Week 3

SPECIES IDENTIFICATION

K-MER FINDER TOOLS & APPLICATION

16s   SMALL-PIECE TO
      REFER TO SPECIE

BREAKING GENOME IN MULTIPLE PIECES (K-MERS)
& IDENTIFY SPECIES ON THOSE SMALL PIECES

K-MERS   SEQUENCE WITH K-BASES

IDENTIFYING SPECIES FROM K-MERS?

COMPARING K-MERS
ARE SIMILAR IF # OF OVERLAPPING K-MERS

KMER DATABASE

LIBRARY OF KMERS OF KNOWN SPECIES

COMPARING   UNKNOWN SPECIE VS KNOWN SPECIES
OUTPUT   BEST MATCHING SPECIE

SPLITTING GENOME IN   16 MERS

ACCTGCCAGCCATATATCCGCAT

TOO MANY   16 MERS

WE KEEP ONLY THE ONES WITH A PARTICULAR PREFIX

LESS REDUNDANCY → SPEED UP COMPUTATIONAL
                   COMPLEXITY IN DATABASE

WHICH SPECIES HAS THE MOST MATCHING K-MERS?

KMA   K-MER ALLIGNMENT

KMER FINDER   BIOINFORMATIC FREE TOOL

SELECT DATABASE
CAN UPLOAD BOTH FASTA & FASTQ
OUTPUT BY EMAIL

SCORE

OF MATCHING KMERS BETWEEN THE 2 GENOMES

QUERY COVERAGE

$$C_q = \frac{\text{\# MATCHING KMERS (IN QUERY)}}{\text{\# KMERS IN QUERY}}$$

TEMPLATE COVERAGE

$$C_t = \frac{\text{\# MATCHING KMERS (IN TEMPLATE)}}{\text{\# KMERS IN TEMPLATE}}$$

DEPTH

$$D = \frac{\text{\# MATCHING KMERS (IN QUERY)}}{\text{\# KMERS IN TEMPLATE}}$$

AS HIGH
AS POSSIBLE

IF NEW SPECIES THE VALUES
WILL BE VERY LOW

MLST TYPING   TOOLS &
                   APPLICATIONS

MULTI LOCUS SEQUENCE TYPING

GOLD STANDARD FOR TYPING
TRADITIONALLY WAS EXPENSIVE & TIME CONSUMING
COST IS DECLINING

MLST TOOL   (Pub MLST)

ST-NUMBER   ID OF UNIQUE COMBINATION   (OUTPUT)
               OF GENE LOCII

BLAS   IF ASSEMBLED GENOME
KMA   IF RAW READS

• SELECTING SPECIE OF STRAIN   (UNKNOWN)
• WE SHOULD HAVE THE   MLS SCHEMA

OUTPUT
          COMBINATIONS OF ⟶ SEQUENCE
          ALLELES TYPES   TYPE

SEQUENCE

GENE

100% COVERAGE MEANS
PERFECT ALIGNMENT

WHAT PART OF SEQUENCE IS
SIMILAR TO GENES IN DATABASE   MISMATCHES?

WEBSERVER AND UNIX TOOLS

RESISTANCE GENE DETECTION

RESFINDER   TOOLS & APPLICATION

IF ISOLATE IS A PATHOGEN /
IS IMPORTANT TO KNOW HOW TO TREAT PATIENT

IDENTIFY IF PATHOGEN IS RESISTENT TO
ANY ANTIBIOTICS

OLD METHOD — PHENOTYPING

DETECTING PRESENCE OF RESISTANCE GENES
AND THE MUTATION THAT IS CAUSING THE RESISTANCE

CONTAINS ALL KNOWN RES- GENOMES

IF NOT ASSEMBLED DATA (RAW READS)
IT USES KMA ALGORITH

90% SIMILARITY (CUT-OFF)
60% MINIMUM LENGTH (CUT-OFF)

IF YOU SELECT SPECIES YOU
CAN STUDY THE MUTATIONS
(CHROMOSOMAL - POINT MUTATION)

>SEQ 1
ATTC -----
>SEQ 2
ATCC.------

OUTPUT (BINARY)

GENE OR MUTATION
RESPONSIBLE FOR
RESISTANCE

ANTIMICROBIAL	CLASS	WGS PREDICTED PHENOTYPE	GENETIC BACKGROUND
TETRACYCLINE	—	RESISTENT	TET A JS567...
FOSFOMYCIN	—	NO RESISTENCE	X
⋮	⋮	⋮	⋮

IDENTITY ⟶ 1,0   IF WE ARE SURE
POSITION LOCI   THAT IS A PERFECT
ALLIGNMENT   MATCH
DEPTH
PHENOTYPE
ALLIGNMENT LENGTH

IF ALLIGNMENT LENGTH < SIZE OF GENE
SOME PART OF THE GENE ARE MISSING

RESFINDER USES BLAST