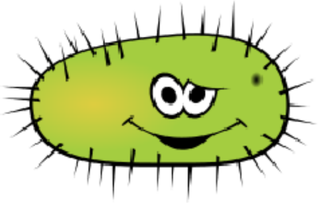
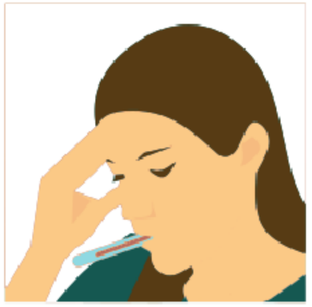
freezeTb; detecting anti-microbal resistance in TB

WHO 2023

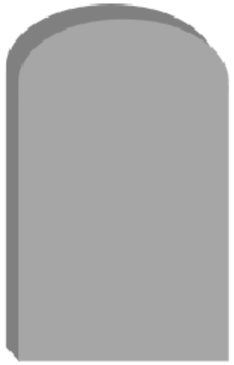


tuberculosis

5-7 million

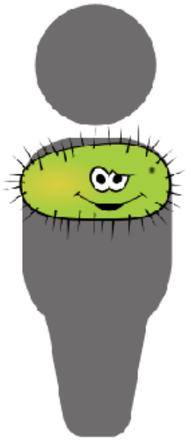


10 million



