In A Different Class?

Whole genome classification of bacterial potato pathogens

The James Hutton Institute

Leighton Pritchard[1,2,3,4]
[1]Information and Computational Sciences,
[2]Weeds, Pests and Diseases,
[3]Centre for Human and Animal Pathogens in the Environment,
[4]Dundee Effector Consortium,
The James Hutton Institute, Invergowrie, Dundee, Scotland, DD2 5DA

Acceptable Use Policy

The James Hutton Institute

Recording of this talk, taking photos, discussing the content using email, Twitter, blogs, etc. is permitted (and encouraged), providing distraction to others is minimised.

These slides will be made available at http://www.slideshare.net/leightonp

Table of Contents

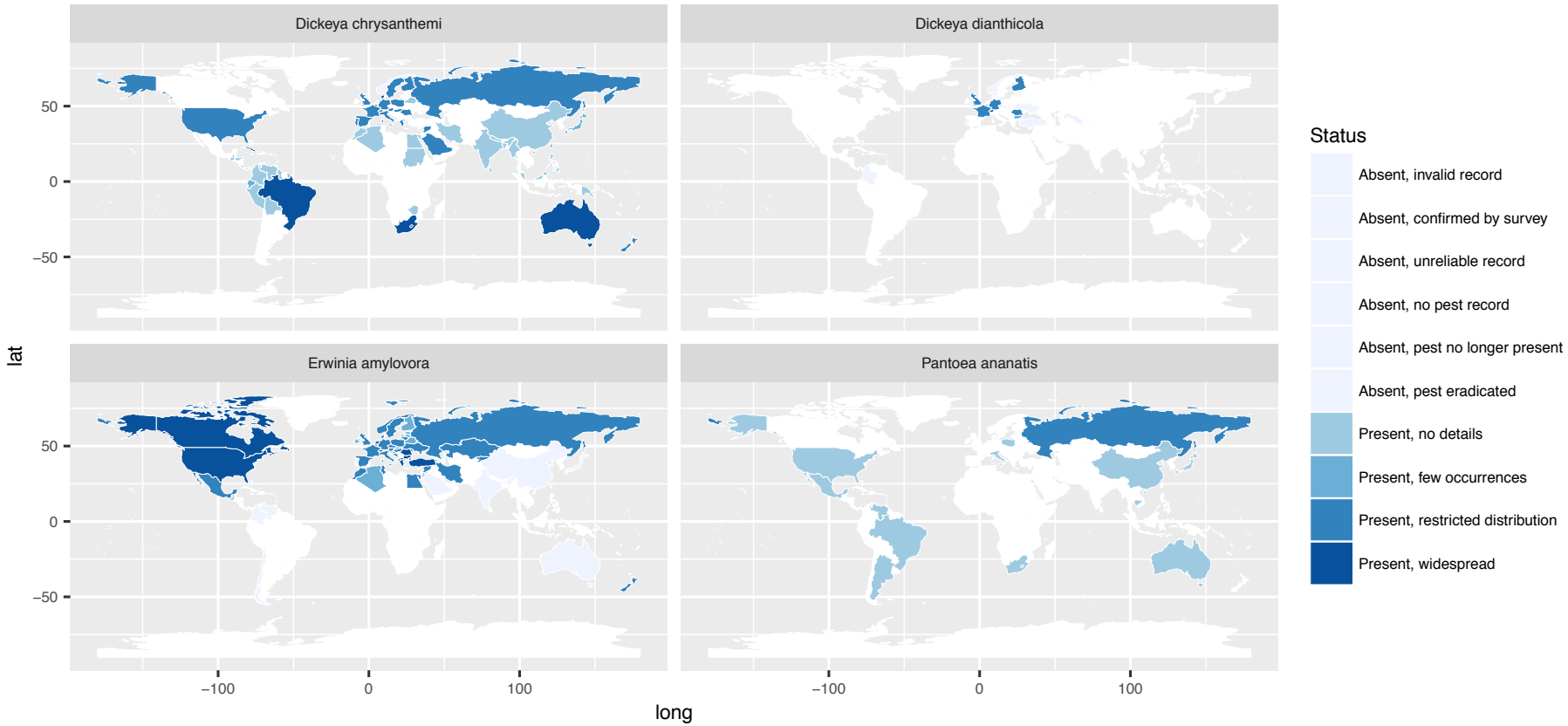
Plant-pathogenic Enterobacteria [a]

[a] Toth *et al.* (2006) *Annu. Rev. Phytopath.* doi:10.1146/annurev.phyto.44.070505.143444

- *Erwinia*, *Dickeya*, *Pantoea* and *Pectobacterium* spp.
- Plant Cell Wall Degrading Enzymes (PCWDEs)

Global threat [a]

[a] EPPO Global Database



A Tangled Taxonomy [a] [b]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h
[b] Williams *et al.* (2010) *J. Bact.* doi:10.1128/JB.01480-09

The James Hutton Institute

Enterobacterial taxonomy difficult to resolve, in general
Soft rot enterobacteria (SRE): *Pectobacterium* and *Dickeya*

Historical classification mostly polyphasic/phenotypic

- SRE originally *Erwinia* spp., now three distinct genera (*Dickeya*, *Pectobacterium*, *Erwinia*)
- *Pectobacterium* spp. used to be *E. carotovora* (and *E. chrysanthemi*)
- *Dickeya* spp. used to be *P. chrysanthemi*
- Not suitable for facultative bacteria?

Binomial nomenclature not designed for large amounts of genome data, or organised metadata curation

Tangled Taxonomy of SRE[a]

[a] Czajkowski *et al.* (2015) *Ann. Appl. Biol.* doi:10.1111/aab.12166

The James Hutton Institute

Old names hold over in the literature, collections, etc.

Name discontinuities affect analysis, databases

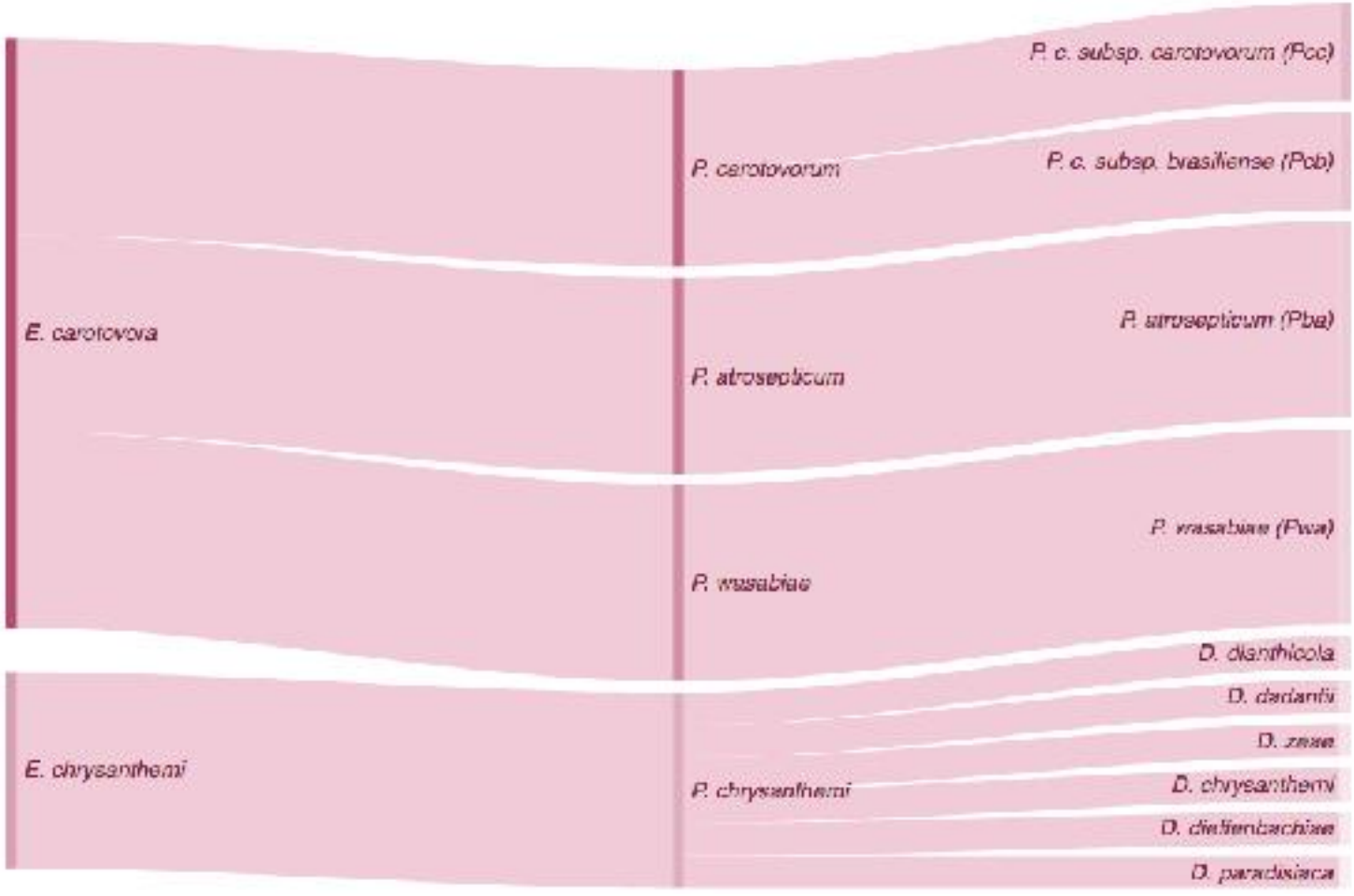


Table of Contents

Dickeya spp. moving across Europe [a] [b]

