CP010405).

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 1: Table S1). 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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

four genes with fewer than 5 homologs found in other *pumilus*/*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 2: Table S2).

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 *safensis/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 *Safensis/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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 [39]. 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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Additional file 3: Table S3). 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.

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 4: Figure S1). Given this and the large number of additional draft genomes it was concluded that a re-analysis would be appropriate. Genome-genome distance analysis (Additional file 5:Figure S4) and "*gyrA*" tree analysis of the 61 available strains were used to examine relationships among the strains. The results using the maximum likelihood method are shown in Figure 8. Alternative tree constructions are provided as Additional file 5: Figure S2 and Additional file 6: Figure S3. The 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. This 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 recently reported *B. pumilus* strains. The second major cluster consists primarily of *B. safensis*

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

isolates but does include several likely misnamed *B. pumilus* strains too. This latter cluster includes both the FO36b and the MERTA8-2 strains.

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 seen may occur due to regulatory differences resulting in key genes associated with resistance being expressed at different levels [60-62]. 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 [63] by contributing to the genetic variability between closely related bacterial strains[64], which are often implicated in the phenotypical differences such as bacterial pathogenesis [64-67]. Bacteriophage-mediated horizontal gene transfer enhances bacterial adaptive responses to environmental changes such as the rapid spread of antibiotic resistance [68]. Furthermore, phages mediate inversions, deletions and chromosomal rearrangements which help shunt genes that could directly impact the phenotype between related strains [64] or between phylogenetically distant strains via horizontal gene transfer (HGT)[69]. All of these evolutionary events have implications for selection and fitness.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The first phage insertion in FO-36b is homologous to the *Bacillus* bacteriophage SPP1. The SPP1 is a 44-kb virulent *Bacillus subtilis* phage, well-known for its ability to mediate generalized transduction, a widespread mechanism for the transfer of any gene from one bacterium to another [70]. The second insertion is homologous to *Brevibacillus* phage Jimmer 1, which is one of several myoviruses that specifically target *Paenibacillus larvae*, a Firmicute bacterium, as a host [71].

The *B. safensis* strain lacks the ICEBs1-like element that was previously found in SAFR-032 and as an incomplete analog in ATCC7061 [40]. As reported earlier [40], the ICEBs1-like element does harbor some SAFR-032 unique genes and thus, their presence was suggested as being possibly responsible for the resistance properties of SAFR-032. The absence of the ICEBs1-like element in the FO-36b genome suggests that this may not be the case. FO-36b has an established phenotype showing spore resistance to peroxide exceeding that of the other JPL-CRF isolates [14]. SAFR-032 spores have been demonstrated to show resistance to UV radiation exceeding that of the other JPL-CRF isolates [17]. Given that both FO-36b and SAFR-032 harbor genes unique to each of them, on their respective phage elements (the two insertion elements in the case of FO-36b that are reported here and the ICEBs1-like element in the case of SAFR-032), a role of these unique genes in their respective unique spore phenotypes cannot be entirely ruled out.

Furthermore, more than one-half of *in silico* predicted phage gene products are hypothetical without any assigned functions [72-76]. Comparative genomic approaches use closely related

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phages from different host organisms and exploit the modular organization of phage genomes [77]. However, these methods are not adequate to address the hypothetical protein coding ORFs that are unique to phage insertions found in a given microbial strain displaying unique phenotypes as in the case of FO-36b and SAFR-032.

Hypothetical phage proteins are considered potential candidates for bacterial detection and antimicrobial target selection. In recent times, efforts towards discovering phage-based antimicrobials has led to the experimental characterization of specific phage proteins [78]. The identification of hypothetical ORFs unique to FO-36b and SAFR-032 phage insertion elements mark them out as potential biomarker candidates for the identification/detection of such strains.

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

Non-phage associated genes

Genes shared by the three resistant spore producing strains but not the non-resistant ATCC7061 strain are candidates for association with thee resistance properties. Of the 65 ORFs we had

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reported earlier to be uniquely shared by SAFR-032 and FO-36b [39], 59 are shared by the MERTA strain (Additional file 4: Table 3). When the analysis is extended to all 61 genomes it was found that in each case at least one additional organism had a homolog to the candidate gene. For example, one of these ORFs (FO-36b locus tag RS87_09285), is found to be shared by *B. safensis* MROC1 (isolated from the feces of *Gallus gallus*) and *B. safensis* RP10 (isolated from soils contaminated with heavy metals in Chile). Most of the strains containing these genes are isolates from environments that have some extreme stress component. However, it is not known if the stress component would include resistance to radiation or peroxide. Based on their names alone, some of these strains, such as *B. altitudinis*, and *B. stratosphericus* may be of special interest for further comparison and investigation of their spore resistance properties.

Highly unique open reading frames

The nine FO36b ORFs (hypothetical proteins) that were found to be absent from all the *safensis/pumilus* (and the *Bacillus sp*. WP8) genomes available in the NCBI database (Table 2A) may be envisioned as possibly contributing to the FO-36b spore resistance. Four of these highly unique ORFs are found on phage elements (one ORF, RS87_03140 on the *Bacillus* bacteriophage SPP1 insertion and three ORFs, viz., RS87_14155, RS87_14285, and RS87_14310 on the *Brevibacillus* phage Jimmer 1 insertion). This is similar to the situation with the ICEBs-1 like element in SAFR-032 that harbors unique SAFR-032 ORFs [40]. Four other ORFs had fewer than 5 homologs found in other *B. pumilus*/*B. safensis* genomes. Two of these four ORFs, are also found on the phage elements and hence could be random remnants of lateral transfer.

