

Student Name: Keith Arora-Williams

Mentor Name: James Nick Benardini

Document: Final Technical Paper

Date: September 9, 2012

Title

Genotypic & Phenotypic Diversity of Microbial Isolates from the Mars Exploration Rovers

Abstract

Mars-bound rovers such as the Mars Exploration Rover (MER) endure strict planetary protection implementation campaigns to assess bioburden. The objective of this study is to identify cultivable microorganisms isolated by the NASA Standard Assay from spacecraft during pre-launch and evaluate their potential to survive conditions on the Martian surface. Of approximately 350 isolates collected from the MER spacecraft archive, 171 microorganisms were reconstituted for characterization via 16S rRNA fingerprinting. Alignment of 16S sequences revealed high levels of sequence similarity to spore-forming species, overwhelmingly of the genera *Bacillus* (73.7%) and *Paenibacillus* (14.0%). Samples underwent phenotype characterization employing multiple carbon sources and ion concentrations in an automated microarray format using the Omnilog system. Working and stock cultures were prepared to address the immediate needs for day-to-day culture utilization and long-term preservation, respectively. Results from this study produced details about the microbes that contaminate surfaces of spacecraft, as well as a preliminary evaluation of a rapid biochemical ID method that also provides a phenotypic assessment of contaminants. The overall outcome of this study will benefit emerging cleaning and sterilization technologies for preventing forward contamination that could negatively impact future life detection or sample return missions.

Introduction

In 1992, the Committee on Space Research (COSPAR) recommended that member nations publish information concerning the estimated biological burden of a spacecraft at launch.¹ Due to limitations of cost and material complexity, it is not practical to sterilize every component of a spacecraft. Current standards limit the biological burden by imposing a threshold levels of raw spores per unit of area.² This format for quantification is a product of the NASA standard spore assay.³ The standard assay protocol is used to recover spores or heat-shock resistant microorganisms from spacecraft surfaces and associated materials, such as tools, paint, lubricant, surfaces, hardware, etc. The standard assay is a cultivation-dependent technique for estimating the amount of aerobic, mesophilic, and heterotrophic spore-forming organisms on a sample. While providing a suitable measure of cleanliness, further characterization of isolated microbes is required to identify each by species and survival capacity. This additional analysis is needed in order to minimize the possibility of forward contamination i.e. extraterrestrial exposure to terrestrial organisms.. Forward contamination

poses a significant threat to interpretation of results of *in situ* experiments designed to search for evidence of extant or fossil Martian microorganisms.²

Archiving of microbial cultures from the NASA Standard Assay is a focus of planetary protection efforts at the Jet Propulsion Laboratory (JPL). Currently, the archive facility at JPL contains over 4500 isolates collected during assembly, testing, and launch operations of pre-flight from Mars-based spacecraft, such as the MER rovers. With the exception of isolates from the Viking and Mars Science Laboratory (MSL), recovered organisms remain frozen in the microbial archive for future study. By analyzing the ecology and evolution of microbes living on spacecraft, we can assess their potential for survival in a myriad of extreme environments.

Five years after planetary protection policies were updated by COSPAR, the Task Group on Issues in Sample Return sought further improvement. They stated that although, “no known terrestrial organisms [can] grow on the Martian surface... the task group strongly recommends that efforts be made to explore current analytical methods for use in bioburden assessment and inventory procedures before spacecraft assembly and launch.”³ Specifically, polymerase chain reaction (PCR) based identification techniques were recommended by the group for assessment and inventory purposes. In addition to employing molecular fingerprinting, this research endeavored to develop an additional method for those two purposes, the assessment and inventorying of the spacecraft-associated bioburden.

The current PCR-related inventory of microbial life consists of a network of databases containing DNA sequences of the 16S ribosomal subunit gene, which contains hypervariable regions that can provide species-specific signature sequences useful for bacterial identification. For example, an organism with a 16S sequence exhibiting <97% or <98.5% sequence homology to all known 16S sequences may be considered a new species or strain, respectively, pending further characterization.⁴ Although much progress has been made over the last two decades, these databases are still routinely expanded and reorganized to make room for new strains and species. As the sensitivity, variety, and resolution of biochemical assays and sensor technology improves, significant differences between members of a species or strain become apparent. In addition, strides in throughput via automation and multiplexing have produced alternative methods to DNA sequencing for identifying environmental samples. Ultimately, the taxonomic boundaries separating microbial life, especially regarding prokaryotes, are being redefined as a result of technologies that allow access to new types and greater volumes of data. One example of such a technology is the Omnilog Automated Microbial Identification System. We focused on developing this tool, procured in May 2012, for purposes of rapid assessment and inventory of spacecraft associated microbes.

