



How to use whole genome sequencing for monitoring of antimicrobial resistance in bacteria

Introduction

Application of whole genome sequencing (WGS) for antimicrobial susceptibility testing (AST) offers the possibility to rapidly transform and improve our way of performing surveillance of antimicrobial resistance (AMR) in view of the extensive information that can be obtained with a single assay.

At present, AST is performed by phenotypic methods by which bacteria in vivo is exposed to different concentrations of selected antimicrobials. This allows definition of either the minimum antimicrobial concentration at which bacterial growth is inhibited (MIC, minimum inhibitory concentration) or the diameter of the zone of inhibition of bacterial growth around an antimicrobial-containing disc. The numerical values obtained (either an MIC-value or an inhibition zone diameter) are interpreted according to a set of criteria to classify isolates as susceptible or resistant to a specific antimicrobial. Although supported in decades by available, comprehensive guidelines for performance and interpretation of AST (EUCAST, CLSI), this approach still suffers of serious drawbacks, for example when defining MIC-values/inhibition zone diameters for some bacteriostatic antimicrobials such as sulphonamides or when MIC-values/inhibition zone diameters are close to breakpoint values for categorization of susceptibility/resistance. Furthermore, due to practical reasons, each bacterial isolate can only be tested for susceptibility to a limited number of antimicrobials.

WGS holds the promise to overcome most of the shortcomings of phenotypic AST and offers a prediction of resistance in bacteria based on presence of known resistance mediating genes. However, definite evidence for feasibility of AST by WGS is lacking for most bacterial species and some scientific and technical hurdles need to be solved. These include knowledge gaps in the genetic basis of AMR and establishment of harmonized quality control metrics in WGS-based AST.

Currently, several databases of AMR genes exist (ResFinder, ARIBA, SRST2, GeneFinder, and CARD etc.). These databases perform differently in detection of AMR determinants but these differences are not immediately evident to all users. Optimally, one single, public database of all known AMR genes and mutations should be established, regularly updated, and strictly curated using minimum standards for the inclusion of new AMR genes. This database should be organized by bacterial species and include predicted phenotypes associated with the AMR genes.

Insufficient knowledge of AMR mechanisms is the major limiting factor in application of WGS for AST and knowledge of the phenotype-genotype relationship differs according to bacterial species and antimicrobial compound. The EU monitoring of AMR in zoonotic and commensal bacteria encompasses *Salmonella* spp., *Escherichia coli*, *Campylobacter jejuni* and *C. coli*, *Enterococcus faecium* and *E. faecalis* and mandates testing of susceptibility to antimicrobials relevant to human and veterinary medicine (Decision 2013/652/EU). According to this legislation, AST results should be interpreted using epidemiological cut-off values (ECOFFs), which divide bacterial populations in two groups based on presence/absence of phenotypically detectable



resistance mechanisms to any given antimicrobial (EUCAST). Overall, results from early reports show a good correlation between genotypes and phenotypes but all these studies emphasize the need of further research including larger number of isolates and better understanding of the molecular basis of resistance to specific antimicrobials before routine application of WGS can be considered as a valid alternative to phenotypic methods in programmes for surveillance of antimicrobial resistance.

Requirements for sequencing

On the laboratory part of WGS, several steps are crucial to obtain DNA of sufficient quantity and quality for sequence analysis. Important measurements are:

WGS process	Impact
DNA quality and quantity	Poor quality or contaminant DNA can be detected. Contamination may originate either from upstream handling of the bacterial isolates and DNA purification or from the preparation and running of the DNA samples on the sequencer.
Raw reads quality and quantity	Used to infer the coverage of the bacterial genome and thereby provide a QC of the sequencing quality.
Reference genome assembly and/or <i>de novo</i> assembly quality	Assembly – the formation of a draft genome sequence from raw reads can be evaluated to provide QC of data processing.

A brief introduction to some of the applied methods for DNA extraction, library preparation and available genomic tools follows below.