[a] Toth *et al.* (2011) *Plant Path.* doi:10.1111/j.1365-3059.2011.02427.x

[b] Parkinson *et al.* (2015) *Eur. J. Plant Path.* doi:10.1007/s10658-014-0523-5

D. dianthicola is established across Europe

D. solani is an emerging, encroaching threat

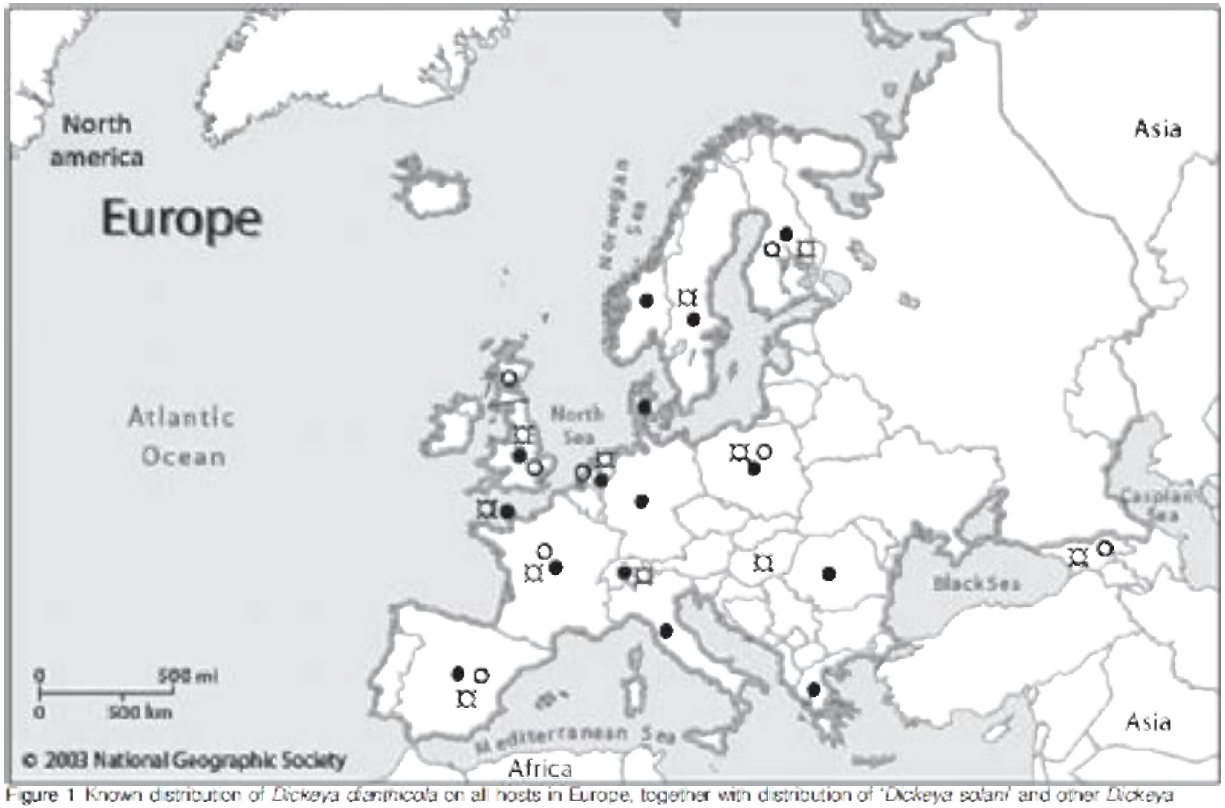


Figure 1 Known distribution of *Dickeya dianthicola* on all hosts in Europe, together with distribution of '*Dickeya solani*' and other *Dickeya* species on potato in Europe. Distribution of *D. dianthicola* on all hosts (updated from CAB International, 2006) (•); *Dickeya* spp. on potato (▫); '*D. solani*' on potato (○).

Legislation [a]

[a] Pritchard *et al.* (2015) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

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 to ensure production of 'clean' (disease-free) seed potato production for export

EUPHRESCO consortium

Euphresco : Control and epidemiology across Europe

Legislation by taxonomy [a]

[a] Pritchard *et al.* (2015) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

"Easy" to incorporate binomial nomenclature into legislation[1]

Assumption: taxonomy can be determined precisely

Taxonomy is a human-imposed hierarchical classification that truly reflects nature

Assumption: taxonomy is a proxy for disease risk

Sharing a common ancestor with another pathogen is the primary factor that influences virulence

[1] (easy for policy-makers to understand)

Issues with legislating by taxonomy [a]

[a] Pritchard *et al.* (2015) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

Is current bacterial taxonomy objective and correct?

- Taxonomy is 'vertical', but pathogenicity may be 'laterally' transferred (plasmid/transposon-borne, etc.) [2]
- Is a species concept even relevant for bacteria?

Mapping from taxonomy to phenotype is not one-to-one [3]

- Testing for disease phenotypes not exhaustive (facultative pathogens)
- Relationship between genome and disease/risk not fully understood

[2] Toth *et al.* (2006) *Ann. Rev. Phyto.* doi:10.1146/annurev.phyto.44.070505.143444

[3] Deans *et al.* (2015) *PLoS Biol.* doi:10.1371/journal.pbio.1002033

Table of Contents

Dickeya qPCR diagnostics [a] [b] [c]

[a] Pritchard *et al.* (2013) *Plant Path.* doi:10.1111/j.1365-3059.2012.02678.x
[b] Pritchard *et al.* (2013) *Genome Ann.* doi:10.1128/genomeA.00087-12
[c] Pritchard *et al.* (2013) *Genome Ann.* doi:10.1128/genomeA.00978-13

The James Hutton Institute

To legislate on or quarantine contaminated materials, one has to be able to identify and discriminate the pathogen

- Having sequenced 25 *Dickeya* isolates, we were approached to develop diagnostics at the species/isolate level
- qPCR is cheaper, quicker and easier than bacterial genome sequencing (for now, anyway...) [4]
- No qPCR primers existed to distinguish among *Dickeya* spp.

[4] Czajkowski *et al.* (2015) *Ann. Appl. Biol.* doi:10.1111/aab.12166

qPCR Primer Design [a]

[a] Pritchard *et al.* (2012) *PLoS One* doi:10.1371/journal.pone.0034498

The James Hutton Institute

1. Bulk predict primer sets on all chromosomes (`Primer3`)
2. Predict cross-amplification *in silico* (`primersearch`)
3. Evaluate *in vitro* against panel of previously "unseen" isolates of known class and report performance metrics

