Functional traces of antibiotic resistance in metagenomes

Meeting Toscano di Bioinformatica e Systems Biology

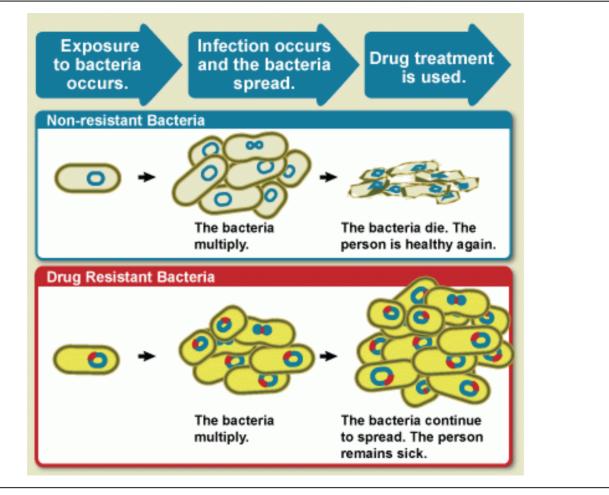
Bioinformatiha 11

14 Ottobre 2024 – Plesso Didattico Morgagni

Matteo Ramazzotti – matteo.ramazzotti@unifi.it

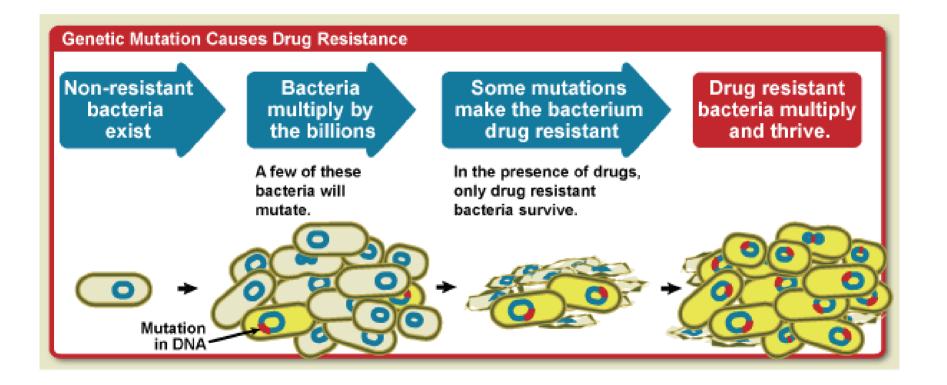


Understanding antibiotic resistance



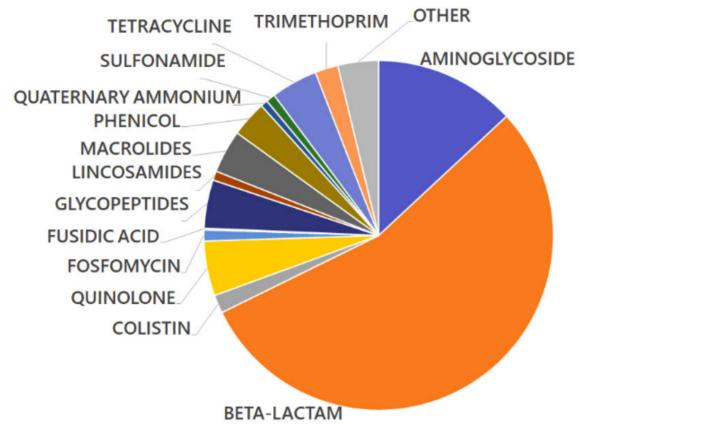
niler



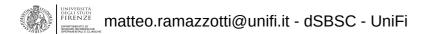


www.niaid.nih.gov

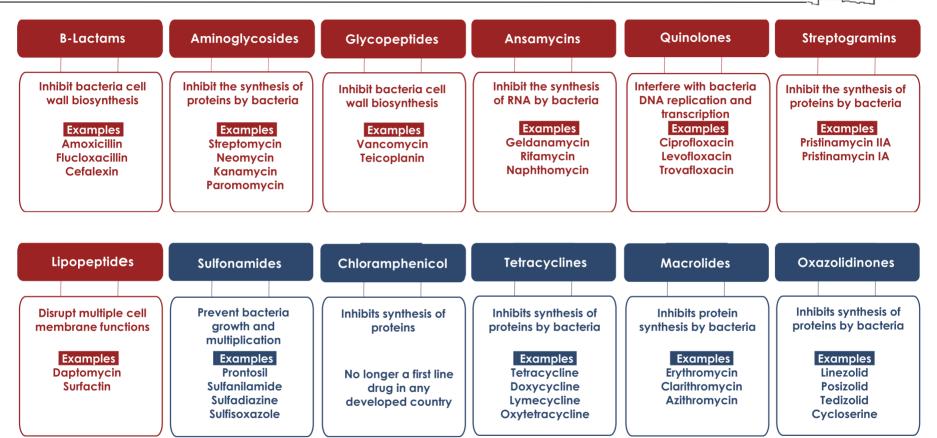




https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/



Major classes of antibiotics

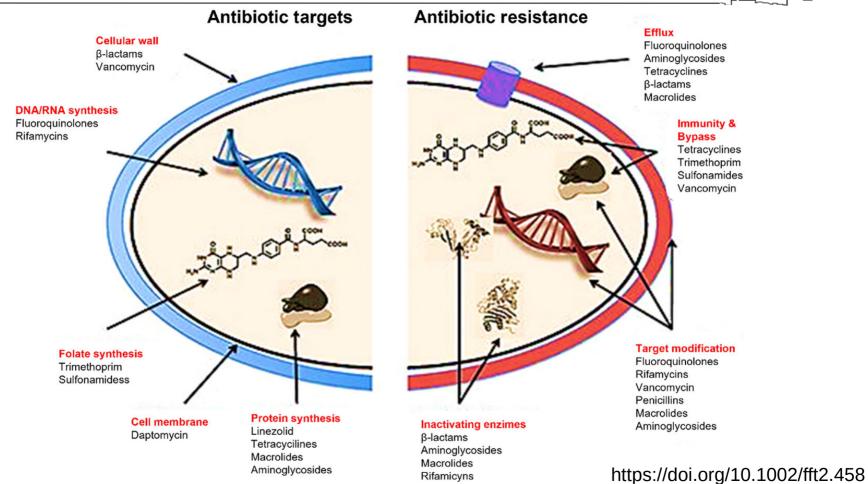


Commonly act as bactericidal agents, causing bacterial cell death 🛛 🛑 Commonly act as bacteriostatic agents, restrict growth & multiplication

https://www.reactgroup.org/toolbox/understand/antibiotics/how-do-antibiotics-work/



Mechanisms of antibiotic resistance





- Mutations
- Gene acquisition
- Membrane impermeability
- Alteration of the binding site
- Enzymatic transformation vph of of S. vinaceus Viomycin kinase P18623
- Efflux pumps NorA of S. aureus Quinolone resistance protein P0A0J7

I mar

Searching for antibiotic resistance in metagenomes

- Identify the candidate protein reference
- Search in Uniprot for similar proteins
- by name?
- by sequence?
- uniprot defined?
- Align all candidates using a multiple sequence aligner (e.g. Clustal)
- Is the alignment reasonable? If not try **removing disturbing proteins**
- **Repeat** until the **alignment** is **optimized**
