Pathogen communities in agriculture:
(devils in the detail)

Leighton Pritchard
Metabarcoding for plant pathogens

Three “pinch points”:

- Are (non-model pathogen) barcode databases comprehensive?
  - riboSeed: mining 16S from Illumina reads
- Are standard barcodes precise enough (for non-model pathogens)?
  - *Pectobacterium atrosepticum* in Scotland
- How do non-standard barcodes perform?
  - How individualizing are oomycete ITS1 sequences in metabarcoding?
riboSeed

Joint-funded PhD

- ≈200 *E. coli* sequenced from lysimeters in Ireland
- Lysimeters contain soil isolated from local environment
- Multiple sites
- What are key differences between *E. coli* from lysimeters and elsewhere?
  - (genome composition, toxins, etc.)
Why riboSeed?

The quest for genome closure

- 16S central to bacterial microbiome (metabarcoding) studies
- Bacterial genomes may contain several (often similar) 16S
- 16S doesn’t assemble well with Illumina
- Many published genomes collapse 16S
  - 16S is longer than most Illumina reads

Is 16S variation underrepresented in bacterial genome assemblies, and therefore in 16S databases?

Better 16S/genome assemblies leads to better 16S databases.
Helpful neighbours

rDNA flanking regions differ within but not between genomes

- Plots of entropy (sequence variation) show flanking regions conserved

1. Identify 16S regions from complete reference genomes
2. Assume that 16S locations are conserved
3. ‘Anchor/pin’ unique reads to flanking regions and reference rDNA
4. ‘Bridge’ by distributing common reads
riboSeed algorithm

de fere novo

de novo

1) Annotated Reference genome
2a) Initial mapping
2b) Subassemblies of partitioned reads
2c) Mapping to concatenated contigs
2d) Subassemblies of partitioned reads
2e) Repeat from 2c
3a) De fere novo assembly
3b) De novo assembly
4) Results
riboSeed: leveraging prokaryotic genomic architecture to assemble across ribosomal regions

Nicholas R. Waters1,2, Florence Abram1, Fiona Brennan1, Ashleigh Holmes1 and Leighton Pritchard2

1Microbiology, School of Natural Sciences, National University of Ireland, Galway, H91 TX33, Ireland. 2Information and Computational Sciences, James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland. 3Bio and Environmental Microbiology, Environmental Research Centre, Teagasc, Johnstown Castle, Wexford, Y35 TC79, Ireland and 'Cell and Molecular Sciences, James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland.

Reassembled November 24, 2017; Revised February 15, 2018; Editorial Decision March 9, 2018; Accepted March 12, 2018.

riboSeed performance on GAGE-B benchmark genomes

<table>
<thead>
<tr>
<th>organism</th>
<th>strain</th>
<th>rDNAs</th>
<th>de novo</th>
<th>riboSeed</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. cereus</td>
<td>NC7401</td>
<td>14</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>S. aureus</td>
<td>M0927</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>V. cholerae</td>
<td>N16961</td>
<td>8</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>X. axonopodis</td>
<td>UA323</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Reassembly of legacy genomes may improve 16S variant representation
**Pectobacterium atrosepticum**

RESAS/SSCR-funded research

- *P. atrosepticum* (*Pba*) is a major pathogen in Scotland
  - One of several Soft-Rot Enterobacteria (SREs)
  - Plant cell wall-degrading enzymes (PCWDEs)
- Detect, classify and track *Pba* in Scottish agriculture
  - Method development, new diagnostics
Historical taxonomy of Pba

- **Erwinia ‘bucket’** (multiple genera)
- **Pectobacterium spp.**
- **P. chrysanthemi → Dickeya**
- Old names hold over in collections and databases
Whole genome classification

- Existing classifications need some revision
- Database errors: not all published genomes annotated with correct taxonomy

https://widdowquinn.github.io/pyani/
Genus-level misclassification

[Image of a heatmap and dendrogram with various species names like Dickeya_parasitica, Dickeya_ellipsis, etc.]
Legislation is taxonomy-based

- European and Mediterranean Plant Protection Organisation (EPPO):
  - “member states should regulate *D. dianthicola* and *E. amylovora* as quarantine pests (A2 list)”

- Seed Potatoes (Scotland) Amendment Regulations (2010):
  - “zero tolerance policy for all *Dickeya* spp. on potatoes in Scotland…”

- Assumes taxonomy is assigned precisely and correctly
  - (also assumes taxonomy is a good proxy for risk...)