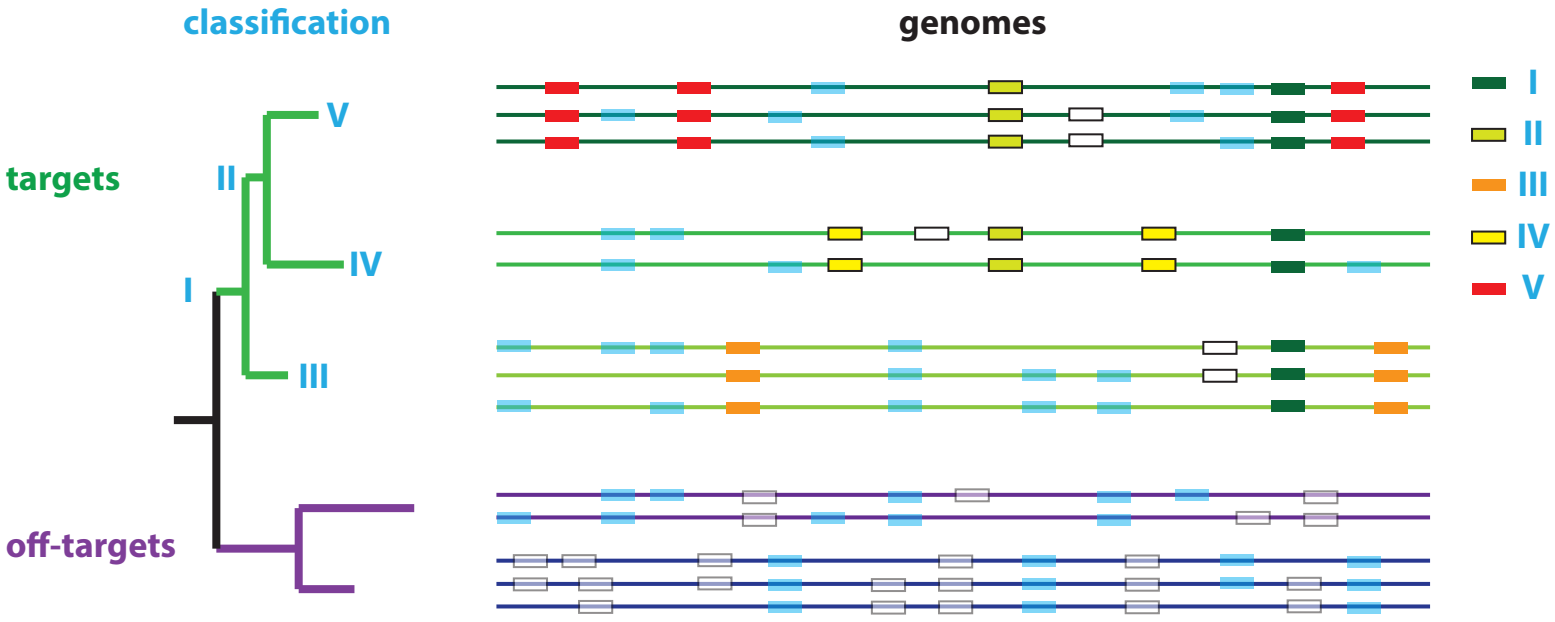


Table of Contents

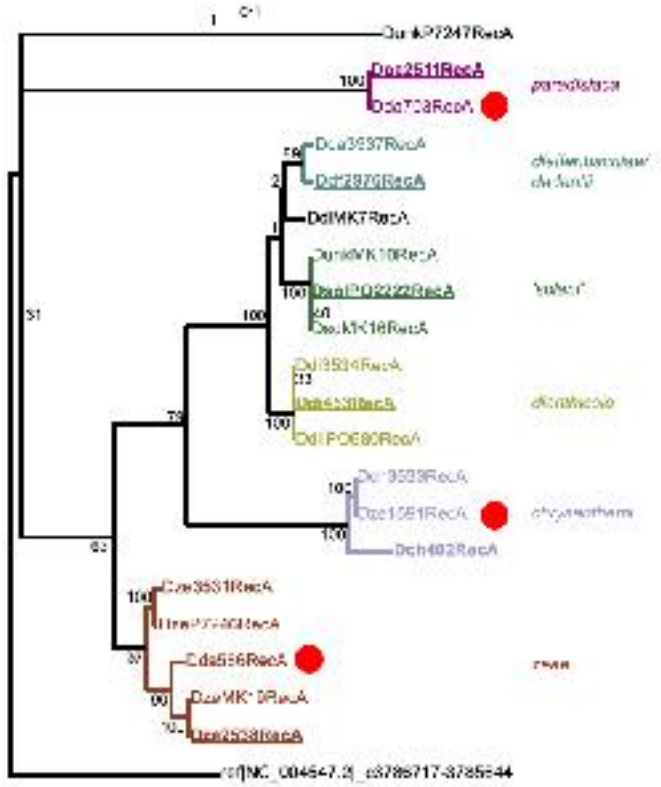
The James Hutton Institute

Classification is a problem! [a]

[a] Pritchard *et al.* (2013) *Plant Path.* doi:10.1111/j.1365-3059.2012.02678.x

The James Hutton Institute

First qPCR design gave no diagnostic primers for several *Dickeya*!



[5]

Misassigned species in GenBank made 'training' impossible.

[5] ML tree, *recA*

Consequences of misclassification [a] [b]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h
[b] Varghese *et al.* (2015) *Nucl. Acids Res.* doi:10.1093/nar/gkv657

The James Hutton Institute

Real-world consequences

- False positives (type I errors):
  clean samples rejected: economic cost
  farms quarantined/close: economic/societal cost
- False negatives (type II errors):
  (irreversible) introduction of infectious material
  potential for novel host jumps and spread

"Gold-standard", correctly classified training and test sets essential to estimate classifier error rates.

MiSI: 18% of NCBI bacterial genomes misclassified at species level

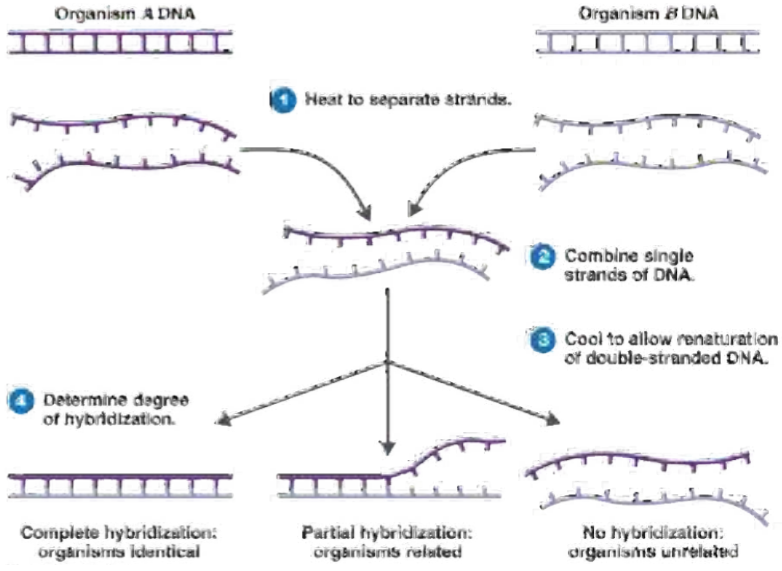
Accurate classification is essential!

Table of Contents

DNA-DNA hybridisation [a] [b]

[a] Morello-Mora and Amann (2001) *FEMS Micro. Rev.* doi:10.1016/S0168-6445(00)00040-1
[b] Chan *et al* (2012) *BMC Microbiol.* doi:10.1186/1471-2180-12-302

- "Gold Standard" for prokaryotic taxonomy, since 1960s. "70% identity ≈ same species."
- Denature DNA from two organisms.
- Allow to anneal. Reassociation ≈ similarity, measured as ΔT of denaturation curves.



Proxy for sequence similarity - replace with genome analysis?

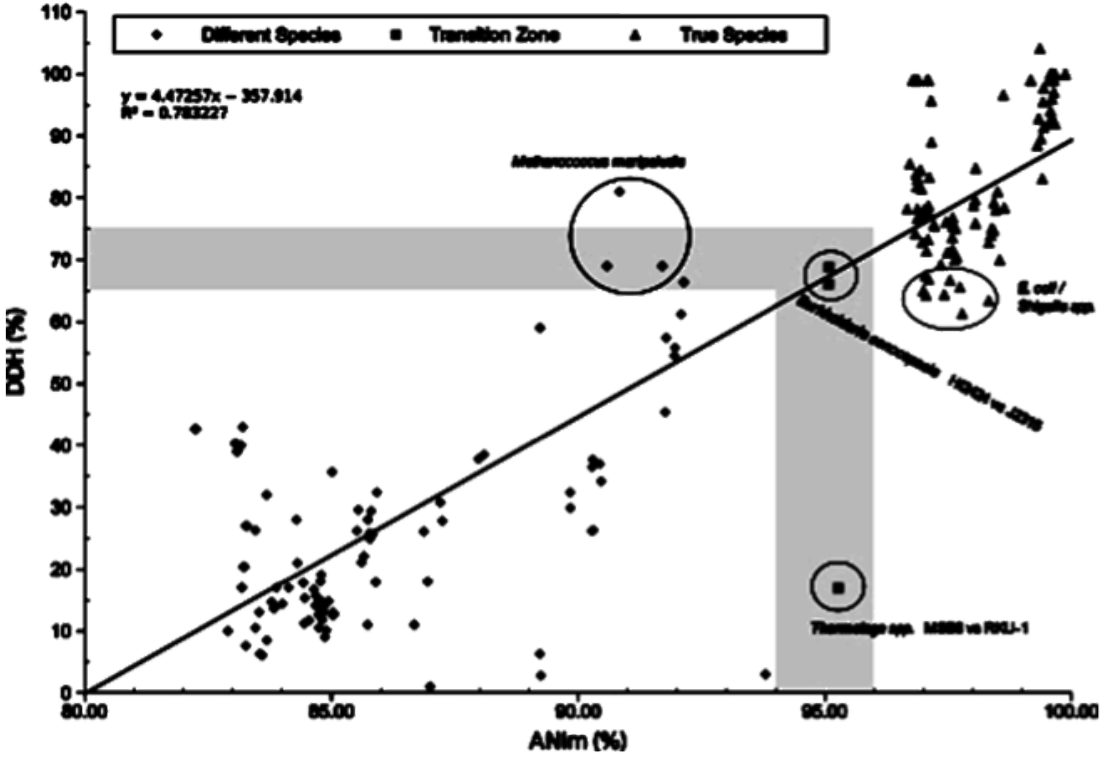
Average Nucleotide Identity (ANIm) [a]

[a] Richter and Rossello-Mora (2009) *Proc. Natl. Acad. Sci. USA* doi:10.1073/pnas.0906412106

The James Hutton Institute

1. Align genomes (MUMmer)
2. **ANIm**: Mean % identity of all homologous region matches

- DDH:ANIm "linear"
- **70%ID ≈ 95%ANIm**



ANIm

The James Hutton Institute

ANIm is. . .

- straightforward to apply to genomes
- average identity of all ‘homologous’ regions
- not dependent on dataset composition (unlike hierarchical clustering)
- (just) another pairwise distance measure
- approximate limiting case of MLST/MLSA/multigene comparisons

Table of Contents

The James Hutton Institute

pyani software [a]

[a] http://widdowquinn.github.io/pyani

The James Hutton Institute

- Python ANI module and script
- Active development
- Parallelises on clusters
- Preprint coming soon (simulated genomes; lots of bacterial taxonomy!)