DNA extraction methods

There are many available methods and kits for DNA extraction. Here are examples of some widely used methods:

MagNA Pure (Roche) boilates for bacterial DNA extraction

The boilate method was developed for PCR templates, but DNA extracted from boilates has proven to be suitable for the preparation of libraries for whole genome sequencing. The advantages of this method include that there are no requirements for special equipment or reagents, preparation is rapid and it is safe to transport pathogenic isolates for sequencing. MagNA Pure is an automated system for DNA extraction. MagNA Pure LC 2.0 Instrument can perform a majority of the extraction steps, such as binding of DNA to magnetic glass particles, washing steps, elution of pure DNA. The purified DNA may be analysed with respect to DNA integrity, recovery, purity and ability to amplify target sequence with LightCycler® 480 and LightCycler® Carousel-Based Instruments.

(<https://lifescience.roche.com/content/dam/RMS/lifescience/Documents/PDF/MagNA-Pure-Versatile-Nucleic-Acid-Purification.pdf>)

Easy-DNA™ Kit (Invitrogen) DNA extraction

This extraction method yields high-quality DNA with an average size between 100 kb and 200 kb, which is suitable for PCR, DNA hybridization, genomic DNA library construction, and other applications. The extraction



procedure contains only 4 steps with no special equipment required (https://assets.thermofisher.com/TFS-Assets/LSG/brochures/713_021456_easydnapr_bro.pdf)

QIASymphony DNA Investigator Kit

DNA extraction can be carried out using an automated method to extract DNA from bacterial cells. The QIASymphony DNA Investigator Kit enables automated purification of genomic DNA from 1–96 samples, such as swabs, filters, casework or crime-scene samples, and blood on the QIASymphony SP. Purification is fast and efficient, and purified DNA performs well in downstream analyses. This method requires QIASymphony SP/AS instrument.

(<https://www.qiagen.com/us/resources/resourcedetail?id=15606127-cc31-4681-a27a-bcec58502f72&lang=en>)

Sequencing method and library preparation

The Illumina sequencing platforms are widely used in many laboratories, and there are specific kits available for library preparation for these platforms. The Illumina NexteraXT DNA library preparation manual is found at:

<https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html>

Genomic tools

A wide range of tools for genomic analysis is available online. A comprehensive list of tools can be found in Appendix 1. Tools used for species identification (kmer-finder), assembly of genomes (SPAdes and Velvet) and identification of AMR genes (ResFinder) are all available on <https://cge.cbs.dtu.dk/services/all.php>

What is needed for quality control of WGS data?

Quality control (QC) is essential to guarantee accuracy and precision of results of any laboratory test, including WGS. Sequences of poor quality can lead to major errors in AST by failing to reveal AMR genes or mutations. Other sources of error may derive from contamination of the DNA and from erroneous data handling.

At present, different QC parameters are available to control and standardize WGS procedures (CDC-PulseNet, 2015; Ellington et al. 2017). Although QC metrics for WGS are widely available, currently, no international standards for QC-thresholds have been set, thus hampering data portability. Agreement on QC metrics and minimum performance standards to ensure good quality of the entire WGS process from DNA isolation to data analysis would facilitate early harmonization of analytical approaches and interpretative criteria for WGS-based predictive AST, and only datasets that pass agreed QC metrics should be used in AST predictions. A set of QC parameters for draft genome assembly and their explanation has been listed by the EUCAST committee (Ellington et al. 2017).

Table 1 QC parameters for draft genome assembly (Ellington et al. 2017)

QC parameter	Explanation
Number of reads	Number of reads refers to sequence yield (the amount sequenced)
Average read length	The average length of all reads, measured in base pairs.

