Finding the appropriate method, with a special focus on: Mapping and alignment

Philip Clausen
Background

Most people choose their methods based on popularity and history, not by reasoning and research.

It is the same thing as going out to buy a car, while the only thing you are sure about is what color you would like it in.
Difference: Alignment vs. Mapping

Mapping:
Find the approximate origin of a sequence.

Alignment:
Find the exact difference between two sequences.
Mapping example:

GTGGTGCAATCTGTTCTCCCCCACAAGGAAAGTA oqxB_EU370913
Alignment example:
Mapping is part of alignment
Why use mapping, if it is included in alignment?

1. Mapping is (alot) faster, and computationally lighter to solve.

2. You might not be interested in the precise differences, but merely wants to know the approximate origin.
Popular mapping methods:

1. KmerFinder
2. Kraken
3. Kallisto (development)
4. Salmon (development)
5. KMA-Sparse
Popular alignment methods:

1. BLAST
2. Bowtie2
3. BWA-MEM (development)
4. GraphMap (development)
5. MiniMap2
6. KMA
Why so many?

1. We have a lot of different analyzes to carry out. Therefore different assumptions and pathways are taken to perform these tasks.

2. One shoe might fit all, if you cut bits of your feet.  
   -Charles Perrault's Cendrillon
Simple example
Features in our sample:

1. Jaws
2. Lungs
3. Feathers
Features in our sample:

1. Jaws
2. Lungs
3. Feathers

100% query coverage
A lot of local alignments, smart for assembly data
BWA* / Bowtie*

Features in our sample:

1. Jaws
2. Lungs
3. Feathers

1/3 query coverage

Based on unique mappings, smart for single genome alignment
Graph-Mappers

Features in our sample:

1. Jaws
2. Lungs
3. Feathers

100% query coverage
Smart when you don’t expect the exact template to be in the database
KMA

Features in our sample:

1. Jaws
2. Lungs
3. Feathers

100% query coverage

Matching only one template, smart for redundant databases
Take home messages

1. Every method has its own assumptions.

2. When you choose a method, you better be sure that your data and hypothesis fits within these assumptions.

3. Most methods have a big variety of settings, which lets you fine tune it to improve predictions or make better use of the computational resources.

4. Read the documentation of your method of choice in order to fulfill #1, #2 and #3.
How is this relevant to AMR?

1. ResFinder-4.0, using an assembly and BLAST approach.

2. ResFinder-4.0, using KMA to map raw reads directly.
1. ResFinder-4.0, using an assembly and BLAST approach.
2. ResFinder-4.0, using KMA to map raw reads directly.

Map reads

Raw reads

Known AMR genes

Call consensus

T: TGGTGCATCTGTCTCCCTCGCGGGAAGTACGAC
Q: TGGTGCATCTGTCTTTGCACCCGAAAGTACGAC
Can the sequencing quality affect this?

1. ResFinder-4.0, using an assembly and BLAST approach.

2. ResFinder-4.0, using KMA to map raw reads directly.
1. ResFinder-4.0, using an assembly and BLAST approach.

Poor WGS data leads to poor assemblies.

Poor assembly is the same as missing data.
2. ResFinder-4.0, using KMA to map raw reads directly.

Map reads

Raw reads

Known AMR genes

Call consensus

T: TGGTGCATCTGTTCCTCCCCCGGCGGGAAGTACGAC
Q: TGGTGCATCTGTTTCCGCCAAACGGTAAGTACGAC
2. ResFinder-4.0, using KMA to map raw reads directly.

Too sensitive methods allow us to detect contamination.

Raw reads

Known AMR genes

T: TGGTGCATCTGTTCTCCCCCGCGGGAAGTACGAC
Q: TGGTGCATCTGTTTCCGCCAAACGGTAAGTACGAC
Best method tomorrow?

When the proper method has been identified, the hard work is not necessarily over.

1. Keep up to date with your current method of choice, check new releases that may include features to boost your analysis.

2. Read articles using the method, and check if they suggest an alternative method or a different set of options.

3. Life as a scientist is a life-long education.
There is no true interpretation of anything; Interpretation is a vehicle in the service of human comprehension. The value of interpretation is in enabling others to fruitfully think about an idea.

-Andreas Buja