Phylogenetic analysis

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The results of the current study using genome-genome distance analysis and "*gyrA*" tree analysis are consistent with, but more detailed than the earlier study [38]. In agreement with the earlier studies the safiensis/pumilus strains form a coherent cluster with three large subclusters (Figure 8). One of these includes the FO-36b, and MERTA strains as well as all other B. safensis strains. *B. pumilus* strains in this grouping are likely misnamed. The FO36b and MERTA strains are both in the same subcluster of the larger *B. pumilus* group.

Conclusions

A recent report [79] 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. Although candidates continue to exist, no specific gene has been identified as likely being responsible for the resistances exhibited by these organisms. The differences in resistance properties can easily be attributed to changes in expression level but of what gene or genes? With a larger phylogenetic tree now available, it should be possible to select a representative subset of strains for further resistance studies as well as sequencing.

List of abbreviations used

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

SAFR-032 – *B. pumilus* SAFR-032.

ATCC7061 – *B. pumilus* ATCC7061^{T}.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

Declarations

1. Ethics approval and consent to participate

Not Applicable

2. Consent for publication

Not applicable

3. Availability of data and material

The datasets used and analyzed within the current study are available from the NCBI Website as referenced in the paper. The sequence of the B. safensis FO-36b strain is being deposited with the NCBI/Genbank with the accession number CP010405. Until the deposit It is 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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. 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

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. Stepanov VG, Tirumalai MR, Montazari S, Checinska A, Venkateswaran K, Fox GE: **Bacillus pumilus SAFR-032-032 Genome Revisited: Sequence Update and Re-Annotation**. *PLoS One* 2016, **11**(6):e0157331.
2. 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.
3. 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.
4. 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.
5. Rummel JD, Race MS, Horneck G: **Ethical considerations for planetary protection in space exploration: a workshop**. *Astrobiology* 2012, **12**(11):1017-1023.
6. Debus A: **The European standard on planetary protection requirements**. *Research in microbiology* 2006, **157**(1):13-18.
7. 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.
8. 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.
9. 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.
10. 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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11. DeVincenzi DL, Stabekis PD: **Revised planetary protection policy for solar system exploration**. *Adv Space Res* 1984, **4**(12):291-295.
12. 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.
13. Race MS: **Anticipating the reaction: public concern about sample return missions**. *Planet Rep* 1994, **14**(4):20-22.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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**

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Space for Planetary Protection Purposes-Experiment PROTECT of the EXPOSE-E Mission. *Astrobiology* 2012, **12**(5):445-456.

25. 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.
26. 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.
27. 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.
28. Moissl-Eichinger C, Cockell C, Rettberg P: **Venturing into new realms? Microorganisms in space**. *FEMS microbiology reviews* 2016, **40**(5):722-737.
29. 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.
30. 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.
31. 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.
32. La Duc MT, Kern R, Venkateswaran K: **Microbial monitoring of spacecraft and associated environments**. *Microb Ecol* 2004, **47**(2):150-158.
33. Opfell JB, Bandaruk W: **Microbial contaminants in the interiors of spacecraft components**. *Life Sci Space Res* 1966, **4**:133-165.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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**

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for the unusual resistances exhibited by Bacillus pumilus SAFR-032 spores. *PLoS One* 2013, **8**(6):e66012.

40. 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.
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. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. 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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

56. 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.
57. 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.
58. 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.
59. Darling AC, Mau B, Blattner FR, Perna NT: **Mauve: multiple alignment of conserved genomic sequence with rearrangements**. *Genome Res* 2004, **14**(7):1394-1403.
60. 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.
61. Nagy GN, Leveles I, Vertessy BG: **Preventive DNA repair by sanitizing the cellular (deoxy)nucleoside triphosphate pool**. *Febs J* 2014, **281**(18):4207-4223.
62. 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.
63. Salmond GP, Fineran PC: **A century of the phage: past, present and future**. *Nature reviews Microbiology* 2015, **13**(12):777-786.
64. 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.
65. 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.
66. 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.
67. 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.
68. Brown-Jaque M, Calero-Caceres W, Muniesa M: **Transfer of antibiotic-resistance genes via phage-related mobile elements**. *Plasmid* 2015, **79**:1-7.
69. Zinder ND, Lederberg J: **Genetic exchange in Salmonella**. *Journal of bacteriology* 1952, **64**(5):679-699.
70. Yasbin RE, Young FE: **Transduction in Bacillus-Subtilis by Bacteriophage Spp1**. *J Virol* 1974, **14**(6):1343-1348.
71. Merrill BD, Grose JH, Breakwell DP, Burnett SH: **Characterization of Paenibacillus larvae bacteriophages and their genomic relationships to firmicute bacteriophages**. *BMC Genomics* 2014, **15**.
72. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W: **Bacteriophage T4 genome**. *Microbiol Mol Biol R* 2003, **67**(1):86-+.
73. 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**.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

74. 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**.
75. 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.
76. 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).
77. Hendrix RW, Hatfull GF, Smith MCM: **Bacteriophages with tails: chasing their origins and evolution**. *Research in Microbiology* 2003, **154**(4):253-257.
78. 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.
79. 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.
80. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Molecular Biology and Evolution* 1987, **4**(4):406-425.
81. 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.
82. 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.
83. Rzhetsky A, Nei M: **A Simple Method for Estimating and Testing Minimum-Evolution Trees**. *Molecular Biology and Evolution* 1992, **9**(5):945-967.
84. Kumar MNaS: **Molecular Evolution and Phylogenetics**: Oxford University Press, New York.; 2000.

