

Intraspecies-Resolution Metabarcoding: automated primer design and a plant pathology case study

Leighton Pritchard^{1,2}, Sonia Humphris², Emma Campbell²,
 Lauren Watts², Ed Haynes³, Hollie Pufal³, Greig Cahill⁴,
 Triona Davey⁴, John Elphinstone³, Ian Toth²

- 1. Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, Scotland, UK
- 2. Cell and Molecular Sciences, James Hutton Institute, Dundee, Scotland, UK
- 3. Fera Science, York, England, UK
- 4. Science & Advice for Scottish Agriculture, Edinburgh, Scotland, UK



1. Tracking *Pectobacterium atrosepticum* below species level

Pectobacterium atrosepticum (Pba) is a bacterial plant pathogen, endemic in Scotland, that causes blackleg/soft-rot in potato.



Figure 1. Soft rot and blackleg on potato, caused by *P. atrosepticum*

To track whether Pba infections are introduced from seed potato, or from environmental reservoirs, we need to be able to discriminate at high-throughput between near-clonal Pba variants. Existing marker-based Pba tracking methods cannot do this well.

VNTR, MLSA: provide **sub-species specificity**, (Eight-gene MLSA identifies eight types of Pba) but **require isolation of individual strains**
16S: **high-throughput, does not require strain isolation**, but ASVs (amplicon sequence variants) **do not provide sub-species resolution**

2. Designing novel marker sequences

Suitable for Illumina
200bp amplicon for high-throughput sequencing and multiplexing

Sequence-variable within Pba
discriminate between sub-species level Pba variants

Primers specific for Pba
recover only Pba from field samples

Even representation across Pba
prefer large variant groups to small

We used the `pdp` diagnostic primer design tool to design candidate primers that were specific to Pba and predicted to amplify a region that is sequence variable across ≈ 200 genome-sequenced Pba isolates.

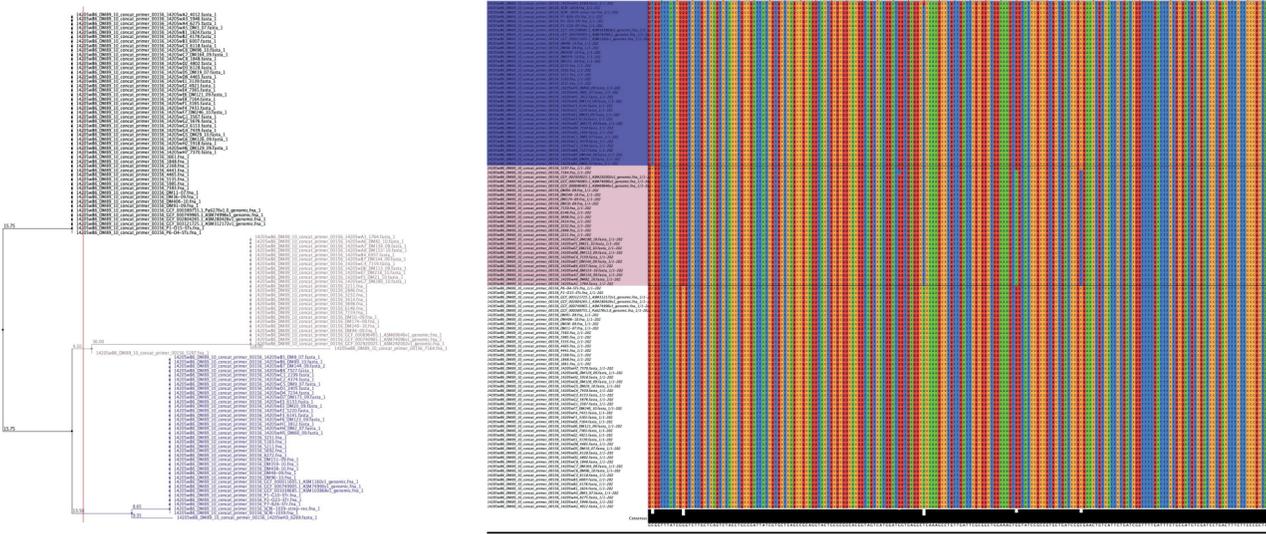


Figure 2. A 202bp amplicon with five SNVs, discriminating six Pba variants, was designed using `pdp`. An alignment of the amplicons is shown (right), alongside a dendrogram of sequence variants produced using Hamming distance.

3. Performance in the field

Eight-gene **MLSA**, applied to 46 Pba isolates from three adjacent fields, identified **six sub-species variants**. Our **novel primers** discriminated **five sub-species** variants in the same data.

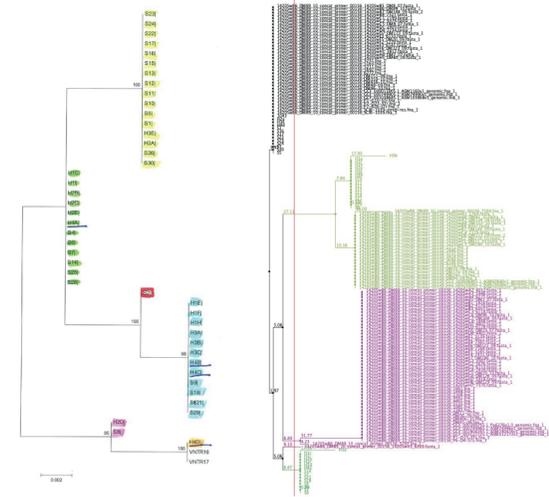


Figure 3. Eight-gene MLSA (left) identifies six variants in 46 Scottish field isolates of Pba. Our novel primers (right) identify five sub-species variants, two of which are unrepresented in known whole-genome sequence data for Pba.

Twelve distinct Pba sub-species variants were identified in 134 symptomatic plants sampled from four fields across two farms in Scotland.

Multiple Pba variants were identified in each field.
A single plant may carry distinct Pba variants on stem and tuber.
A single lesion may contain multiple Pba variants.

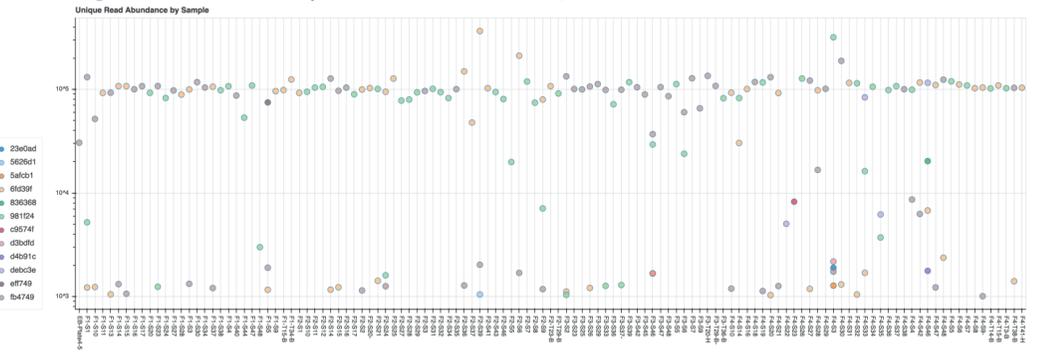


Figure 4. Amplicon counts, coloured by amplicon variant, from 134 samples across four fields (F1-F4). Samples labelled S or T are from stem and tuber respectively. Image generated using `pymetabc`.

The authors gratefully acknowledge funding support from The Scottish Government, and from the Scottish Society for Crop Research