Following a similar 16S sequencing pilot project conducted in 2010 on MSL isolates in collaboration with the University of Idaho, we aimed to achieve the following objectives for MER isolates

- i. Identify cultivable microbes collected from spacecraft surfaces
- ii. Develop working and replicate stocks of isolates for future use

- iii. Develop the Omnilog system, and associated database, as a rapid method for biochemical characterization of spacecraft-associated microbes

The purpose of the study is to assist those involved in planning future Mars missions, such a proposed Mars Sample Return Campaign, by allowing them to draw from a vast inventory of genetic and biochemical data of terrestrial organisms recovered by the NASA Standard Assays.

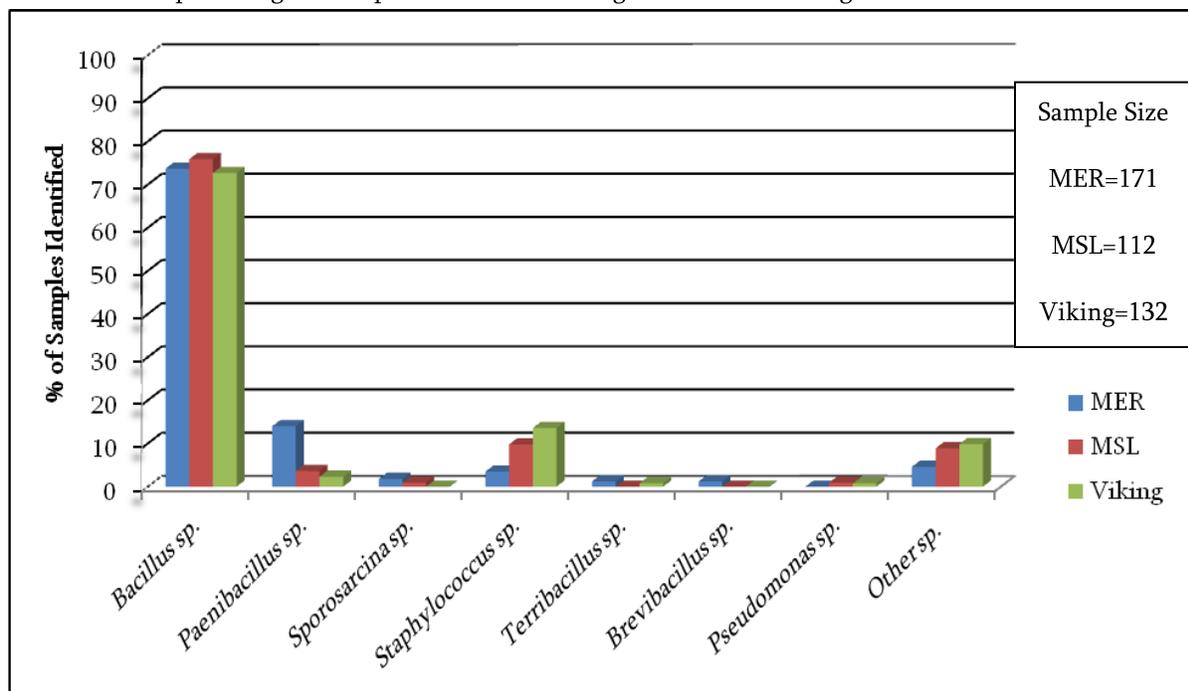
Based on the consistency of sequencing data between MSL and MER samples, we concluded that while the majority of microorganisms isolated from cleanroom conditions are known, a significant degree of diversity remains unfamiliar. We also determined that, at present, the Omnilog system shows promise as a method of identifying spacecraft associated microbes in a fraction of the time required for 16S sequencing. However, many more samples need to be tested before a supplementary database of spacecraft-associated microbes can be linked to the default database of profiled organisms.

