

Bacillus pumilus SAFR-032: A model for planetary protection

George E. Fox

ABSTRACT—In order to prevent forward contamination of Mars, it is necessary to minimize the bioburden on spacecraft that will contact the Martian surface. *Bacillus pumilus* strains whose spores are unusually resistant to ultraviolet light and other means of sterilization have been consistently found in the clean rooms used in spacecraft assembly as well as on spacecraft. In order to better understand the origin of these organisms and to develop effective means to eliminate them, the complete genome of *Bacillus pumilus* SAFR-032 and *B. safiensis* F036b were determined in collaboration with the Baylor College of Medicine Human Genome Sequencing Center (HGSC). These two genomes are being compared to each other and to the closely related and previously determined genome of *B. licheniformis* and *B. subtilis*. In particular, genes involved in DNA repair, response to oxidative stress, and sporulation are being intercompared to obtain a list of genes likely associated with dramatic changes in resistance.

METHODOLOGY

The complete genome of *B. pumilus* SAFR-032 was determined in collaboration with the Human Genome Sequencing Center at Baylor College of Medicine (HGSC). It was compared in detail with its closest sequenced relatives, *B. licheniformis* and *B. subtilis*.¹ In addition, a set of 454 sequence data was obtained for *B. safiensis* F036b. Possible open reading frames were initially predicted by Glimmer² and GeneMarkS³ and then automatically examined for similarity to known COGS, protein domains, and other annotated genes (i.e., pre-run BLAST searches). The annotators looked at all the data and in some cases literature references before assigning a name and, if possible, a likely function for the open reading frame.⁴ The annotation team was made up of a variety of individuals from the HGSC group and UH. Each gene was annotated separately by two annotators and the resulting annotations compared automatically. Differences were subsequently resolved by conference between the annotators with input from the group as a whole when needed. Using the Projector 2 software, *B. safiensis* contigs were mapped on the *B. pumilus* genome. Thus, it was possible to compare gene lists and, to a lesser extent, gene order in these three organisms.

RESULTS

The comparison between *B. pumilus* SAFR-032/*B. licheniformis* has been described in detail with respect to genes likely to be involved in DNA repair and regulation of the oxidative stress response¹ and will be summarized here. The primary observation was that there are 13 genes uniquely missing from



George E. Fox

the SAFR-032 genome relating to the DNA repair and regulation of the oxidative stress response, any of which might have been replaced with a component leading to elevated resistance. The comparison also revealed the importance of looking at sequence variability. Changes in function typically correlate with rapid acceptance of sequence changes. Most genes in a three-way comparison between *B. subtilis*, *B. licheniformis*, and *B. pumilus* have similarities in the 70-90 percent range. However, in a few cases, the *B. pumilus* gene has considerably more variation. An example is the catalase gene, KatX2, whose sequence similarity falls to 48 percent for comparisons between SAFR-032 and the other two genomes

but is above 70 percent in the *B. licheniformis*/*B. subtilis* comparison. Another example is the splAB operon. The translated SplB sequence is highly conserved in *B. pumilus*, but SplA has substantial sequence diversity when compared to the presumed *B. subtilis* and *B. licheniformis* homologs, indicating possible differences in SP lyase genetic regulation.

With regard to sporulation, 46 proteins are known to be components of the spore coat in *B. subtilis*. These proteins play a prominent role in the resistance of the spores to chemical and physical agents. Detailed comparisons of these proteins are ongoing and will be covered in a second paper now in preparation. Multiple interesting results have been obtained. For example, the operon cot JA(BC) is missing completely from both *B. pumilus* SAFR-032 and *B. safiensis* F036b, while it is found in *B. subtilis* and *B. licheniformis*. Is it present or absent in the reduced-resistance, *B. pumilus*-type strain (ATCC 7061)? With the *B. safiensis* F036b contigs mapped to the SAFR-032 genome, searches for each SAFR-032 gene in F036b were conducted. Most genes, in fact, had homologs that

were positioned correctly in the contig map. A total of 301 SAFR-032 genes were found to be missing in F036b. Of these, 115 are genes that have been previously assigned function. The remaining 186 include 20 ORFs that have homologs only in other *Bacillus* species and 87 ORFs that had homologs in various bacterial genomes in addition to the *Bacillus* species. The remaining 87 ORFs are hypothetical proteins found exclusively in SAFR-032. The missing genes of known function that might relate to the resistance differences seen between the two strains include genes coding for a small acid soluble protein, two cell division-related proteins, and a sigma factor. Efforts to identify genes that are unique to F036b are being conducted by searching the unmapped contigs for novel genes. Most of these are very small and contain obvious assembly errors or encode rRNA fragments.

