Escherichia coli community diversity – hitch-hiking for the solution.

Adrian Cookson

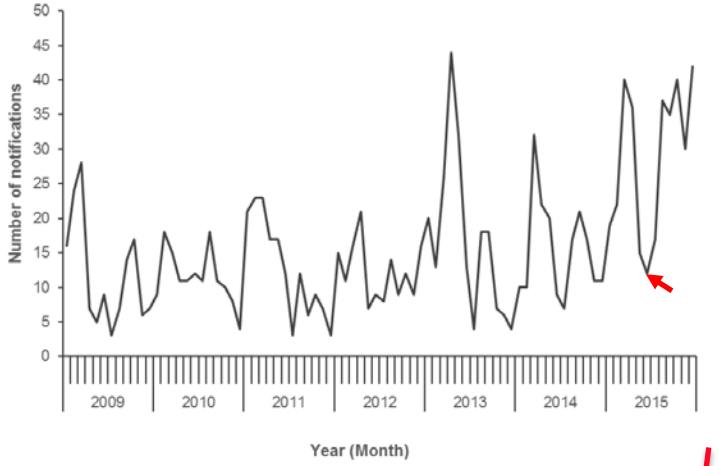
Hopkirk Institute, Palmerston North.

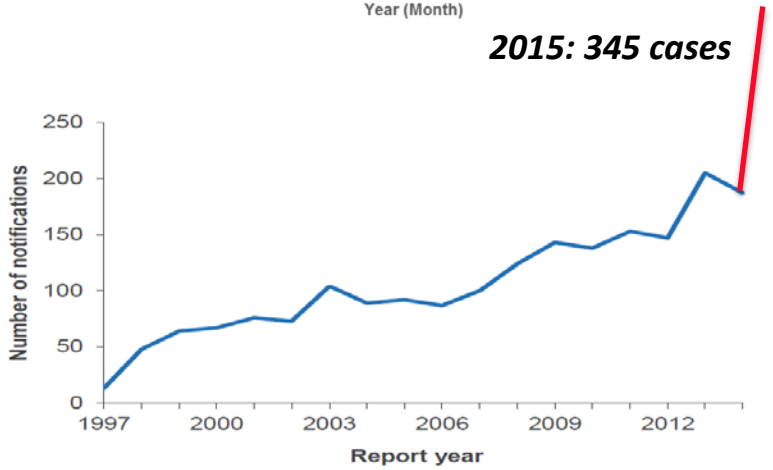
IDReC Symposium, Otago University, 22-23 March 2016.

OIE Collaborating Centre for Veterinary Epidemiology and Public Health

Shiga toxin-producing *Escherichia coli* (STEC) notifications in New Zealand

- STEC – zoonotic pathogen
  - Ruminant reservoir
  - Human symptoms of infection: diarrhoea and haemolytic uraemic syndrome (HUS)
  - Cause of large foodborne outbreaks of diarrhoea and HUS overseas
  - Seasonal pattern
  - Many cases missed

