

# Project report for SSCR-Potatoes

## ***Project title:***

A Genomic Atlas of Scottish *Pectobacterium* (GASP): An underpinning resource for *Pectobacterium atrosepticum* diagnostics and outbreak detection.

## ***Applicant(s):***

Dr Leighton Pritchard; Dr Sonia Humphris

## ***Background to the project***

*Pectobacterium atrosepticum* (Pba) is a bacterial pathogen of potato causing blackleg on infected plants, having significant economic impact in Scotland, and globally.

We sequenced genomes of ~40 Pba strains recently isolated (2009-2014) from diseased potato plants on Aberdeenshire farms. Reconstructing their evolutionary history revealed that Pba is divided into four distinct subgroups not distinguished by current diagnostics. These subgroups are represented in recent Scottish infections but no subgroup is unique to Scotland, and we do not have a long-term account of their presence. It is possible that further subgroups remain to be discovered.

Several alternative scenarios could explain this observed pattern of Pba subgroups. For example, all Pba subgroups may coexist naturally alongside each other (perhaps worldwide). At the other extreme, distinct populations of pathogens that evolved in geographical isolation may have been mixed and spread by recent agriculture and plant trade. It would be possible to distinguish between these and other alternative scenarios by sequencing historical (1950-present) Scottish Pba isolates from JHI collections. This would provide insight into the spread and stability of Pba in Scotland, directly enabling development of more accurate diagnostics that could monitor Pba presence and transfer in greater detail, to inform control measures.

## ***Aims and objectives***

We aim to produce a map of Pba diversity and evolution across Scotland, over the last ~50 years. This will establish earliest known dates of observation and location for each Pba subgroup in Scotland, and distinguish between environmentally persistent/endemic and recently introduced groups. We will determine the extent of evidence for gene transfer within Pba population groups, to understand the risk of spread of novel virulence genes (e.g. with new introductions of Pba). These goals will help establish whether the prevalence and diversity of Pba in Scotland is relatively constant, or whether it has arisen through agricultural/ornamental trade and imports.

## Research results

### Identification and sequencing of Scottish *Pectobacterium* isolates

The JHI bacterial collection contains over 1000 isolates sampled since ≈1950. The metadata associated with these isolates had originally been recorded as hand-written notes in a set of notebooks (figure 1) and, although transcribed to Excel format prior to the project, required additional manual renormalisation and rewriting in order to ensure data integrity for digitisation. A local database (SREdb) with a modern browser interface was written to hold the collection's metadata (figures 2-4) to aid selection of Scottish *Pectobacterium* isolates with broad geographic coverage, and coverage of the range of hosts and material sources available for sequencing.

In total 130 candidate isolates (viable at last extraction) were identified. 119 of these candidates had not previously been sequenced (111 "*Erwinia carotovora atroseptica*"; 8 *P. atrosepticum*). Seven were derived from field trials (source not recorded), 104 from potato (43 from stem, 55 from tuber; the remainder unknown), three from soil and two from water sources. Eighty-five of these isolates were originally sampled between 1960 and 1990.

The isolates were batched into two blocks for sequencing, selected from the collection and grown up from stocks. The isolates were sequenced at the MicrobesNG commercial service.

### Bioinformatic analyses

The sequenced genomes were assembled and annotated. Initial analysis indicated that seven collection isolates were not *Pectobacterium*. The remaining isolates were combined with previously sequenced Scottish isolates (obtained from SASA) and 150 publicly-available *Pectobacterium* genomes for classification analyses, including ANIm, in silico MLST sequence typing, and SNP tree analyses. The isolates were also used collectively to aid design of sub-species level metabarcoding primers in a complementary Scottish Government project, for outbreak and contamination detection and tracking.

ANIm analysis, taking a species-level threshold of 95% identity and at least 50% genome coverage, indicates that the remaining Scottish isolates are all *P. atrosepticum* (figure 5). SNP tree analysis also places the sequenced Scottish isolates within *P. atrosepticum*.

In silico MLST analysis places the Scottish *Pba* isolates under five sequence types, only one of which (ST7) was previously recorded in the PubMLST reference database. SNP tree analysis of all available *Pba* isolates divides them into five major clades (figure 6) broadly corresponding to MLST type. A novel clade consisting of five historical Scottish isolates dating from central Scotland in the period 1966-1981 was found, in addition to the four clades previously known.