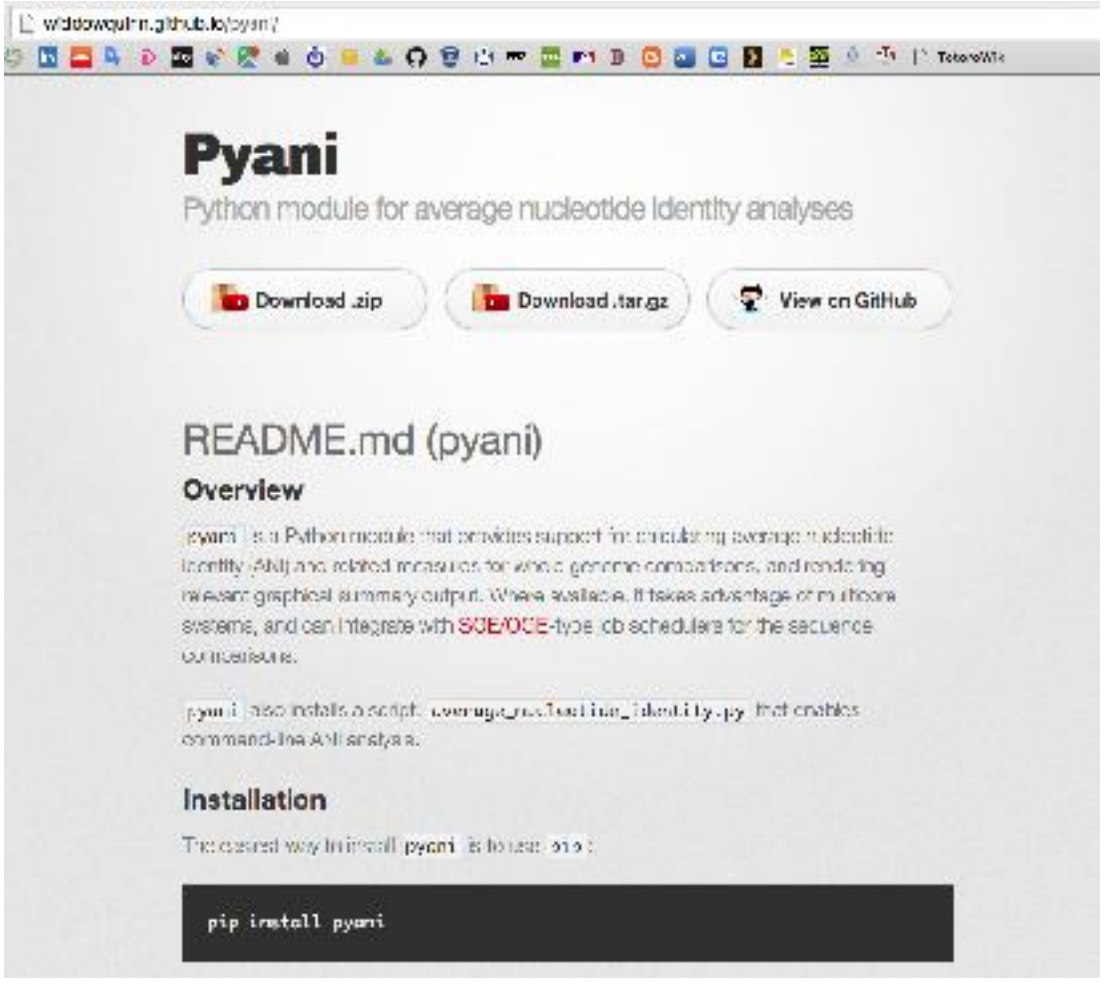


Table of Contents

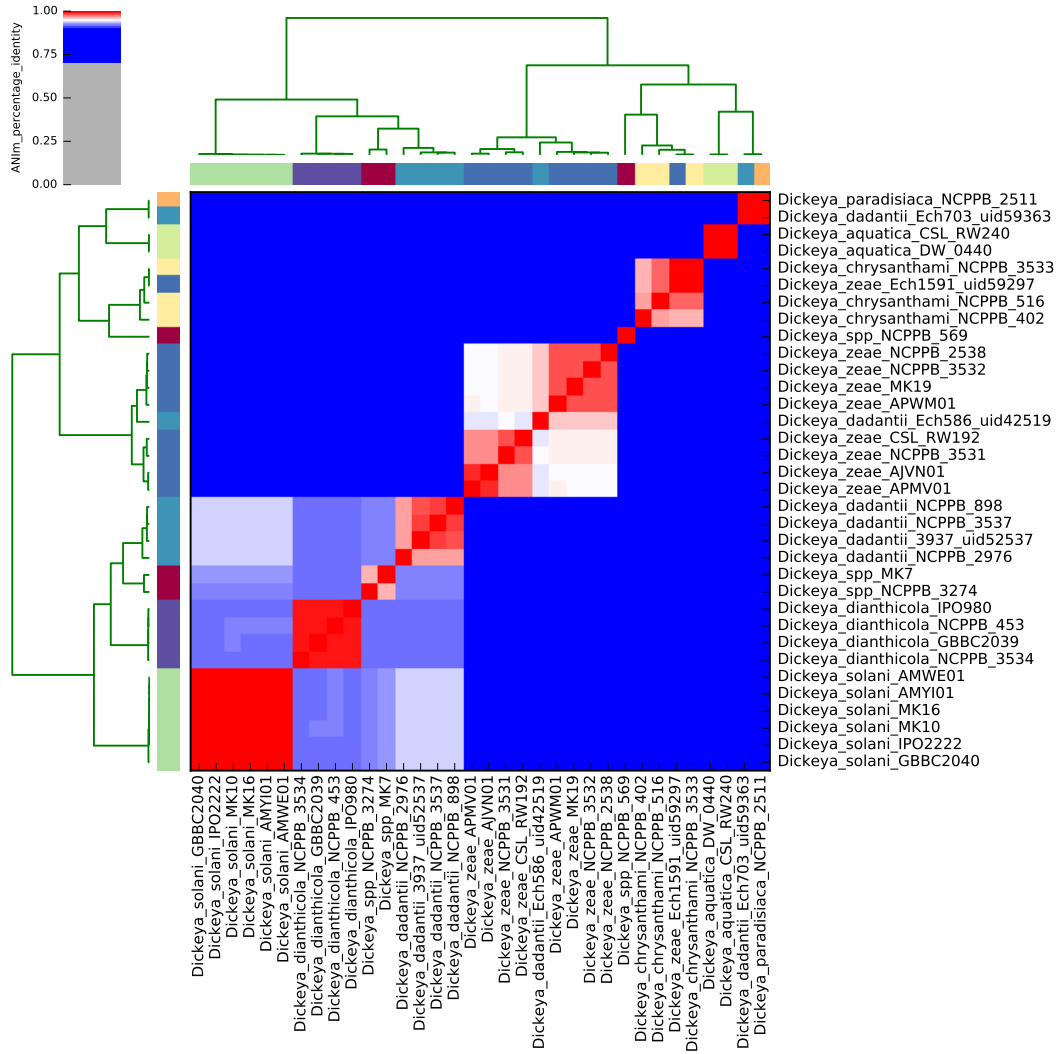
The James Hutton Institute

34 Dickeya spp. ANIm [a]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

- Nine species-level groups (two novel)
- Correctly places three species misidentified in GenBank

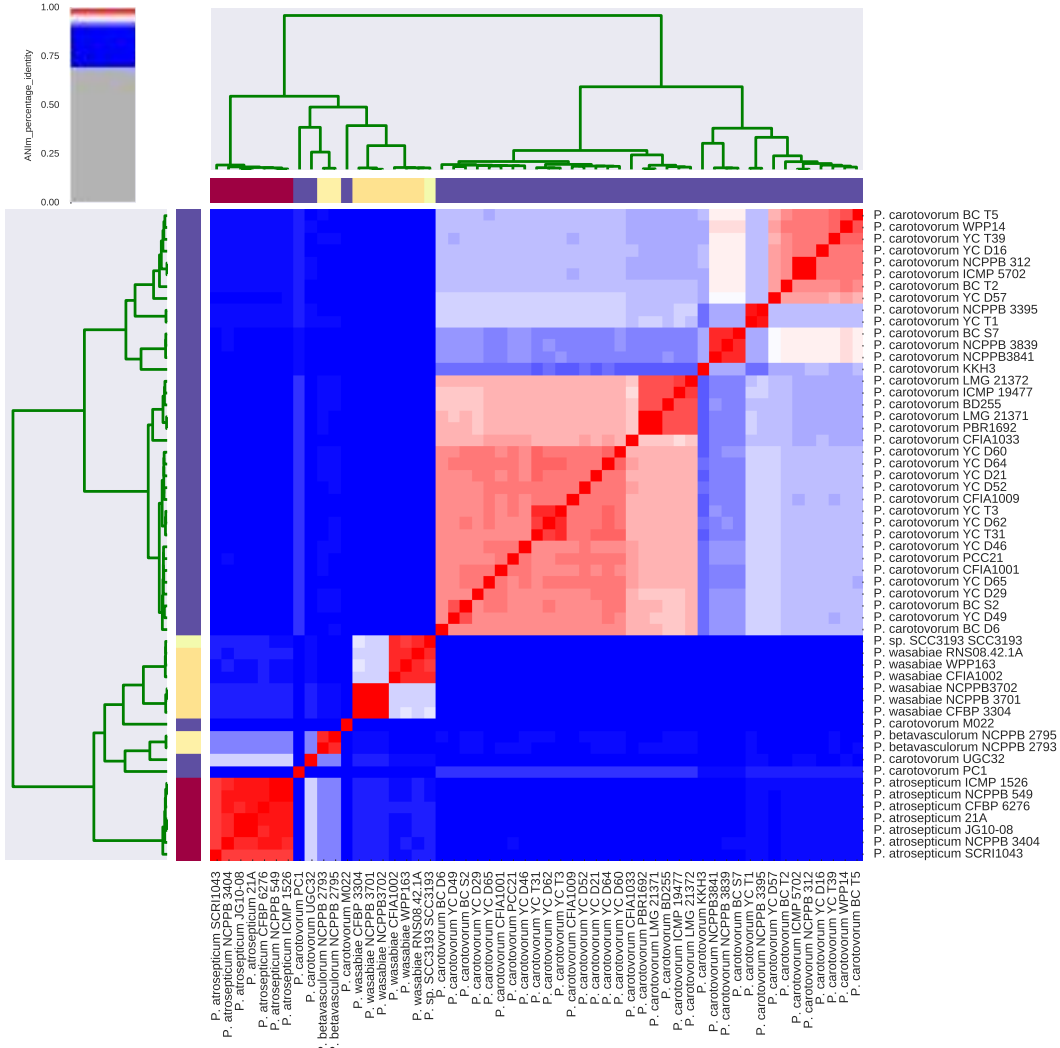


55 Pectobacterium spp. ANIm [a]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

- Ten species-level groups (four novel)
- *P. carotovorum* split: several species - no clean *Pcb*/*Pcc* split
- *P. wasabiae* split: two species



Interpreting ANIm [a]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

Criticisms of ANIm

- 95% threshold 'arbitrary'
- Similarity classification, not phylogenetic reconstruction
- No functional (or gene-based) interpretation of risk (cf. pangenome classification and analysis)

ANIm only considers 'homologous' regions

- define 'homologous'
- be misled by HGT/LGT - low total extent of homology?
- is homology phenotypically significant for risk assessment?

