Abstract

With the genome data in sequence data available, the need for novel methods to organize and visualize this data becomes ever more important. We’ve made improvements to two existing tools (OrtholugeDB and IslandViewer) to improve usability and expand their dataset coverage, and have released a novel visualization tool (IslandPlot) to allow the interactive visualization of genomic data.

Collectively, these tools facilitate more robust, yet user friendly and flexible, comparative analysis and data visualization for microbial genomes. These tools are built upon the Ortholuge platform and allow for the visualization of orthologous genes, and island data sharing model intended to facilitate communication between provincial and federal public health institutions.

IslandViewer2 is a web-accessible application for the computational discovery and analysis of genomic islands (GIs) in bacterial genomes. GI clusters of genes with probable horizontal origin, are of interest because they disproportionately contain genes encoding notable adaptations including antibiotic resistance and virulence1. GI predictions are integrated from Islandpath- DIMOB, SIGI-HMM2, and IslandPICK results. Recent changes have improved the functionality and usability of IslandViewer, including a complete redesign of the server backend to build a more efficient system. A custom visualization tool, IslandPlot, has been developed as a more flexible interface and to accommodate the growth in the number of genomes. Virulence Factor Database (VFDB), Antibiotic Resistance Genes Database3,3 (ARGDB), Comprehensive Antibiotic Resistance Database (CARD) annotations, and pathogen-associated genes can be overlaid on GI predictions. A module to identify virulence or resistance gene homologues in genomes missing such annotations has been incorporated. The ability to handle unfounded genomes is being incorporated, which involves alignment of contigs to a reference and accuracy assessment, to allow for rapid GI detection before genome finishing. Overall, IslandViewer improvements provide a much more flexible interface, coupled with additional highly relevant annotation information, to aid analysis of these regions of particular medical and adaptive interest in bacterial genomes.

IslandPlot is a light weight visualization library written in javascript using the D3 library. It provides a rich API to allow the rapid visualization of complex genomic data using a convenient standards-based JSON configuration file. When integrated in to an existing web service IslandPlot allows researchers to interact with data, dynamically alter the view, or even page or resize the visualization in their browser window. In addition IslandPlot has built in functionality to export any visualization in png or svg format for easy inclusion in manuscripts or presentations.

OrtholugeDB is a database of orthologous predictions for microbial genomes

Currently there are over 2000 completely sequenced bacterial and archaean genomes and this number is continuing to grow. There is a need for a continuously updated ortholog database for microbial species, that preferably includes further assessment of orthologs. We have created OrtholugeDB (www.pathogenomics.ca/ortholugeDB): a web-accessible database of ortholog predictions for all completely sequenced bacterial and archaean species. In addition to the RBBH-based ortholog predictions, this database provides:

- Ortholog4 assessments of ortholog quality
- Inparanoid-based5 predictions of inparalogs (recent gene duplications)
- Pre-computed multi-species ortholog groups at several different taxonomic levels

This data will be updated routinely. All steps in the Ortholuge pipeline have been automated, including the selection of the outgroup comparison species.

Web Interface

OrtholugeDB has the ability to:

- Retrieve all orthologs shared between two genomes
- Retrieve orthologs for a gene
- Conduct comparative analysis, identifying genotype-phenotype associations through analysis such as identifying genes found in one or more user-defined reference genome but not found in genomes of a different phenotype

References


Acknowledgements

IRIDa and IslandViewer funding is gratefully provided by Genome Canada, Genome BC and the federal Genomics R&D Initiative (GRID), with additional support by Simon Fraser University, the CIHR/MSFHR Bioinformatics Training Program, and Michael Smith Foundation for Health Research Trainee Awards. We thank additional project advisors for constructive comments.

Ortholuge and OrtholugeDB were funded in part by Genome Canada/Genome BC with the support of Cystic Fibrosis Foundation Therapeutics Inc. and Michael Smith Foundation for Health Research (MSFHR).