QC parameter	Explanation
Number of reads mapped to reference sequence	The number of reads that map to a closed (finished) genome (same strain).
Proportion of reads mapped to reference sequence (%)	The proportion of reads that map to a closed genome (same strain).
Number of reads mapped to reference chromosome	The number of reads that map to a closed chromosome (same strain).
Proportion of reads mapped to reference chromosome (%)	The proportion of reads that map to a closed genome's chromosome (same strain). This cannot exceed 100%.
Reads mapped to reference plasmids	The number of reads that map to plasmids, if present.
Proportion of reads mapped to reference plasmids (%)	The proportion of reads that map to plasmids (if present) of the closed genomes. This cannot exceed 100%.
Depth of coverage, total DNA sequence	Describes the number of times the sequenced base pairs cover the reference DNA. Number of base pairs sequenced divided by the total size (both chromosome and plasmids) of the closed genome (same strain), often expressed with an "x" (e.g. 30x). A minimum depth of 30x is usually preferred.
Depth of coverage: chromosome	As for total DNA coverage, but describes the number of base pairs sequenced divided by the total size of the closed chromosome (same strain).
Depth of coverage: plasmid	As for total DNA coverage, but describes the number of base pairs sequenced divided by the total size of the closed plasmid (same strain).
Size of assembled genome	Often used to identify contamination. If the calculated size of all the contigs in base pairs exceeds that expected it could indicate more than one genome.
Size of assembled genome per total size of DNA sequence (%)	The proportion of contigs that map directly to the closed genome (same strain). This cannot exceed 100%.
Total number of contigs	Generally, the total number of contigs assembled, <1000 contigs indicates good quality. For organisms with genomes 5- 6 Mb in size then <100 contigs is (generally) realistic.
Longest contig length	The length of the longest contig.
Mean, median and standard deviation	Mean, median and standard deviation of the contigs, used to evaluate quality
N50	The length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs, and for which the collection of all contigs of that length or shorter also contains at least half of the sum of the lengths of all contigs. N50 >15 000 normally indicates good quality, but minimum size of 30 000 bp is often preferred.
NG50	Helpful for comparisons between assemblies. As N50, except that 50% of the genome size must be of the NG50 length or longer. Where the assembly size ≤ the genome size then NG50 cannot exceed N50.

References

CLSI, Clinical and Laboratory Standards Institute. Available at: <http://clsi.org/>

CARD, the Comprehensive Antibiotic Resistance Database. Available at <https://card.mcmaster.ca/home>

CDC-PulseNet, Laboratory standard operating procedure for Pulse-Net Nextera XT library prep and run setup for the Illumina MiSeq and Pulse-Net standard operating procedure for Illumina MiSeq data quality control. Available at: <https://www.cdc.gov/pulsenet/pdf/pnl32-miseq-nextera-xt.pdf> and <http://www.pulsenetinternational.org/assets/PulseNet/uploads/wgs/PN07-Illumina-MiSeq-Data-QC.PDF>



Decision 2013/652/EU. Commission implementing decision of 12 November 2013 on the monitoring and reporting of antimicrobial resistance in zoonotic and commensal bacteria.

Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, Grundman H, Hasman H, Holden MT, Hopkins KL, Iredell J, Kahlmeter G, Köser CU, MacGowan A, Mevius D, Mulvey M, Naas T, Peto T, Rolain JM, Samuelsen Ø, Woodford N; The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect.* 2017 Jan;23(1):2-22. doi: 10.1016/j.cmi.2016.11.012.

ENGAGE, Establishing next generation sequencing ability for genomic analysis in Europe. Further information: <http://engage-europe.eu/>

EUCAST, the European Committee on Antimicrobial Susceptibility Testing. <http://eucast.org/>

NCBI, the Bacterial Antimicrobial Resistance Reference Gene Database. Available at: <http://www.ncbi.nlm.nih.gov/bioproject/313047>

Appendix 1

Guideline to bioinformatics tools – ‘Availability, what to use – and when’

This document describes the most commonly used software and algorithms for processing whole genome sequencing. It is divided into categories which describe the key processes for analysing short read data.