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" =

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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*

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. Molecular Phylogenetic analysis by the Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [82]. The tree with the highest log likelihood (-18689.4597) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value.

Tables and captions

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Additional file 2: Table S2. *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.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additional file 3: Table S3. *B. safensis* F0-36b genes reported earlier as shared by *B. pumilus* SAFR-032 and not found in the *B. pumilus* ATCC7061^{T} strain [39], compared with the *B. safensis* JPL-MERTA-8-2 strain, and the other *pumilus/safensis* genomes.

Additional file 4: 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 [59].

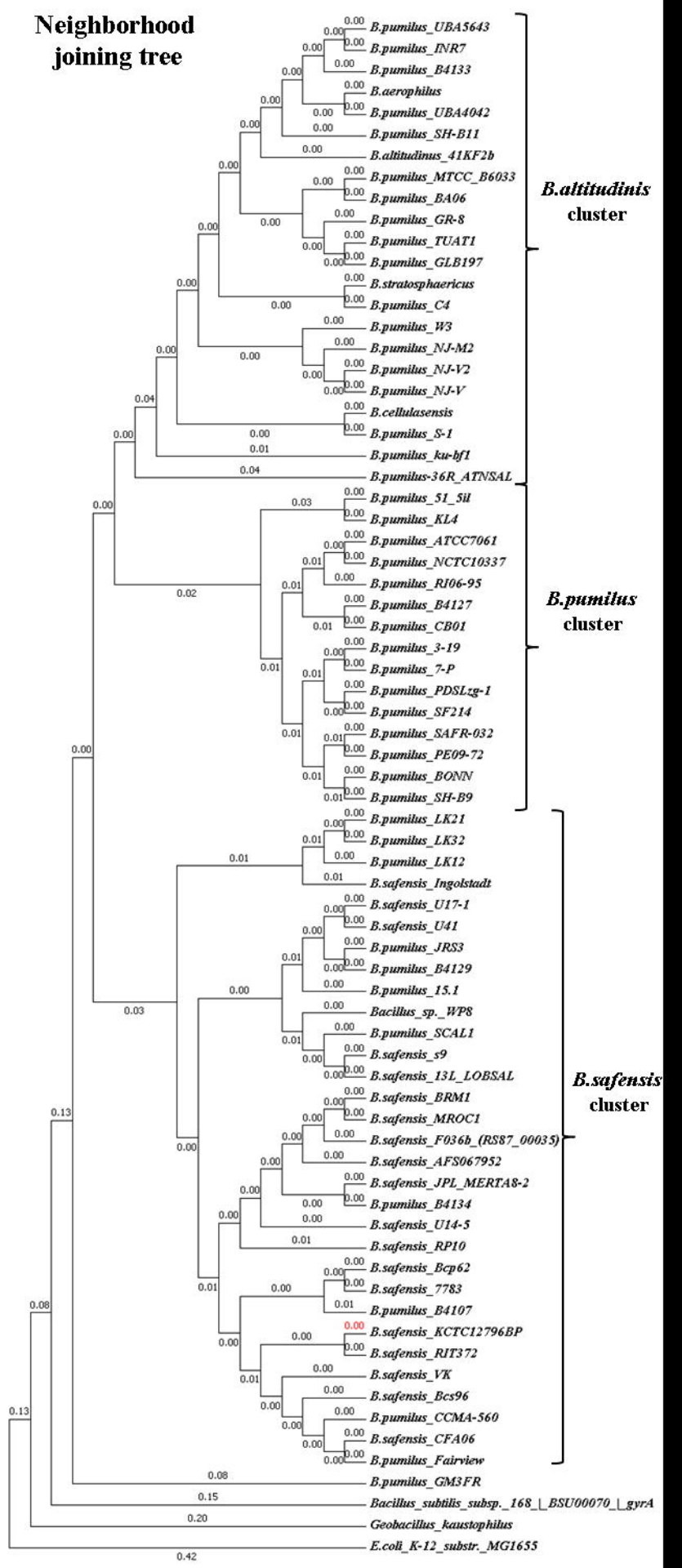
Additional file 5: Table S4. In silico DNA-DNA hybridization (DDH) values showing Genome-genome distance [49] relationship values for the genomes of various *B. pumilus, B. safensis, B. altitudinis* strains. The genomes of *Geobacillus kaustophilus, and B. subtilis subsp. subtilis str.* 168 serving as outliers in the *Firmicutes* group and that of gram-negative *E.coli* MG1655, as a non-*Firmicutes* outlier.

Additional file 6: Figure S2. Molecular Phylogenetic analysis by the Neighbor-Joining method. The evolutionary history was inferred using the Neighbor-Joining method [80]. The optimal tree with the sum of branch length = 1.63399956 is shown. (next to the branches). The evolutionary distances were computed using the Maximum Composite Likelihood method [81] and are in the units of the number of base substitutions per site.

