

Day 7. Bridging the gaps in bioinformatics

Annotation and AMR



- Annotation
 - Full genome annotation
 - Functional annotation
 - How do we predict?
 - Kmer
 - · Seed and extend
 - Identification of genes of interest
 - Database considerations
 - AMR
 - How to make your own database
 - Replicons







Annotation

- "The annotation process infers the structure and function of the assembled sequences. Protein-coding genes are often annotated first, but other features, such as non-coding RNAs or presence of regulatory or repetitive sequences, can also be annotated." (Dominguez Del Angel et al, 2018)
- Once we have the assembled the genome annotation is straight forward:
 - Conformation of species
 - Prediction of genes
 - Prediction of function
 - Conserved modules
 - Typing
 - Other features
- If only certain features are of interest, we can use raw sequence data directly





ATGCGCGAT

DTU

Title

5



Full genome annotation

- With Whole genome sequencing (WGS) we capture (almost) everything in the cell
 - Prokka: rapid prokaryotic genome annotation (<u>GitHub tseemann/prokka: Rapid</u> prokaryotic genome annotation)
 - ANNOVAR: Higher organisms (ANNOVAR Documentation (openbioinformatics.org))
 - NCBI-PGAP: Prokaryotic annotation (<u>NCBI Prokaryotic Genome Annotation Pipeline</u> (<u>nih.gov</u>))
 - Predictive annotation: eggNOG-mapper (eggNOG-mapper (embl.de))
- These pipelines usually generate multiple output files, which can be used for further data handling or visualization
- There are multiple visualization tools, e.g. IGV, which can be installed locally or used online.



E. Coli strain K-12 from NCBI visualized with gff in IGV (IGV)

6

Title

DTU



Data is everything

- We can produce a lot of data, but is it useful?
 - Bioinformatics relies heavily on findings produced in the wet lab
 - We cannot prove biological functions purely in silico
 - · We can make prediction based on previous findings
 - We can track changes down to the nucleotide level
 - We are limited by our databases
 - Some genes have no annotation information
 - Gene with unknown function, but recognized modules or motives
 - Pseudo genes
 - We are only interested in a small subset of information
 - · If we are interested in a given subject, we can reduce the amount of data
 - In E. coli K-12 there are 4639 predicted genes of which 3 are related to AMR
 - Specialized databases
 - · Simplifies analysis and clarifies message

DTU

ecoc

Data is everything

- Translating classic biochemical methods to bioinformatics
 - Serotyping of *E. coli* is based on testing the O- and Hantigens with antisera
 - 186 O-antigens and 53 H-flagellar antigens determines the serotype
 - Time consuming and sometimes inconclusive
- O- and H- genes have been extracted and their associated phenotype determined
 - Predicting the Serotype based on genetic profile showed
 >98% concordance with phenotype
 - Analysis can be conducted in minutes rather than hours
 - It is still only an estimate

(Joensen et al., 2015, Fratamico et al., 2016)



E. coli in scanning electron microscopic image, CDC/ Evangeline Sowers, Janice Haney Carr, 2005, Public domain image, <u>https://phil.cdc.gov/Details.aspx?pid=10042</u>

DTU

DTU



Prediction of species

- 16s rRNA gene formed the basis as the first method for sequenced based taxonomy
- Other approaches:
 - gyrB gene, rMLST, species-specific functional domain profiles
 - Shortcomings:
 - Only represents a small fraction of the entire genome
 - Today we have WGS data, better to ultilize the whole genome instead of just one gene for identification



DTU



Prediction of species - Kmerfinder

- With WGS we can use all the genetic information to predict the species
- Kmerfinder works by breaking a genome into little pieces (k-mers) and identifying the species from these pieces (k-mers)





k-mers

DTU

- A k-mer is a continuous sequence of k bases
 - e.g a certain length of DNA, RNA or protein
- k is any positive integer
- Sequences with high similarity must share k-mers
- Consider the nucleotide sequence 'ACTCCGTAACG'.