1 million

Treatment success

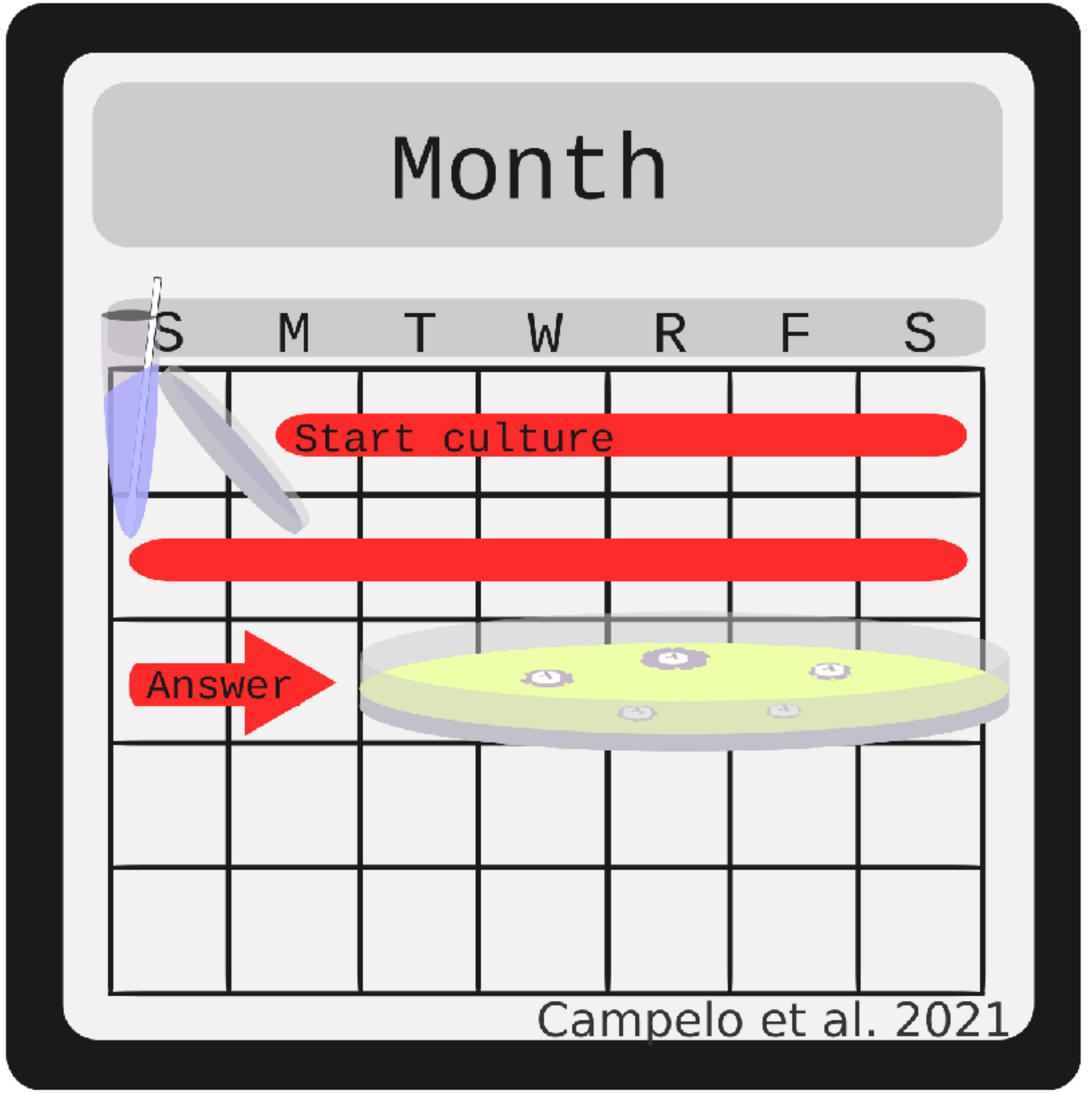


85%

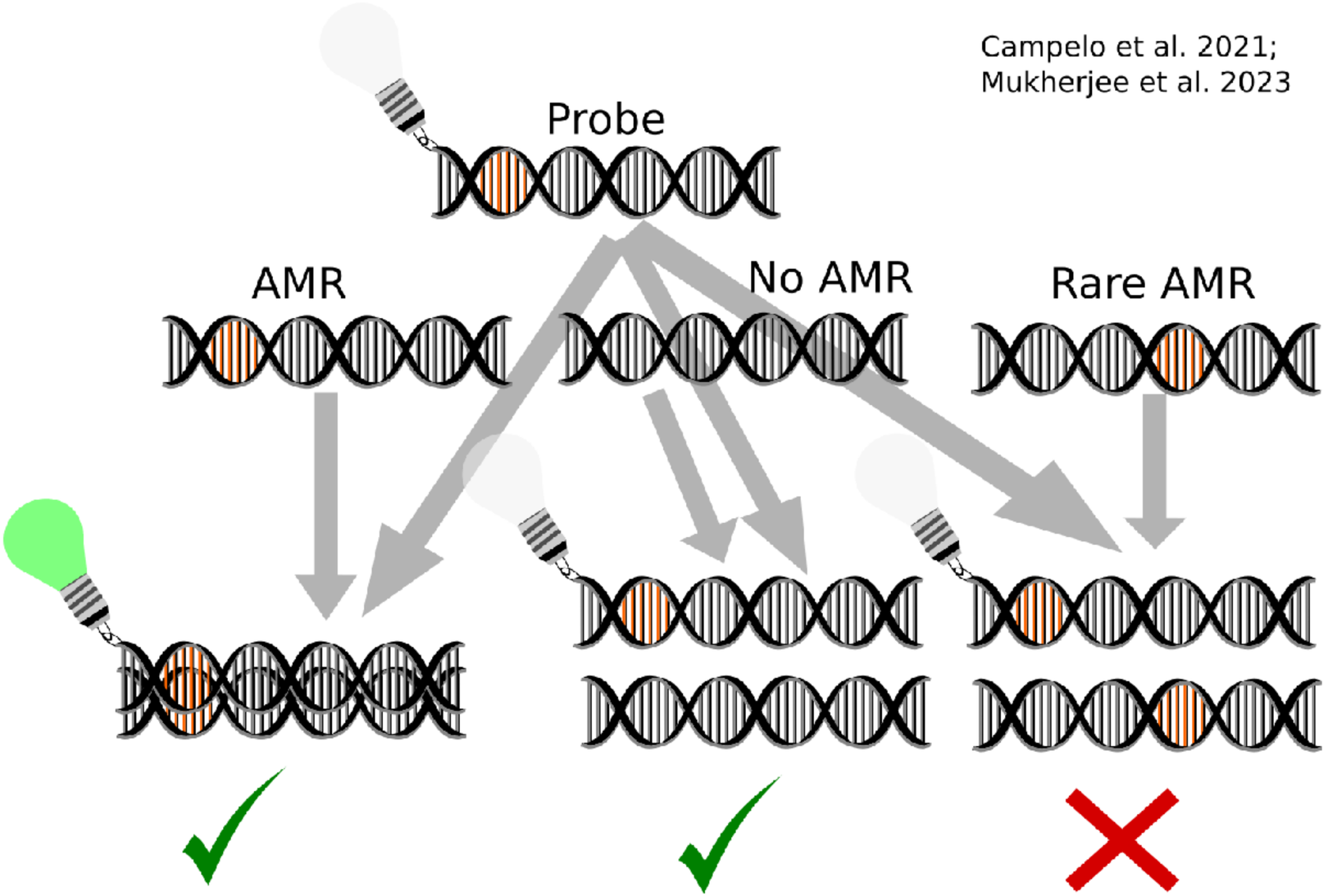
63%

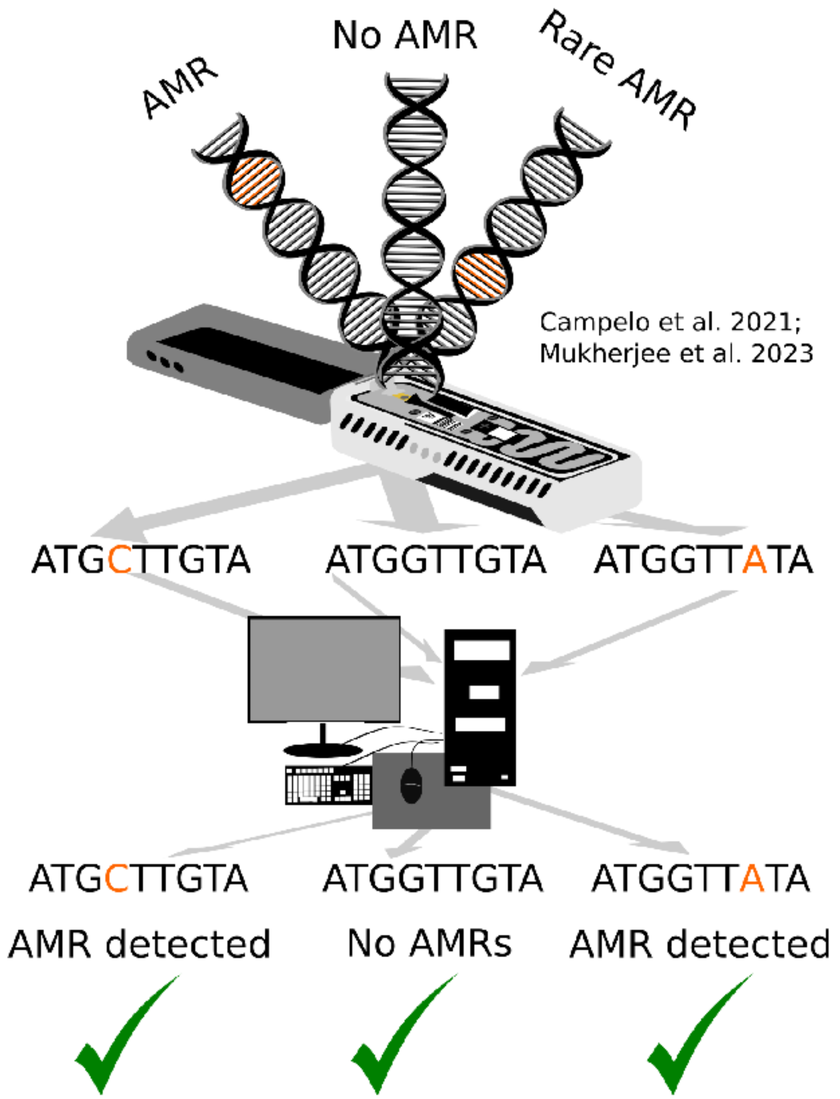
WHO 2023

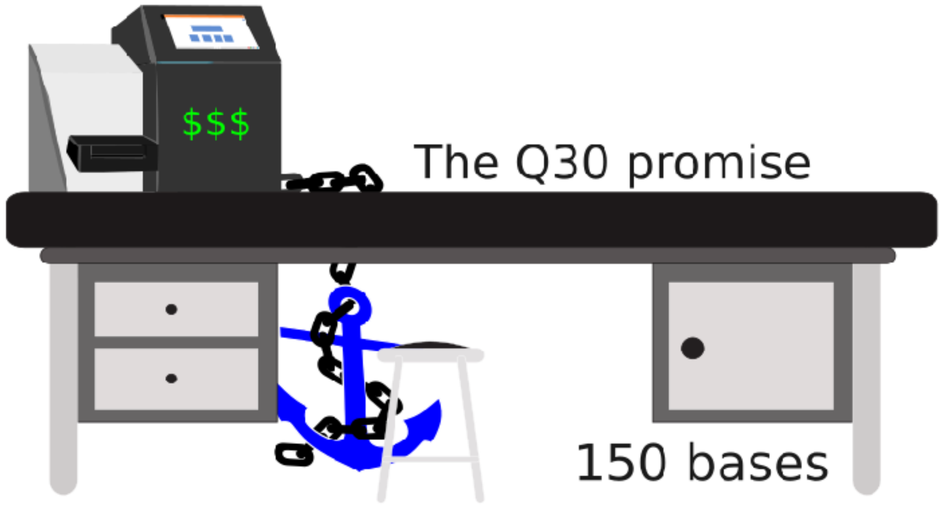
Drug resistant detection

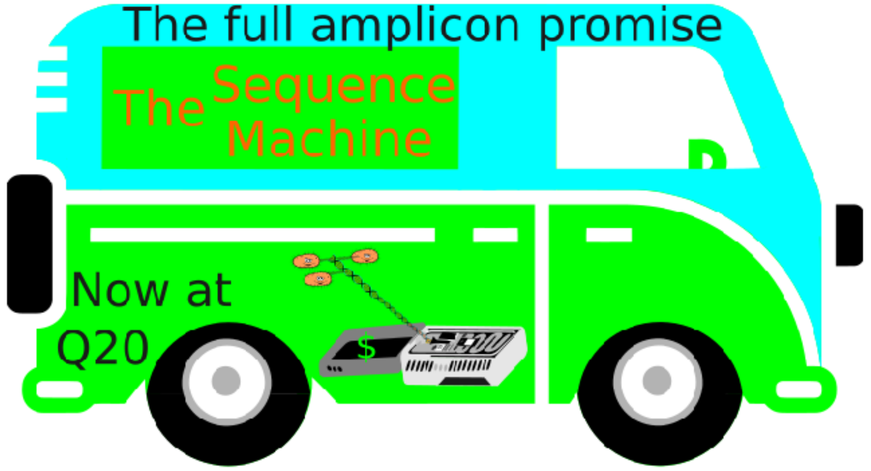

Month
S M T W R F S
Start culture
Answer
Campelo et al. 2021

Campelo et al. 2021;
Mukherjee et al. 2023




AMR
No AMR
Rare AMR
Campelo et al. 2021;
Mukherjee et al. 2023
ATGCTTGTA
ATGGTTGTA
ATGGTTATA
ATGCTTGTA
ATGGTTGTA
ATGGTTATA
AMR detected
No AMRs
AMR detected



