Pathogen Detection and the role of Microbial Communities in the phyllosphere during fungal infection of wheat.
Introduction

• Fungal diseases are responsible for major losses in crop production, especially wheat.

Microbial communities influence disease outcomes at infection sites.

Microbiome
1. Can we detect fungal pathogens and associated microbiomes through a metagenomics approach?

2. How can we improve the species classification?

- A proof-of-concept study
- Benchmarking taxonomic classification strategies using mock communities.
Detection of fungal wheat pathogen from field samples

- Sampling wheat leaves with confirmed phenotypes
- Blind DNA extraction
- Portable MinION sequencer
- Two-step BLAST search

Data analysis

Visualization

nucleotide database

blasting BLAST NCB

Pathogen genomes

Wheat Genomes

Trimming

Quality control

DNA sequencing

barcodes: barcodes2, barcodes3, barcodes4

Healthy wheat leaves

and stripe rust

Disease 4: Septoria tritici blotch

Disease 3: Yellow spot

Disease 2: Septoria tritici blotch

Disease 1: Stripe rust

Sampling
Detection of fungal wheat pathogen from field samples


Blast against reference database from 5 species:

- P. nodorum
- P. striiformis f. sp. tritici
- Z. tritici
- P. triticola
- P. triticina

Major wheat disease identification:
- Healthy wheat: 91.4%
- Wheat stripe rust: 0.3%
- Yellow spot: 0.9%
- Septoria tritici blotch: 0.7%
- Puccinia striiformis

QC & trimmed reads


Table of species:

<table>
<thead>
<tr>
<th>Common name</th>
<th>Disease name</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. nodorum</td>
<td>Septoria nodorum</td>
<td>Pseudocercosporella graminicola</td>
</tr>
<tr>
<td>P. striiformis</td>
<td>Puccinia striiformis</td>
<td>Triticum aestivum</td>
</tr>
<tr>
<td>Z. tritici</td>
<td>Septoria tritici blotch</td>
<td>Zymoseptoria tritici</td>
</tr>
<tr>
<td>P. triticola</td>
<td>Yellow spot</td>
<td>Septoria nodorum</td>
</tr>
<tr>
<td>P. triticina</td>
<td>Wheat stripe rust</td>
<td>Parastagonospora nodorum</td>
</tr>
</tbody>
</table>
Microbiome profiles are pathogen specific.

Reads did not hit nucleotide database BLAST against NCBI Reference database

Species:
- Leptospheurina
- Erwinia, Sphingomonas
- Alternaria, Pseudomonas
We can detect fungal wheat pathogen from field samples using nanopore metagenomics shotgun sequencing.

- Microbiome profiles are pathogen specific.
Benchmarking taxonomic classification strategies

- Success rate: % of identified reads that belong to the mock community
- Recall rate: % of identified reads

- 2 mock fungal communities: pooled DNA & pooled tissues
- Metagenomics
- Shotgun sequencing (Nanopore)
- 2 algorithms: K-mer (Kraken2) & Alignment (BLASTn)
- 2 databases: nt & Refseq fungi

How to improve the classification?
• Alignment + specific database is the "best approach" for classification.

• Choice of database affect the result more than the choice of algorithms.

• blastn against Refseq Fungi result in the highest species level success rate.
Applying cut-offs on alignment proportion improves classification.

Genera completeness = \frac{\# of genera identified belongs to the mock}{\# of genera in the mock}

Alignement proportion = \frac{\# of identical matches}{Read length}

Cut-offs on alignment proportion works the best compared to e-value, read length, read quality and percentage of identity.
Optimizing community composition analysis:

• Create a ‘gold standard’ classification and community compositions:
  • Using pairwise alignment algorithms (minimap2)
  • Using database with only genomes from the species in the mock community
  • To maximize the success rate (100%)

Compare other broadly applicable classifications to the ‘gold standard’.
Benchmarking community composition using 'gold standard' composition analysis.

### Candida rugosa

<table>
<thead>
<tr>
<th></th>
<th>BUSCO</th>
<th>N/L50 (Mb)</th>
<th>Contig number</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUSCO</td>
<td>87.6%</td>
<td>12/0.8</td>
<td>188</td>
<td>26.6</td>
</tr>
</tbody>
</table>

### Cryptococcus magnus

<table>
<thead>
<tr>
<th></th>
<th>BUSCO</th>
<th>N/L50 (Mb)</th>
<th>Contig number</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUSCO</td>
<td>91.6%</td>
<td>2/3.6</td>
<td>13</td>
<td>16.6</td>
</tr>
</tbody>
</table>

### Cryptococcus mesorugosa

<table>
<thead>
<tr>
<th></th>
<th>BUSCO</th>
<th>N/L50 (Mb)</th>
<th>Contig number</th>
<th>Size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUSCO</td>
<td>92.3%</td>
<td>2/3.2</td>
<td>15</td>
<td>13.9</td>
</tr>
</tbody>
</table>

- Constructed reference database with only genomes in the mock.
- Using pairwise alignment (minimap2) to construct the 'gold standard' classification and composition.
- Compare different strategies with 'gold standard' for similarities.
- Apply cut-offs on alignment proportion and access the similarities' change.

- Benchmarking community composition analysis.
Blastn against Refseq_fungi database has the closest B distance with the gold standard community composition. Comparing community composition analysis using Bhattacharyya (B) distance between two lists of probabilities a measurement of absolute distance B distance (angle of):

Gold standard community composition.

Blast against Refseq_fungi database has the closest B distance with the
Comparing community composition analysis using Bhattecharya (B) distance

Running hypotheses:

• No need to restrict the dataset for better community composition analysis.

B distance (angle o):

A measurement of absolute distance between two lists of probabilities.
Next step: tracking the quantitative abundance of fungal pathogens and associated microbiomes during infection.

**Hypothesis:**

The pathogen development is associated with the quantity of microbial communities.

**Experimental design:**

1. Sampling of wheat disease trial four times per growing season for three years.
2. Quantifying the abundance of major pathogen species and their associated microbiomes.

Collaboration Welcome!
1. We can detect fungal pathogens and describe their associated microbiomes through a metagenomics approach.

2. Database affects classification more than the classification algorithms, and alignment + specific database is the best approach.

3. Applying cut-offs on alignment proportions can further improve the classification.

Take-home messages
Acknowledgements

Supervisors:
John Rathjen
Benjamin Schwessinger
Eric Stone
Tony Goldthorpe
Michael McCaig
Andrew Milgate

Primary Industry:
NSW Department of Primary Industry:

The University of Sydney:
Wieland Meyer
Laszlo Irinyi
The Rathjen lab
The Schwessinger lab

Thank you for listening!