- We have this long sequence that we can cut into smaller pieces or k-mers
- We can extract all the 4-mers (substrings of length 4) in this DNA sequence





k-mers

- We can extract all the 4-mers (substrings of length 4) in this DNA sequence.
- We created a window of length 4 and slide it from left to right, shifting one character at a time.
- If the length of a given DNA sequence is N, we end up with N k+1 k-mers
- Total number of k-mers = N k + 1
- In the previous slide, the DNA sequence is 11 characters long (N = 11) and k = 4, so we get eight k-mers (11-4 + 1)





Species prediction with k-mers

- Sequences with high simlarity must share k-mers
- We can break genomes up into k-mers and compare them







Species prediction with KmerFinder





Species prediction with KmerFinder

- Genomes are spilt into 16-mers
- But all these k-mers won't be in the database
- Tool will take too long to compute
- So redundant data is reduced
- Only 16-mers with specific prefixes are kept e.g ATGAG
- Reduced redundancy speeds up the tool
- But how does the tool the compare k-mers?



Adapted from: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon



KmerFinder webtool

\$

Select database

Bacteria organisms

Upload file(s)

To input the sequences, upload a single FASTA file, or one/two FASTQ file(s), or one interleaved FASTQ file on your local disk by using the applet below. Both assembled genome (in FASTA format) and raw reads single end or paired end (in FASTQ format) are supported. Gzipped FASTA/FASTQ files are also supported.

If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking here.

Choose File(s)			
ime	Size	Progress	Status
Upload Remove			

https://cge.food.dtu.dk/services/KmerFinder/

DTU



Kmer webtool output

Hit	Score	z- score	Query Coverage [%]	Template Coverage [%]	Depth	Total Query Coverage [%]	Total Template Coverage [%]	Total Depth
Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 <u>get</u> sequence	10879	527.6	97.08	99.79	1.00	97.08	99.79	1.00
Escherichia coli, Escherichia coli NA114 g <u>et</u> sequence	16	9.3	0.14	0.15	0.00	54.87	58.92	0.60
Salmonella enterica, Salmonella enterica subsp. enterica subsp. enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence	12	6.9	0.11	0.11	0.00	6.59	6.87	0.08

Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon



Kmer webtool output

Hit	Score	Sco num	o re iber of k-r	mers te	Depth	Total Query Coverage [%]	Total Template Coverage [%]	Total Depth
Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 <u>get</u> sequence	10879	527.6	97.08	99.79	1.00	97.08	99.79	1.00
Escherichia coli, Escherichia coli NA114 <u>get</u> sequence	16	9.3	0.14	0.15	0.00	54.87	58.92	0.60
Salmonella enterica, Salmonella enterica subsp. enterica subsp. enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence	12	6.9	0.11	0.11	0.00	6.59	6.87	0.08

Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon



Kmer webtool output

Hit	Score	z- score	z-scor statistic	e cal signific	ant	Total Query Coverage [%]	Total Template Coverage [%]	Total Depth
Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence	10879	527.6	97.08	99.79	1.00	97.08	99.79	1.00
Escherichia coli, Escherichia coli NA114 <u>get</u> <u>sequence</u>	16	9.3	0.14	0.15	0.00	54.87	58.92	0.60
Salmonella enterica, Salmonella enterica subsp. enterica subsp. enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence	12	6.9	0.11	0.11	0.00	6.59	6.87	0.08

Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon



Kmer webtool output

- 00 005	Query Coverage							
Hit	Score	z- score	Query Coverage [%]	perce unkn	entage own sa	of k-mers mple that	from map to a	otal pth
Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence	10879	527.6	97.08	99.79	1.00	97.08	99.79	1.00
Escherichia coli, Escherichia coli NA114 get sequence	16	9.3	0.14	0.15	0.00	54.87	58.92	0.60
Salmonella enterica, Salmonella enterica subsp. enterica subsp. enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence	12	6.9	0.11	0.11	0.00	6.59	6.87	0.08

Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon

DTU



Kmer webtool output

				1		emplate (Coverage	
Hit	Score	z- score	Query Coverage [%]	Template Coverage [%]	pe ter	rcentage mplate ge	of k-mers nome that	from t covered by
Escherichia coli, Escherichia coli O104:H4, Escherichia coli O104:H4 str. 2011C-3493 get sequence	10879	527.6	97.08	99.79	1.00	97.08	99.79	1.00
Escherichia coli, Escherichia coli NA114 <u>get</u> <u>sequence</u>	16	9.3	0.14	0.15	0.00	54.87	58.92	0.60
Salmonella enterica, Salmonella enterica subsp. enterica subsp. enterica subsp. enterica serovar Typhimurium, Salmonella enterica subsp. enterica serovar Typhimurium str. T000240 get sequence	12	6.9	0.11	0.11	0.00	6.59	6.87	0.08

Source: Introduction to CGE tools, presentation by Pimpalas Leekitcharoenphon



Functional annotation AMR

- Attaching biological, chemical or otherwise functional information to a DNA sequence
- Often you are only interested in a limited set of genes, we will look further into AMR
- AMR is a large threat to public health
 - Carried on mobile genetic elements (MGE) -> horizontal gene transfer
 - Estimated 1.27 million people died due to AMR in 2019 and estimated up to 10 million deaths by 2050 (Murray et al., 2019)
 - Development of new drugs is slow (Norrby et al., 2005)

DTU



- AMR is conferred by different mechanisms:
 - Acquired resistance genes
 - Mutation
 - (Copy numbers)
- MGEs can transfer resistance genes between isolates closely or distantly related
- Resistance genes tend to aggregate, meaning MGEs often confer resistance to multiple classes
- May integrate into host chromosome

- Point mutations can confer resistance by various mechanisms:
 - Change the target of a drug, making the strain resistant
 - Upregulate the expression of a gene
 - Downregulate the expression of a gene
 - Change target specificity of protein
 - Usually species specific





Transfer of plasmid with resistance gene

Genetic basis of AMR

- AMR is conferred by different mechanisms:
 - Acquired resistance genes
 - Mutation

Date

DTU

- (Copy numbers)
- MGEs can transfer resistance genes between isolates closely or distantly related
- Resistance genes tend to aggregate, meaning MGEs often confer resistance to multiple classes
- May integrate into host chromosome



Transfer of plasmid with resistance gene

Note!

We also have intrinsic resistance in certain species, e.g. *Mycobacterium tuberculosis* inherently possess erm(37) protecting against macrolides, lincosamide and streptogramin



- Change the target of a drug, making the strain resistant
- Upregulate the expression of a gene
- Downregulate the expression of a gene
- Change target specificity of protein
- Usually species specific





AMR tools and databases

- There are multiple tools which all utilize their own and/or each others databases for predicting antimicrobial resistance
 - Resfinder (<u>ResFinder 4.1 (dtu.dk</u>)), AMRfinderplus (<u>Releases · ncbi/amr (github.com</u>)), CARD (<u>https://card.mcmaster.ca/home</u>), KmerResistance, ARIBA
 - Differences exists due to
 - · How the database is created and curated
 - How the database is maintained
 - · How the tool conducts its search
 - The correct tool/database will likely depend on the type of analysis or workflow you are using
 - Approach results from tools with a critical mindset!

