HIVE genomic Comparator: Mycoplasma Use case

Alena Debrazhynetskaya1, Alin Voskanian1, Ekaterina Osipova1, Konstantinos Karagiannis1, Vahan Simonyan1

1Food and Drug Administration, Center for Biologics Evaluation and Research, Silver Spring, MD, 20993

Genomic DNA Sample Preparation
- M.hyorhinis DBS 1050 gDNA was isolated using Genomic tip 100G kit (Qiagen). Briefly, bacterial cells, collected at the early stationary phase of growth, were treated by 200 ng/ml lysozyme, followed by subsequent SDS- RNase cocktail and
- Proteinase K solution (masterPure DNA purification kit, Epicentre). Purified cell lysate was used for gDNA extraction on silicagel G-100 column chromatography and eluted with dh2O(Qiagen).
- The purity of isolated DNA was confirmed by electrophoresis on 1.0% agarose gel. Sample DNA was quantified by using fluorescence PicoGreen assay (Invitrogen).

Regions of high genomic similarity occur among many organism or within a single genome during such biological phenomenon as horizontal gene transfer, copy number variation, transposition of moveable elements, etc. HIVE has developed a tool allowing identification of internal or cross-genome similarity.

In this use-case the genome comparator was used to analyze internal structure of Mycoplasma hiorhinis and two assemblies available DBS-1050 and Pac Bio assembly.

The PacBio assembly provided was 40kb longer than that available on genebank and with the use of HIVE-genome comparator we have discovered that the denovo assembly by PacBio seemed to have failed to circularize the genome and instead has introduced computational artifacts. This tool will be used for various human genome assemblies as well as for discovering relationships between viral quasispecies.

Conclusion
Mycoplasma is a group of bacteria known as the smallest and simplest self-replicating life forms. A massive genome reduction, associated with a high number of recombinant variations and multiple repeated sequences changing mosaic and antigenic structures of the mycoplasma cells surface, provides with advantages during evolution. Different mobile genetic elements constitute up to 40% of mycoplasma genome size and mediate large-scale genome rearrangements with multiple short and long simple sequence repeats (SSRs), deletions, insertions, inversions, variable lipoprotein segments, etc. As result, mycoplasma genomes sequencing as well as their de novo assembly represent a quite challenging task, even when verified with the other closely related genomes.

The figures above are showing the chord diagrams produced by the HIVE genome comparator. Figure A shows self similarity between insilico DBS 1050 genebank reads and genome. Figure B shows self similarity between the PacBio reads and genome. And figure C shows similarity across both genomes with the left(green) representing PacBio and the right(red) representing genebank. Each line that extends across the circle indicates similarity for the corresponding coordinates.

The table above shows the level of similarity between different genetic regions of DBS 1050 and PacBio assembly.