Australian National University

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

Yiheng Hu

PhD candidate

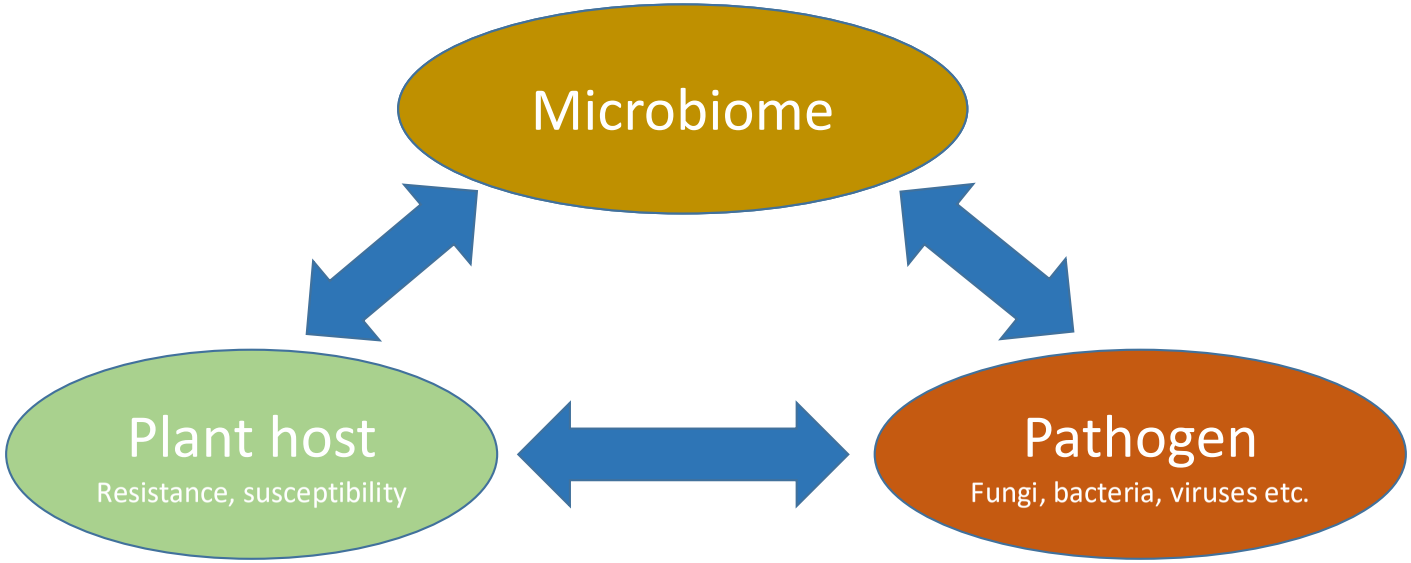
Research School of Biology, The Australian National University

International Plant & Animal Genome XXVIII Congress, 15/01/2020

Adapted from van Gogh (1888), *Wheat field*.

Introduction

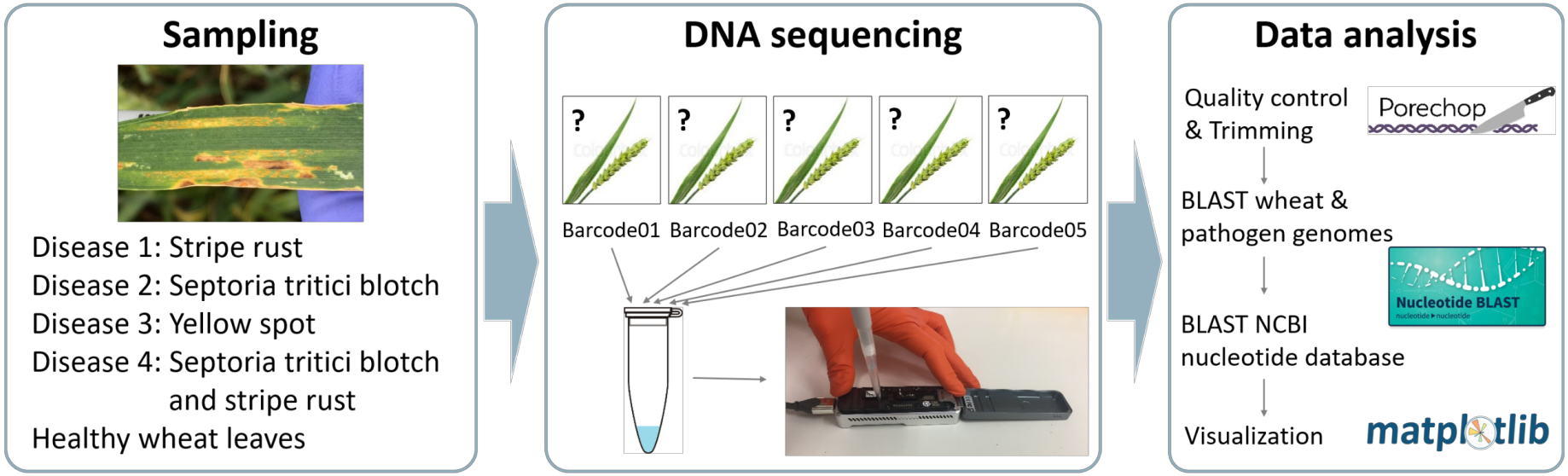
- Fungal diseases are responsible for major losses in crop production, especially wheat.
- Microbial communities influences disease outcomes at infection sites.



