We are interested in your strain typing method and challenges!

Application of Whole Genome Sequences for Microbial Strain Typing

Sujan Timilsina, Ph.D. Ryan Cox, Ruvim Kolosey, Bindhu Verghese, Ph.D. Charles River Laboratories, Newark, DE

sujan.timilsina@crl.com
+1-302-602-7341

1 Abstract

Microbial identification and strain typing are critical in pharmaceutical microbiology for ensuring product safety, controlling contaminants, and meeting regulatory requirements. Traditional strain typing methods like pulsed-field gel electrophoresis and multi-locus sequence typing have their limitations in scalability and resolution. Increased use of next-generation sequencing (NGS) techniques for microbial identification and characterization, coupled with the availability of bacterial whole genomes, advanced sequence analysis approaches such as single nucleotide polymorphisms (SNPs), core and pan-genome analysis, is opening the door for more modern approaches to microbial strain typing. However, several of these approaches are species-specific and require optimization per species or project. In this study, we analyzed conserved genes from several bacterial species to analyze the possibility of using these genes as a set of universal genes for prokaryotic strain typing. We compared the strain typing results using these universal genes to results from species-specific housekeeping genes and with species with no established strain typing methods. Our data demonstrated that these conserved genes successfully distinguished bacterial strains when compared to the results from established species-specific strain typing schemes, providing a path forward in the development of a rapid, universal strain typing solution using bacterial genomes that is not limited to a single species.

2 Methods

For the study, we pulled whole genome sequences for representative species and their strains and confirmed with 16S sequences.

1. *rps* genes were extracted based on GenBank annotation of reference strains.
2. The ST genes from the type strains in Accugenix® and PubMLST databases were used as references for strain comparisons.
3. Sequences were concatenated and phylogenetically analysed.
4. More than 25 unique species were compared, and we are sharing results from the three species each representing different category:
   a. validated Accugenix® ST service available,
   b. publicly available ST schemes, and
   c. organisms phylogenetically analysed using whole genomes beyond ribosomal genes.

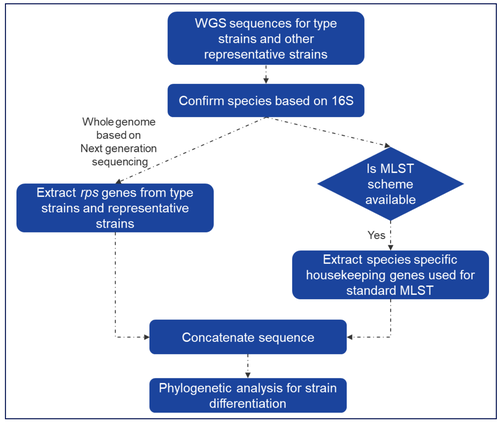

Figure 1. Schematic representation of strain typing comparison with *rps* genes extracted from whole genomes of bacteria with species specific MLST analysis. The results were compared based on the number of strain types and the resolution observed from each result.

3 Results

Each species used in this study encompassed the type strain along with additional representative strains. The concatenated sequence length for *rps* gene sequences for each species hovered around ~22,000 bp while the sequence length varied depending on the housekeeping MLST genes used for individual species.

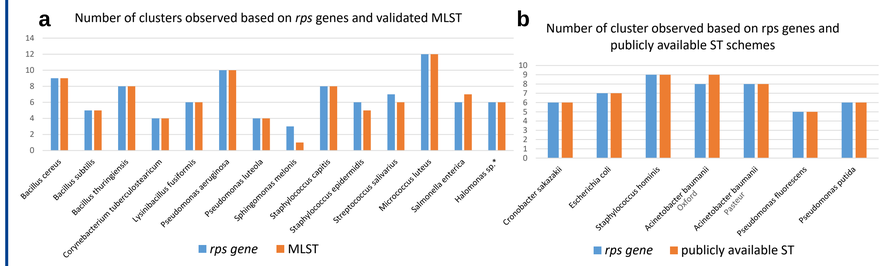

Figure 2. Number of strain type clusters observed when using *rps* genes for each representative species with at least 10 unique strains compared.(A) shows the comparison between the number of clusters observed using *rps* gene and Accugenix® validated strain typing method. (B) shows the similar comparison of clusters observed using *rps* genes and publicly available ST schemes that are not always part of Accugenix® validated libraries. In both cases number of strain types and their relatedness were similar when using either *rps* genes or MLST genes that are unique to each species.

3 Results

The number of strain typing clusters identified using housekeeping genes that are specific to a particular species were similar to the clusters observed from the universal set of *rps* genes (Figure 2). Figure 3 shows the strain types of *Bacillus subtilis* as determined based on the sequences of (a) *rps* genes, and (b) four housekeeping genes (*pycA*, *pur, glpF,* and *tpi*) used for validated strain typing purposes. The phylogenetic distribution from the *rps* genes is consistent with the GMP verified assay for the organism.