- Build an Hidden Markov Model of the alignment

Now we have a mathematical model to search in genomes - or reads !!!



Viomycin is a member of the tuberactinomycin family, a group of nonribosomal **peptide antibiotics** exhibiting **anti-tuberculosis** activity

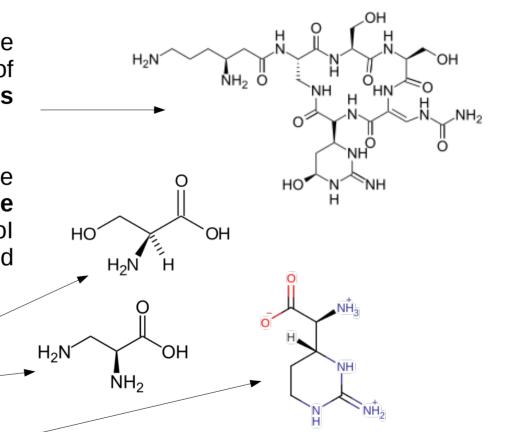
It consists of a central cyclic pentapeptide produced by a **nonribosomal peptide synthetase** composed by VioA, VioF, VioI and VioG genes that condense and cyclize

2 x L-2,3-diaminopropionate (L-Dap)

2 x L-serine (L-Ser)

1 x (2S,3R)-capreomycidine (L-Cam)

https://en.wikipedia.org/wiki/Viomycin



- VioA: NRPS (A-PCP-C-A-PCP-C)
- VioH: Type II thioesterase
- VioO: NRPS (A-PCP)-β-lysine activation
- VioB: 2,3-diaminopropionate synthase
- Viol: NRPS (PCP-C)
- VioP: Lysine 2,3-aminomutase
- VioC: L-Arg hydroxylase
- VIoJ: 2,3-diaminopropionyl α , β -desaturase
- vph: Viomycin phosphotransferase
- VioD: Capreomycidine synthase
- VioK: Ornithine cyclodeaminase
- VioQ: Capreomycidine hydroxylase
- VioE: Permease
- VioL: Carbamoyltransferase

In Streptomyces puniceus

the viomycin gene cluster includes

36.3 kb of contiguous DNA that encodes

20 open reading frames (ORFs) that

are involved in the

biosynthesis, regulation and activation

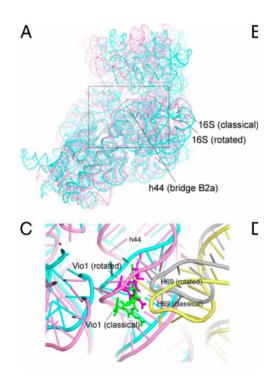
of viomycin

https://en.wikipedia.org/wiki/Viomycin



Viomycin blocks protein synthesis

Viomycin blocks the translocation step of protein synthesis, stabilizing the ribosome in a state of intersubunit rotation, resembling the hybrid-state intermediate.