Results

Figure 1: 16S rDNA BLAST Data. Of the 171 sequences obtained, those considered to be matched, a novel species, or a novel strain contain more than 98.5%, between 98.5 and 97%, or less than 97% sequence homology respectively. Samples classified as non-spore formers by sequence were positively matched.

Matched Sequences	<i>Nov. Sp.</i> Candidates	<i>Nov. Sp. Subsp.</i> Candidates	Non-spore-forming <i>Sp.</i>
148	5	18	8

Figure 2: The graph is a comparison of isolates from the MSL, MER, and Viking missions showing the percentage of samples distributed among the most common genera identified.

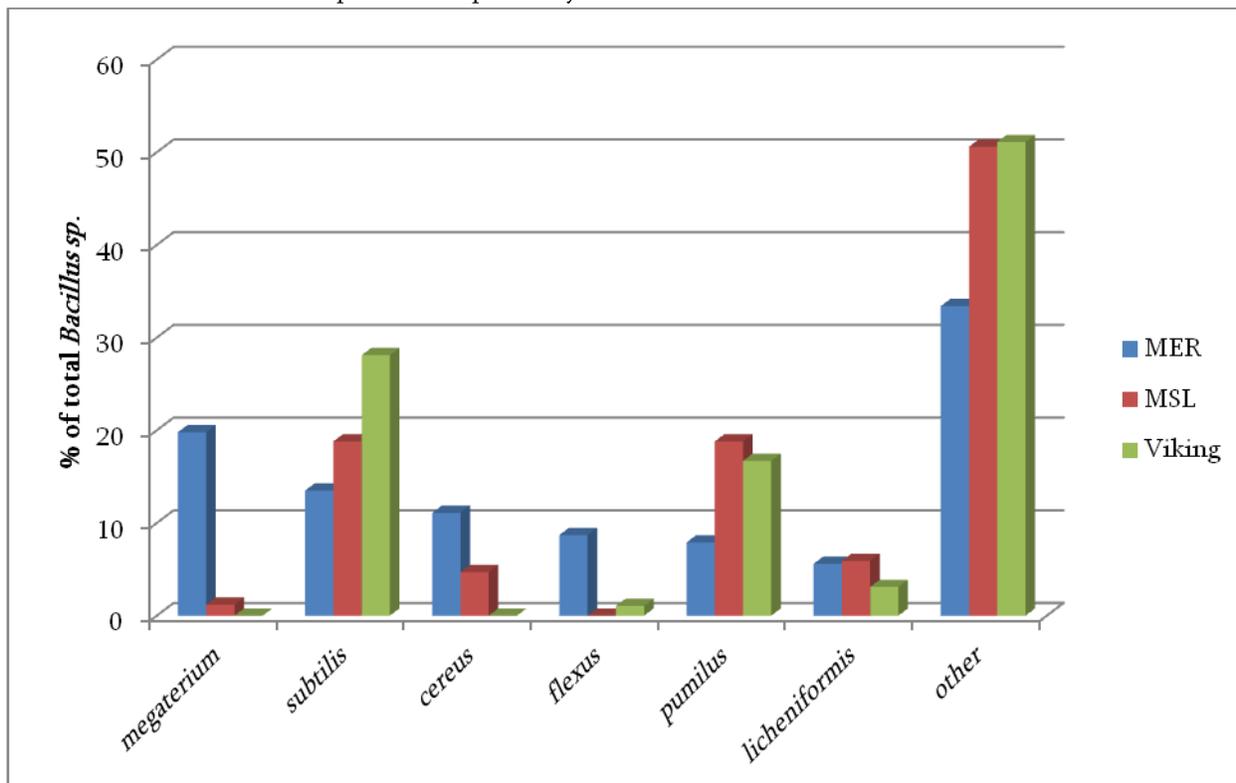


All of the 188 MER isolates provided for the study from the archive were well-isolated and most survived extended storage at -80°C. Only 5 of the submitted cryovials showed no growth after a minimum of 3 attempts at culturing on TSA plates (MER63, MER68, MER176, MER190, and

MER196). 9 samples failed to be amplified by PCR (MER96, MER121, MER131, MER14, MERTA31, MERTA44, MERTA45, and MERTA95). 3 samples that were successfully amplified during PCR failed during the sequencing reaction (MER14, MER60, and MER-TA58), resulting either in truncated transcripts of nonsensical sequence or unacceptably noisy chromatograms. The remaining 171 samples returned high quality sequences. The shortest of these sequences fell between 146 to 478 nucleotides short of the expected length of 1465 bp, all showed sequence homology above 98.5%.