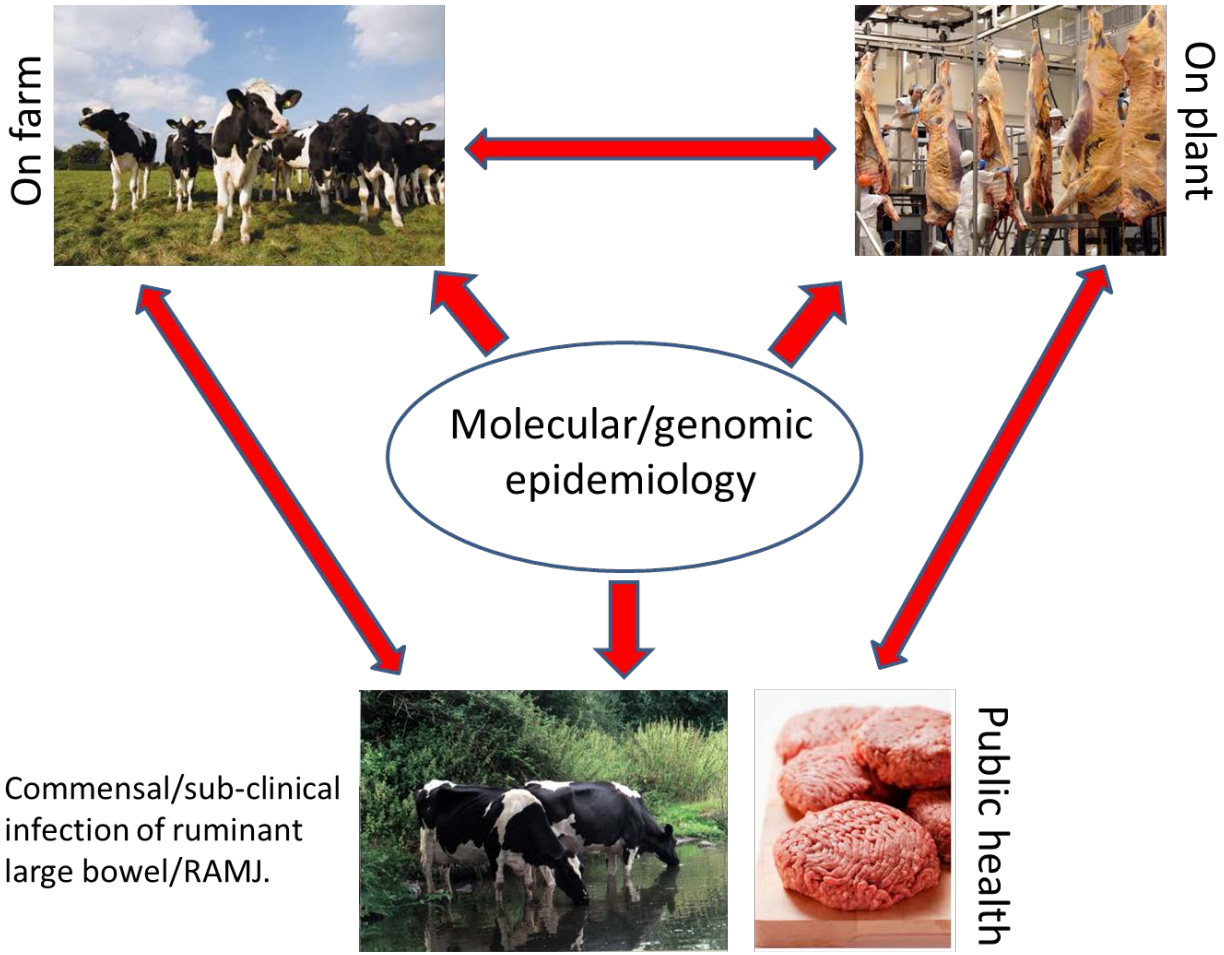




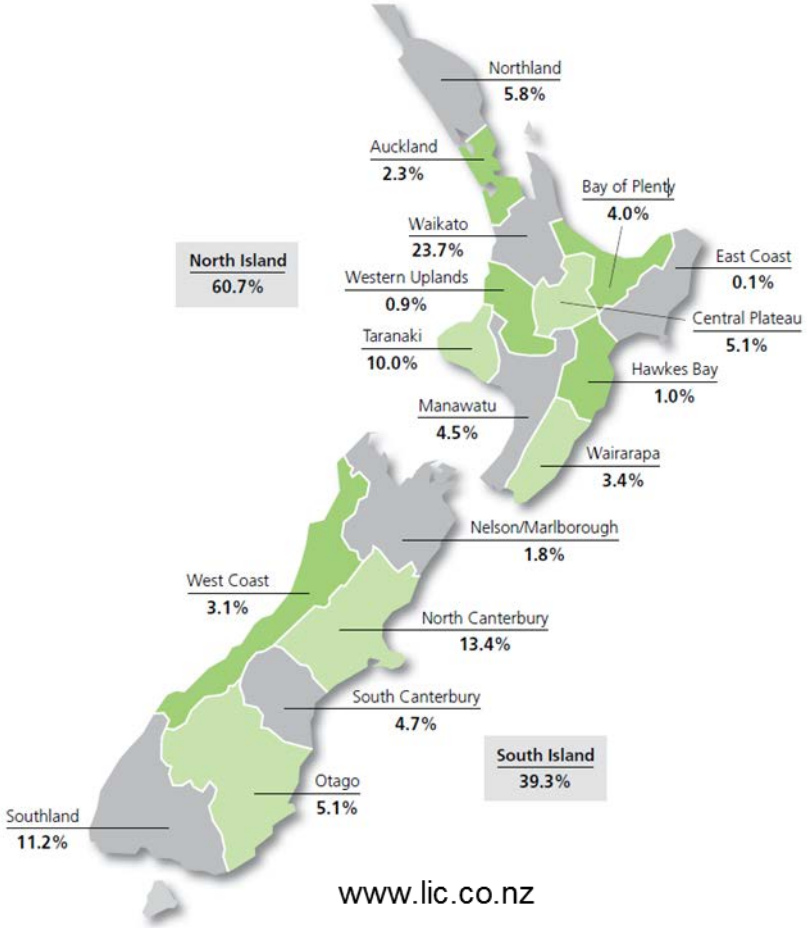
2015. Incidence of 7.7 STEC cases per 100,000 population. (US: 1.1 per 100,000; Australia: 0.4 per 100,000; Ireland: 6.1 per 100,000; Scotland: 4.5 per 100,000).