Viomycin is toxic for Streptomyces ribosomes, so it is: 1. phosphorylated (Vph) - inactivation

2. exported (VioE)

3. dephosphorylated (VioS) - reactivation

The presence of Viomycin phosphotransferase homologues in bacteria may indicate potential drug resistance



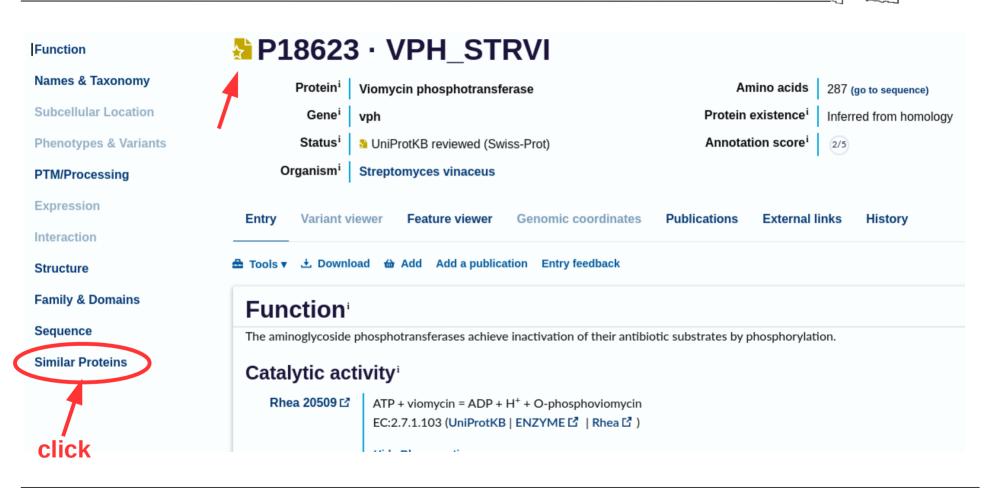
Protein sequences: viomycin phosphotrasferase



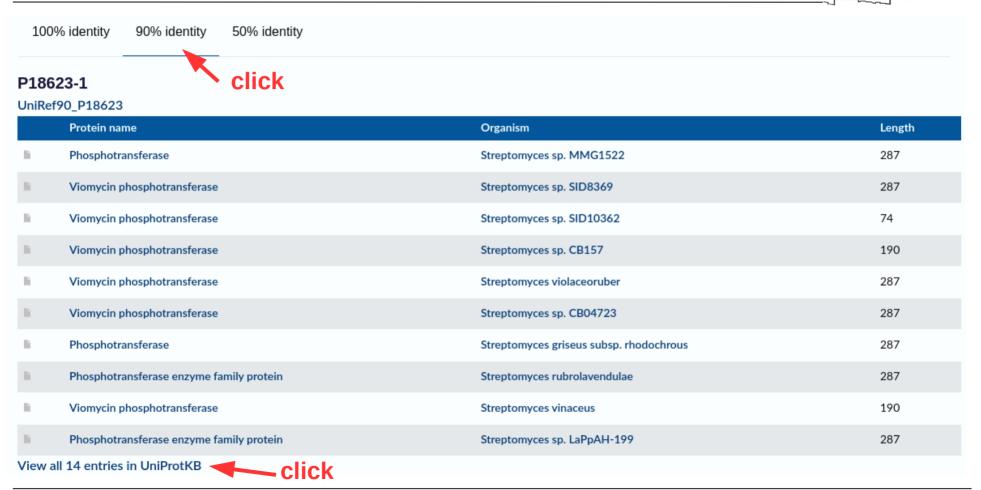
Status	U	niProt	KB 70 result	S		
 Reviewed (Swiss-Prot) (1) Unreviewed (TrEMBL) (69) 	a 1	Γools▼ 土 Down	load (70) 命 Add View:	Cards O Table •	🖉 Customize columns 🛛 📽 Share 🔻	
Taxonomy	X	Entry 🔺	Entry Name 🔺	Gene Names 🔺	Protein Names 🔺	Organism 🔺
Filter by taxonomy		P18623	VPH_STRVI	vph	Viomycin phosphotransferase[]	Streptomyces vinaceus
Group by		D3K8V4	D3K8V4_STRGR	vph	Viomycin phosphotransferase	Streptomyces griseus
Taxonomy		D3K8V1	D3K8V1_9ACTN	vph	Viomycin phosphotransferase	Streptomyces sp. CB157
Keywords		D3K8W2	D3K8W2_9ACTN	vph	Viomycin phosphotransferase	Streptomyces griseobrunneus
Gene Ontology Enzyme Class		A0A1M4DY32	A0A1M4DY32_9ACTN	BN4615_P997	Viomycin phosphotransferase (Viomycin kinase)[]	Nonomuraea gerenzanensis
Proteins with		D3K8W5	D3K8W5_9ACTN	vph	Viomycin phosphotransferase	Streptomyces roseochromogenus
3D structure (1)		D3K8V0	D3K8V0_9ACTN	vph	Viomycin phosphotransferase	Streptomyces californicus
Active site (1)		D3K8V2	D3K8V2_9ACTN	vph	Viomycin phosphotransferase	Streptomyces californicus
Binding site (1)		D3K8V3	D3K8V3_STRGR	vph	Viomycin phosphotransferase	Streptomyces griseus
Catalytic activity (1)		D3K8W3	D3K8W3_9ACTN	vph	Viomycin phosphotransferase	Streptomyces bacillaris
Chain (1) More items		D3K8W4	D3K8W4_9ACTN	vph	Viomycin phosphotransferase	Kitasatospora mediocidica