Isolates were classified into the genera listed in Fig. 2. The samples included in the “Other” category of Fig. 2 were matched to members of the genera *Oceanobacillus*, *Micrococcus*, *Streptomyces*, *Hydrogenophaga*, *Thermoactinomyces*, and *Carnobacterium*. The non-spore formers identified include *H. intermedia*, *C. inhibens*, *S. epidermis*, *S. caprae*, and *S. lugdunensis*.⁵ All but *C. inhibens* are common human-based contaminants, the exception being associated with marine life.⁶ A notable difference between this and the MSL study is percentage of spore formers in the former (4.6%), as seen in Fig. 1, is lower than in the latter (18.6%). The overwhelming majority of isolates were of the genera *Bacillus*. The most common *Bacillus* sp. are listed in Fig. 3 and a partial phylogenetic tree showing selected species within other genera can be found in Appendix B.

Figure 3: Distribution of isolates from the MSL, MER, and Viking spacecraft within major *Bacillus* sp.: 21 other species were positively identified in the MER mission.



The 5 isolates with less than 97% sequence homology merit further investigation. The table in Fig. 4 highlights the quality of the sequence and the irregularity of the species matched.

Figure 4: Novel Species Candidates: The matched homologous regions are all within 100 bp of the expected sequence length (1465 bp), aside from MER-9, all species matched are unique among the samples considered.

Isolate Number	Closest Species Match	Bases Matched / Homologous Region (bp)	Identity	Gaps	Contig Length (bp)
MER 91	<i>Paenibacillus daejeonensis</i>	1380 / 1434	96.23%	3	1479
MER 9	<i>Sporosarcina luteola</i>	1437 / 1492	96.31%	17	1496
MER-TA 12	<i>Thermoactinomyces sanguinis</i>	1337 / 1387	96.40%	2	1498
MER 111	<i>Paenibacillus sabinae</i>	1366 / 1413	96.67%	5	1466
MER 57	<i>Cohnella phaseoli</i>	1403 / 1451	96.69%	1	1478

The Omnilog system was tested on 99 organisms, profiled in 14,880 reactions, monitored over a total of 18,088 plate images. 32 of the organisms profiled were classified by the system as either Low Similarity or No Match to members of the database. Due to the complexity of the system's internal algorithms responsible for data processing and species matching, the results from these experiments are still being analyzed. Preliminary observations of internally processed results showing positive/negative calls for a select group of wells are shown in Fig 5.

Figure 5: Selected Omnilog Plate Data: The percentages are drawn from 46 of the 99 organisms profiled. 12 of the 96 wells are noted. The leftmost column provides a description of the reagent indicated in the middle column. The rightmost column shows the percentage of organisms able to utilize a carbon source or metabolize in the presence of the reagent. This was determined by comparing the relative pixel value of a well to the negative control well. Rows 8 and 10 contain two reagents each that shared the same survival rate.

Gram Positive Antibiotic	Vancomycin	0%
Salinity Tolerance Assays	1% NaCl	63.2%
	5% NaCl	94.7%
	8% NaCl	86.8%
Facultative Anaerobic Organism Metabolite	1% Sodium Lactate	97.4%
Antibiotic / Antimicrobial Agents	Fusidic Acid	0%
	Lincomycin	0%
	Rifamycin SV or Minocycline	5.3%
Reactive Oxygen Species Production	Tetrazolium Blue	55.3%
Selective Media	Potassium Tellurite or Sodium Butyrate	100%

Discussion

Microbial diversity is constantly changing with environmental conditions. While a number of models have been proposed to account for this diversity, they share certain themes. An ecotype is defined as a population of prokaryotes that are genetically cohesive and ecologically distinct. In the event that a mutant or recombination event allows an adaptation to emerge in a species, an ecotype is

redefined and all clonal descendants outcompete nearly all other variants within the particular environmental niche. Ecotypes continue to be redefined as organisms pass, via airborne spore dispersal for example, into a new ecological niche. Horizontal gene transfer to adjacent organisms further separates this divergence in ecotypes.⁷ These distinctions are often unresolvable by 16S rRNA sequences. It is likely that species pairs with greater than 99% 16S sequence homology such as *Burkholderia pseudomallei* to *B. mallei*, as well as, *Bacillus anthracis* to *B. cereus* would have never been distinguished had the two former not developed into pathogens.⁴