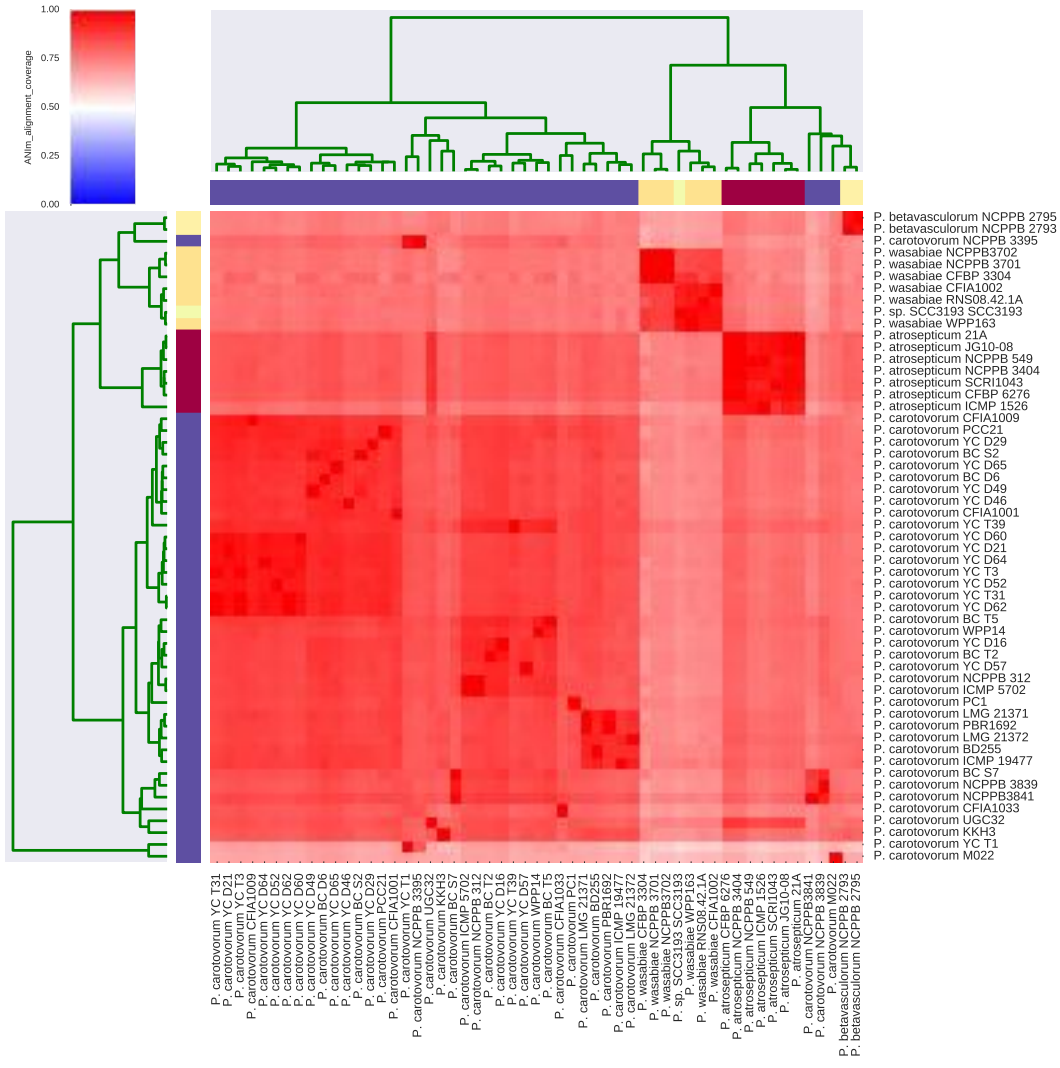
Coverage plots help interpretation: exclude HGT/LGT bias.

55 Pectobacterium spp. ANIm[a]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

- All isolates align over >50% of whole genome



34 Dickeya spp. ANIm[a]

[a] Pritchard *et al.* (2016) *Anal. Methods* doi:10.1039/c5ay02550h

The James Hutton Institute

- Most isolates align over >50% of whole genome
- Community: two outlier species are questionable assignments as *Dickeya*

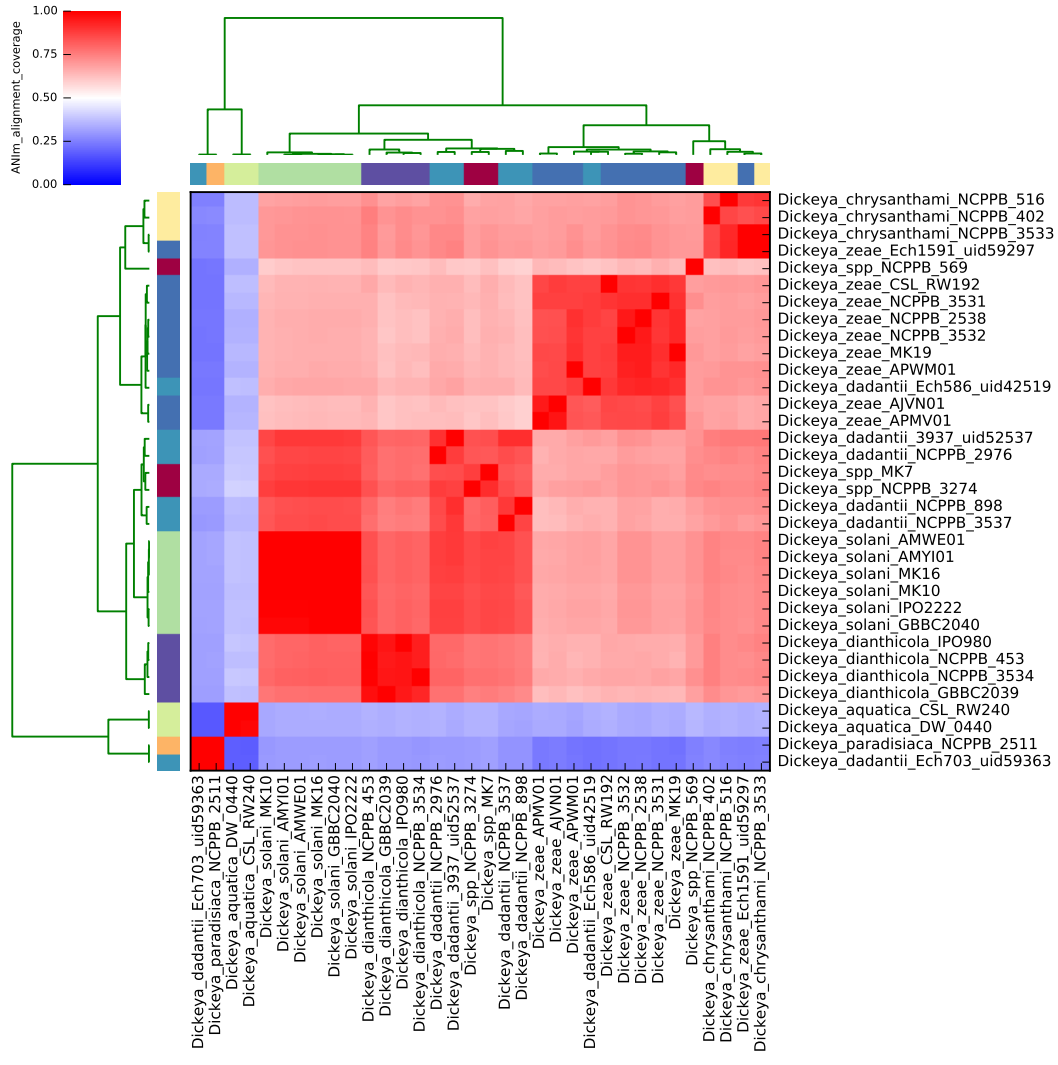
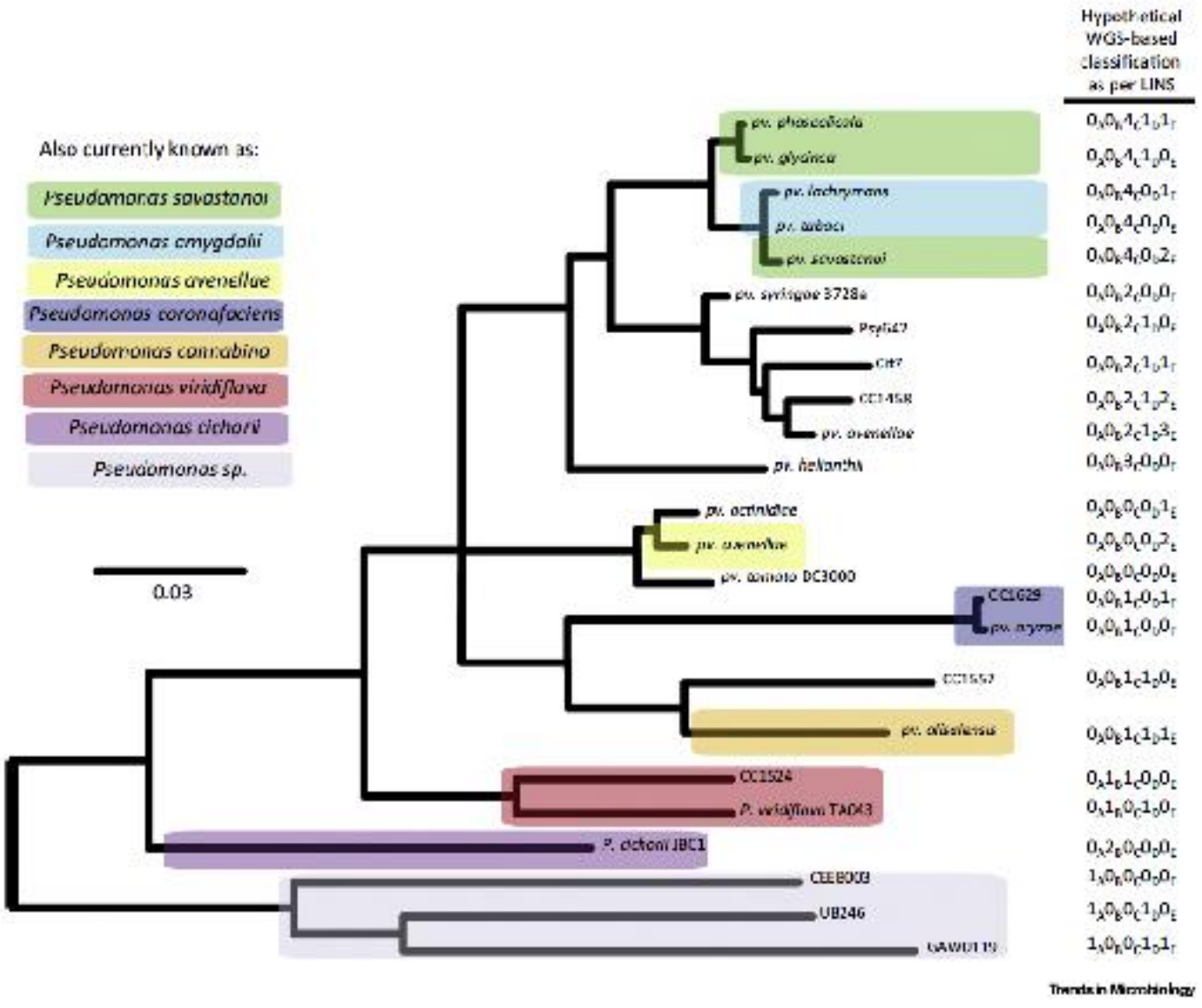