DTU

חדח	RGI Criteria	ARO Term	SNP	Detection Criteria	AMR Gene Family	Drug Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
	Perfect	acrB		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, perum, tetracycline antibiotic, rifarnycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	100.0	100.00
#	Perfect	Escherichia coli acrA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	fluoroquinolone antibiotic, cephalosporin, glycylcycline, penam, tetracycline antibiotic, rfarnycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics	antibiotic efflux	100.0	100.00
	Perfect	Escherichia coli emrE		protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	macrolide antibiotic	antibiotic efflux	100.0	100.00
	Perfect	kdpE		protein homolog model	kdpDE	aminoglycoside antibiotic	antibiotic efflux	100.0	100.00
	Perfect	Adam		protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	nitroimidazole antibiotic	antibiotic efflux	100.0	100.00
CARD output:	Perfect	mdtG		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	phosphonic acid antibiotic	antibiotic efflux	100.0	100.00
	Perfect	mdtH		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinalone antibiotic	antibiotic efflux	100.0	100.00
Data was	Perfect	HNS		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, cephalosporin, cephamycin, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
complete	Perfect	marA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump, General Bacterial Porin with reduced permeability to beta- lactame.	fluoroquinolone antibiotic, monobactam, carbapenem, cophalosporin, glycylcydine, cephramydin, penam, tetracycline antibiotic, rifamydin antibiotic, phenicol antibiotic, penem, disinfecting agents and antiseptics	antibiotic efflux, reduced permeability to antibiotic	100.0	100.00
	Perfect	ugd		protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	100.0	100.00
Coli strain	Perfect	ndA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
	Perfect	mdtB		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
44 hits in	Perfect	mdtC		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
totall	Perfect	baeS		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
	Perfect	baoR		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic, aminocoumarin antibiotic	antibiotic efflux	100.0	100.00
	Perfect	Yoji		protein homolog model	ATP-binding cassette (ABC) antibiotic efflux pump	peptide antibiotic	antibiotic efflux	100.0	100.00
Let us take a	Perfect	PmrF		protein homolog model	pmr phosphoethanolamine transferase	peptide antibiotic	antibiotic target alteration	100.0	100.00
closer look	Perfect	errrY		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	100.0	100.00
	Perfect	елтК		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	tetracycline antibiotic	antibiotic efflux	100.0	110.26
	Perfect	evgA		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
	Perfect	evgS		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic	antibiotic efflux	100.0	100.00
	Perfect	az D		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminoglycoside antibiotic	antibiotic efflux	100.0	100.00
	Perfect	emrR		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fuoroquinolone antibiotic	antibiotic efflux	100.0	100.00
	Perfect	ептА		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00
	Perfect	enrB		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	fluoroquinolone antibiotic	antbiolic efflux	100.0	100.00
Date	DTU							Titl	e 29

EXAMPLE CARD output:

- EmrY, emrK and emrB
- Perfect hits!

DTU

 Expect for emrK, ID and COV are 100%

Filename

GCF_000005845.2_ASM584v2_genomic

 Should we expect resistance to tetracycline and fluoroquinolones in this isolate?

RGI [▲] Criteria	¢ ARO Term	\$ Detection ∲ Criteria	AMR Gene Family
Perfect	emrY	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump
Perfect	emrK	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump
Perfect	emrB	protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump
			anabioue enfox pump

Drug ∳ Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
tetracycline antibiotic	antibiotic efflux	100.0	100.00
tetracycline antibiotic	antibiotic efflux	100.0	110.26
fluoroquinolone antibiotic	antibiotic efflux	100.0	100.00



Lets try a different tool for the strain: ResFinder

No resistance at all? ٠

ResFinder-4.1	Server -	Results
---------------	----------	---------

Input Files: GCF_000005845.2_ASM584v2_genomic.fna

Warning: One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

escherichia coli com

Antimicrobial	Class	WGS-predicted phenotype	
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	



DTU Ħ

Lets try a different tool for the strain: ResFinder

- No resistance at all? ٠
- No resistance to ٠ tetracycline or quinolones?

ResFinder-4.1 Server - Results

Input Files: GCF_000005845.2_ASM584v2_genomic.fna

Warning:

One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

Antimicrobial	Class	WGS-predicted phenotype	Genetic backgro
mikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	



Lets try a different tool for the strain: ResFinder

- No resistance at all?
- No resistance to tetracycline or quinolones?
- One tool gives 44 hits, another gives 0 what is the truth?

ResFinder-4.1 Server - Results

Input Files: GCF_000005845.2_ASM584v2_genomic.fna

Warning:

One or more resistance genes does not exist in the phenotype database. The Summary table does not take this into account.

escherichia coli complete

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
mikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	No resistance	
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	No resistance	
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	No resistance	
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	No resistance	
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	No resistance	
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	No resistance	
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	No resistance	
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	





Differences in output example

- The strain run in this example is a standard laboratory strain E. coli K-12 substrain MG1655
- It is not expected to have any phenotypic resistance to tetracycline (Zhang et al., 2022)
 - Not actually expected to have any particular phenotypic resistance different from wild-type
 e. coli
- If run on AMR finderplus, no resistance genes are found either.
- Approach databases with care and select based on your scope
 - How does results translate to the laboratory, genotypic =/= phenotypic
 - How much expertise is demanded to utilize findings
 - What is the aim of your analysis

Resfinder

https://cge.food.dtu.dk/services/ResFinder/

 The tool has (3) databases:

Refinder is a tool

and database for detection of genes and point mutations conferring AMR.