A widely used method of broaching the problem of differentiating strains within a species, aside from biochemical phenotyping, is Multi-Locus Sequence Analysis. The method involves PCR amplification of >7 single-copy genes, usually responsible for housekeeping or unique to the particular taxon. However, this method takes double the time and money per sample. It also provides less information about the isolate.⁸

Another alternate method for microbial identification that is still being developed is Matrix-Assisted Laser Desorption/ Ionization Time-of-Flight Mass Spectrometry. MALDI-TOF is used primarily as a research tool for protein analysis; however it is currently being developed a faster, cheaper alternative to 16S sequencing that requires no culturing. The major disadvantages of this method are the cost of the instrument, the lack of a reference database to which samples can be compared, and the variability in results between different labs and makes of mass spectrometers. Despite these limitations, a recent publication on >600 clinical samples managed to identify >99% of samples and reserved lengthier, costlier biochemical assays for confirming the remaining low confidence IDs.⁹ As it has been shown that most prokaryotes in the biosphere are non-culturable, the employment of this method in the future is anticipated.¹⁴

At present, the method of metabolic profiling in automated microarrays is therefore an ideal partner to 16S sequence fingerprinting for generating a rapid, high-confidence inventory of spacecraft-associated microbes. Sequence homology gives a relative measure of the evolutionary distance from known organisms and Omnilog-based phenotyping adds sufficient qualification. Each assay within the Omnilog plate interrogates a separate network of genetic information. Some examples include genes encoding enzymes responsible for ion transport, catabolism, protein folding, as well as those coordinating their production and degradation. Genomes of *Bacillus* sp. have been collected from a wide range of environments including hydrothermal vents, tidal flats, soil, alkaline environments, shallow marine water, and a shallow water column from an oligotrophic environment. The ability to thrive in such varied conditions demonstrates a widely distributed degree of metabolic capabilities within the genus. It has been shown that less than one third of genes present in *B. subtilis* are conserved across the genus.¹⁰ It is therefore unlikely that all of these networks probed by the Omnilog system will be conserved across distinct strains.

Omnilog's data analysis software, Retrospect 2.0, allows for custom metabolic profiles to be created for organisms that show little, if any, similarity to those in the database. The process involves aggregating multiple sets of data obtained for the same organism at different starting cell densities. For this purpose, the system employs a spectrophotometric method of estimating inoculation density

that produces highly variable results. Further development of the ID system will require either finding an alternative method for obtaining initial inoculation density or determining the optimal number of profiles necessary for each unknown sample, such that a reliable aggregate custom ID can be created. Another feature of Retrospect software that requires development is the ability to cluster profiles into phenotypic cladograms. This is an extremely valuable feature that allows for the side-by-side visual comparison of isolates with high 16S homology, as well as for discerning common features of extremophile metabolism.

The comparison of *Bacillus* sp. between the Viking, MER, and MSL spacecraft show a consistent distribution consisting predominately *Bacillus* sp.. *B. megaterium*, *B. flexus*, and *B. cereus* appeared significantly more frequently in MER samples than in the future and past missions.¹¹ A potential explanation for this is that the more partial 16S sequences associated with strains of these species were accepted into the public databases in recent years.¹² However, this explanation fails to explain the more frequent appearance of *B. lentus* in the profile of Viking-associated microorganisms, one which relied solely on biochemical assay-based classification.¹³

Metabolic profiles generated for the subset of organisms showed a striking distinction from those microbes isolated during a study on spacecraft assembly facilities (SAF). Microbes isolated in our study were 100% gram positive, compared to approximately 50% of SAF isolates. In general, the degree of salinity tolerance at 8% NaCl was much higher than the 50% of SAF isolates, but an intentional selection bias prevents conclusions from being drawn.¹³

The high proportion of non-spore formers isolated from the SAF cleanrooms (75%), as compared to MSL (18.6%) and MER (4.6%) is indicative of improvements in aseptic technique during analysis or contamination control during spacecraft assembly. In summary, the findings of this work both confirm and qualify our prior understanding of the microbial population associated with spacecraft cleanrooms.