1. Can we detect fungal pathogens and associated microbiomes through a metagenomics approach?
   - A proof-of-concept study

2. How can we improve the species classification?
   - 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

Hu et al., (2019) *Phytobiomes J*, 3:92–101.

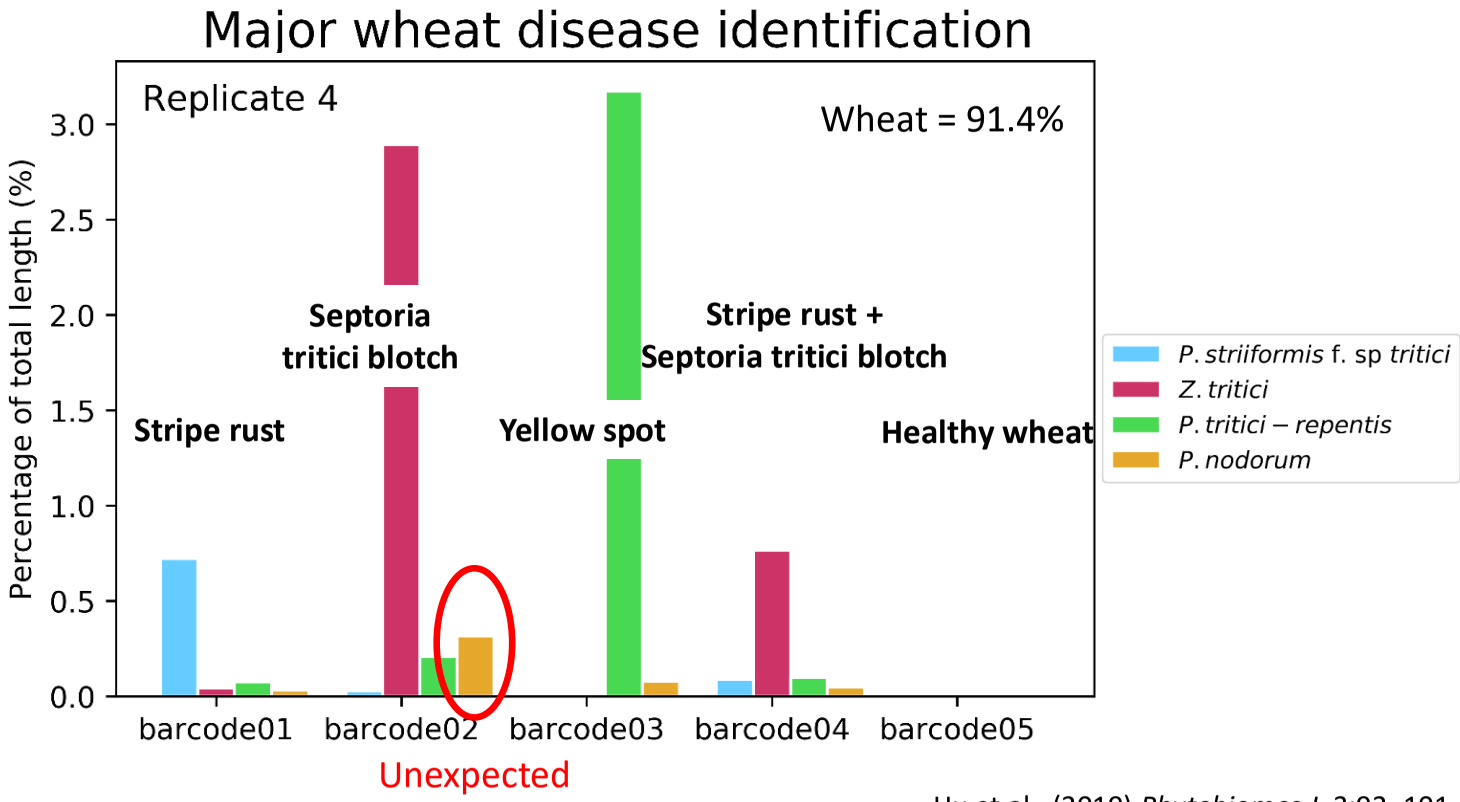
Detection of fungal wheat pathogen from field samples