The uniprot reference is in Streptomyces vinaceus



The most similar proteins can be used to build a model





UniProtKB 15 results

🚔 Tools 🔻 📩 Download (15) 🛍 Add 🛛 View: Cards 🔿 Table 💿 💆 Customize columns 👒 Share 🔻 15 rows selected

	BLAST (15)	Entry Name 🔺	Gene Names 🔺	Protein Names 🔺	Organism 🔺
	Align (15)	VPH_STRVI	vph	Viomycin phosphotransferase[]	Streptomyces vinaceus
CIICK -	Map IDs (15)	1 A0A0M8SW11_9ACTN	ADK56_29900	Phosphotransferase	Streptomyces sp. MMG1522
	A0A6G2W9k	7 A0A6G2W9K7_9ACTN	vph , GTY38_08995	Viomycin phosphotransferase	Streptomyces sp. SID8369
	A0A6B2RDI3	A0A6B2RDI3_9ACTN	G3I36_21830	Viomycin phosphotransferase	Streptomyces sp. SID10362
	D3K8V1	D3K8V1_9ACTN	vph	Viomycin phosphotransferase	Streptomyces sp. CB157
	A0A1D8G7S	5 A0A1D8G7S5_9ACTN	A4G23_04374	Phosphotransferase enzyme family protein	Streptomyces rubrolavendulae
	D3K8U9	D3K8U9_STRVI	vph	Viomycin phosphotransferase	Streptomyces vinaceus
		2 A0A1V0UK02_STRVN	B1H20_31865	Viomycin phosphotransferase	Streptomyces violaceoruber
	A0A7H8ZKS	A0A7H8ZKS9_9ACTN	vph , HXS80_02400	Viomycin phosphotransferase	Streptomyces sp. CB04723
		A0AA89JTE8_STRGR	ADK33_27280	Phosphotransferase	Streptomyces griseus subsp. rhodochrous
	A0A1G6NDG	2 A0A1G6NDG2_9ACTN	F610DRAFT_02420	Phosphotransferase enzyme fam	
	A0A6G3PQN	2 A0A6G3PQN2_9ACTN	vph , G3l27_23705	Viomycin phosphotransferase	eset Align 15 sequences
			ho novt n		

on the next page, click -



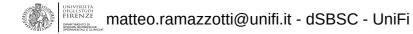
the reflect

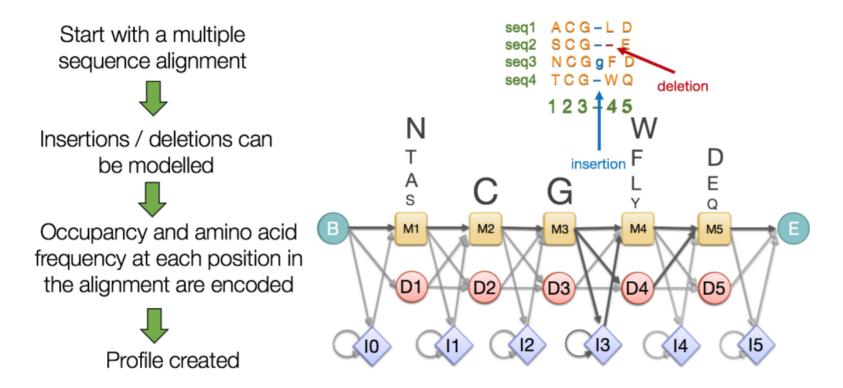
The model must be carefully observed and refined

					Name			Created	Status	
	Align results				sp P18623 VPH_	_STRVI +14	_	2024-10-13	Completed	
	Overview Trees Percent Id		kt Output Inp	ut Parameters	s API Request		А	fter con	pletion	
check to add	A	Resubmit View: O C	Overview 🖲 Wrap	oped						
	tr A0A1D8G7S5 A0A1D8G7S5_9ACTN tr A0A7H8ZKS9 A0A7H8ZKS9_9ACTN tr A0A66B2RDI3 A0A6B2RDI3_9ACTN tr A0A662W9K7 A0A6G2W9K7_9ACTN tr A0A1G6NDG2 A0A1G6NDG2_9ACTN tr D3K8V2 D3K8V2_9ACTN tr D3K8V2 D3K8V2_9ACTN tr A0A663PQN2 A0A463PUT8_STRGR tr A0A663PQN2 A0A6G3PQN2_9ACTN tr A0A90M8SW11 A0A0M8SW11_9ACTN tr A0A90M8SW11 A0A0M8SW11_9ACTN tr D3K8U9 D3K8U9_STRVI tr D3K8U9 D3K8U9_STRVI tr D3K8U9 D3K8U9_STRVI	MG MR MR MR MR MR MR MR MR	I E T H R D L I E T H R D L I E T H R D L I E T H R D L T E T H R D L I E T H R D L	L S R L L P L S R L L P	GDTVGGLA GDTVGGLA GDTVGGLA GDTVGGLA GDTVGGLA GDTVGGLA GDTVGGLA	VHEGQFHH VHEGQFHH VHEGQFHH VHEGQFHH VHEGQFHH VHEGQFHH VHEGQFHH VHEGQFHH		G S OR V V O G S H R V V O	CFART 4 CFART 4 CFART 4 CFART 1 CFART 4 CFART 1 CFART 1 CFART 1	16 23 16 14 16 14 16 16 16 16 16 16 16 16 16 16 16 16 16
	A0A1D8G7S5:Domain		1		1					

they are all good: all in !!