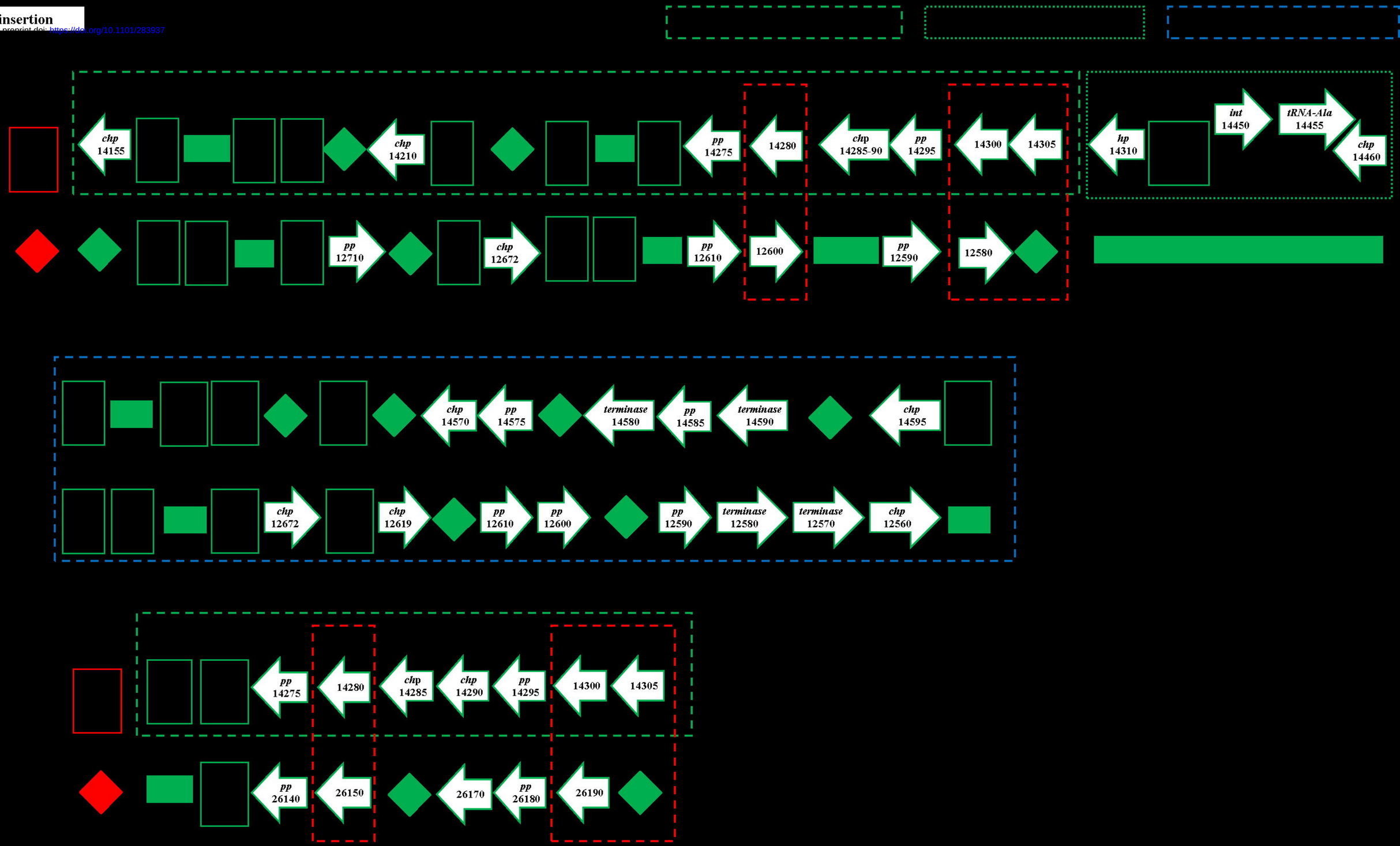
Additional file 7: Figure S3. Phylogenetic analysis using the Minimum Evolution method.

bioRxiv preprint doi: https://doi.org/10.1101/283937; this version posted March 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