Methods

Isolate Cultures

Archived samples were thawed in ice at room temperature, briefly vortexed, and streaked on a Tryptic Soy Agar gel in petri dish using plastic disposable loops. Cultures were then moved from the biosafety cabinet and incubated at 33°C until single colonies with distinct morphologies were apparent.

Archive Expansion

Archived isolates were originally stored in 1 mL of a 30% glycerol solution stored at -80 degrees Celsius. After ensuring their purity, multiple colonies were harvested from TSA plates and placed into 5 new cryovials containing 1 mL of 30% glycerol for archiving and an additional working stock was stored on cryobeads (bioMérieux sa, Marcy l'Etoile, France).

DNA Extraction

After isolate purity was confirmed, a single colony from each sample was used to inoculate 15 mL of Tryptic Soy Broth in 50 mL Falcon Tubes. The tubes were shaken for at least 24 hours and allowed to settle for 10 minutes. 1 mL harvested from the bottom of each tube was pipetted into a Tissue LEV Total RNA Purification Kit cartridge for automated extraction using the Maxwell 16 MDx system (Promega, Madison, WI). The RNA recovery protocol was used, as it showed a higher yield from low biomass samples.

PCR & Sequencing

Extracted DNA was used as the template for PCR amplification of the 16S rDNA gene. 8F and 1512R primers were used along with GoTaq (Promega) and MolBio reagents. Gel electrophoresis was used to confirm the reaction. 1 µL of the each PCR product was illuminated on a 1% agarose gel containing SYBR Green dye (Life Technologies, Carlsbad, CA) using UV radiation. The remaining PCR product was purified for sequencing using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

Sanger sequencing was performed by Macrogen Inc. (Rockville, MD). The three sequencing primers used were 27F, 512F, and 1492R. The three reads were stitched together using Vector NTI (Life Technologies) and aligned to the NCBI database using an online batch BLAST tool (<http://greengene.uml.edu/>).

Phylogenetic Analysis

Type Strain sequences were selected based on the top hits from the NCBI database and aligned with sample sequences using MEGA 5 software (<http://www.megasoftware.net/>). The software was also used to calculate evolutionary distances and generate phylogenetic trees.

Biochemical Characterization

The Omnilog system (Biolog, Hayward, CA) is composed of an incubator that can hold 48, 96-well plates, an automated USB camera, and a computer containing software for monitoring the data as it is logged in real-time and analyzing the kinetic data. Each sample is plated across the 96 wells, each of which contains a different reagent (Appendix A) and a colorimetric dye. The organisms ability to metabolize in the presence of a given carbon source or selective medium is monitored by the camera as the dye darkens. This kinetic metabolic profile is then compared to a reference database of >2650 species. The pixel values are calculated by the software and normalized to the negative control well.

Acknowledgements

The described research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. The author would like to acknowledge members of the Biotechnology and Planetary Protection Group at JPL especially J. Nick Benardini, P. Vaishampayan, M. T. La Duc, and W. Schubert. Additional thanks to K. Buxbaum from the JPL Mars Program Office and to S. Smith from the University of Idaho for their support.

Funding for the research efforts was provided by the National Aeronautics and Space Administration through the Experimental Program to Stimulate Competitive Research Grant # NNX11AQ30A, as well as by the Amgen Scholar Program. Additional funding and research support was provided by the University of Idaho. All materials are copyright 2012 California Institute of Technology.