QC & trimmed reads

↓

BLAST against reference database from 5 species:

Common name / disease name	Species
Wheat	*Triticum aestivum*
Wheat stripe rust	*Puccinia striiformis*
Septoria tritici blotch	*Zymoseptoria tritici*
Yellow spot	*Pyrenophora tritici-repentis*
Septoria nodorum blotch	*Parastagonospora nodorum*



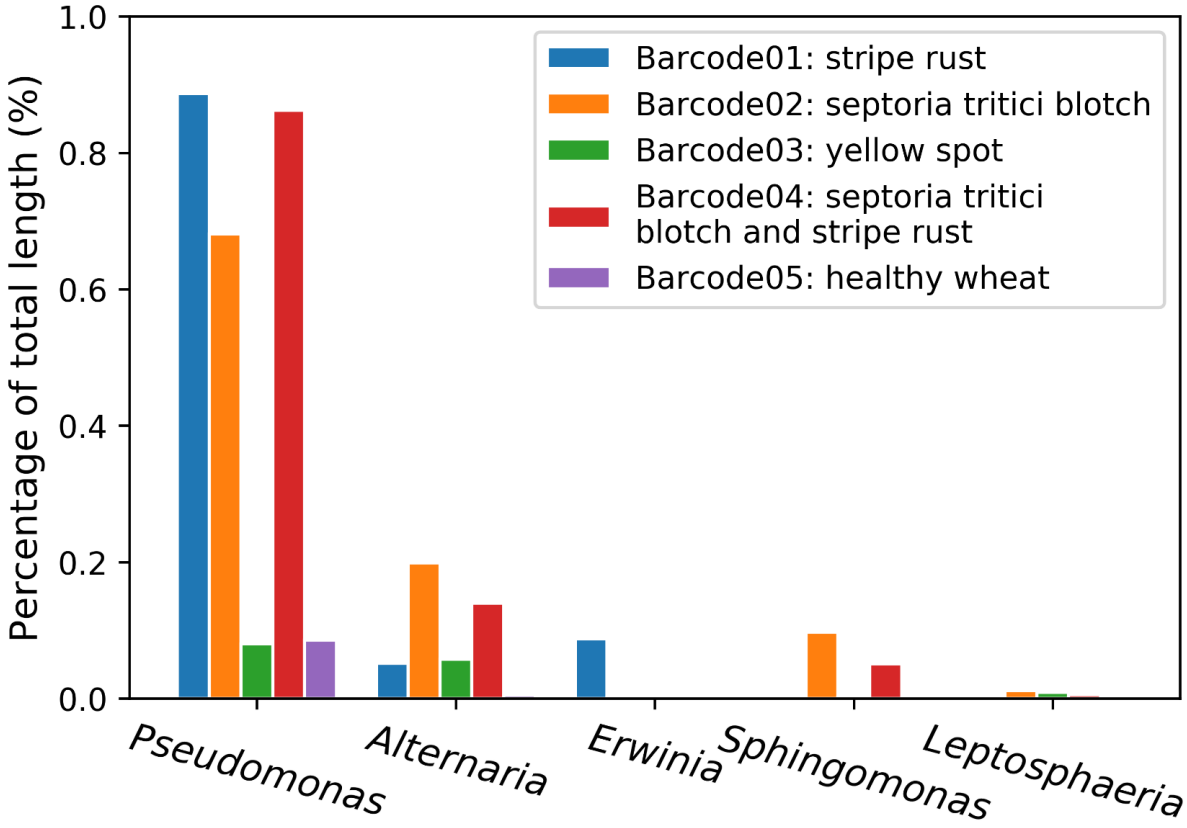
Hu et al., (2019) *Phytobiomes J*, 3:92–101.

Microbiome profiles are pathogen specific

Reads did not hit reference database

↓

BLAST against NCBI nucleotide database

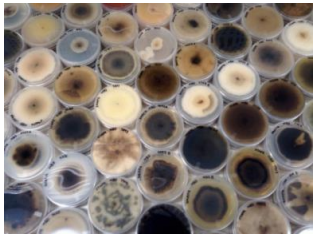


- We can detect fungal wheat pathogen from field samples using nanopore metagenomics shotgun sequencing.

- Microbiome profiles are pathogen specific.

How to improve the classification?

