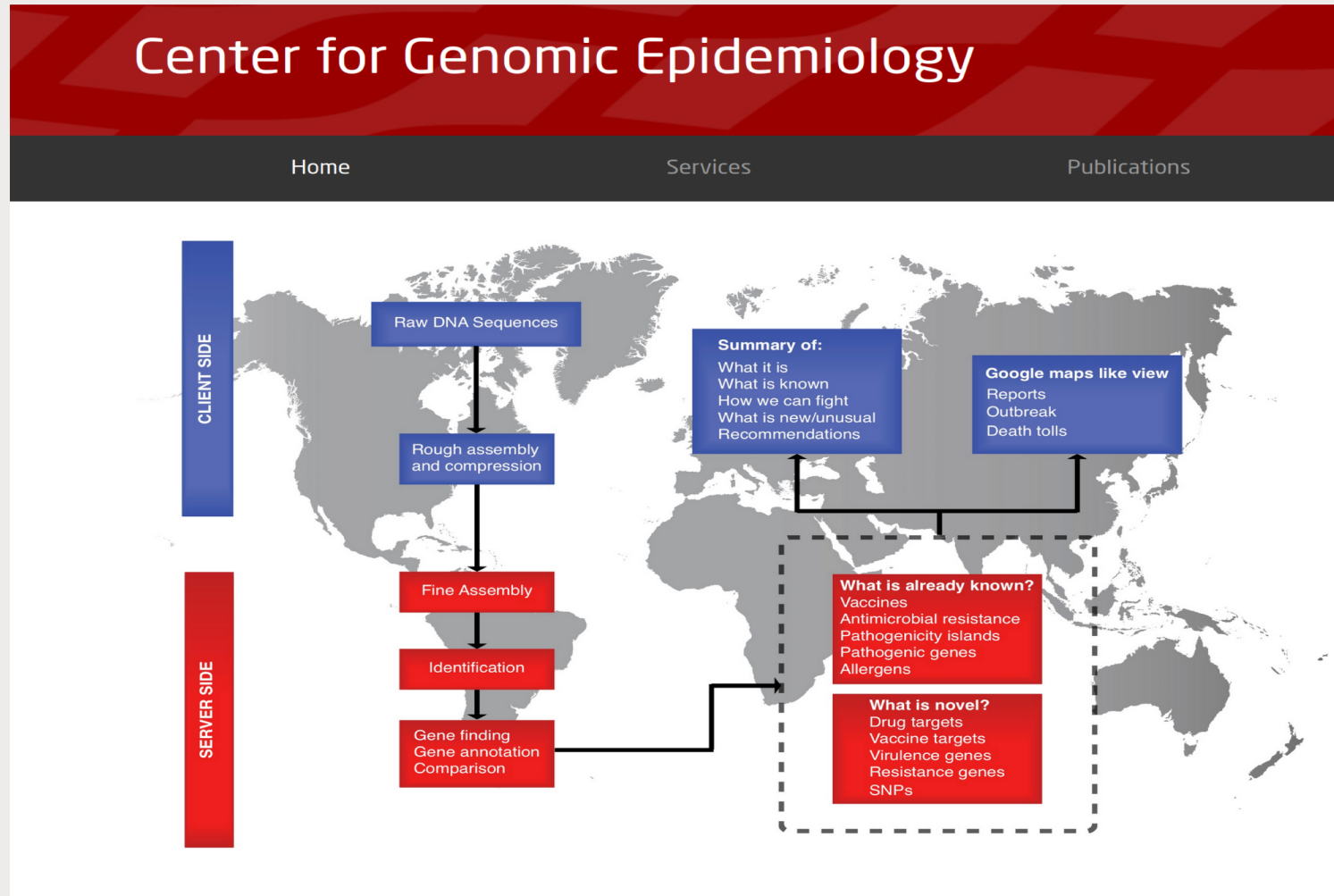# CGE Online Bioinformatics Tools

Tolbert Sonda, t.sonda@kcri.ac.tz

Data Management and Bioinformatics Unit

Kilimanjaro Clinical Research Institute, Tanzania

Learning objectives:

- To learn how to access CGE Tools

- To use KmerFinder in identification of bacterial species

- To use ResFinder in determination of AMR genes

- To retrieve and interpret outputs from KmerFinder and ResFinder

CGE home page

www.genomicepidemiology.org/

Overview of CGE Tools

http://www.genomicepidemiology.org/services/

Typing

MLST
Multi Locus Sequence Typing (MLST) from an assembled genome or from a set of reads.