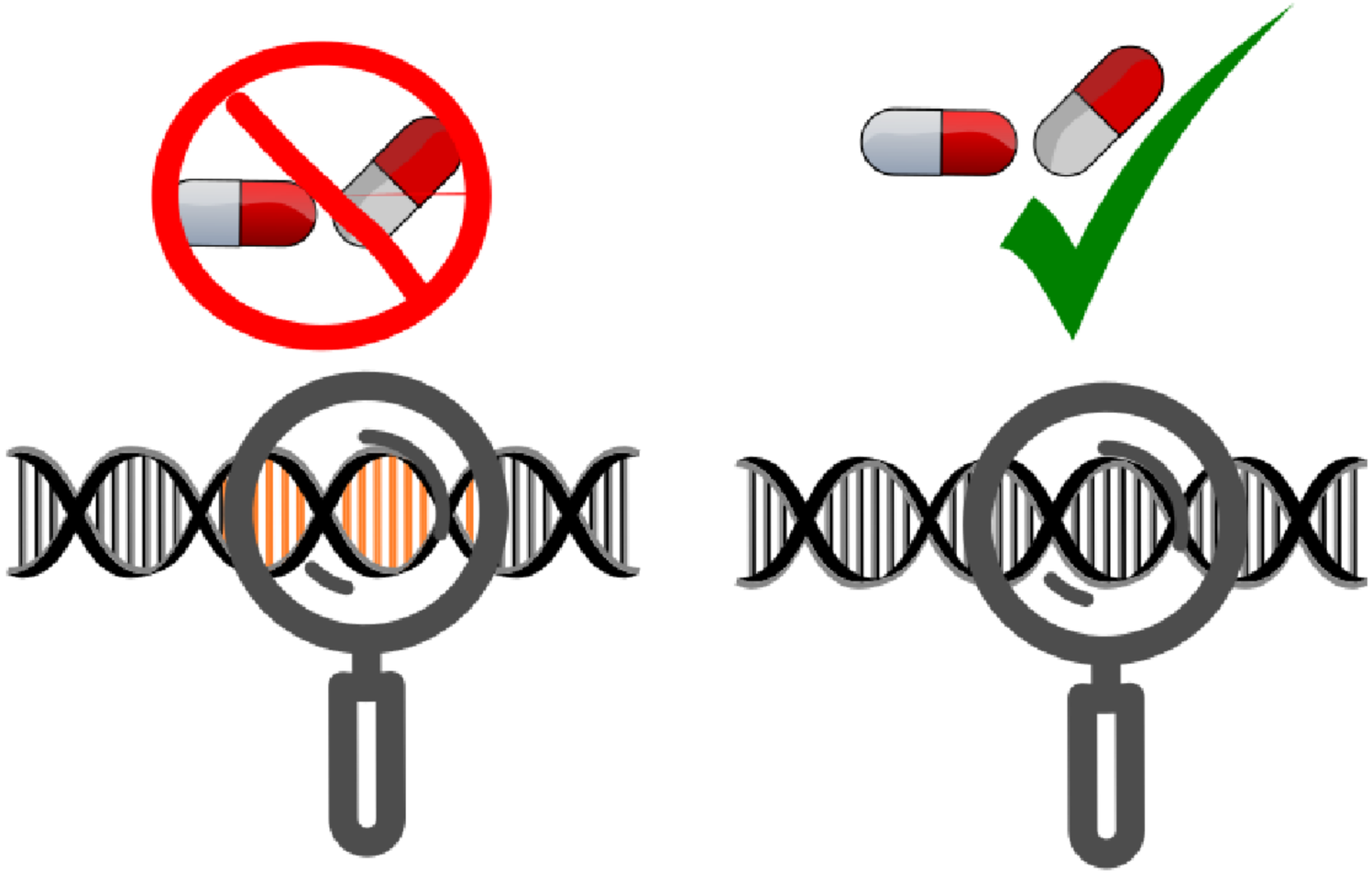


Does not include van or great danes

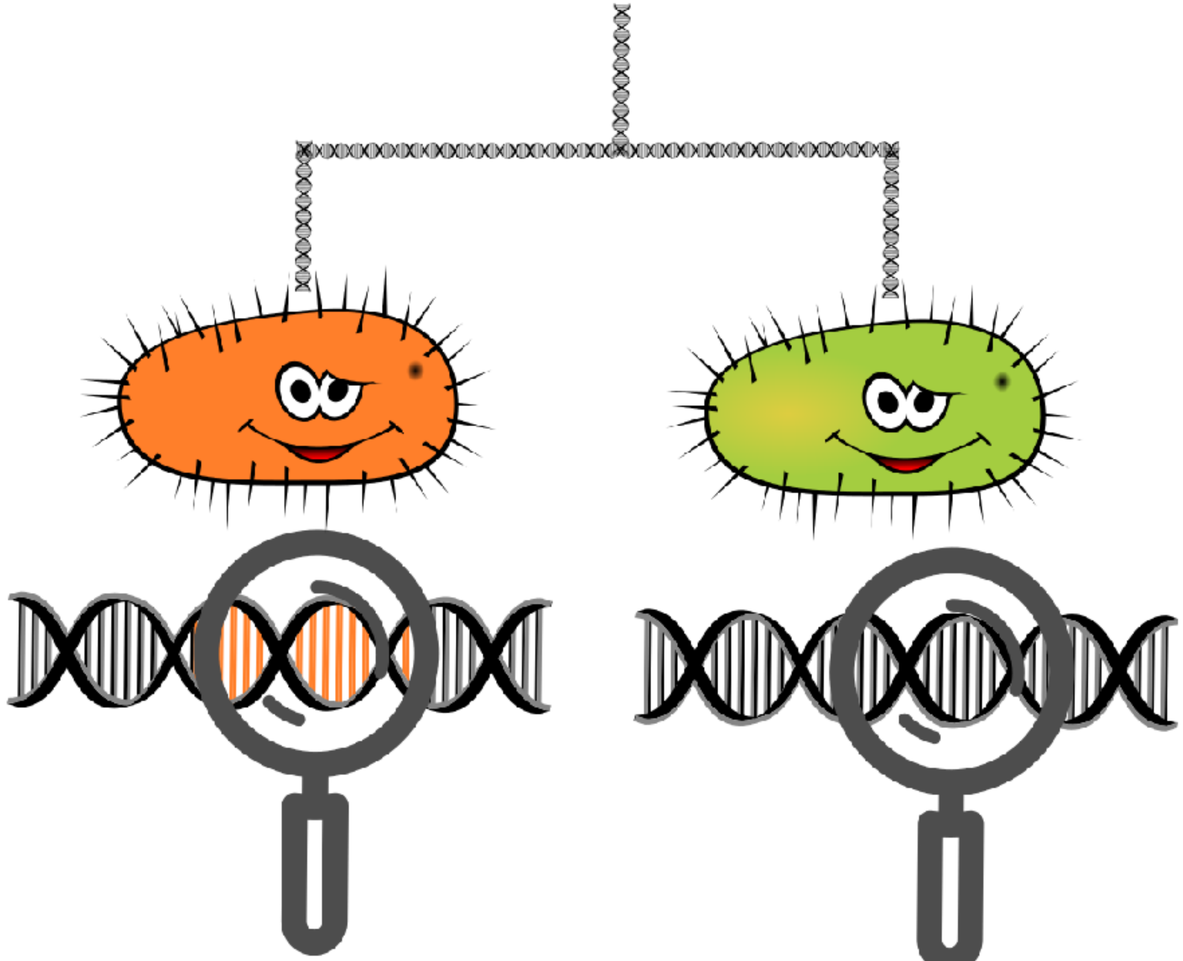
Xia et al. 2023

freezeTb; the software side

Expected output



Check for AMRs



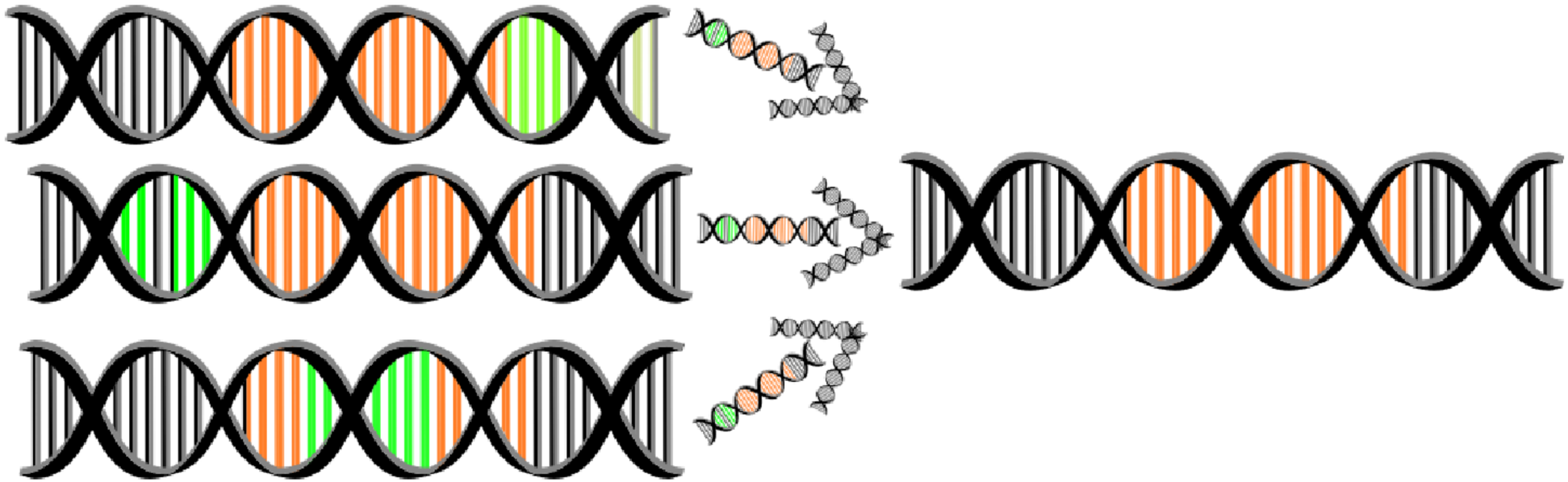
Detect Lineages

2. a. MIRU-VNTR

- Amplicons target 24 repeat regions across the genome
- Amplicons lineage found by Length of amplicon
- Current gold standard

2. b. Spoligotyping

- Amplicon targets direct repeat region (1 primer)
- Lineage found by detecting 43 internal spaces
- Lineage id is barcode with presence/absence of spacer



Build consensus

TBProfiler; the current resource

TBProfiler; pros

- Detects AMRs uses WHO's 2023 catalog
- Detects spoligotypes and SNP lineages (Nanopore?)
- Outputs an vcf files

TBProfiler; cons

- More focused on Illumina
- > 5 dependencies or conda; not fun to install

Other solutions;

ONTs: Have not tried

- Detects spoligotypes
- Likely more to add

Plentaly of small programs that add in more dependencies

AMR detection; no small program I know of

Lineage detection is simple

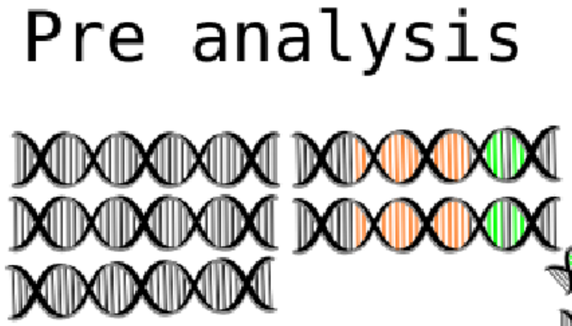
Consensus building; non-model methods are simple

Most of it is simple, why not make our own?