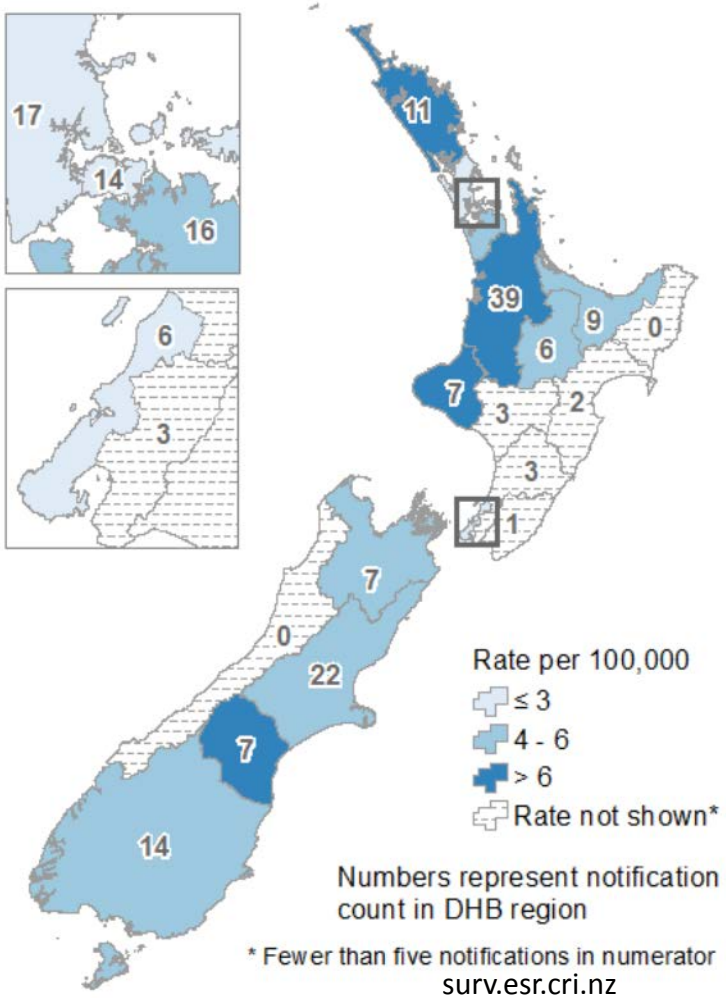
STEC contamination and transmission pathways

- O157 serogroup associated with 80-90% STEC clinical cases in NZ
- Isolated cases or sporadic household-associated outbreaks
- Risk factors – contact with animals/animal faeces/environment
- ***Contaminant of export meat products***
- ***Meat inspections require absence of seven clinically important serogroups (STEC7)***



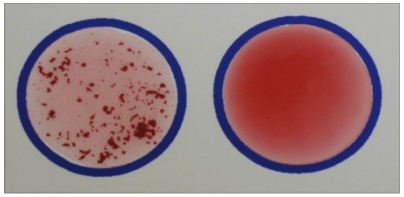
Regional distribution of dairy cows (2013/14) and STEC infection notification (2014)