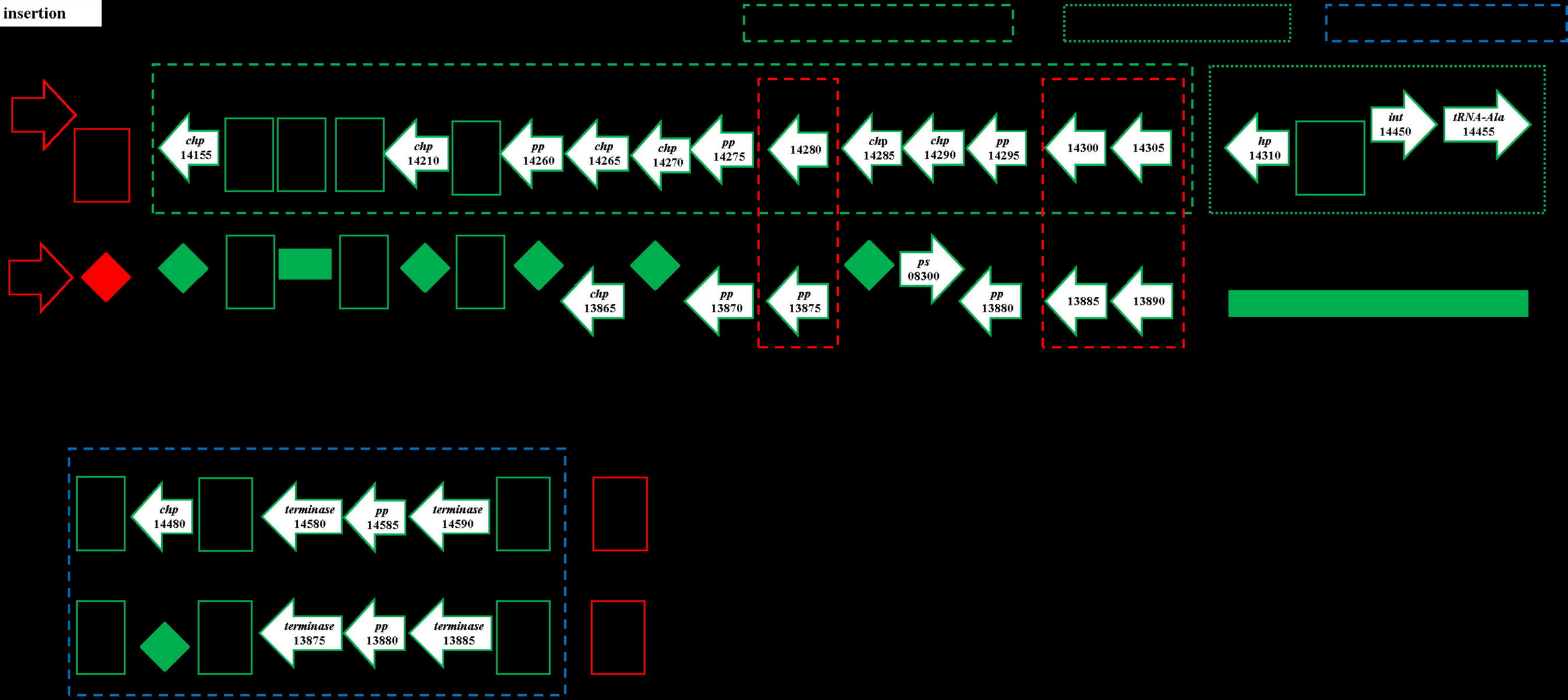
The evolutionary history was inferred using the Minimum Evolution method [83]. The optimal tree with the sum of branch length = 1.63399956 is shown. (next to the branches). The evolutionary distances were computed using the Maximum Composite Likelihood method [81] and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm [84] at a search level of 1. The Neighbor-joining algorithm [80] was used to generate the initial tree.