FreezeTb; the software side

- CLI; requires sam file
- R GUI; fastq file only, but requires minimap2

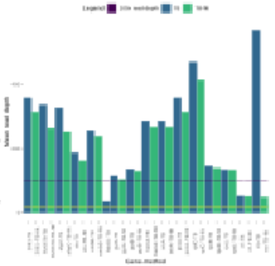
Pre analysis



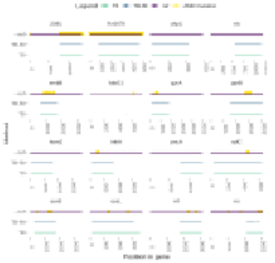
Mapped reads

Length < 200
Q-score < 7
Mapq < 15

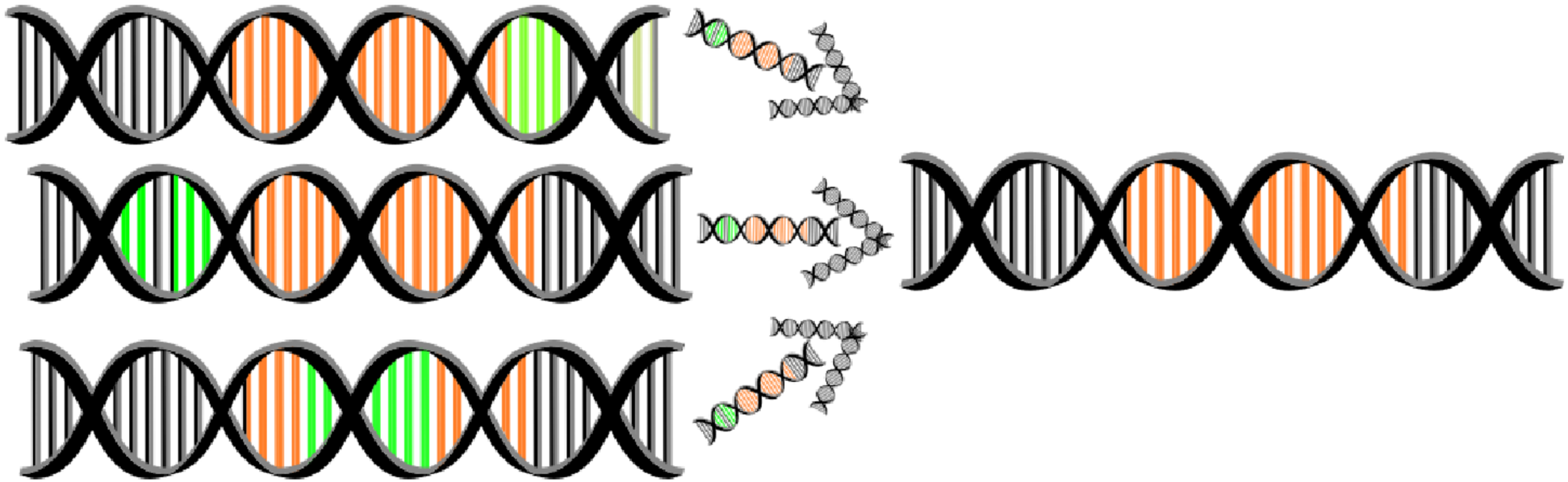
Fitler reads



Read depth



Coverage



Build consensus

Filters out bases with Q-scores under 7

Break up consensuses

- Fragments assemblies under 20x read depth
- Remove fragments less than 200 bases long