DTU

•

- Poinfinder db
- Resfinder_db
- (DesinFinder_db)

Databases with antimicrobial resistance genes and chromosomal point mutations

ResFinder database

PointFinder database

https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/

https://bitbucket.org/genomicepidemiology/pointfinder_db/src/master/

Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo



Resfinder

The tool and • databases are freely available github

DTU

- The main page ٠ contains instructions for installation of the tool and how to download and install the associated databases
- We will be using this tool today

Date

EARS-Net	Hadea Seq4AMR I To	pols and sites 📙 References 🚯 Task 3.b Report on t 🚯 Task 3.c d	WGS met XII Overview of activit	🚯 Templates_Ha	sDEA2 🐂 myCWT Mine Rejser 📕 New folder	
Bitbucket					Q Search	
resfinder	db Geno res	omic Epidemiology / Databases sfinder_db				Clone
<> Source	2s	9 master • Files • Filter files Q				
¢ Commits						
Branches	Nam	/ ne	Size	Last commit	Message	
Pull reque	sts	.gitignore	37 B	2018-12-14	Add install script to install database for KMA indexing	
Deployme	nts 🕞	CHECK-entries.sh	2.33 KB	2019-01-23	CHECK-entries: make sure to escape regex chars	
Issues	Ð	INSTALLpy	3.79 KB	2020-04-24	Fixed version of KMA	
Jira issues	Ð	README.md	5.37 KB	2021-04-20	Added hsitory file to content overview	
Security	Ð	aminoglycoside.fsa	196.86 KB	2022-04-21	Adds genes dfrE and bleO	
Download	s 🕒	antibiotic_classes.txt	2.71 KB	2022-04-21	Adds genes dfrE and bleO	
	₽	beta-lactam.fsa	1.78 MB	2022-02-03	delete duplicates inside same fsa file	
	₽	colistin.fsa	91.6 KB	2021-03-11	added gar1.fosi1,erm50.qnrb89.catt.qnrb91.aac6.qnrb90.mcr126.mcr127	
	Ð	config	912 B	2021-03-09	added aac(3)-lla_6_CP023555, blaCMY-150_2_NG_060513, blaCARB-4_1_U14749, mupA_1_X75439, mupA_2_GU23	7136
	Ð	disinfectant.fsa	24.15 KB	2021-02-19	added disinfectant db	
	Ð	fosfomycin.fsa	18.68 KB	2021-03-11	added gar1.fosl1,erm50,qnrb89.catt.qnrb91,aac6,qnrb90,mcr126,mcr127	
	Ð	fusidicacid.fsa	1.96 KB	2019-02-20	Update fusidic acid db	
		Source: Geno	typic detection o	f AMR in S	almonella, presentation by Ana Rita Rebelo	



Resfinder

- The databases contain:
 - Nucleotide references for AMR elements
 - Curation for each genes
 - Phenotypes associated with specific species

• We can get:

- Class of AMR gene
- Specific phenotypes
- Supporting sources
- Mechanism
- Other notes

	phenotype_panels.txt	2.55 KB 202	21-10-06
	phenotypes.txt	502.55 KB 202	2-04-21
Genomic Er pheno	Epidemiology / Databases / resfinder_db otypes.txt	Pull reque	sts Check out
Source	✓ ŷ master ✓ ∮ d504dca ✓ Full commit		
C res	sfinder_db / phenotypes.txt 📙		Edit •••
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	<pre>Dem_gattersJoint: Class Print Prior Prior Prior Period President Prior Period President Prior Period Prior Period Prior Period Per</pre>	<pre></pre>	Ih-əac(6')-Iid

Source: Genotypic detection of AMR in Salmonella, presentation by Ana Rita Rebelo