Table of Contents

The James Hutton Institute

A new classification scheme [a]

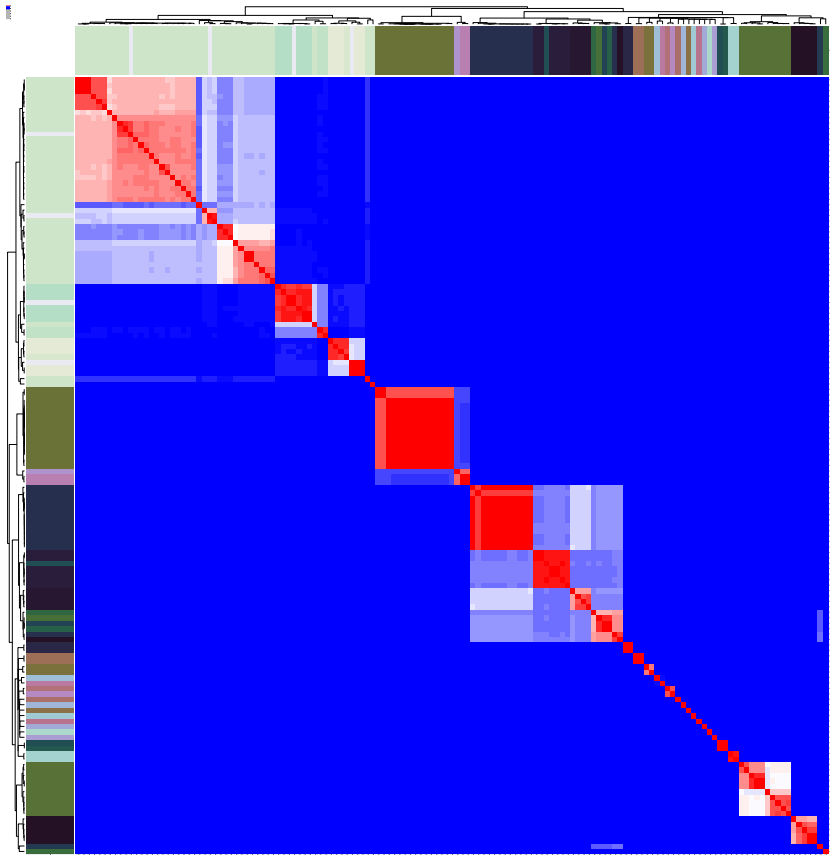
[a] Baltrus (2016) *Trends Microbiol.* doi:10.1016/j.tim.2016.02.004

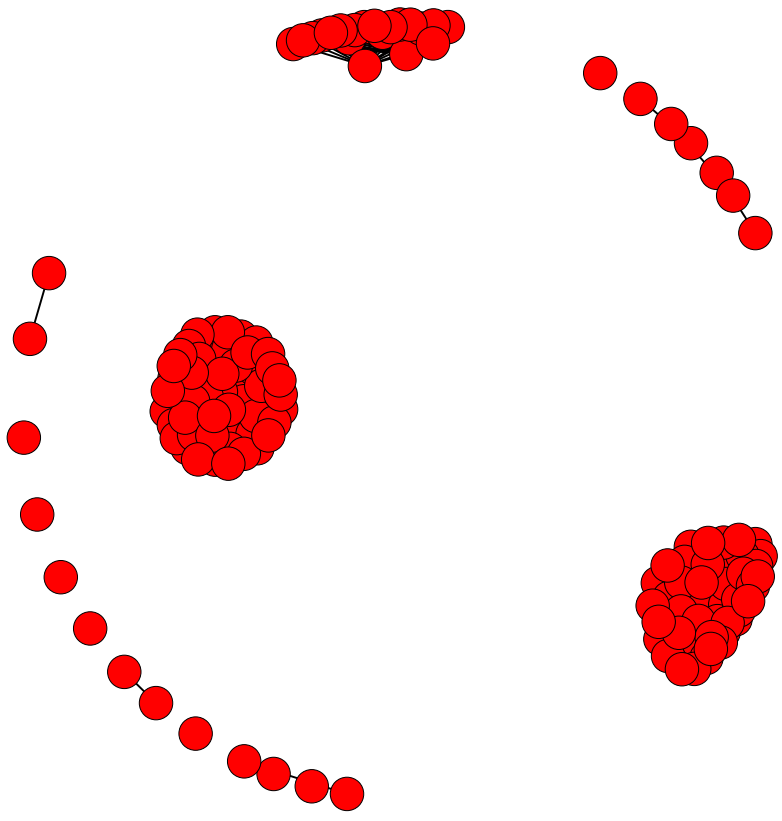


ANIm graphs

The James Hutton Institute

ANIm identity/coverage scores define networks (143 genomes)



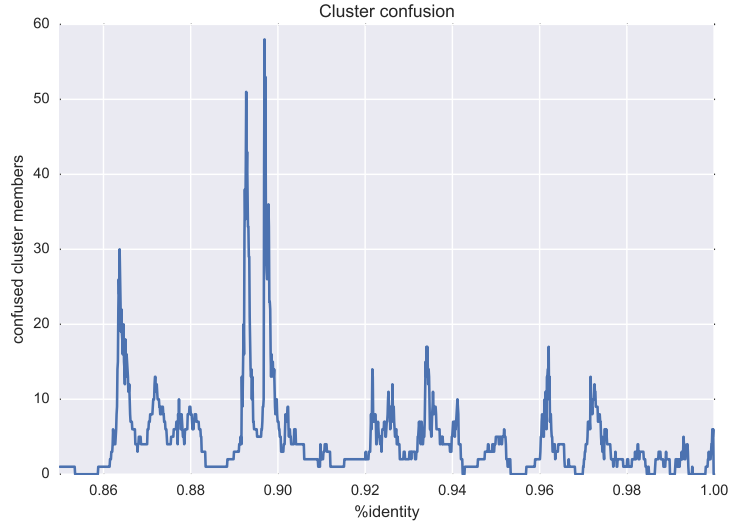


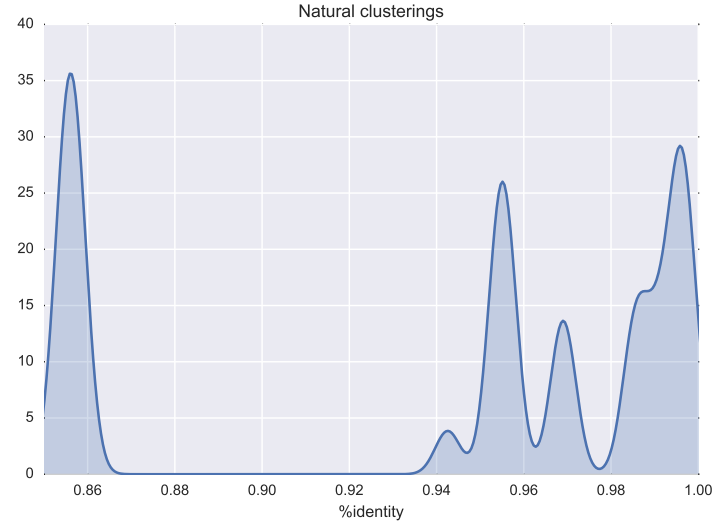
Natural clusterings

Natural clusterings occur in the data - ‘cliques’