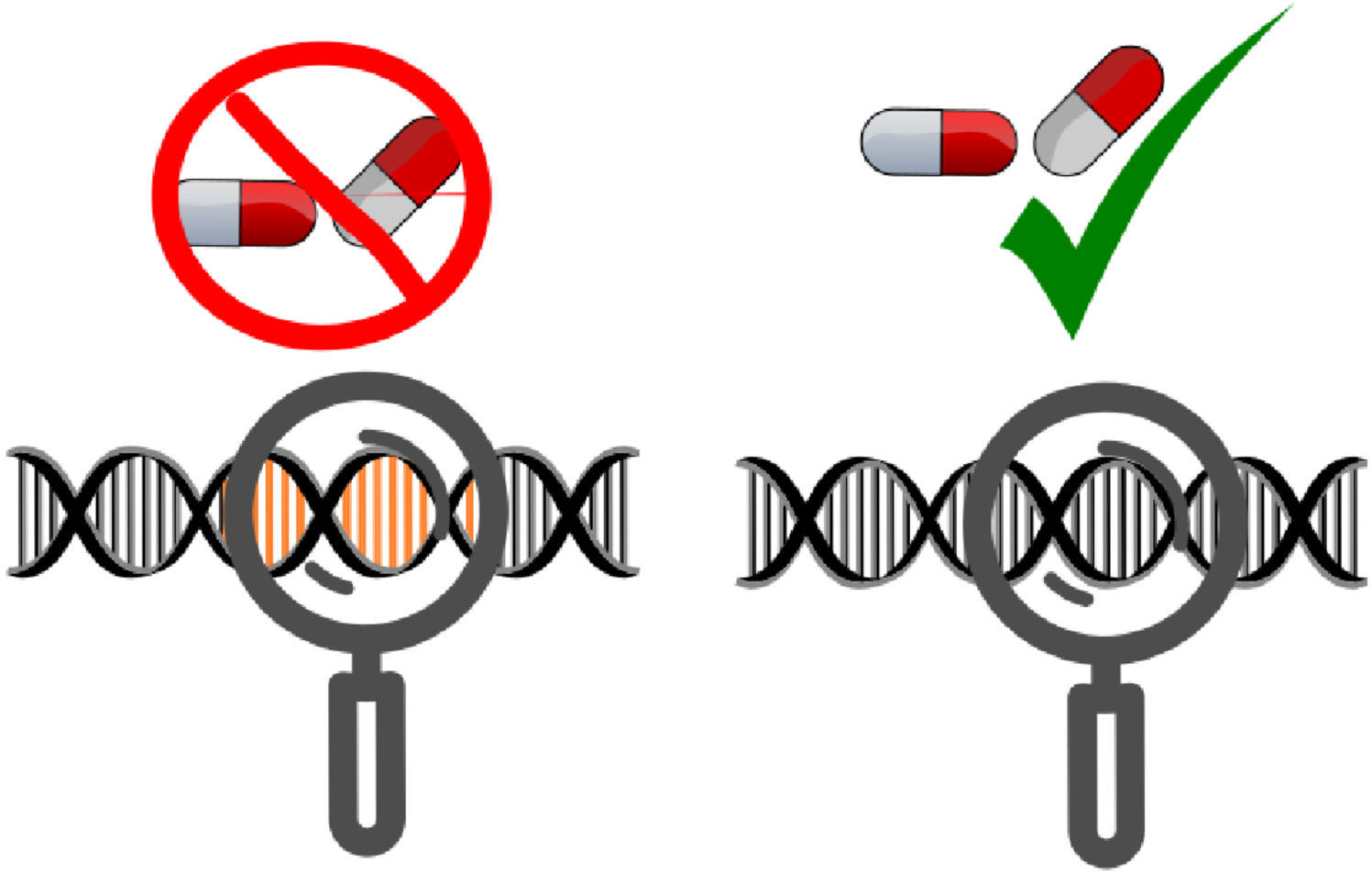
Masking:

- Select nucleotides with most support
- Mask nucleotides with less than 50% of mapped reads

Indel removal:

- Remove indels with less than 70% of mapped reads

Ouput and sam file



Check for AMRs

Using 2023 WHO tuberculosis AMR mutation catalog

Gene deletions are ignored

Frames shifts / LoF's treated as normal AMR

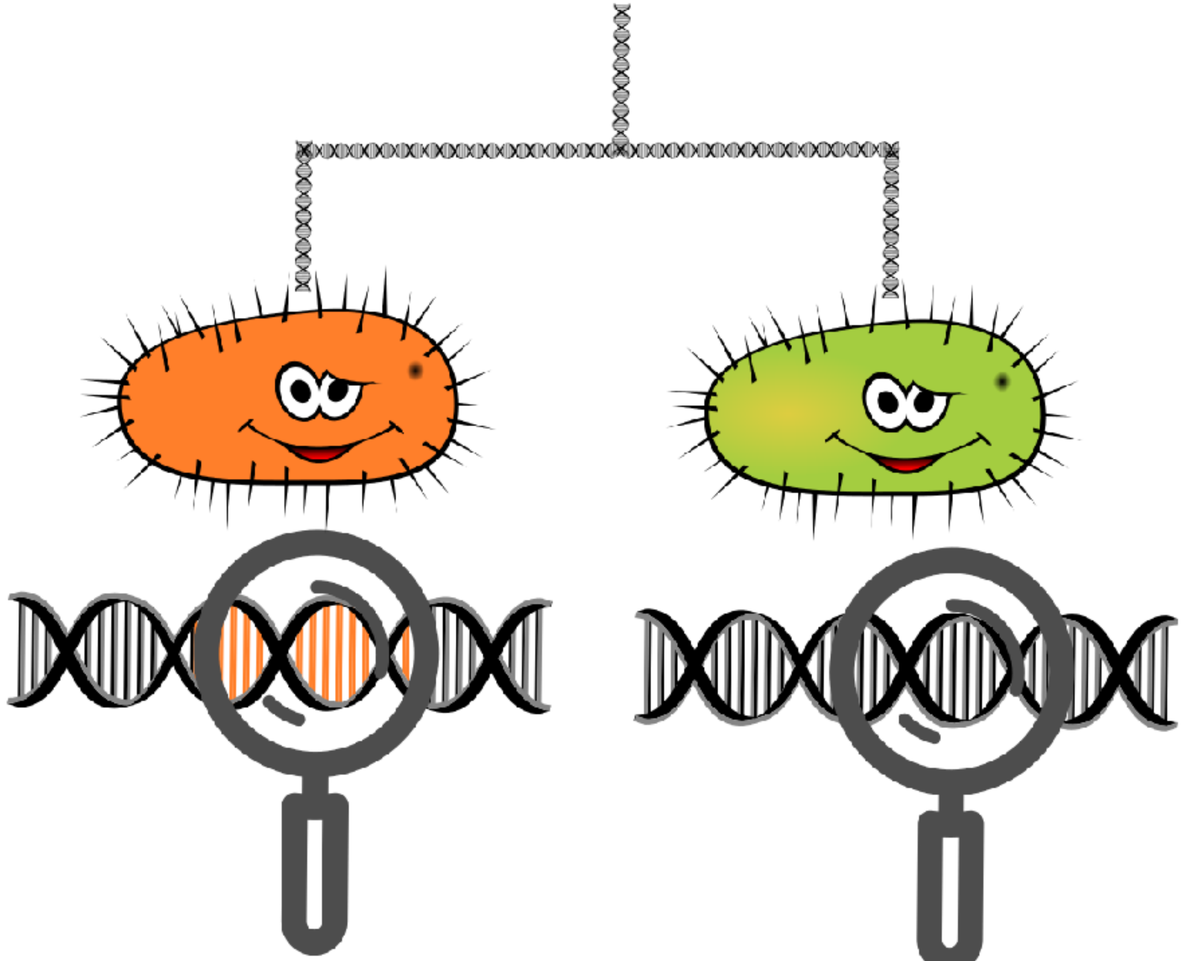
- Less false positives due to indel errors