the alignment continues down here





The probabilistic **model** considers **observed frequencies** of the amino acids in each position as well as **their transitions** from the observed occupancy of each position in a multiple sequence alignment.

www.ebi.ac.uk



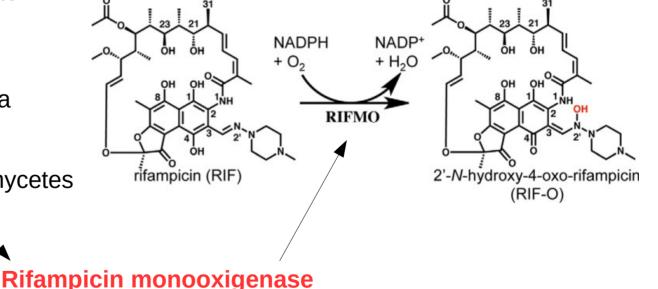
matteo.ramazzotti@unifi.it - dSBSC - UniFi

Deactivation by modification

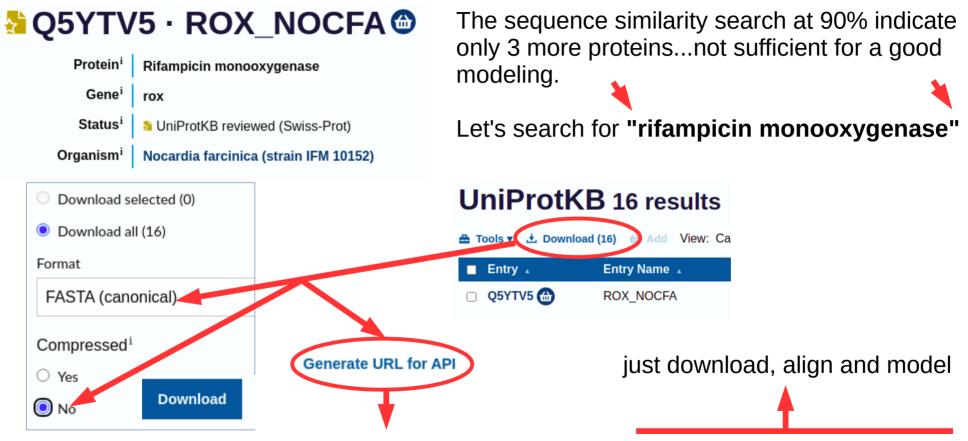
Rifampicin is a potent frontline antibiotic against tuberculosis and other mycobacterial infections. It is produced in *Amycolaptosis rifamycinica* and then modified in vitro. It acst by **blocking RNA Polymerase**

Two mechanisms are known to deactivate it:

- point mutations in RNA polymerase in mycobacteria
- enzyme-mediated inactivation in soil actinomycetes

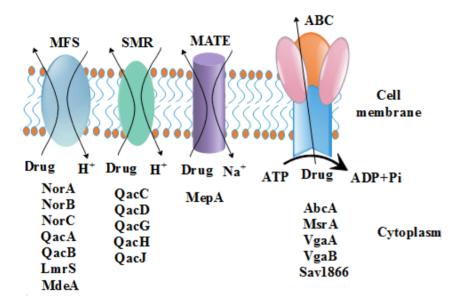


The uniprot reference is in Nocardia farcinica



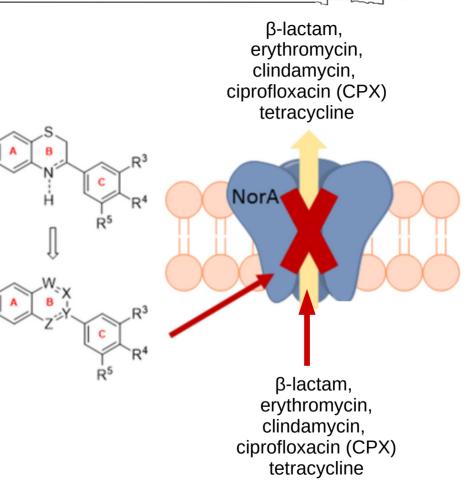
wget -O rifampicin.fasta "https://rest.uniprot.org/uniprotkb/stream?format=fasta&query=%28%22rifampicin+monooxygenase%22%29"