REFERENCES

- Gioia, J., Yerrapragada, S., Qin, X., Jiang, H., Igboeli, O.C., Muzny, D., Dugan-Rocha, S., Ding, Y., Hawes, A., Liu, W., Perez, L., Kovar, C., Dinh, H., Lee, S., Nazarath, L., Blyth, P., Holder, M., Buhay, C., Tirumalai, T., Liu, Y., Dasgupta, I., Bokhetache, L., Fujita, M., Karouia, F., Moorthy, P.E., Siefert, Jo., Uzman, A., Buzumbo, P., Verma, A., Zwiya, H., McWilliams, B.D., Olowu, A., Clinkenbeard, K.D., Newcombe, D., Golebiewski, L., Petrosino, J.F., Highlander, S.K., Nicholson, W.L., Fox, G.E., Venkateswaran, K., and Weinstock, G.M. Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus*. *PLoS One* **2**, e928 (2007).
- Delcher, A. L., Harmon, D., Kasif, S., White, O., Salzberg, S.L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636-4641 (1999).
- Besemer, J., Lomsadze, A., and Borodovsky, M. GeneMark a self-training method for regulatory regions. *Nucleic Acids Res.* **29**, 2607-2618 (2001).
- McLeod, M.P., Karpathy, S.E., Gioia, J., Qin, X., Highlander, S.K., Fox, G.E., McNeil, T.Z., Jiang, H., Muzny, D., Jacob, L.S., Hawes, A.C., Sodergren, E., Gill, R., Hume, J., Morgan, M., Hong, C., Yu, X., Walker, D.H., and Weinstock, G.M. The complete genome of *Rickettsia typhi* and comparison with *R. prowazekii* and *R. Conorii*. *J. Bacteriol.* **186**, 5842-5855 (2004).

PUBLICATIONS

- Gioia, J., Yerrapragada, S., Qin, X., Jiang, H., Igboeli, O.C., Muzny, D., Dugan-Rocha, S., Ding, Y., Hawes, A., Liu, W., Perez, L., Kovar, C., Dinh, H., Lee, S., Nazarath, L., Blyth, P., Holder, M., Buhay, C., Tirumalai, T., Liu, Y., Dasgupta, I., Bokhetache, L., Fujita, M., Karouia, F., Moorthy, P.E., Siefert, Jo., Uzman, A., Buzumbo, P., Verma, A., Zwiya, H., McWilliams, B.D., Olowu, A., Clinkenbeard, K.D., Newcombe, D., Golebiewski, L., Petrosino, J.F., Highlander, S.K., Nicholson, W.L., Fox, G.E., Venkateswaran, K., and Weinstock, G.M. Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus*. *PLoS One* **2**, e928 (2007).
- Highlander, S.K., Hulten, K.G., Qin, X., Jiang, H., Yerrapragada, Y., Mason, Jr., E.O., Shang, Y., Williams, T., Fortunov, R.M., Liu, Y., Igboeli, O., Petrosino, J., Tirumalai, M., Uzman, A., Fox G., Cardenas, A.M., Muzny, D., Hemphill, L., Ding, Y., Dugan, S., Blyth, P.R., Buhay, C.J., Dinh, H.H., Hawes, A. C. F., Holder, M. K., Kovar, C.L., Lee, S.L., Liu, W., Nazareth, L.V., Wang, Q., Zhou, J., Kaplan, S.L., and Weinstock, G.M. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC-Microbiology* **7**, 99 (2007).
- Karpathy, S.E., Qin, X., Gioia, J., Jiang, H., Liu, Y., Petrosino, J.F., Yerrapragada, Y., Fox, G.E., Haake, S.K., Weinstock, G.M., and Highlander, S.K. Genome sequence of *Fusobacterium nucleatum* subspecies polymorphum—a genetically tractable fusobacterium. *PLoS ONE* **2**, e659 (2007).
- Viswanath, L., Lu, Y., and Fox, G.E. Genome display tool: Visualizing possible correlations in complex data sets. *Source Code Biol. Med.* **2**, 1 (2007).