The list was initially generated by ENGAGE (Establishing Next Generation Sequencing Ability for Genomic Analysis in Europe, has received funding from European Food Safety Authority (EFSA), grant agreement GP/EFSA/AFSCO/2015/01 (New approaches in identifying and characterizing microbiological and chemical hazards)) and is adapted from <http://www.engage-europe.eu/resources/protocols-and-training>

1. Quality Assessment and Trimming

This is the process by which the quality of fastq files is determined and subsequent optional trimming of the data to trim or remove poor quality reads.

1.1. Trimmomatic

A flexible read trimming tool that will remove Illumina adapters, reads below a certain length and low quality ends of the read (Windows, Mac OS X and Linux).

<http://www.usadellab.org/cms/index.php?page=trimmomatic>

Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bolger, Lohse & Usadel (2014). Bioinformatics, btu170.

1.2. Seqtk

Tool for processing sequences in the FASTA or FASTQ that can be used for adapter removal and trimming of low-quality bases (Windows, Mac OS X and Linux).

<https://github.com/lh3/seqtk>

1.3. FastX

Toolkit for FASTQ and FASTA preprocessing that can be used for trimming, clipping, barcode splitting, formatting and quality trimming (Windows, Mac OS X and Linux).

http://hannonlab.cshl.edu/fastx_toolkit/

1.4. FastQC

A quality control tool for assessing the quality of NGS data (Windows, Mac OS X and Linux)

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

2. Assembly

This is the process of joining short reads into longer contigs (contiguous lengths of DNA) without the need for a reference sequence.

2.1. VelvetK

Perl script to estimate best k-mer size to use for your Velvet de novo assembly (Windows, Mac OS X and Linux).

<http://bioinformatics.net.au/software/velvetk.shtml>

2.2. KmerGenie

Best k-mer length estimator for single-k genome assemblers like velvet (Windows, Mac OS X and Linux).

Informed and Automated k-Mer Size Selection for Genome Assembly. Chikhi & Medvedev HiTSeq 2013.

<http://kmergenie.bx.psu.edu/>

2.3. Khmer

Set of command-line tools for dealing with large and noisy datasets to normalise and scale the data for more efficient genome assembly (Linux and Mac OS X)

The khmer software package: enabling efficient nucleotide sequence analysis. Crusoe et al., 2015. F1000 <http://dx.doi.org/10.12688/f1000research.6924.1>

<http://khmer.readthedocs.io/en/v2.0/>

2.4. Minia

Short-read assembler based on a de Bruijn graph for low-memory assembly (Windows, Mac OS X and Linux).

Space-efficient and exact de Bruijn graph representation based on a Bloom filter. Chikhi & Rizk. Algorithms for Molecular Biology, BioMed Central, 2013, 8 (1), pp.22.

<http://minia.genouest.org/>

2.5. SPAdes

Short and hybrid-long read assembler based on a de Bruijn graph that also performs error correction and is a multi-k genome assembler (Mac OS X and Linux).

SPAdes: A New Genome Assembly Algorithm and Its Applications to Single- Cell Sequencing. Bankevich, Nurk, Antipov, Gurevich, Dvorkin, Kulikov, Lesin, Nikolenko, Pham, Prjibelski, Pyshkin, Sirotkin, Vyahhi, Tesler, Alekseyev, and Pevzner. Journal of Computational Biology 19(5) (2012), 455-477. doi:10.1089/cmb.2012.0021

<http://cab.spbu.ru/software/spades/>

2.6. Velvet

De novo short read genome assembler with error correction to produce high quality unique contigs (Linux).

Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Zerbino and Birney. Genome Res. May 2008 18: 821-829; Published in Advance March 18, 2008, doi:10.1101/gr.074492.107

<https://www.ebi.ac.uk/~zerbino/velvet/>

2.7. Canu

Long-read assembler designed for high-noise data such as that generated by PacBio or Oxford Nanopore MinION. Canu also performs error correction (Windows, Mac OS X and Linux).

Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Koren, Walenz, Berlin, Miller & Phillippy. doi:http://dx.doi.org/10.1101/071282 <http://canu.readthedocs.io/en/stable/index.html>

2.8. Bandage

Program for visualising de novo assembly graphs by displaying connection which are not present in the contigs file for assembly assessment (Linux and Mac OS X).

Bandage: interactive visualization of de novo genome assemblies. Wick, Schultz, Zobel, and Holt. Bioinformatics (2015) 31 (20): 3350-3352 doi:10.1093/bioinformatics/btv383 <http://rrwick.github.io/Bandage/>

3. Annotation

The process which takes the raw sequence of contigs resulting from assembly and marks it with features such as gene names and putative functions.

3.1. Prokka

Software tool for the rapid annotation of prokaryotic genomes (Windows, Mac OS X and Linux).

Prokka: rapid prokaryotic genome annotation. Seemann T. Bioinformatics. 2014 Jul 15;30(14):2068-9. PMID:24642063

<http://www.vicbioinformatics.com/software.prokka.shtml>

3.2. RAST

Fully-automated service for annotating complete or nearly complete bacterial and archeal genomes (Online tool).

The RAST Server: Rapid Annotations using Subsystems Technology. Aziz, Bartels, Best, DeJongh, Disz, Edwards, Formsma, Gerdes, Glass, Kubal, Meyer, Olsen, Olson, Osterman, Overbeek, McNeil, Paarmann, Paczian, Parrello, Pusch, Reich, Stevens, Vassieva, Vonstein, Wilke & Zagnitko. BMC Genomics, 2008

<http://rast.nmpdr.org/>

3.3. Genix

Fully automated pipeline for bacterial genome annotation (Online tool).

http://labbioinfo.ufpel.edu.br/cgi-bin/genix_index.py

4. Alignment or sequence searching

Tools to align a sequence to other sequences locally or against publically available nucleotide or protein archives.

4.1. BLAST

Search tool to find regions of similarity between biological sequences through alignment and calculating statistical significance (Windows, Mac OS X and Linux)

Basic local alignment search tool. Altschul, Gish, Miller, Myers & Lipman. Journal of Molecular Biology, Volume 215, Issue 3, 1990, Pages 403-410

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

4.2. MUMmer

A system for rapidly aligning entire genomes and finding matches in DNA sequences (Windows, Mac OS X and Linux).

Versatile and open software for comparing large genomes. A.L. Delcher, A. Phillippy, J. Carlton, and S.L. Salzberg, Nucleic Acids Research (2002), Vol. 30, No. 11 2478-2483.

<http://mummer.sourceforge.net/>

4.3. Mega

Sophisticated and user-friendly software suite for analysing DNA and protein sequence data from species and populations (Windows, Mac OS X and Linux)

MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Kumar, Stecher & Tamura (2016) Molecular Biology and Evolution 33:1870-1874

<http://www.megasoftware.net/>

5. Mapping

Alignment of short reads against a reference sequence so that amount of coverage or variations compared to the reference can be assessed.

5.1. BWA

Software package for mapping low-divergent sequences against a large reference genome using the Burrows-Wheeler transform algorithm (Windows, Mac OS X and Linux).

Fast and accurate short read alignment with Burrows-Wheeler Transform. Li H. and Durbin R. (2009) Bioinformatics, 25:1754-60. [PMID: 19451168]

<http://bio-bwa.sourceforge.net/>

5.2. Bowtie 2

Tool for aligning sequencing reads to long reference genomes also based on the Burrows-Wheeler transform algorithm (Windows, Mac OS X and Linux).

Fast gapped-read alignment with Bowtie 2. Langmead & Salzberg. Nature Methods. 2012, 9:357-359.

<http://bowtie-bio.sourceforge.net/index.shtml>

5.3. Tablet

Lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments that can be used to view mapping (Windows, Mac OS X and Linux).

Using Tablet for visual exploration of second-generation sequencing data. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD and Marshall D. 2013. Briefings in Bioinformatics 14(2), 193-202.