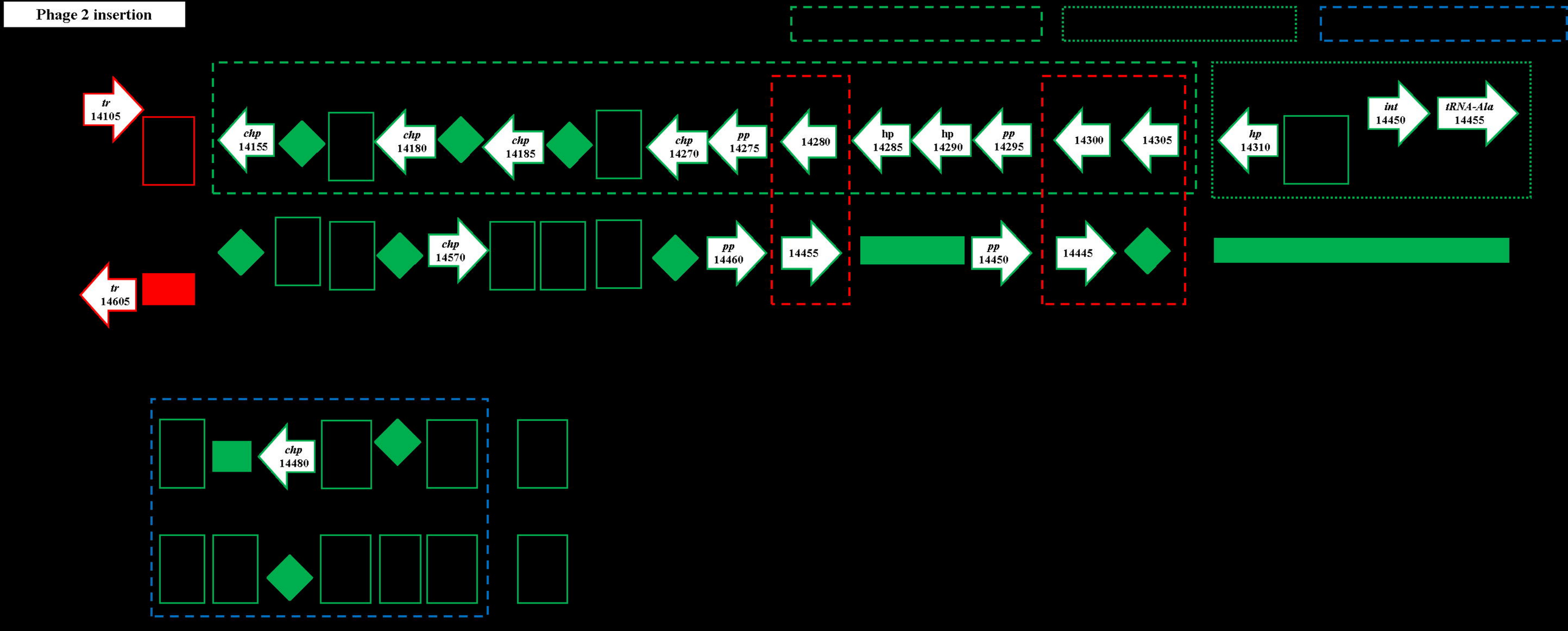

Neighborhood joining tree
B.pumilus_UBA5643
B.pumilus_INR7
B.pumilus_B4133
B.aerophilus
B.pumilus_UBA4042
B.pumilus_SH-B11
B.altitudinus_41KF2b
B.pumilus_MTCC_B6033
B.pumilus_BA06
B.pumilus_GR-8
B.pumilus_TUAT1
B.pumilus_GLB197
B.stratosphaericus
B.pumilus_C4
B.pumilus_W3
B.pumilus_NJ-M2
B.pumilus_NJ-V2
B.pumilus_NJ-V
B.cellulasensis
B.pumilus_S-1
B.pumilus_ku-bf1
B.pumilus-36R_ATNSAL
B.altitudinis cluster
B.pumilus_51_5il
B.pumilus_KL4
B.pumilus_ATCC7061
B.pumilus_NCTC10337
B.pumilus_RI06-95
B.pumilus_B4127
B.pumilus_CB01
B.pumilus_3-19
B.pumilus_7-P
B.pumilus_PDSLzg-1
B.pumilus_SF214
B.pumilus_SAFR-032
B.pumilus_PE09-72
B.pumilus_BONN
B.pumilus_SH-B9
B.pumilus cluster
B.pumilus_LK21
B.pumilus_LK32