Deactivation by extrusion with efflux pumps



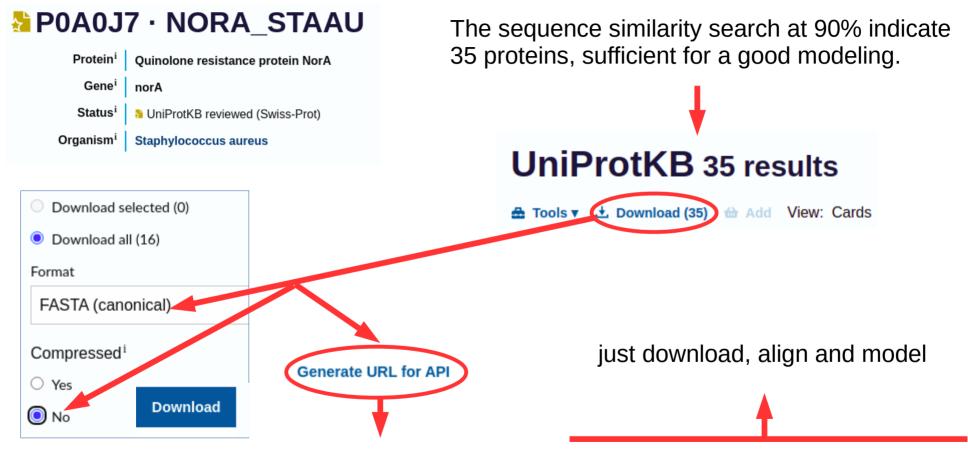
Efflux pumps overexpression is a major facilitator of multidrug resistance.

Active research is ongoing to block such aspecific pumps



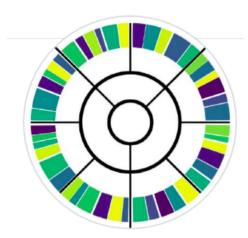


The uniprot reference is in Staphylococcus aureus



wget -O quinolone.fasta "https://rest.uniprot.org/uniprotkb/stream?format=fasta&query=%28uniref_cluster_90%3AUniRef90_P0A0J4%29

CARD



Comprehensive Antibiotic Resistance Database

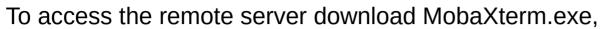
Species	New Isolates	Total Isolates
Salmonella enterica	<u>606</u>	<u>677,770</u>
E.coli and Shigella	27	<u>428,411</u>
<u>Campylobacter jejuni</u>	<u>24</u>	<u>129,117</u>
Listeria monocytogenes	22	<u>69,389</u>

See more organisms...

https://doi.org/10.1128/aac.00419-13

https://www.ncbi.nlm.nih.gov/pathogens/





 click on new session/ssh ip address: 150.217.159.17 username: bioinfo11

- click connect

password: #14ott24# --- you won't see anything during password typing

Once in type Is [enter]

You'll find the following folders:

reads : paired-end reads of 3+3 samples
 data : processed reads – assembled and transformd to potential proteins
 tax: processed reads – taxonimized and quantified

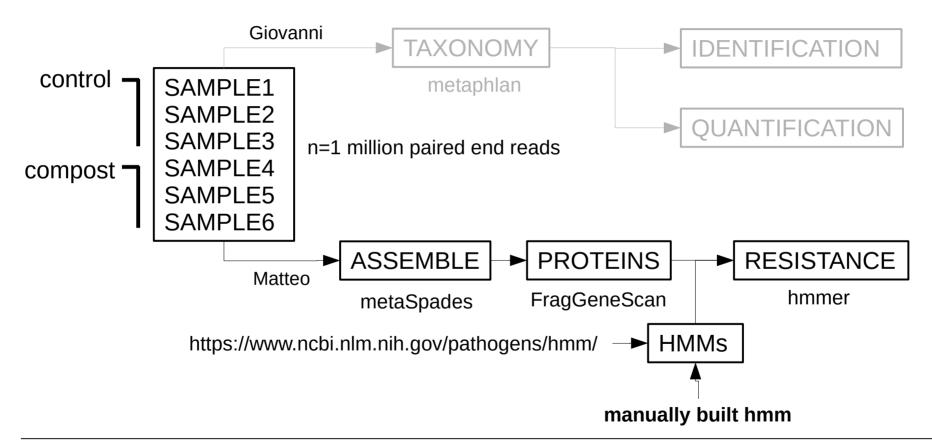
stud01 - 20 : once in the computer we will us assign you a number N

- cd studN this will move you into the folder N
- stay there = never type "cd .."



Lab workflow

data from SRR7414903 - PMC9541739





Input dataset - a matter of antibiotic resistance

Environ Microbiol. 2022 Aug; 24(8): 3705–3721.

Published online 2022 May 18. doi: <u>10.1111/1462-2920.16022</u>

PMCID: PMC9541739 PMID: 35466491

Metagenomic tracking of antibiotic resistance genes through a preharvest vegetable production system: an integrated lab-, microcosm- and greenhouse-scale analysis

Ishi Keenum, ¹ Lauren Wind, ² Partha Ray, ³ Giselle Guron, ⁴ Chaoqi Chen, ⁵ Katharine Knowlton, ⁶ Monica Ponder, ⁴ and Amy Pruden^{II}

Prior research demonstrated the potential for **agricultural production systems** to **contribute to the environmental spread of antibiotic resistance genes** (ARGs)