‘Clique’ membership varies with %identity

‘Clique’ membership at given %identity a permanent classification



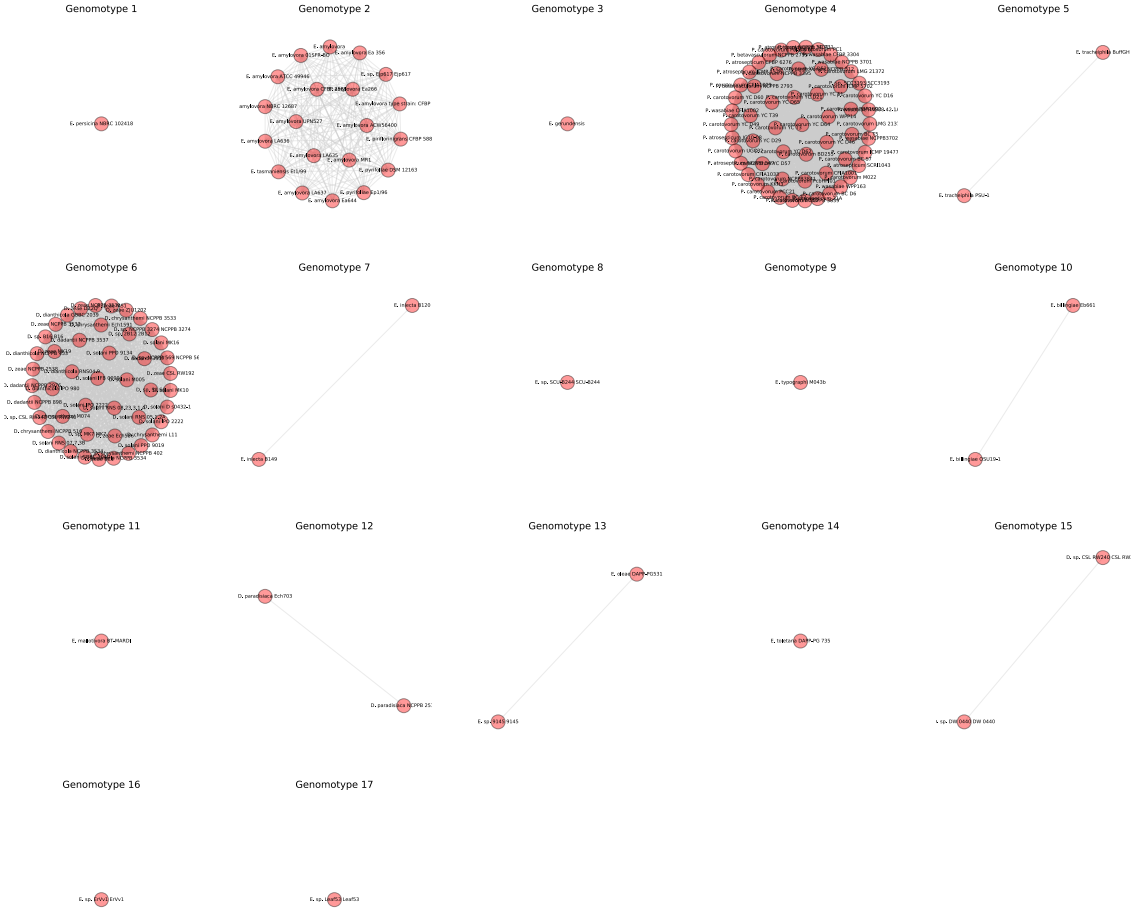


‘Genus’, ‘species’, ‘subspecies’ and ‘clonal’ genomotypes indicated

Genus genomotype

The James Hutton Institute

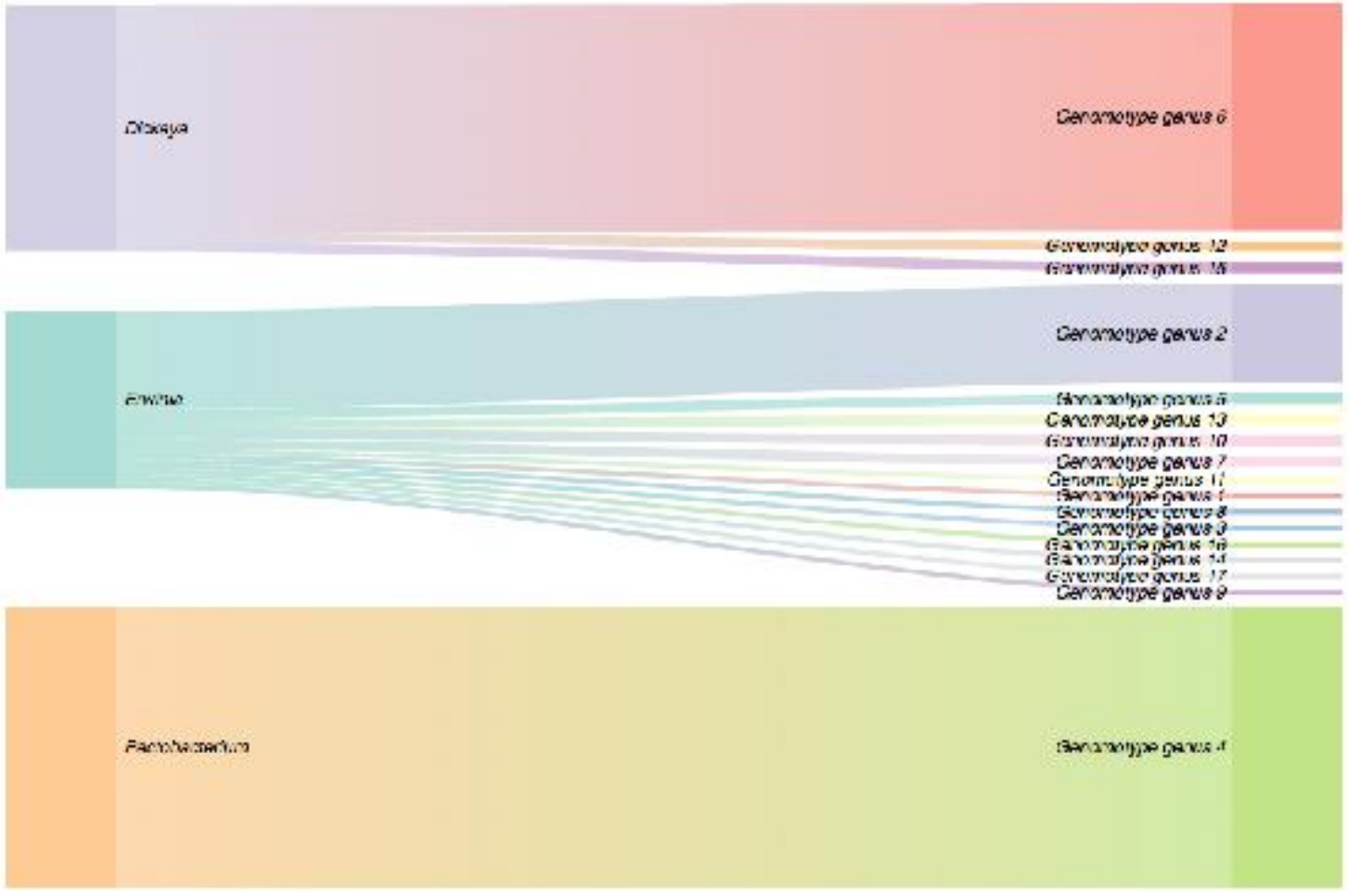
17 genus-level genomotypes indicated



Genus reclassification

Erwinia splits into 13 genomotypes

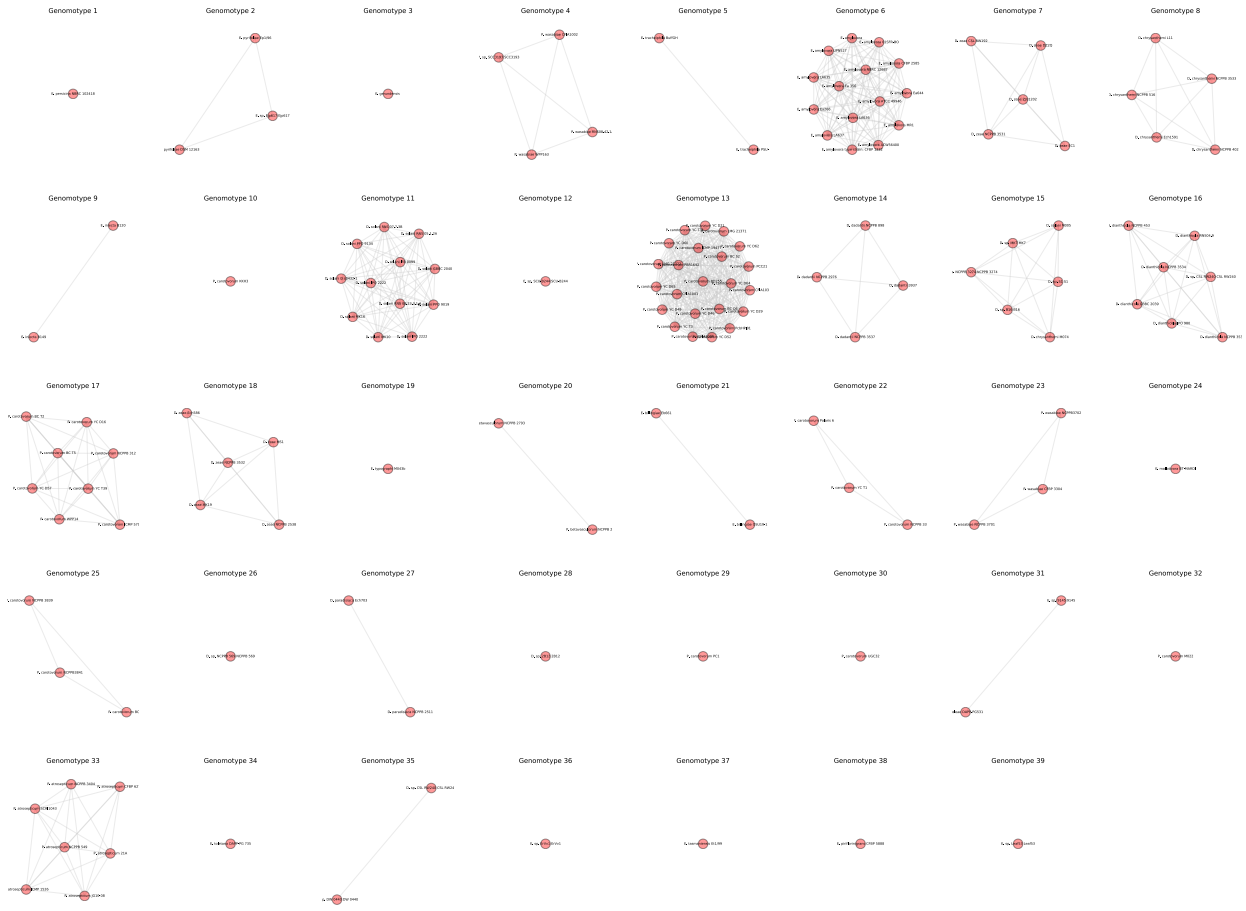
Dickeya splits into 3 genomotypes



Species genomotype

The James Hutton Institute

39 species-level genomotypes indicated



Species genomotype

The James Hutton Institute

Most species classifications remain unaffected

P. carotovorum subspecies confusion/splits

Suggested reclassifications

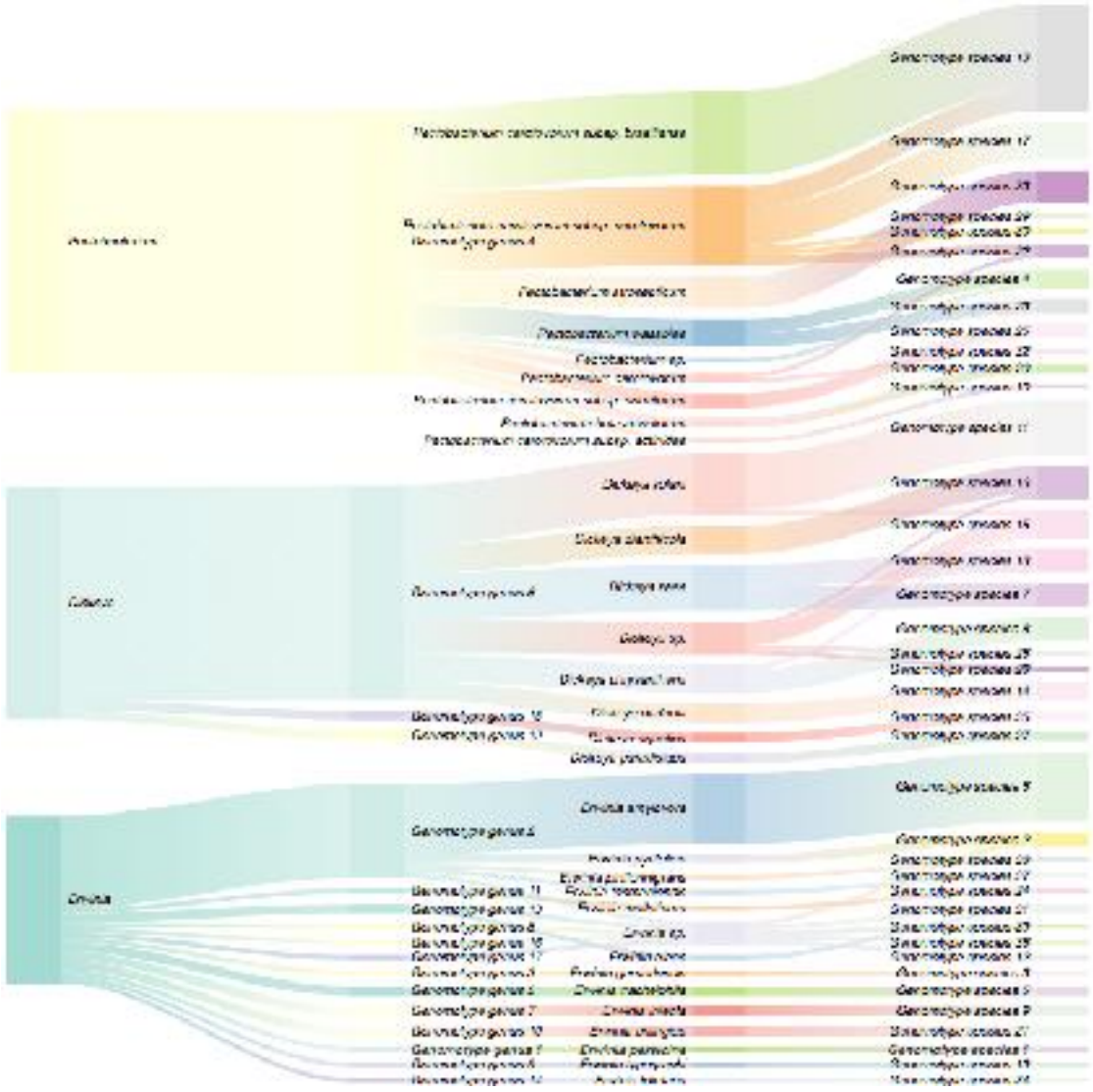


Table of Contents

The James Hutton Institute

Final thoughts

The James Hutton Institute

Diagnostics and large-scale analyses need accurate classification

- Historical collections/public databases have inaccuracies
- Bacterial taxonomy can be messy!

Whole-genome classification with ANI works

- Retrospective sequencing and classification: clean things up
- Misclassification and hidden diversity may be difficult news for metagenomics...
- Accurate MinION classification in-the-field with ANIm is possible?

Table of Contents

Acknowledgements

Dickeya/Erwinia/Pectobacterium
Steve Baeyen (ILVO)
Emma Campbell (JHI)
Sean Chapman (JHI)
John Elphinstone (Fera)
Tracey Gloster (St Andrews)
Rachel Glover (Fera)
Sonia Humphris (JHI)
Martine Maes (ILVO)
Katrin Mackenzie (BioSS)
Neil Parkinson (Fera)
Minna Pirhonen (Helsinki)
Gerry Saddler (SASA)