Benchmarking taxonomic classification strategies



2 mock fungal communities: pooled DNA & pooled tissues

Metagenomics shotgun sequencing (Nanopore)

- 2 algorithms: K-mer (Kraken2) & Alignment (BLASTn)
- 2 databases: nt & Refseq fungi

- Recall rate: % of identified reads

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

https://marketsci.org/market-science-blog/page/2/

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

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

$$\text{Alignment proportion} = \frac{\text{\# of identical matches}}{\text{Read length}}$$

- Cut-offs on **alignment proportion** works the best compare 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’ to optimize the community composition analysis.

Benchmarking community composition analysis using ‘gold standard’ composition

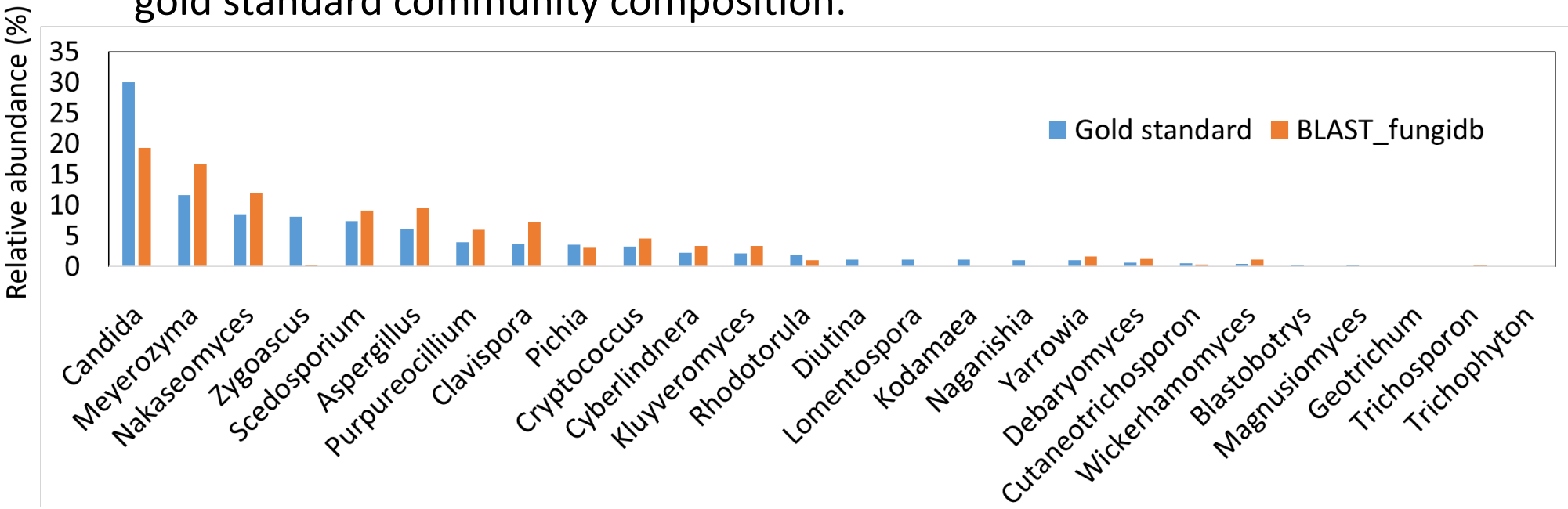
- Constructed reference database with only genomes in the mock.
- Using pairwise alignment (minimap2) to construct the ‘gold standard’ classification and composition.
- Compare different strategies with ‘gold standard’ for similarities.
- Apply cut-offs on alignment proportion and access the similarities’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

Candida rugosa	
Size (Mb)	13.9
Contig number	15
N/L50 (Mb)	2/3.2
BUSCO	92.3%
Candida mesorugosa	
Size (Mb)	16.6
Contig number	13
N/L50 (Mb)	2/3.6
BUSCO	91.6%
Cryptococcus magnus	
Size (Mb)	26.6
Contig number	188
N/L50 (Mb)	12/0.8
BUSCO	87.6%

Comparing community composition analysis using Bhattacharyya (B) distance

B distance (angle ^{o}):
A measurement of absolute distance between two lists of probabilities

- Blastn against Refseq_fungi database has the closest B distance with the gold standard community composition.



Comparing community composition analysis using Bhattacharyya (B) distance

B distance (angle ^{o}):
A measurement of absolute distance between two lists of probabilities

Running hypothesis:

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

Next step: tracing 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!

Take-home messages

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.

Acknowledgements

Supervisors:
John Rathjen
Benjamin Schwessinger
Eric Stone

NSW Department of Primary Industry:
Andrew Milgate
Michael McCaig
Tony Goldthorpe

The University of Sydney:
Wieland Meyer
Laszlo Irinyi

The Rathjen Lab
The Schwessinger lab

Thank you for listening!