PlasmidFinder
PlasmidFinder identifies plasmids in total or partial sequenced isolates of bacteria.

pMLST
Multi Locus Sequence Typing (MLST) from an assembled plasmid or from a set of reads.

cgMLSTFinder
Core genome Multi Locus Sequence Typing (cgMLST) from a set of reads.

KmerFinder
Prediction of bacterial species using a fast K-mer algorithm.

MGE
Identification of mobile genetic elements and their relation to antimicrobial resistance genes and virulence factors.

SpeciesFinder
Prediction of bacterial species using the S16 ribosomal DNA sequence.

SeroTypeFinder
Prediction of serotypes in total or partial sequenced isolates of E. coli.

SeqSero
SeqSero predicts the Salmonella serotype of either the pre-assembled or raw read sequence data provided to the service.

PCR-tools

RUCS
RUCS: Rapid Identification of PCR Primers Pairs for Unique Core Sequences.

Other

MyKMAfinder
MyKMAfinder performs typing or pheno typing using KMA based on a user defined database.

MyDbFinder
MyDbFinder performs typing or pheno typing using blast based on a user defined database.

MyKmerFinder
MyKmerFinder performs typing or pheno typing using Kmers based on a user defined database.

DeHumanizer
The DeHumanizer web-server is a tool for human filtering based on the method described by Zhang et al.

HostPhinder
HostPhinder identifies the bacterial host of a query phage genome based on its genomic similarity to a database of phage genomes with known host.

MetaPhinder
MetaPhinder: Identifying Bacteriophage Sequences in Metagenomic Data Sets.

Phenotyping

ResFinder
Identifcation of acquired antibiotic resistance genes.

ResFinderFG
Identifcation of functional metagenomic antibiotic resistance determinants.

LRE-finder
Identifcation of genes and mutations leading to linezolid resistance.

KmerResistance
Identifcation of acquired antibiotic resistance genes using Kmers.

PathogenFinder
Prediction of a bacteria's pathogenicity towards human hosts.

VirulenceFinder
Identifcation of acquired virulence genes.

Restriction-ModificationFinder
Determination of Restriction-Modification sites (based on REBASE.)

SPIFinder
SPIFinder identifies Salmonella Pathogenicity Islands.

ToxFinder
ToxFinder identifies genes involved in mycotoxin synthesis.

Phylogeny

MINTyper
Identification of SNPs with automatic filtering, masking and site validation together with inferred phylogeny based on both long and short sequencing data.

CSIPhylogeny
CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality* SNPs.

NDtree
NDtree constructs phylogenetic trees from Single-End or Pair-End FASTQ files.

Evergreen
Evergreen generates a forest of constantly updated phylogenetic trees with publicly available whole-genome sequencing data from foodborne, bacterial isolates that were deposited in the short sequencing read archives (NCBI SRA/ENA).

TreeViewer
Phylogeny Tree Viewer.

Metagenomics

CCMetagen
CCMetagen: Comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data.

CGE offers a lot, but your question matters the most

Tools on demand

We want to know bacteria species

We want to know AMR genes

I will demo the use of these tools to analyse bacterial sequence data

Uploading your sequences

Accessing your results

Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.

To get notified by email: | t.sonda@kcri.ac.tz | Notify me via email

This page will update itself automatically.

In case the page does not reload, press here to refresh the status page.

ResFinder-4.1: Job 602268C100000E771ECA5425 finished ➤ Inbox ×

CGEwebface@cbs.dtu.dk
to me ▾

Dear sir/madam,
Your job 602268C100000E771ECA5425 has finished on our ResFinder-4.1 server.
You may visit the following link to study the results:
https://cge.cbs.dtu.dk//cgi-bin/webface.fcgi?jobid=602268C100000E771ECA5425

The data will expire at 2021-02-16 11:49:37.