Elaine Shemilt (Duncan of Jordanstone)
Ian Simpson (Edinburgh)
Ian Toth (JHI)
Jan van der Wolf (Wageningen)
Johan van Vaerenbergh (ILVO)
Frank Wright (BioSS)
Eirini Xemantilotou (St Andrews/JHI)

Nematodes
Peter Cock (JHI)
John Jones (JHI/St Andrews)
Robbie Rae (John Moores)
Peter Thorpe (JHI)

Phytophthora
Miles Armstrong (Dundee)
Anna Avrova (JHI)
Jim Beynon (Warwick)
Paul Birch (Dundee)
David Cooke (JHI)
Kath Denby (York)
Eleanor Gilroy (JHI)
Sarah Green (Forest Research)
Edgar Huitema (Dundee)
Rory McLean (Dundee)
Hazel McLellan (Dundee)
Sophien Kamoun (TSL)
Gail Preston (Oxford)
Paul Sharp (Edinburgh)
Jens Steinbrenner (Warwick)
Pieter van West (Aberdeen)
Steve Whisson (JHI)

Computational and Systems Biology
David Broadhurst (Edith Cowan)
Peter Cock (JHI)
Mark Dufton (Strathclyde)
Roy Goodacre (Manchester)
Douglas Kell (Manchester)
David Martin (Dundee)
Iain Milne (JHI)
Pedro Mendes (Manchester)

E. coli/other bacteria
Florence Abram (Galway)
Martina Bielaszewska (Muenster)
Fiona Brennan (Galway)
Ken Forbes (Aberdeen)
Nicola Holden (JHI)
Ashleigh Holmes (JHI)
Paul Hoskisson (Strathclyde)
Helge Karch (Muenster)
Norval Strachan (Aberdeen)
David Studholme (Exeter)
Nick Waters (Galway)

Metagenomics/Communities
Natalie Ferry (Salford)
Thomas Freitag (JHI)
Ryan Joynson (Liverpool)
John Mitchell (St Andrews)
Les Noble (Aberdeen)
Jim Prosser (Aberdeen)
V Anne Smith (St Andrews)

Potato
Micha Bayer (JHI)
Glenn Bryan (JHI)
Graham Etherington (TSL)
Ingo Hein (JHI)
Florian Jupe (JHI)
Jonathan Jones (TSL)
Dan Maclean (TSL)
. . .and many others. . .

Licence: CC-BY-SA

The James Hutton Institute

By: Leighton Pritchard