The SNP trees indicate interesting features of the dataset. Each major clade contains representatives sampled over several decades in a range of locations, suggesting that the clade groupings, and strains, are distinct and stable populations coexisting geographically in Scotland over decades. Clades representing geographically circumscribed isolates (e.g. from the Black Isle) are also identifiable. One clade contains inoculated samples recovered from a single 2014 field trial, which can be assumed to represent a measure of relative clonality. A similar level of genomic variation is seen in isolates spanning a five-year period on the East coast from Black Isle to Fife, indicating that whole genome sequencing would be a feasible means of tracking outbreaks and inferring transfer chains. Analysis of the data will continue under complementary projects.

## *Outcomes*

1. Provide an unprecedented historical genomic atlas for analysis of a major cause of potato disease in Scotland, spanning the nation over six decades. It will deliver a 'background map' of the diversity and evolution of this pathogen in Scotland, against which the potential impacts and origins of recent isolates can be assessed and interpreted.
2. Improve our ability to design accurate and specific diagnostics for rapid in-field detection and identification of pathogens, and deliver greater understanding of the evolution of *Pba*, to help inform control measures.

We developed a database cataloguing the Hutton isolate collection, and identified ≈100 isolates spanning six decades, covering Scotland from the Black Isle to the Borders. We grew and extracted these isolates for sequencing at the MicrobesNG sequencing service and combined the data with other *Pectobacterium* genomes, including public information, to conduct a comprehensive whole-genome classification of each isolate using ANIm, in silico MLST, and SNP trees. This allowed us to produce a background map of the geographical and genomic variation, and evolutionary diversity, of *P. atrosepticum* in Scotland. These maps will inform future work, as a baseline of historical genomic variation in *Pba* across Scotland.

The sequences also enabled design of sub-species level diagnostic molecular metabarcoding tools for *Pectobacterium* variants in complementary work funded by the Scottish Government to determine the relative influence of seedborne and environmental pathogen load. The diagnostic tools have been validated against recent SASA field isolates, and collection isolates held at Fera, confirming their specificity and discriminatory capacity.

We anticipate that, taken together, the diagnostic tools and survey of historical genomic variation informed by this work will aid growers, industry and scientists better appraise and respond to future outbreaks of *Pba*.

## *Next steps*

With further funding and development, the SREdb database developed for this work could become a persistent culture collection database dedicated to plant pathogens held in the JHI collections.

Both the comparative genomic and diagnostic tool outputs will be the topic of future outputs (papers, posters, presentations), and data and tools released to the community.

The sequence data generated with this funding will feed into applications for future funding for the phylogenomic classification software tool `pyani`, and the diagnostic primer/marker design software tool `pdp`. Funding will be sought with collaborative groups to test and validate the diagnostic tools that are generated.

# Appendix

174	HPV mutant	Pba SCR1 1043	Tn5 QM1	Kan <sup>R</sup>
175	<i>Erwinia amylovora</i> 1430	<i>E. amylovora</i> - France	PI8 H21 +ve for HR in potato	clone from reversion
176	Pba 2109 + pGEM-Teasy	Pba SCR1 1043	pGEM-Teasy	Amp <sup>R</sup>
177	Pba 4430 transduced mutant	Pba SCR1 1043 PI G-15	Tn5 QM1	Kan <sup>R</sup>
178	Pba ehpa transduced mutant	Pba SCR1 1043	Tn5 QM1	Kan <sup>R</sup>
179	Pba ehpa transduced mutant	Pba SCR1 1043	Tn5 QM1	Kan <sup>R</sup>
180	Pba 4430 complemented mutant	Pba SCR1 1043	pGEM-Teasy-4430	Amp <sup>R</sup> Kan <sup>R</sup>
181	Pba 4430 pGEM-T	Pba SCR1 1043	pGEM-Teasy EMPTY VECTOR	Amp <sup>R</sup> Kan <sup>R</sup>
182	Pba 0482 pGEM-32	Pba SCR1 1043	pGEM-32 EMPTY VECTOR	Amp <sup>R</sup> Kan <sup>R</sup>
183	<i>Dickeya</i> IPO 2222	<i>Dickeya</i>	New clade for from Jan Van der Wolf Netherlands	
184	<i>E. coli</i> EPI100™ TIR containing cosmid 642 with Pba TISS cluster	<i>E. coli</i> EPI100™-TIR	Phage T1-resistant EPI100™-TIR E.coli. PWEB-TNCT™ epizyme cosmid	Amp <sup>R</sup> TC <sup>R</sup> 2.5
185	<i>E. coli</i> EPI100™ TIR containing	<i>E. coli</i> EPI100™-TIR	Phage T1-resistant EPI100™-TIR E.coli	Amp <sup>R</sup> TC <sup>R</sup> 2.5