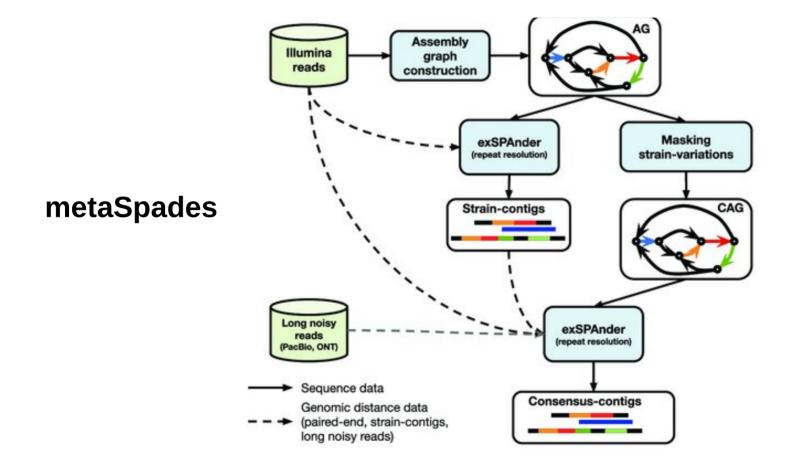
1. control compost x 3
 2. compost with antibiotics x 3

3. control manure

4. manure with antibiotics

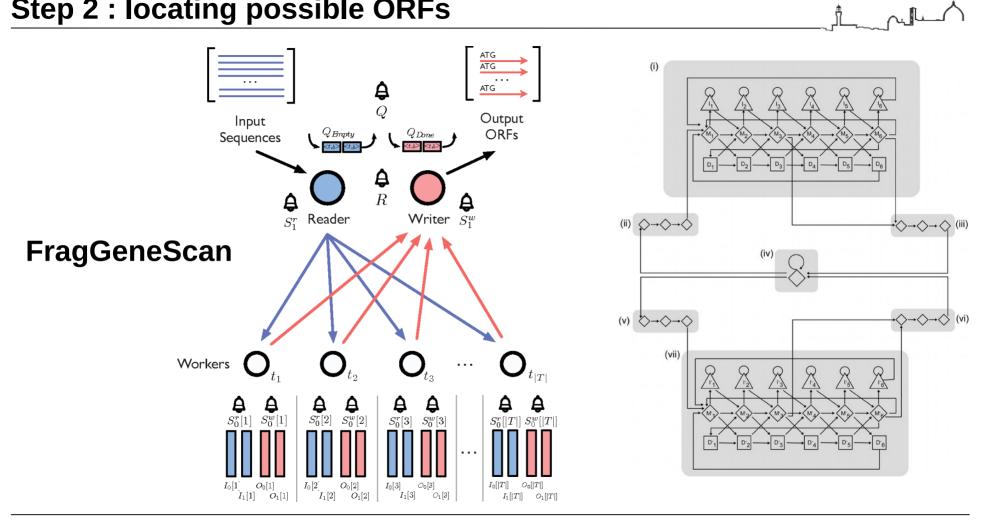


Step 1 : from short reads to scaffolds and contigs





Step 2 : locating possible ORFs





Viomycin phosphotrasferase HMM preparation protocol

go to www.uniprot.org

observe	 # viomycin resistance has been studied in Streptomyces vinaceus # go to https://www.uniprot.org/uniprotkb/P18623/entry # last section is "similar proteins" 						
compile	<pre># click on 90%identity # we see the preformed uniref_cluster_90:UniRef90_P18623 with # click on View all # thick the selector to select all proteins # click on tools -> align</pre>						
refine	# we notice a warning, indicating to remove some proteins because they are identical # go back and remove protein 8(A0A1V0UK02),11(A0A1G6NDG2),12(A0A6G3PQN2),13(D3K8V0) # click on tools -> align. No more warning, we can go on with 11 proteins						
download	# paste the copied text in the terminal						
align	<pre># type ctrl-c # now you should have a file names viomycin.fasta in your space # let's realign sequences using a local cluslalo # clustalo -i viomycin.fasta -o viomycin.aln # inspect the alignment by typing</pre>						
build	 #/bin/alan viomycin.aln (press q to exit) # build the hidden markov model # hmmbuild -n viomycin viomycin.hmm viomycin.aln # stamp the model in binary format (makes calculatio faster) # hmmpress viomycin.hmm 	scan	for i in \$(ls/assembled/sample*/scaffolds.fasta.prot.faa) do hmmsearch rifampicin.hmm \$i grep -c ">>" done				



مرالب

go to www.uniprot.org

observe	 # rifampicin resistance has been studied in Nocardia farcinica # go to https://www.uniprot.org/uniprotkb/Q5YTV5/entry # using the strategy seen before (90% similarity) just returns 4 proteins form different strain of the same genus # we can then switch to the query "rifampicin monooxygenase": we obtain 16 valid proteins, with no redundancy
download	 # select all entries and click Download: this time click on "Generate URL for API" # this returns a link that can be used to directly download proteins to our terminal wget -O rifampicin.fasta "https://rest.uniprot.org/uniprotkb/stream?format=fasta&query=%28%22rifampicin+monooxygenase%22%29" # we now have all 16 proteins in our space, ready to be aligned and checked
align	clustalo -i rifampicin.fasta -o rifampicin.aln /bin/alan viomycin.aln # it seems consistent, let's then build the hidden Markov model
-	hmmbuild -n rifampicin rifampicin.hmm rifampicin.aln hmmpress rifampicin.hmm
scan	for i in \$(ls/assembled/sample*/scaffolds.fasta.prot.faa) do hmmsearch rifampicin.hmm \$i grep -c ">>" done