<https://ics.hutton.ac.uk/tablet/>

6. Variant Calling

6.1. SAMtools

Toolkit that provides various utilities for manipulating alignments in the SAM format and also can be used generating consensus sequences and variant calling (Windows, Mac OS X and Linux).

The Sequence alignment/map (SAM) format and SAMtools. Li*, Handsaker*, Wysoker, Fennell, Ruan, Homer, Marth, Abecasis, Durbin and 1000 Genome Project Data Processing Subgroup (2009) Bioinformatics, 25, 2078-9. [PMID: 19505943]

<http://samtools.sourceforge.net/>

6.2. GATK

Toolkit with a primary focus on variant discovery and genotyping (Windows, Mac OS X and Linux).

The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. McKenna, Hanna, Banks, Sivachenko, Cibulskis, Kernytsky, Garimella, Altshuler, Gabriel, Daly & DePristo. Genome Res. September 2010 20: 1297-1303; doi:10.1101/gr.107524.110

<https://software.broadinstitute.org/gatk/>

6.3. Picard

A set of command line tools (in Java) for manipulating high-throughput sequencing data and formats (Windows, Mac OS X and Linux).

<http://broadinstitute.github.io/picard/>

7. Phylogenetic analysis

Assessment of the evolutionary relationship between samples using either distance-based or Bayesian methodologies.

7.1. RaxML

Randomized Axelerated Maximum Likelihood program for sequential and parallel Maximum Likelihood based inference of large phylogenetic trees (Windows, Mac OS X and Linux).

RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. A. Stamatakis. *Bioinformatics* (2014) 30 (9): 1312-1313.

<https://sco.h-its.org/exelixis/software.html>

FastTree

Faster tool for speedy inference of approximately-maximum-likelihood phylogenetic trees from alignments of nucleotide or protein sequences (Windows, Mac OS X and Linux).

FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. Price, Dehal & Arkin (2009). *Molecular Biology and Evolution* 26:1641-1650, doi:10.1093/molbev/msp077.

<http://www.microbesonline.org/fasttree>

7.2. CSI Phylogeny

Tool to call SNPS, filters the SNPs, do site validation and inference of phylogeny through a graphical user interface (Online tool).

Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms.

Kaas, Leekitcharoenphon, Aarestrup & Lund. *PLoS ONE* 2014; 9(8): e104984.

<https://cge.cbs.dtu.dk/services/CSIPhylogeny/>

7.3. Harvest

Suite of core-genome alignment and visualization tools for quickly analysing thousands of intraspecific microbial genomes, including variant calls, recombination detection, and phylogenetic trees (Windows, Mac OS X and Linux).

The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Treangen, Ondov, Koren & Phillippy. *Genome Biology*, 15 (11), 1-15

<https://www.cbcb.umd.edu/software/harvest>

7.4. Gubbins

Gubbins (Genealogies Unbiased By recombinations In Nucleotide Sequences) is an algorithm that iteratively identifies loci containing elevated densities of base substitutions while concurrently constructing a phylogeny based on the putative point mutations outside of these regions (Windows, Mac OS X and Linux).

Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Croucher, Page, Connor, Delaney, Keane, Bentley, Parkhill, Harris, Nucleic Acids Research, 2014. doi:10.1093/nar/gku1196

<https://github.com/sanger-pathogens/gubbins>

7.5. BEAST

Cross-platform program for Bayesian analysis of molecular sequences using MCMC (Windows, Mac OS X and Linux).

Bayesian phylogenetics with BEAUti and the BEAST 1.7. Drummond, Suchard, Xie & Rambaut (2012) Molecular Biology And Evolution 29: 1969-1973.

<http://beast.bio.ed.ac.uk/>

7.6. FigTree

A graphical viewer of phylogenetic trees and program for producing publication-ready figures of trees (Windows, Mac OS X and Linux).