Figure 1: Handwritten Hutton isolate collection records.

SREdb Home   Isolates

## Welcome to the James Hutton Institute SRE collection database

This database contains records of soft-rot Enterobacterial (SRE) isolates held by the James Hutton Institute.

- Welcome to the James Hutton Institute SRE collection database
  - How to use the database
  - The isolate table
  - The Accession table
  - Georegions

### How to use the database

This database provides navigation and inspection of metadata concerning the isolates held in the Hutton plant pathogenic bacteria isolate collection.

The top menu bar provides navigation options:

- SREdb Home**: will return you to this welcome page
- Isolates**: will take you to a sortable table of all isolates held in the collection

### The Isolate table

The main **isolate table** provides summary information about each isolate held in the collection, concerning accession information, current taxonomic status and known viability, and the source material and location from which it was sampled.

SREdb Home   Isolates

1073 isolates returned

Accession	Taxon class	NCBI Taxon ID	Viable?	Mutants	Resistances	Source Country	Source Location	Source	Source NCBI Taxon ID	Sample Date	Submit Date
SCR 0001	<i>Erwinia carotovora atroseptica</i>	29471	Yes	-	-	Israel	None	potato	4113	None	None
SCR 0002	<i>Erwinia carotovora atroseptica</i>	29471	Yes	-	-	USA	None	potato	4113	1950	None

Figure 2: Hutton SREdb landing page

### Geographical region bounded by (54.4, -14.0), (61.0, -0.3)

274 isolates returned



Accession	Taxon class	NCBI Taxon ID	Viable?	Mutants	Resistances	Source Country	Source Location	Source	Source NCBI Taxon ID	Sample Date	Submit Date
SCRI 0009	<i>Erwinia carotovora atroseptica</i>	29471	Yes	-	-	UK	None	tomato	4081	None	None
SCRI 0010	<i>Erwinia carotovora carotovora</i>	555	Yes	-	-	UK	None	brassica	3705	None	None
SCRI 0011	<i>Erwinia carotovora atroseptica</i>	29471	No	-	-	Scotland	None	tomato	4081	None	None
SCRI 0012	<i>Erwinia carotovora atroseptica</i>	29471	Yes	-	-	Scotland	None	potato	4113	1985	None
SCRI 0013	<i>Erwinia carotovora atroseptica</i>	29471	Yes	-	-	Northern Ireland	Belfast	potato	4113	1961	None

Figure 3: Hutton SREdb interactive map of Scottish isolates

### SCRI 1043 *Erwinia carotovora atroseptica*

Isolate Information

Property	Value
Accession:	SCRI 1043
Taxonomic class:	<i>Erwinia carotovora atroseptica</i>
NCBI Taxonomic ID:	29471
Viable?:	Yes
Licensed?:	No
Synonyms:	Marshall B/L 1a ATCC BAA 672
Test notes:	Seq strain, Mol strain Warwick Univ., PCR(DaB), a ms gluc +, O/F +/+, Cat +, Ck +, CVP 33+, No HR in N.benth and N.clevis; Fixes N2
Source:	potato
Source NCBI taxon ID:	4113
Source tissue:	stem
Source symptomatic?:	Yes
Source country:	Scotland
Source location:	Dunning
Source address (approximate):	Dunning, Perth and Kinross, Scotland, PH2 0PS, UK
Source lat/lon (approximate):	56.3120026 -3.5861107
Sampling date:	1985
Sampler:	L. Hymán
Submission date:	1985
Submitter:	L. Hymán
Submitting country:	Scotland
Submitting location:	Invergowrie