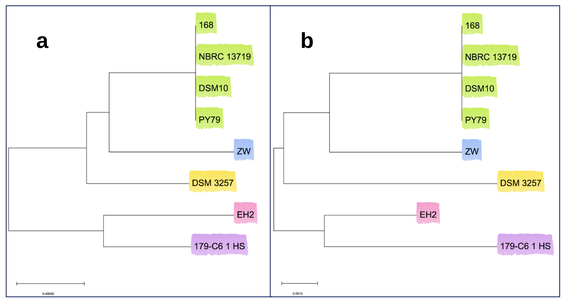

Figure 3. Phylogenetic distribution of 8 different strains representing *Bacillus subtilis*. The *rps* and the MLST associated genes were extracted from the whole genome sequences of the strains downloaded from NCBI (www.ncbi.nih.gov). The sequences were concatenated, and neighbor-joining phylogenetic trees were constructed based on the ribosomal protein subunit genes (A) and the four housekeeping genes used for GMP-validated strain typing assay (B).

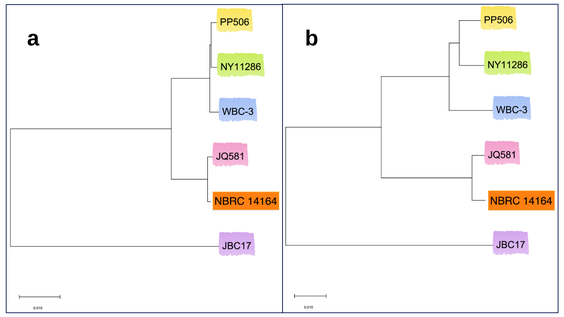

Figure 4. Phylogenetic distribution of 6 different strains representing *Pseudomonas putida.* The *rps* and the MLST associated genes were extracted and concatenated. Neighbor-joining phylogenetic trees were constructed based on the ribosomal protein subunit genes (A) and the eight housekeeping genes publicly available from PubMLST database for strain typing (B). The phylogenetic distribution and the strain types were clustered similarly using both approach of *rps* and MLST associated genes.

Figure 5. Neighbour Joining phylogenetic distribution of 8 different strains of *Xanthomonas perforans* based on the *rps* genes. The variations observed in the strains using *rps* genes were consistent with the phylogenetic groups described in Front. Microbiol. 10:448 based on the core genome analysis of >1,300 genes conserved within the 58 strains analyzed during that study.

Similar strain types were observed for organisms with use of *rps* genes and a publicly available MLST schemes for *Pseudomonas putida* (Figure 4). The MLST genes for *P. putida* includes *argS, gryB, ileS, nuoC, ppsA, recA, rpoB,* and *rpoD* – described by Ogura et al. 2019.

Furthermore, the strain typing based on the use of *rps* genes was analysed for *Xanthomonas perforans* that has been previously studied using a core-genome phylogenetic assay. With the core-genome MLST (cgMLST) three phylogenetic groups were determined with variations within the strains (Timilsina et al. 2019). Consistent strain distribution was observed with the approach used to extract and phylogenetically analyse *rps* genes used in this study.

4 Conclusion

Upon review and consideration of the results, *rps* genes provided robust strain differentiation, valuable insights into relatedness, and easy data sharing and collaboration in all aspects. The study findings can be concluded with three points:

1. The *rps* genes can be effectively pulled from the whole genome sequences to perform a subsequent strain typing comparison. The approach resulted in strain typing resolution at equal or better compared to the currently available species-specific strain typing methods based on 4-7 housekeeping genes.
2. The strain typing method was applicable for all species where a whole genome sequence is available, thus removing the need for development and validation time required during the implementation of strain typing assay for individual species.
3. The approach is reproducible and can be curated in-house into a central database system to compile ribosomal genes from type strains of representative species.

5 References

1. Jolley, K. A., Bliss, C. M., Bennett, J. S., Bratcher, H. B., Brehony, C., Colles, F. M., Wimalarathna, H., Harrison, O. B., Sheppard, S. K., Cody, A. J., & Maiden, M. C. (2012). Ribosomal multilocus sequence typing: Universal characterization of bacteria from domain to strain. Microbiology, 158(4), 1005–1015.
2. Ogura, K., Shimada, K., & Miyoshi-Akiyama, T. (2019). A multilocus sequence typing scheme of Pseudomonas putida for clinical and environmental isolates. Scientific Reports, 9(1), 13980.
3. Timilsina, S., Pereira-Martin, J.A., Minsavage, G.V., Iruegas-Bocardo, F., Abrahamina, P., Potnis, N., Kolaczkowski, B., Vallad, G.E., Goss, E.M., & Jones, J.B. (2019). Multiple recombination events drive the current genetic structure of *Xanthomonas perforans* in Florida. Front. Microbiol. 10:448.

We are interested in your strain typing needs and challenges! Please scan the QR code for a quick survey on microbial strain typing