Yours sincerely,
    CGE webserver

For help and technical assistance contact cgehelp@cbs.dtu.dk

Bacterial species identified

KmerFinder-3.2 Server - Results

KmerFinder 3.2 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth	q_value	p_value
NC_017628.1 Escherichia coli IHE3034, complete genome	1569	✅ 130699	13	169895	✅ 75.61	✅ 77.38	0.77	75.61	77.38	0.77	130658.16	✅ 1.0e-26
NZ_CP060061.1 Escherichia coli strain G6/7 chromosome, complete genome	19495	13351	55	177007	7.72	7.36	0.08	54.57	54.60	0.53	13184.14	1.0e-26
NZ_CP022279.1 Escherichia coli strain STEC299 chromosome, complete genome	18745	7042	54	163145	4.07	4.38	0.04	74.16	79.98	0.79	6879.21	1.0e-26
NZ_CP048304.1 Escherichia coli strain 9 chromosome, complete genome	19524	2892	57	165353	1.67	1.77	0.02	71.33	75.26	0.75	2724.89	1.0e-26
NZ_CP032204.1 Escherichia coli strain AR_0013 chromosome, complete genome	10619	1868	61	179738	1.08	1.06	0.01	58.16	57.40	0.56	1692.27	1.0e-26
NZ_CP035755.1 Escherichia coli E110019 extrachromosomal	1634	38	0	864	0.02	4.40	0.04	0.21	42.25	0.42	37.08	1.6e-09

More explanations available

The following contains a brief explanation of all columns of the output including the columns in the extended output

Template: shows the accession numbers or name of the template sequences
Assembly: RefSeq assembly accession ID
Num: is the sequence number of accession entry in the KmerFinder database
Score: is the total number of matching Kmers between the query and the template
Expected: is the expected score, i.e.the expected total number of matching Kmers between query and template (randomly selected).
Template length: is the number of Kmers in the template
Query_Coverage [%]: is the percentage of input query Kmers that match the template.
Template_Coverage [%]: is the template coverage.
Depth: is the number of matched kmers in the query sequence divided by the total number of Kmers in the template. For read files this estimates the sequencing depth.
tot_query_Coverage [%]: is calculated based on the ratio of the score and the number of kmers in the query sequence, where the score includes kmers matched before.
tot_template_Coverage [%]: is calculated based on ratio of the score and the number of unique kmers in the template sequence, where the score includes kmers matched before.
tot_depth: depth value based on all query kmers that can be found in the template sequence .
q_value: is the quantile in a standard Pearson Chi-square test, to test whether the current template is a significant hit.
p_value: is the p-value corresponding to the obtained q_value.
Accession number: accession number of entry ID in fasta file.
Description: additional descriptions available in fasta file, or in the case of organism databases the identifier lines of fasta files.
TAXID: NCBI's TaxID number of the hit
Taxonomy: complete taxonomy of the hit
TAXID Species: NCBI's species TaxID number of the hit (sometimes bacterial strain or substrain TaxIDs can be given above)
Species: Species name

The extended output can be **downloaded** as a text-file by clicking the "RESULTS" button on the bottom of the page.

AMR genes identification

Accessing your results

Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.

To get notified by email: | t.sonda@kcri.ac.tz | Notify me via email

This page will update itself automatically.

In case the page does not reload, press here to refresh the status page.

ResFinder-4.1: Job 602268C100000E771ECA5425 finished   Inbox ×

CGEwebface@cbs.dtu.dk
to me ▾

Dear sir/madam,
    Your job 602268C100000E771ECA5425 has finished on our ResFinder-4.1 server.
    You may visit the following link to study the results:
        https://cge.cbs.dtu.dk//cgi-bin/webface.fcgi?jobid=602268C100000E771ECA5425

    The data will expire at 2021-02-16 11:49:37.