Figure 4: SREdb accession page for SCRI 1043

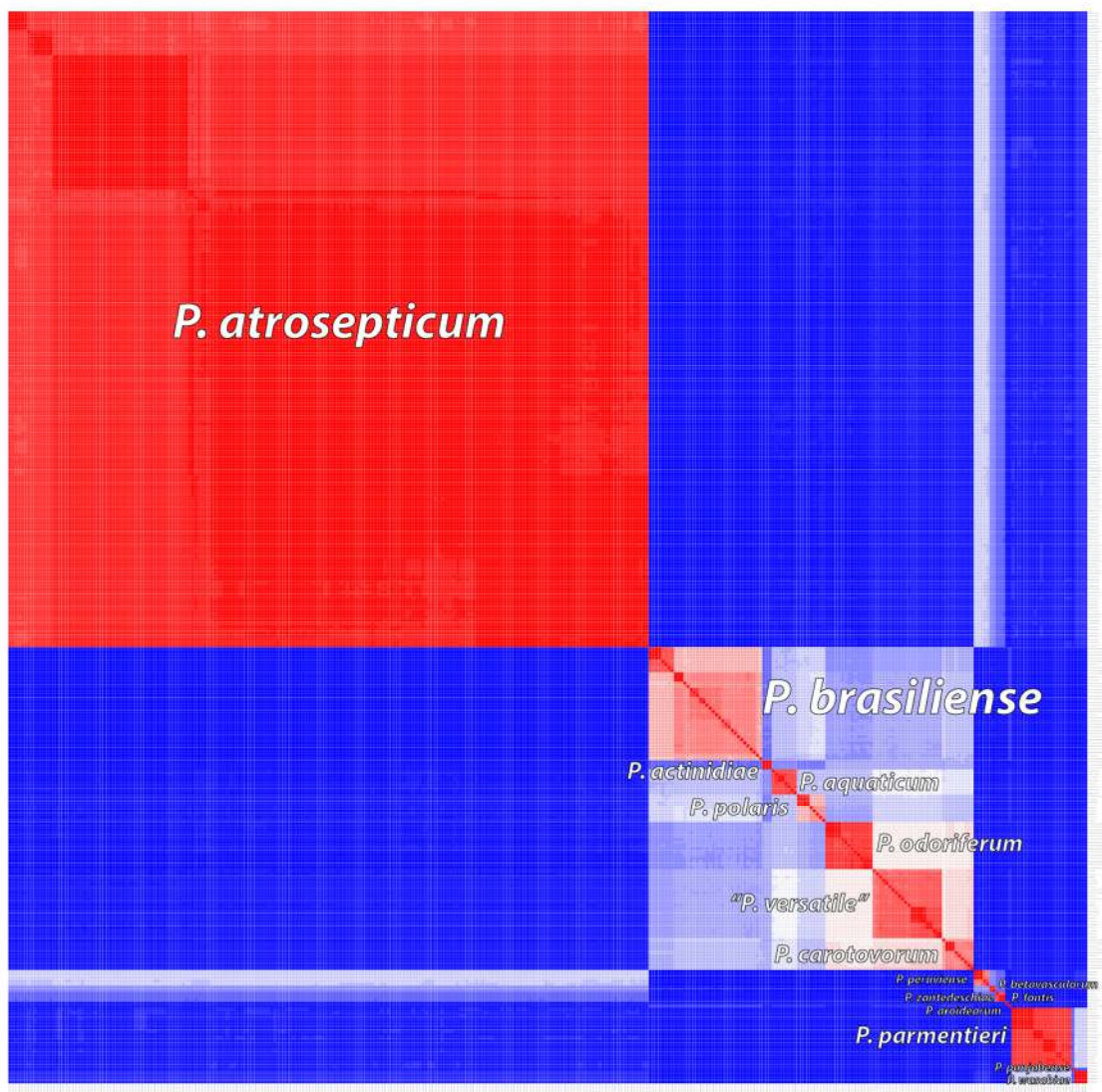


Figure 5: ANIm analysis of over 300 sequenced *Pectobacterium* genomes. Red boxes along the diagonal indicate species determined by whole genome comparison. Fourteen *Pectobacterium* species are indicated. All Scottish historical isolates are found in the *P. atrosepticum* grouping.