مرالب

go to www.uniprot.org

observe	 # quinolone resistance has been studied in Staphilococcus aureus # go to https://www.uniprot.org/uniprotkb/P0A0J7/entry # using the strategy seen before (90% similarity) returns 36 proteins form different strain of the same genus
download	# select all entries and click Download : click on " Generate URL for API" # this returns a link that can be used to directly download proteins to our terminal wget -O qinolone.fasta "https://rest.uniprot.org/uniprotkb/stream?format=fasta&query=%28uniref_cluster_90%3AUniRef90_P0A0J4%29 # we now have all 16 proteins in our space, ready to be aligned and checked
align	clustalo -i quinolone.fasta -o quinolone.aln /bin/alan quinolone.aln # it seems consistent, let's then build the hidden Markov model
build	hmmbuild -n quinolne quinolone.hmm quinolone.aln hmmpress qinolone.hmm
scan	for i in \$(ls/assembled/sample*/scaffolds.fasta.prot.faa) do hmmsearch qinolone.hmm \$i grep -c ">>" done



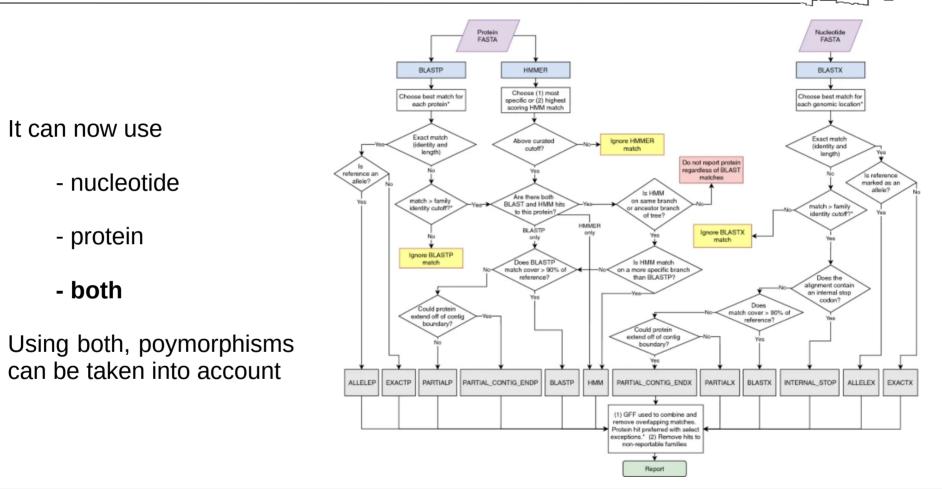
 $1 \sim 10^{-1}$

```
hmms from the NCBI models
                                                       and reports the model name (so
#from user's home
for j in $(ls ../hmm/db/hmm_NCBI/HMM/)
                                                       the resistance type) and the
do
                                                       counts of scaffolds potentially
cp ../hmm/db/hmm_NCBI/HMM/$j tmp
                                                       containing such resistance.
hmmpress -f tmp
cat tmp | grep -B1 DESC
for i in $(ls ../assembled/sample*/scaffolds.fasta.prot.faa)
do
echo $i | perl -ne '/.+?(sample\d+_\w+)\//;print $1,": "'
hmmsearch tmp $i | grep -c ">>"
done
rm tmp*
                                                       You'll see that samples 1,2,3
read
                                                       (untreated controls) are usually
done
                                                       less affected by resistance than
                                                       4.5.6 (treated with antibiotics)
```

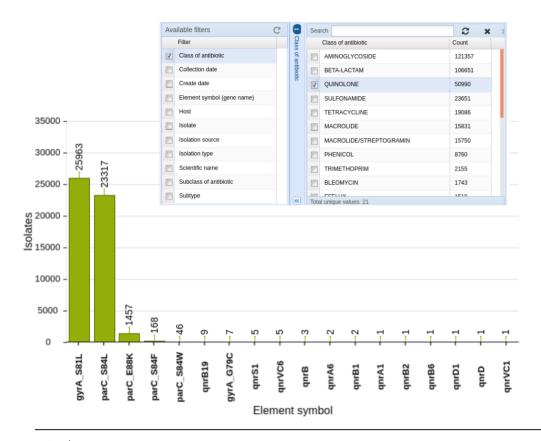


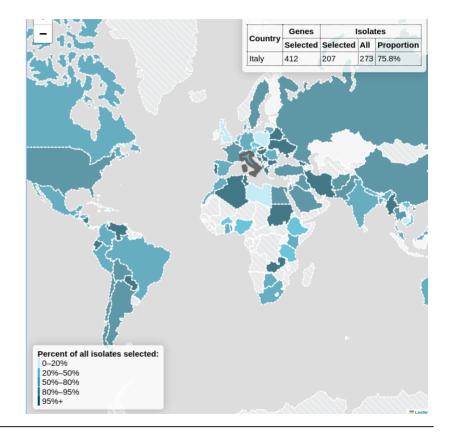
This bash script streams all

The latest tool: NCBI AMRFinderPlus



https://www.ncbi.nlm.nih.gov/pathogens/microbigge_map/







1_ ~_l_