For amino acid catalog variants

- AMR region in read converted to amino acids

AMR called if

- Sequences match
- Read AMR region length = catalog AMR length



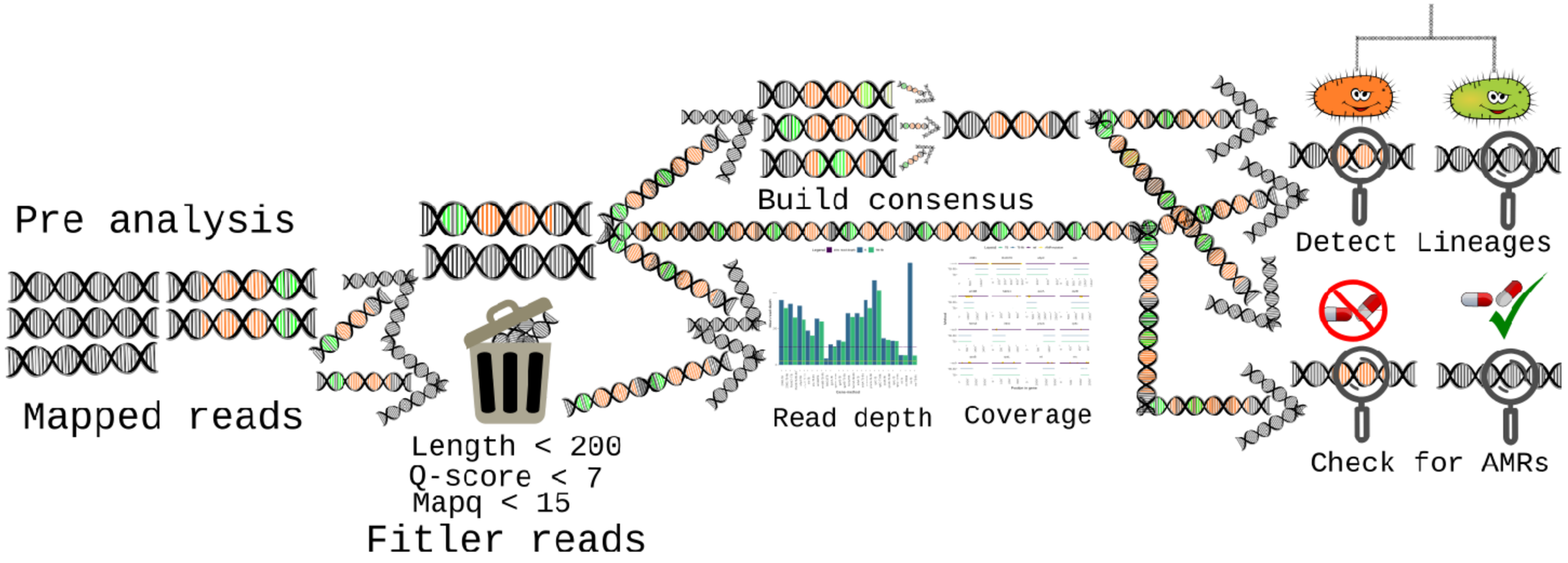
Detect Lineages

Miru lineages:

- See if read has primer coordinates (from sam file)
- Check reads repeat length

Spoligotyping:

- Get direct repeat (DR) region from consensus
- Map spacer sequences to DR with Waterman alignment
- Make barcode from mapped spacers and find lineage


Pre analysis
Mapped reads
Length < 200
Q-score < 7
Mapq < 15
Fitler reads
Build consensus
Read depth
Coverage
Detect Lineages
Check for AMRs

Each step in freezeTb is modular:

- You like an module; you can get an module
- Each module takes an sam file

FreezeTb; works, but not an replacement

Future steps:

- Finish and integrate spoligotyping
- Maybe add in fast kmer seeding for spoligotyping

Fast kmer seeding is close to an read mapper

Why not add an read mapper? Ok probrably not.

?