<http://tree.bio.ed.ac.uk/software/figtree/>

8. **Virulence and antimicrobial resistance gene prediction**

Inference of potential for a virulent phenotype or resistance to an antimicrobial based on nucleotide sequence

8.1. PathogenFinder

Web-server for the prediction of bacterial pathogenicity by analysing the input proteome, genome, or raw reads provided by the user (Online tool).

PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. Cosentino, Larsen, Aarestrup & Lund (2013) PLoS ONE 8(10): e77302.

<https://cge.cbs.dtu.dk/services/PathogenFinder/>

8.2. ResFinder

Web-server that identifies acquired antimicrobial resistance genes in total or partial sequenced isolates of bacteria (Online tool).

Identification of acquired antimicrobial resistance genes. Zankari, Hasman, Cosentino, Vestergaard, Rasmussen, Lund, Aarestrup & Larsen (2012) J Antimicrob Chemother.

<https://cge.cbs.dtu.dk/services/ResFinder/>

9. Species identification

9.1. Kraken

System for assigning taxonomic labels to short DNA sequences, usually obtained through metagenomics studies (Linux).

Kraken: ultrafast metagenomic sequence classification using exact alignments. Wood & Salzberg (2014) Genome Biology, 15:R46.

<https://ccb.jhu.edu/software/kraken/>

10. Comparative genomic tools

Comparison of multiple genomes to determine regions of similarity or difference either on a gene-by-gene basis or across the whole genome.

10.1. BEDTools

Toolkit for the manipulation of genome data for genomic analysis tasks on genomic intervals from multiple files (Mac OS X and Linux)

BEDTools: a flexible suite of utilities for comparing genomic features. Quinlan & Hall (2010) Bioinformatics 26 (6) doi:10.1093/bioinformatics/btq033

<http://bedtools.readthedocs.io/en/latest/index.html>

10.2. Roary

High speed stand-alone pan genome pipeline, which takes annotated assemblies in GFF3 format and calculates the pan genome (Windows, Mac OS X and Linux).

Roary: Rapid large-scale prokaryote pan genome analysis. Page, Cummins, Hunt, Wong, Reuter, Holden, Fookes, Falush, Keane & Parkhill (2015) Bioinformatics; 31(22):3691-3693. doi:10.1093/bioinformatics/btv421

<https://github.com/sanger-pathogens/Roary>

10.3. Mauve

Interactive genome alignment software that allows for easy browsing of multiple genomes to look for similarities and differences (Windows, Mac OS X and Linux).

Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Darling, Mau, Blattner, & Perna (2004) Genome Res. 1394-1403; doi:10.1101/gr.2289704

<http://darlinglab.org/mauve/mauve.html>

10.4. ACT

Java application for displaying pairwise comparisons between two or more DNA sequences and allowing browsing of detailed annotation (Windows, Mac OS X and Linux).

ACT: the Artemis Comparison Tool. Carver, Rutherford, Berriman, Rajandream, Barrell & Parkhill (2005) *Bioinformatics*;21;16;3422-3 DOI: 10.1093/bioinformatics/bti553

<https://www.sanger.ac.uk/science/tools/artemis>

10.5. BRIG

Image generating software that displays circular blast comparisons between a large number of genomes or DNA sequences (Windows, Mac OS X and Linux).

BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons (2011) Alikhan, Petty, Zakour & Beatson. *BMC Genomics*, 12:402. PMID: 21824423

<http://brig.sourceforge.net/>

10.6. EasyFig

Python application for creating linear comparison figures of multiple genomic loci with an easy-to-use graphical user interface (GUI) (Windows, Mac OS X and Linux).

Easyfig: a genome comparison visualiser. Sullivan, Petty & Beatson (2011) *Bioinformatics*; 27 (7): 1009-1010. PMID: 21278367

<http://mjsull.github.io/Easyfig/>

10.7. SeqFindR

Tool to easily create informative genomic feature plots by detecting the presence or absence of genomic features from a database in a set of genomes. If infrastructure is not available the cloud based services are worth considering (Mac OS X and Linux)

<https://github.com/mscook/SeqFindR>

--- --- ---