- Detection, tracking, and people’s livelihoods depend on accurate classification, and good-quality training and test sets
Establishing a baseline

- Sequence and classify *Pba* across Scotland
- Improve on existing MLST/VNTR/16S/qPCR schemes
  - Trained on ‘old’ classifications, not genome data
- Hutton holds historical Scottish pathogen isolates
  - Geographical/long-term variation
- Contemporary isolates from SASA Scotland
  - Current spread: historical vs introduction
- Develop sequence markers to identify/track *Pba* intra-species
50 *Pba* isolates (2009-2015)

- 50 *Pba* isolates sequenced from Scottish potato infections (2009-2015)
  - Whole-genome SNP tree (parSNP)
Pba SNP distribution
Geographical distribution

- All isolates sampled from Aberdeenshire
  - More variation in one county than previously known globally?
- Stable coexistence of variants (same 16S...)?
  - Why such variation?
- Sequencing >100 more isolates
Alternative genomic markers

- Adapting diagnostic primer design software
  - https://widdowquinn.github.io/find_differential_primers/
Vertical or horizontal transmission?

- *Pba* acquired from seed, environment, or both?
  - **Can** bulk sample seed before planting
  - **Can** sample from tubers (diseased/healthy) in field
  - **Cannot** bulk sample the field

- Use Bayesian networks to infer “causality”
  - Relative virulence: $P(\text{symptoms} | A, \text{not } B \text{ on tuber})$
  - Vertical transfer: $P(A \text{ on tuber} | A \text{ on seed})$
  - Environmental transfer: $P(A \text{ on tuber} | A \text{ not on seed})$
  - Virulence: $P(\text{disease} | A \text{ on tuber})$
Phytothreats

BBSRC/RESAS-funded collaboration

- Phytophthora spp. cause disease on crops, ornamentals
  - Many Phytophthoras invasive, and/or notorious e.g. *P. ramorum, P. austrocedri, P. infestans*
- Track and trace Phytophthora spp. through plant nurseries
  - Accreditation scheme
  - Advice on “clean practice”
  - Stakes for business if quarantine organisms discovered
ITS1 Identification

- ITS1 used for *Phytophthora* phylogenetics
  - “One species, one ITS1”
  - Some species separated by a single SNP
ITS1 Metabarcoding

- A popular approach

**Molecular Ecology Resources**

**RESEARCH ARTICLE**

Pyrosequencing of environmental soil samples reveals biodiversity of the *Phytophthora* resident community in chestnut forests

Andrea Vannini, Natalia Bruni, Alessia Tomassini, Selma Franceschinii & Anna Maria Vettraino

Department for Innovation in Biological, Agricultural and Forest Systems (DBAF) - University of Tuscia, Viterbo, Italy

**Distribution and diversity of Phytophthora across Australia**


**Techniques**

Metabarcoding Analysis of Phytophthora Diversity Using Genus-Specific Primers and 454 Pyrosequencing


**PLOS ONE**

**RESEARCH ARTICLE**

The Use of Genus-Specific Amplicon Pyrosequencing to Assess *Phytophthora* Species Diversity Using eDNA from Soil and Water in Northern Spain

Santiago Catala, Ana Pérez-Gierra, Paloma Abad-Campos

1 Instituto Agrofísico Mediterráneo, Universal PO Avinguda de València, Comarca de Vena s/n, Valencia, Spain.
2 Forest Research, Alice Holt Lodge, Farnham, Surrey, United Kingdom

**ORIGINAL ARTICLE**

ITS1 metabarcoding highlights low specificity of lichen mycobiomes at a local scale

Fernando Fernández-Mendoza, Antonia Fleischhacker, Theodora Kopun, Martin Grube, Lucia Muggia
Synthetic “ITS1” control sequences

- Four artificial, unique sequences
  - Same base composition as *Phytophthora* ITS1
    - random sequence
  - “Crosstalk”/contamination assumed if:
    - Synthetics found in sample channels
    - *Phytophthora* sequence found in synthetic channels

- Estimate sequencing error, and potential for (cross-) contamination.
Synthetic controls

Many extra sequences seen!