B.pumilus_LK12
B.safensis_Ingolstadt
B.safensis_U17-1
B.safensis_U41
B.pumilus_JRS3
B.pumilus_B4129
B.pumilus_15.1
Bacillus_sp._WP8
B.pumilus_SCAL1
B.safensis_s9
B.safensis_13L_LOBSAL
B.safensis_BRM1
B.safensis_MROC1
B.safensis_F036b_(RS87_00035)
B.safensis_AFS067952
B.safensis_JPL_MERTA8-2
B.pumilus_B4134
B.safensis_U14-5
B.safensis_RP10
B.safensis_Bcp62
B.safensis_7783
B.pumilus_B4107
B.safensis_KCTC12796BP
B.safensis_RIT372
B.safensis_VK
B.safensis_Bcs96
B.pumilus_CCMA-560
B.safensis_CFA06
B.pumilus_Fairview
B.safensis cluster
B.pumilus_GM3FR
Bacillus_subtilis_subsp._168_|_BSU00070_|_gyrA
Geobacillus_kaustophilus
E.coli_K-12_substr._MG1655

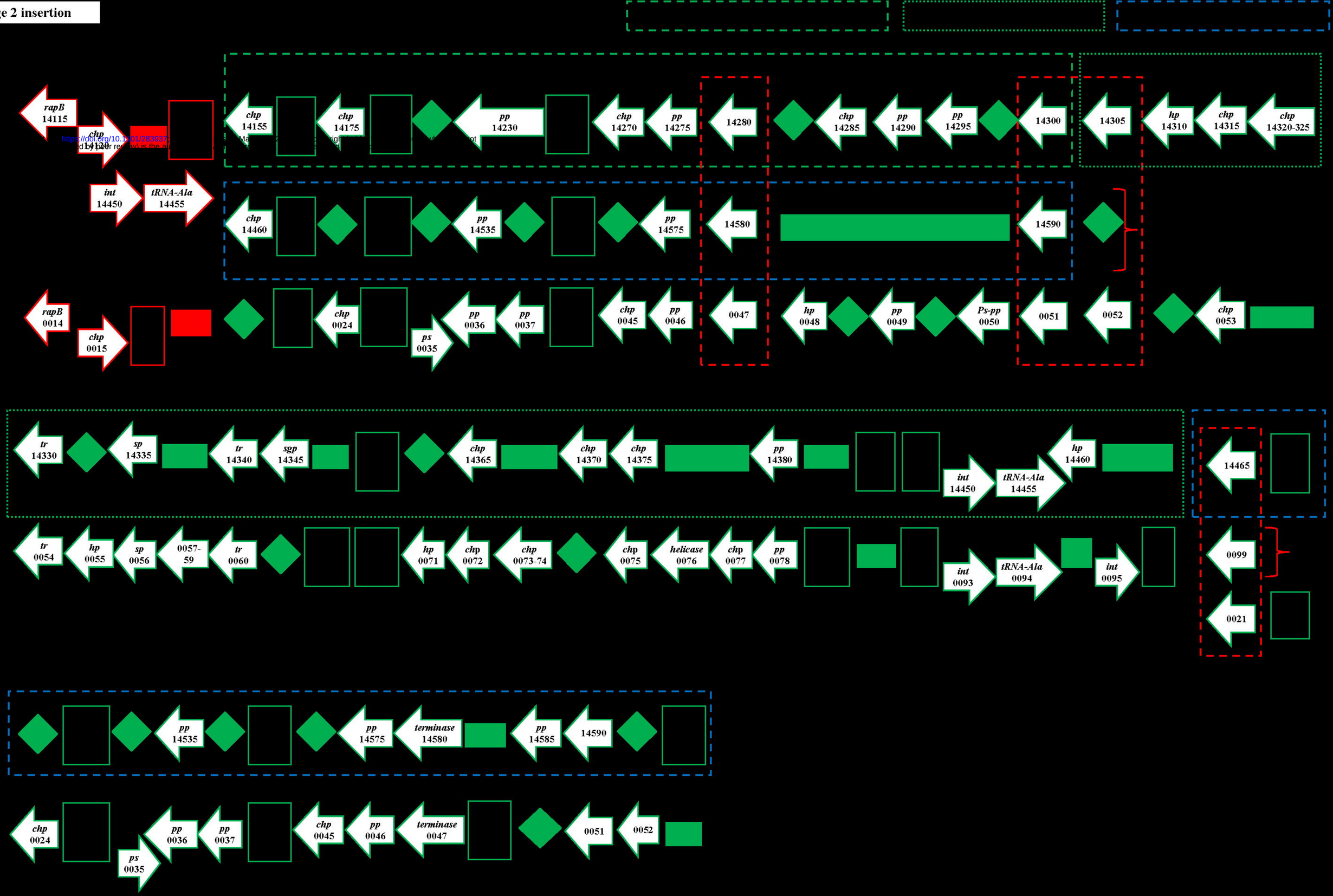
BioRxiv preprint doi: https://doi.org/10.1101/283937

