Week 2

Friday, 30 July 2021 16:21

APPLICATION OF GENOMIC TOOLS

ONE TECH TAKES IT ALL

- EVOLUTION IN TECNIQUES
- TOOL BOX OF CGE

GENOMIC EPIDEMIOLOGY

SUB-TYPING TECHNOLOGIES 1920

SEROTYPING , PHAGE & PLASMID TYPING

PULSED-FIELD GEL ELECTROPHORESIS PFGE

MULTI-LOCUS SEQUENCE TYPING MLST

NEXT-GENERATION SEQUENCING NGS

2 WEEKS - 1 MONTH ~6 DAYS

OLD	NEW
ISOLATION	ISOLATION DNA & LIB PREP NGS
CARACTHERIZATION & SUB TYPING	BIOINFORMATICS PLUG & PLAY TOOLS
INTERPRETATION	

ONE TECH THAT TAKES IT ALL SINCE IT WORKS FOR ALL KINDS OF BACTERIAS

SPECIES
VIRULENCE
PHYLOGENY
RESISTANCE
} WHAT TO STUDY

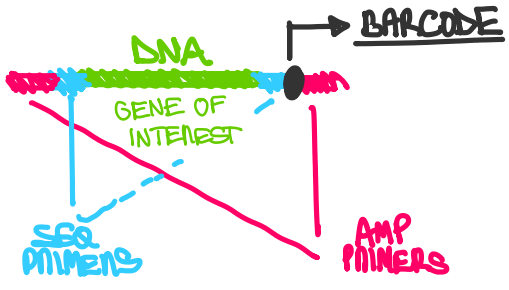
INTRODUCTION OF NGS

NEXT GENERATION SEQUENCING

LIBRARY PREPARATION

- FRAGMENT DNA
- LIGATE ADAPTORS (CONTAIN AMPLIFICATION & SEQUENCING PRIMERS)

- AMPLIFICATION
- SEQUENCING



IF IN ONE RUN YOU HAVE MORE THAN 1 SAMPLE THEN CREATE A SEQUENCE SPECIFIC FOR EACH SAMPLE

SEQUENCE DATA

STD DATA STORES → .FASTA FILES

1) HEADER + ID
2) SEQUENCE DATA (YOU CAN PUT MORE THAN ONE HERE)

>GI | 297332897 | REF ESCHERICCHIA COLI 1247 CHROMOSOME, COMPLETE

RAW-READS

SEQUENCER OUTPUT IS NOT FASTA (2 LINES) BUT FASTQ (4 LINES)

FAST-Q = FASTA + QUALITY SCORES

- HEADER
- DNA SEQ
- NAME FIELD
- QUALITY SCORES

} 4 LINES PER READ

DE-MULTIPLEXING DONE BY SEQUENCERS

SINGLE-END READS

PAIR-END READ → FRAGMENT GAP

INSERT SIZE — ALL FRAGMENTS WITH NO ADAPTERS

DATA QUALITY

- COVERAGE
- DEPTH
- BREADTH OF COVERAGE

TRIMMING DATA "N" NOT APPLICABLE

REMOVING LOW-QUALITY DATA

DATA SPLITTING
DATA CLIPPING
DATA TRIMMING

COVERAGE

$$C = N \frac{L}{G}$$

L — READ LENGTH
G — GENOME LENGTH
N — # OF READS

BREADTH OF COVERAGE

$$C_b = \frac{\text{ASSEMBLY SIZE}}{\text{TARGET SIZE}}$$

$C_b \in [0,1]$

P(TO RECONSTRUCT WHOLE GENOME)

EXAMPLE

$N = 5 \cdot 10^6$ READS
$L = 100$ bp
$G = 5 \cdot 10^6$ bp

COVERAGE
$C = 100\ X$

100 READS COVER EACH POSITION IN THE GENOME

ASSEMBLY

RAW READS ARE VERY SHORT, SO WE MERGE THEM IN ORDER TO MAKE LONGER SEQUENCES

- MAPPING TO REF. (W/ REF. GENOME)
  STUDYING & ANALYZING GENE, SNP ?

- DE-NOVO ASSEMBLY (W/OUT REF. GENOME)
  DISCOVERING & IDENTIFYING NEW GENES

ANALYSIS - MOST PROGRAMS IN UNIX
IMPORTANT TO KNOW COMMAND-LINE TOOLS

NEW WEB BASE TOOLS
EASIER TO USE

DIFFERENT SEQUENCERS
NEEDING DIFFERENT ASSEMBLERS

DE-BRUIJN GRAPHS

FROM READS TO CONTIGS
(FASTQ) (FASTA)

GOAL — IDENTIFY GENE OR GENETIC MARKER THAT YOU ARE LOOKING FOR.

GOOD CONTIGS — IF CONTIGS ARE LARGE ENOUGH TO CONTAIN A GENE OR COMPLETE MARKER THAT YOU ARE STUDYING

- # OF CONTIGS — LONGER ARE BETTER, EASY TO ASSEMBLE ↓ YOU SHOULD WANT AS FEW AS POSSIBLE
- N50

MEDIAN SIZE OF YOUR CONTIGS

↑ THE BIGGER THE BETTER

SPLITTING BARCODE
TRIM & LOOK FOR QUALITY
ASSEMBLY (WITH OR WITHOUT REFERENCE)

DE NOVO ASSEMBLY
FROM RAW READS TO CONTIGS
TOOL DESCRIPTION & APPLICATIONS

N50 - HOW GOOD IS YOUR ASSEMBLY

MOST CGE ARE ASSEMBLY BASED