Yours sincerely,
    CGE webserver

For help and technical assistance contact cgehelp@cbs.dtu.dk

ResFinder has lots of output data

ResFinder-4.1 Server - Results

Input Files: *R0003_17_Kpneumoniae.fna*

Warning:
One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account. **Warning:** One or more resistance genes have unknown phenotypes. The species specific summary table does not take this into account.

complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
vancomycin	glycopeptide	No resistance	
mupirocin	pseudomonic acid	No resistance	
tobramycin	aminoglycoside	Resistant	aac(3)-IIa (aac(3)-IIa_X51534), aac(6')-Ib-cr (aac(6')-Ib-cr_DQ303918)
hygromycin	aminoglycoside	No resistance	
isepamicin	aminoglycoside	No resistance	
virginiamycin s	streptogramin b	No resistance	
hydrogen peroxide	peroxides	No resistance	
butirosin	aminoglycoside	No resistance	
ampicillin	beta-lactam	Resistant	blaCTX-M-15 (blaCTX-M-15_AY044436), blaTEM-1B (blaTEM-1B_AY458016)
astromicin	aminoglycoside	No resistance	
lividomycin	aminoglycoside	No resistance	
sulfamethoxazole	folate pathway antagonist	Resistant	sul2 (sul2_AY034138), sul1 (sul1_U12338)
temocillin	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	Resistant	oqxA (oqxA_EU370913), oqxB (oqxB_EU370913), dfrA27 (dfrA27_FJ459817)
oleandomycin	macrolide	No resistance	
unknown macrolide	macrolide	No resistance	
unknown fluoroquinolone	fluoroquinolone	No resistance	
florfenicol	phenicol	No resistance	
clindamycin	lincosamide	No resistance	
quinupristin	streptogramin b	No resistance	
ceftriaxone	beta-lactam	Resistant	blaCTX-M-15 (blaCTX-M-15_AY044436)
cephalothin	beta-lactam	Resistant	blaTEM-1B (blaTEM-1B_AY458016)
lincomycin	lincosamide	No resistance	
butiromycin	aminoglycoside	No resistance	

Fluoroquinolone

Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
oqxA	99.4047619048	1176/1176	1..1176	NODE_13_length_129855_cov_8.6232_ID_25	2239..3414	chloramphenicol, benzylkonium chloride,cetylpyridinium chloride,nalidixic acid,ciprofloxacin ,trimethoprim	18440636	EU370913	Must be in an operon with oqxB,phenotype differs based on genomic location of the operon PMID 25801572,also nitrofurantoin resistance PMID 26552976
aac(6')-Ib-cr	100.0	600/600	1..600	NODE_59_length_9497_cov_4.91057_ID_117	7868..8467	ciprofloxacin	unpublished	DQ303918	MIC of ciprofloxacin does not always increase above ECOFF PMID 16369542
oqxB	99.0168093879	3153/3153	1..3153	NODE_13_length_129855_cov_8.6232_ID_25	3438..6590	chloramphenicol, benzylkonium chloride,cetylpyridinium chloride,nalidixic acid,ciprofloxacin ,trimethoprim	18440636	EU370913	
qnrB6	100.0	645/645	1..645	NODE_59_length_9497_cov_4.91	320..964	ciprofloxacin	unpublished	EF523819	

| Beta-lactam									
Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
blaSHV-172	99.7677119628	861/861	1..861	NODE_3_length_307310_cov_7.97062_ID_5	146084..146944	unknown beta-lactam	unpublished	KF513177	Class A
blaCTX-M-15	100.0	876/876	1..876	NODE_55_length_11691_cov_8.28943_ID_109	3349..4224	amoxicillin,ampicillin,aztreonam,cefepime,cefotaxime,ceftazidime,ceftriaxone,piperacillin,ticarcillin	11470367, 26169409	AY044436	Class A
blaTEM-1B	100.0	861/861	1..861	NODE_55_length_11691_cov_8.28943_ID_109	7046..7906	amoxicillin,ampicillin,cephalothin,piperacillin,ticarcillin	15388431	AY458016	Class A

Conclusion

You should be able:

- To access KmeFinder & ResFinder and upload your sequences

- To access results or outputs from KmeFinder & ResFinder

- To interpret output tables from KmeFinder & ResFinder

Thank you
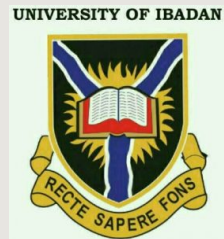
This programme is being funded by the UK Department of Health and Social Care.
The views expressed do not necessarily reflect the UK Government's official policies.