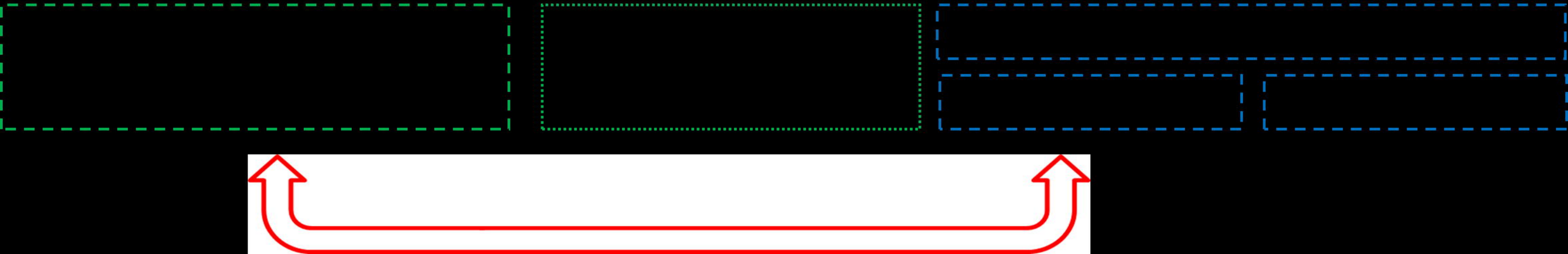


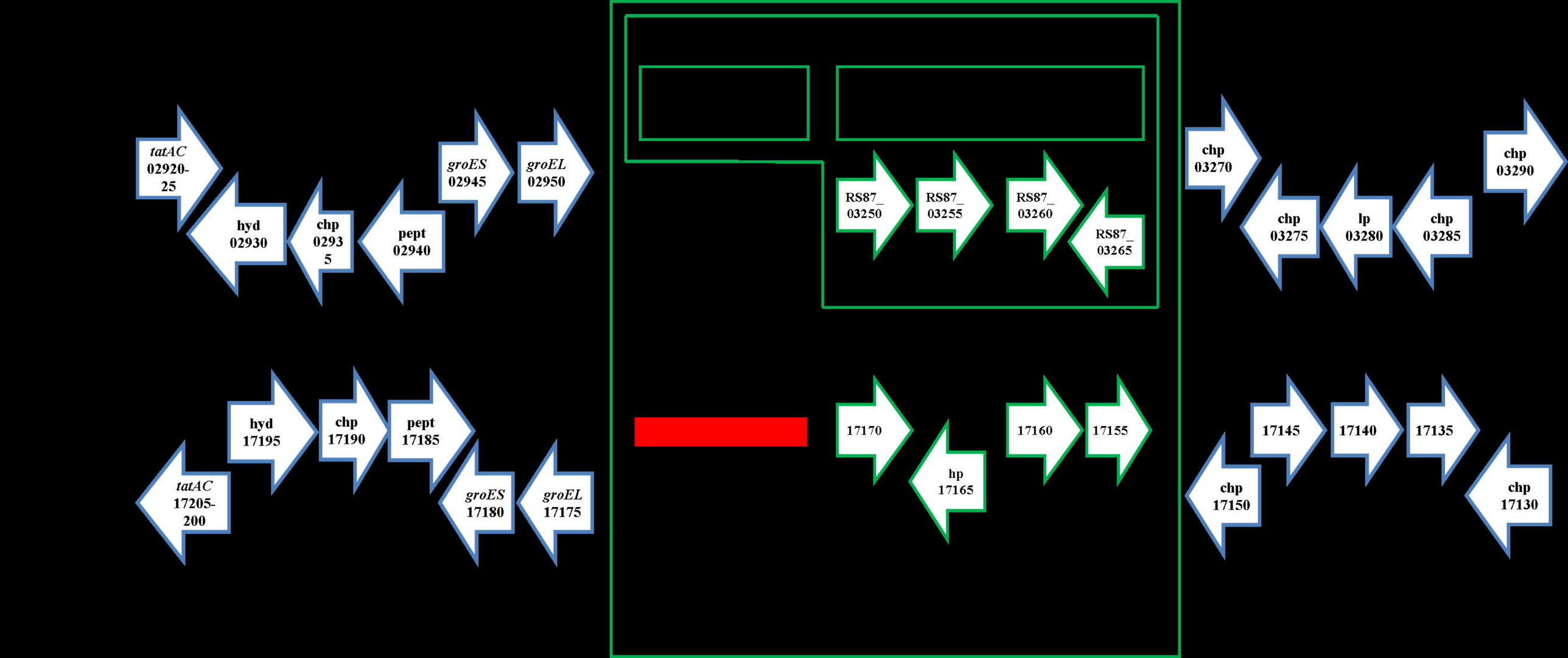
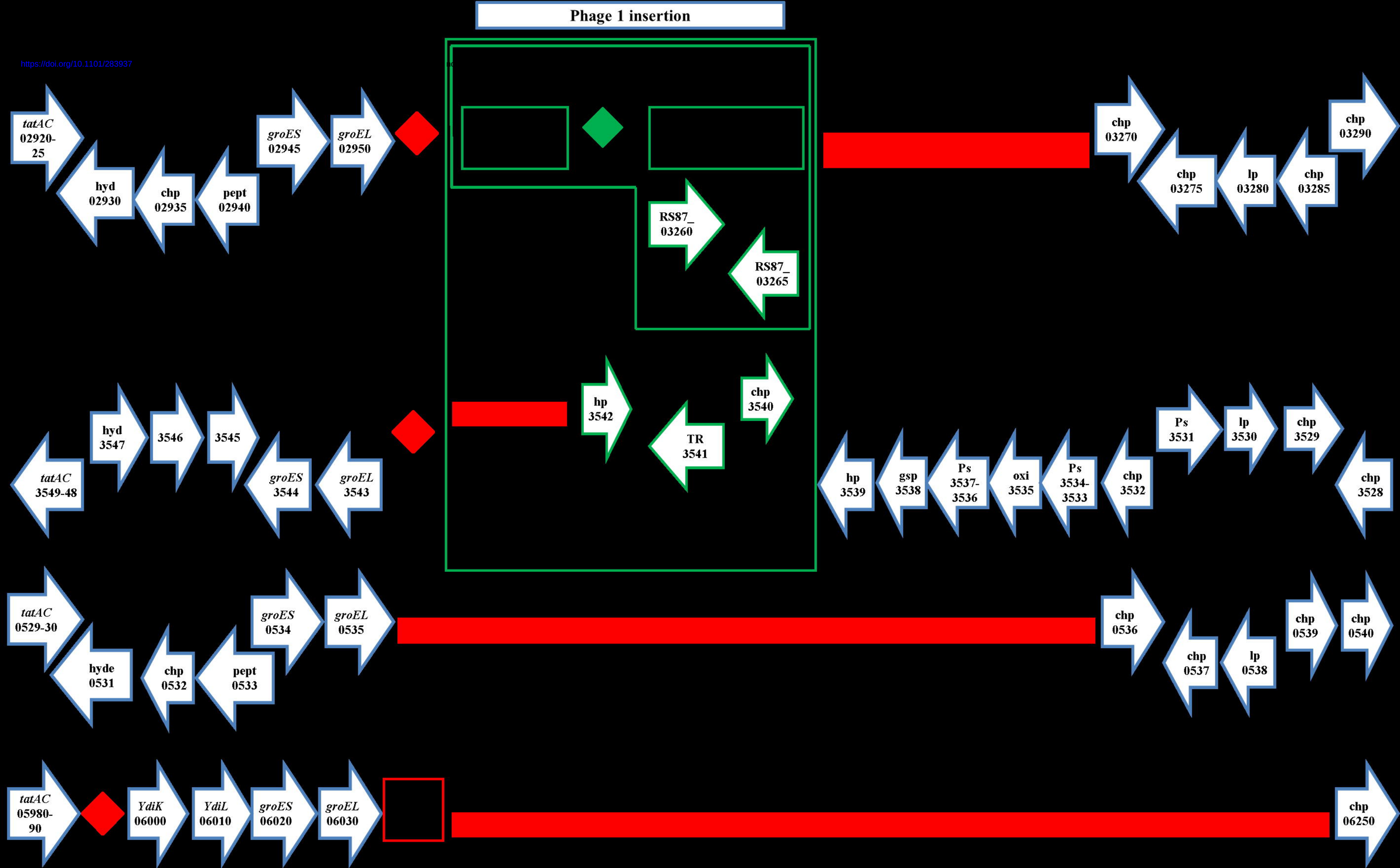
Phage 2 insertion




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
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-Ala 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-Ala 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




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
tatAC 17205-200
hyd 17195
chp 17190
pept 17185
groES 17180
groEL 17175
17170
hp 17165
17160
17155
chp 17150
17145
17140
17135
chp 17130

https://doi.org/10.1101/283937