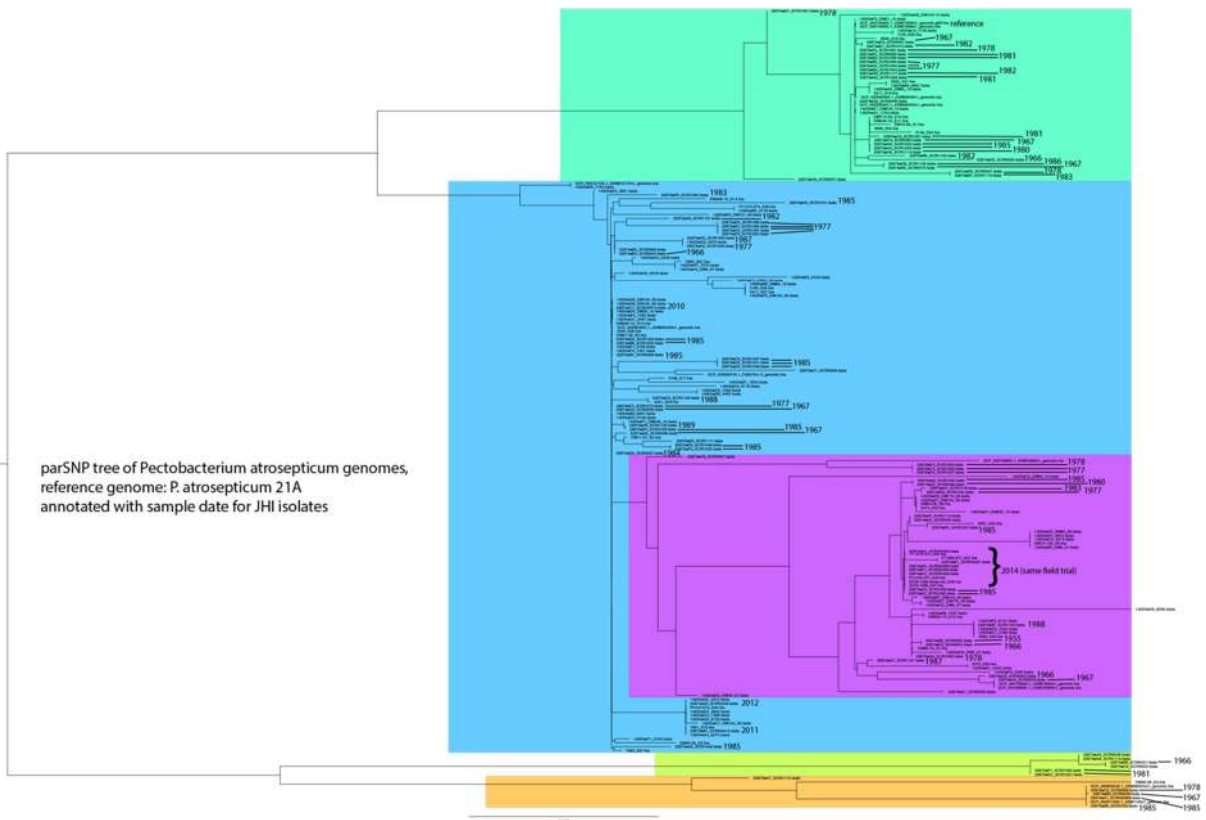


Figure 6: parSNP tree of 203 *Pba* genomes. Background colours indicate major SNP variant clades. Year of isolate sampling is indicated for historical isolates. There are five major clades (each indicated in a different colour). These major clades correspond broadly to *in silico* MLST classifications. Although most isolates are from Scotland, the publicly available sequences from New Zealand, Belorussia, Canada, Russia and China are also found within each of these clades. There is considerably more genomic variation in these Scottish isolates than was previously available in public databases. Essentially clonal groupings are observed for inoculated bacteria in a field trial (indicated), some isolates identified over several decades in a single location (indicating stable local persistence), and some isolates identified over a larger geographical region in a short time period (suggestive of transfer).