- Synthetic mixture – four controls only:
  - Target read counts: $\approx 1e6$, $\approx 1e4$, $\approx 1e3$, $\approx 1e2$
  - Each mixture at three dilutions (sensitivity):
    - 1x, 10x, 100x
  - Six combinations of synthetics (GC1-3, GL4-6)
  - Long tail of amplicons that not expected in the original mixture
  - How many (different) sequence variants are there?
PCR amplification introduces variation

- Most abundant control has most variants
  - Sequence variants can be more abundant than sample sequences!
- Variants differ from control sequence by up to \( \approx 5-8\% (~10-15\text{bp}) \)
- Most abundant control equally abundant at all dilutions
  - Other controls become undetectable with dilution: NOT QUANTITATIVE
- “unmatched” non-control sequences present up to \( \approx 1e2 \) abundance
- The most abundant sequence may not be in the sample!
Synthetic controls: DADA2
- Most variation is excluded, but...
  - ITS1 sequences truncated
  - PCR variants still come through
  - PCR variants still more abundant than controls
Cross-contamination

Estimate from synthetic controls

- BLASTN+ non-control sequences vs NCBI nt
- Identifiable oomycete sequences at up to ≈25 abundance
  - Possible cross-contamination filtering threshold?
- ≈50% of variants unique to each lane
  - Variation introduced in PCR amplification, not in original sample?
“One species, one ITS1”? 

- Genomics: “One isolate, ≈40-170 ITS1” 
  - (and they can be quite diverse, as in *P. infestans...*)
Single-isolate controls

- One *Phytophthora* isolate per well/lane/barcode
  - Also synthetic controls and environmental samples
  - Some controls amplified with different PCR cycle numbers

Estimate within-isolate variation

- How many OTUs per individual *Phytophthora* isolate?
- Do similar OTUs make individualization difficult?
Many unique ITS1s in all samples

- Identify ITS1 matches:
  - Profile HMM of known *Phytophthora* ITS1 sequences
    - Contamination abundance up to $\approx 10^3$!

**Distinct amplicons by lane/barcode**

- Single isolates: up to $\approx 2000$
- Environmental isolates: up to $\approx 3000$
- Synthetic controls: up to $\approx 2000$

**Log 10 abundance**
Several OTUs (swarm) for each isolate

- Are OTUs good representations of diversity?

Swarms/OTUs by lane/barcode

**Single isolates**: up to \( \approx 150 \)

**Environmental isolates**: up to \( \approx 450 \)

**Synthetic controls**: up to \( \approx 20 \)

**Most Phytophthora isolates**: \( \approx 40-60 \)
Remove contamination

- Reduces OTUs per isolate
  - Remove all amplicons <25 abundance
  - Use synthetic controls on every plate to threshold?

Swarms/OTUs by lane/barcode

- Single isolates: up to $\approx 6$
- Environmental isolates: up to $\approx 40$
- Synthetic controls: up to $\approx 4$

Most Phytophthora isolates: $\approx 1-2$
Are the reduced OTUs individualizing?

- Jaccard distance between swarm/OTU members
  - Several OTUs unique to *Phytophthora* species
  - Some OTUs common to *Phytophthora* species
- ITS1 can be individualising:
  1. Threshold for contamination/low abundance sequence variants
  2. Filter against ITS1 HMM profile
Summary

- Improving non-model pathogen barcode coverage
  - **riboSeed**: improve 16S assembly from (legacy) Illumina genomes

- Assessing performance of alternative barcodes
  - **pdp**: automate design of alternative barcodes
  - **synthetic controls**: cross-contamination/methodological artefacts
  - **single-isolate controls**: individualization/potential for confusion
Acknowledgements

Emma Campbell
Greig Cahill
Beatrix Clark
John Elphinstone
Pete Hedley
Ewan Mollison
Eva Randall

Paul Sharp
Lauren Watts

Microbiology Society