DTU =

Resfinder

- The tool is also available ٠ online, where isolates can uploaded
- The stand-alone tool utilize ٠ the same inputs, as we wil find out in the exercises
- Select the databases to us ٠
- Choose thresholds for ID a • COV
- Provide species ٠
- Specify input format ٠

Date

DTU

Give input files ٠

ResFinder identifies acquired genes and/or finds chro in total or partial DNA sequence of bacteria. ResFinder and PointFinder software: (2022-03-10) ResFinder database: (2022-02-04) PointFinder database: (2021-02-01) For analysis part of EFSA, go to ResFinder-EFSA	mosomal mutations mediating antimicrobial resistance	The database is curated by: Frank Meller Aarestrup (dick to contact)	Chromosomal poin Select thresho	nt mutations 🗹 old for %ID	
Chromosomal point mutations			90 %	•	~
Acquired antimicrobial resistance genes 🗆			Select minime	um length	~
Select species Campylobacter spp.* Compromotional point musicion database exists			Show unkno	own mutations, not found i	n the database
Select species Campylobacter spp.* Commonwel poor mucture statemer estate Select type of your reads Assembled Genome/Contigs			Show unkno	wn mutations, not found i	n the database
Select species Campylobacter spp.* Commonwel poor mustor database estas Select type of your reads Assembled Genome/Contigs Type prior Wates bridden. Env. 407: More sive the start of the web adves: Michaese File(s)	i https and net just http: Fir it by clinking here		Show unkno	wn mutations, not found i	n the database
Select species Campylobacter spp.* Campylobacter spp.* Select type of your reads Select type of your reads Assembled Genome/Contigs flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the sort of the web adves a flyeu get or 'Acres bridden. Error 407: More sure the sort of the sort o	i https: and net just http: first by citating here Size Progress	Status	Show unkno	wn mutations, not found i	n the database
Select species Campylobacter spp.* 'Oromanout giver matche database edus Select type of your reads Assembled Genome/Contigs Type pro * Keres finlades for 407: Male sue the sort of the web abres: More of the coses File(s) Name Subset (Coses File(s)) Name Subset (Coses File(s)) Name Subset (Coses File(s)) Name Subset (Coses File(s)) Name Subset (Coses File(s)) Subset (Coses File(s)) S	r Hages and Hear Just Hags First by eliciting News Size Progress	Status	Show unkno	wn mutations, not found i	n the database

38



Resfinder

- The tool is also available online, where isolates can be uploaded
- The stand-alone tool utilize the same inputs, as we will find out in the exercises
- Select the databases to use
- Choose thresholds for ID and COV
- Provide species
- Specify input format
- Give input files
- For AMR genes, specific class can be chosen

	lesFinder 4.1		
	Service Instructions Output Article abstract Citations Overview of genes Database history	1	
e	ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria. ResFinder database: (2022-02-04) PointFinder database: (2021-02-01) For analysis part of FFSA, on to BerEinder/FFSA	The database is curated by: Frank Moller Aarestrup (dick to contact)	
	Chromosomal point mutations 🗆		Acquired antimicrobial resistance genes Select Antimicrobial configuration
	Acquired antimicrobial resistance genes 🗆		Select multiple items, with Ctri-Click (or Cmd-Click on Mac) - as default all databases are selected Aminoglycoside Beta-lactam Collictin
d	Select species Campylobacter spp.* "Orenegamed joint material distribute exists		Disinfectant Fluoroquinolone Fosfomycin
	Select type of your reads Assembled Genome/Contigs		Select threshold for %ID
	If you get an 'Access forbidden. Error 403': Make sure the start of the web adress is https: and not just http: Fir it by clicking New		90 %
	H ⊂voorse Fiel(s) Name Size Progress	Status	Select minimum length
	© Ujsload and Bernove		
SS			
	Confidentiality: The sequences are kept confidential and will be deleted after 48 hours.	Source: Genotyp	ic detection of AMR in Salmonella, presentation by Ana Rita Rebelo
			Тир 20