References

1. **National Research Council.** 1997. Mars Sample Return: Issues and Recommendations. 1:2-7. *National Academy Press*, Washington D.C.
2. **National Aeronautics and Space Administration.** 2011. Planetary Protection Provisions for Robotic Extraterrestrial Missions. NPR 8020.12D, April 2011. *NASA*, Washington, DC.
3. **National Aeronautics and Space Administration.** 2010. Handbook for the Microbial Examination of Space Hardware. HDBK 6022, *NASA*, Washington, DC.
4. **Gevers, D., Thompson, F., Swings, J., Vandamme, P., Peer, Y. d., Stackebrandt, E., et al.** 2005. Re-evaluating prokaryotic species. *Nature Reviews Microbiology*. 3(9), 733-739.
5. **Nordstrom, J.M., et al.** 2012. Comparison of bacteria on new, disposable, laundered, and unlaundered hospital scrubs. *Am J Infect Control*. 40(6):539-43.
6. **Jöborn, A., et al.** 1999. Carnobacterium inhibens sp. nov., isolated from the intestine of Atlantic salmon (*Salmo salar*). *Int J System Bacteriol*. 49:1891-1898.
7. **Koepfel, A., Nevo, E., Perry, E., Cohan, F., Ratcliff, R., Connor, N., et al.** 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7): 2504-2509.
8. **Maiden, M.C., et al.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 95(6): 3140-3145
9. **Cherkaoui, A., et al.** 2010. Comparison of Two Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Methods with Conventional Phenotypic Identification for Routine Identification of Bacteria to the Species Level. *J. Clin. Microbiol*. 48(4): 1169-1175
10. **Alcaraz, L. D., et al.** 2010. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics," *BMC Genomics*, 11(332):
11. **Venkateswaran, K., et al.** 2001. Molecular microbial diversity of a spacecraft assembly facility. *Syst Appl Microbiol*. 24(2): 311-20.
12. **Maughan, H., & Van der Auwera, G.** 2011. *Bacillus* taxonomy in the genomic era finds phenotypes to be essential though often misleading. *Infect Genet Evol*. 11(5):789-97.
13. **Puleo, J. R., et al.** 1977. Microbiological profiles of the Viking spacecraft. *Appl. Environ. Microbiol*. 33(2): 379-384.

14. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169 (1995).

Appendix A: Omnilog Reagents

The following table shows the arrangement of reagents on the Gen III Microplate used in the Omnilog system. The top row of cells correspond to wells A1 (left) to H1 (right). Orange cells indicate the control wells, white cells show carbon utilization tests, and grey cells show chemical resistance tests.

Negative Control	D-Raffinose	α -D-Glucose	D-Sorbitol	Gelatin	Pectin	p-Hydroxy-Phenylacetic Acid	Tween 40
Dextrin	α -D-Lactose	D-Mannose	D-Mannitol	Glycyl-L-Proline	D-Galacturonic Acid	Methyl Pyruvate	γ -Amino-Butyric Acid
D-Maltose	D-Melibiose	D-Fructose	D-Arabitol	L-Alanine	L-Galactonic Acid Lactone	D-Lactic Acid Methyl Ester	α -Hydroxy-Butyric Acid
D-Trehalose	β -Methyl-D-Glucoside	D-Galactose	myo-Inositol	L-Arginine	D-Gluconic Acid	L-Lactic Acid	β -Hydroxy-D, L-Butyric Acid
D-Cellobiose	D-Salicin	3-Methyl Glucose	Glycerol	L-Aspartic Acid	D-Glucuronic Acid	Citric Acid	α -Keto-Butyric Acid
Gentiobiose	N-Acetyl-D-Glucosamine	D-Fucose	D-Glucose- 6-PO ₄	L-Glutamic Acid	Glucuronamide	α -Keto-Glutaric Acid	Acetoacetic Acid
Sucrose	N-Acetyl- β -D-Mannosamine	L-Fucose	D-Fructose-6-PO ₄	L-Histidine	Mucic Acid	D-Malic Acid	Propionic Acid
D-Turanose	N-Acetyl-D-Galactosamine	L-Rhamnose	D-Aspartic Acid	L-Pyroglutamic Acid	Quinic Acid	L-Malic Acid	Acetic Acid
Stachyose	N-Acetyl Neuraminic Acid	Inosine	D-Serine	L-Serine	D-Saccharic Acid	Bromo-Succinic Acid	Formic Acid
Positive Control	1% NaCl	1% Sodium Lactate	Troleandomycin	Lincomycin	Vancomycin	Nalidixic Acid	Aztreonam
pH 6	4% NaCl	Fusidic Acid	Rifamycin SV	Guanidine HCl	Tetrazolium Violet	Lithium Chloride	Sodium Butyrate
pH 5	8% NaCl	D-Serine	Minocycline	Niaproof 4	Tetrazolium Blue	Potassium Tellurite	Sodium Bromate

Appendix B: Phylogeny of Select Species

The following figure is a cladogram of isolates and reference organisms based on 16S sequence. The isolates selected include those most divergent from *Bacilli* sp. and the scale is given in units of evolutionary distance.