Resfinder

• Species are neccesary for point mutations

- Not all species are present
- "Other" allows us to search for • AMR genes

esFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance total or partial DNA sequence of bacteria.	The database is curated by: Frank Moller Aarestrup	
esFinder and PointFinder software: (2022-03-10) esFinder database: (2022-02-04) ointFinder database: (2021-02-01)	(click to contact)	
or analysis part of EFSA, go to ResFinder-EFSA		
hromesomal point mutations		Solart sparing
		Computebactor con *
cquired antimicrobial resistance genes 🗆		Campylobacter spp."
		Campylobacter spp."
		Campylobacter Jejunin
Campylobacter spp."		Campyiobacter coll*
Promotomol point mutation databage exists		Escherichia colin
elect type of your reads		Salmonella spp.*
Assembled Genome/Contigs		Plasmodium falciparum*
		Neisseria gonorrhoeae*
you get an "Access paralables, prior 442". Have sure the part of the web bankss is right and hot just righ, his it by culosing here.		Mycobacterium tuberculosis*
Ht Choose File(s)		Enterococcus faecalis*
Name Size Progress	Status	Enterococcus faecium*
		Klebsiella*
O Upload Remove		Helicobacter pylori*
		Staphylococcus aureus*
		Other
		*Chromosomal point mutation database
		Chroniosoniai point mutation uatabase
ionfidentiality:		and the



ID and **COV**

- What is identity (ID)?
 - Proportion of matching nucleotides:



9/10 bp align = 90% ID

- What is coverage/length (COV)?
 - Proportion of nucleotides covered:



7/10 bp covered = 70% COV, 6/10 bp align = 60% ID

Note on databases

- Make your own database!
 - MyDBFinder (https://cge.food.dtu.dk/services/MyKMAfinder/)
 - MyKMAFinder (https://cge.food.dtu.dk/services/MyDbFinder/)
 - NCBI-Blast (BLAST+ executables BLASTHelp documentation (nih.gov))
- · Organize genes of interest into a fasta file
 - Upload to online solution
 - Make blast database with makeblastdb (part of standalone blast tools)
- Large range of customization
- Better options likely exist for larger datasets

DTU



Acknowledgements

The creation of this training material was commissioned by ECDC to Statens Serum Institut (SSI) and produced by The National Food Institute at the Technical University of Denmark (DTU) with the direct involvement of Lauge Holm Sørensen and Niamh Lacy-Roberts



References

- Dominguez Del Angel V, Hjerde E, Sterck L et al. Ten steps to get started in Genome Assembly and Annotation [version 1; peer review: 2 approved]. F1000Research 2018, 7(ELIXIR):148 (https://doi.org/10.12688/f1000research.13598.1)
- Murray, Christopher J. L., et al. "Global Burden of Bacterial Antimicrobial Resistance in 2019: a Systematic Analysis." Lancet, vol. 399, no. 10325, Elsevier B.V., 2022, pp. 629–55, doi:10.1016/S0140-6736(21)02724-0.
- Norrby, S. Ragnar, et al. "Lack of Development of New Antimicrobial Drugs: A Potential Serious Threat to Public Health." Lancet Infectious Diseases, vol. 5, no. 2, Lancet Publishing Group, 2005, pp. 115–19, doi:10.1016/S1473-3099(05)70086-4.
- Zhang, P., Mao, D., Gao, H. et al. Colonization of gut microbiota by plasmid-carrying bacteria is facilitated by evolutionary adaptation to antibiotic treatment. ISME J 16, 1284–1293 (2022). https://doi.org/10.1038/s41396-021-01171-x
- Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in Molecular Serotyping and Subtyping of Escherichia coli. Front Microbiol. 2016 May 3;7:644. doi: 10.3389/fmicb.2016.00644. PMID: 27199968; PMCID: PMC4853403.
- Joensen, Katrine Grimstrup, et al. "Rapid and Easy In Silico Serotyping of Escherichia Coli Isolates by Use of Whole-Genome Sequencing Data." Journal of Clinical Microbiology, edited by K. C. Carroll, vol. 53, no. 8, American Society for Microbiology, 2015, pp. 2410–26, doi:10.1128/JCM.00008-15.

DTU