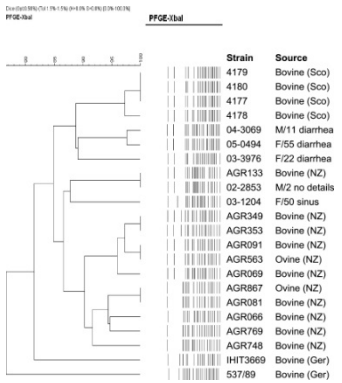


Project background – *E. coli* differentiation

- Serology – O (lipopolysaccharide), H (flagella) and K (capsule) grouping
- Pathotypes – EPEC, STEC, ETEC, EAEC, EIEC
- Subtyping
  - Pulsed field gel electrophoresis (PFGE)
  - Multi locus sequence typing (MLST)
  - Insertion sequence (IS) typing
  - Genome sequencing – single nucleotide polymorphisms (SNPs)



E. coli differentiation using pure cultures



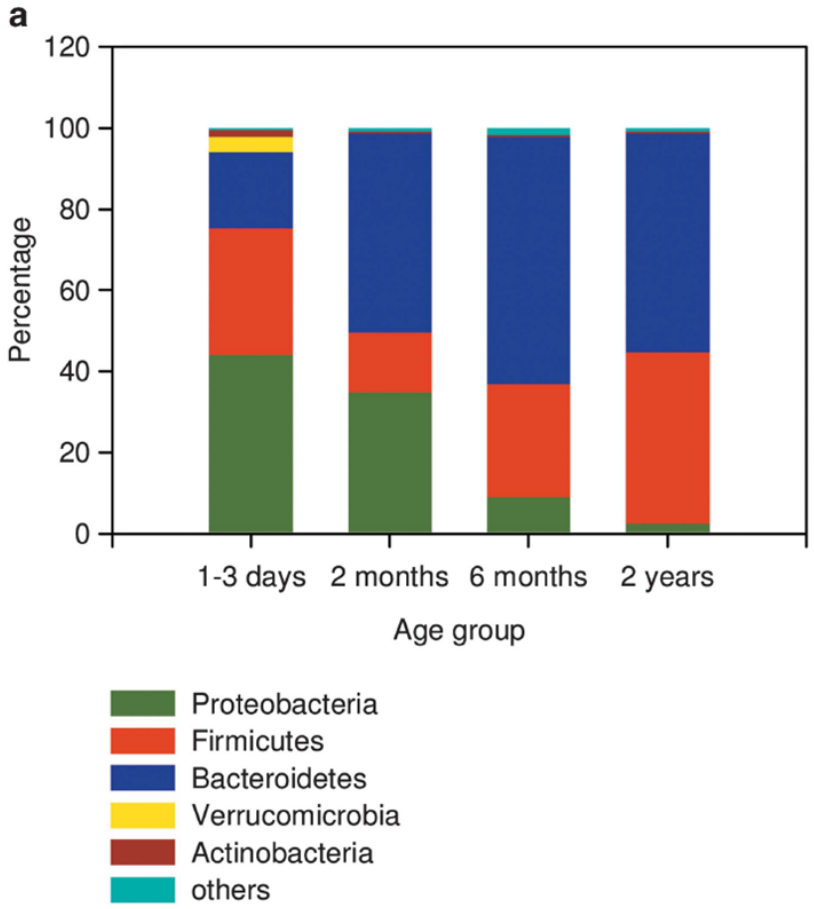
Project background – *E. coli* diversity

- Assessed using culture-based methods
- Beef cattle faecal samples:[1]
  - 30 serotypes from 10 animals fed roughage & molasses
  - 21 serotypes from 11 animals fed roughage
  - 17 serotypes from 9 animals fed grain
- Human faecal samples:[2]
  - 1 - 15 (av. 5) biotypes from 9 healthy humans over 6 weeks
- ***Can we assess* E. coli *diversity of the intestine using culture independent methods?***

[1]Bettelheim et al., 2005. J Appl Micro. 98. 699-709
[2]Apperloo-Renkema et al., 1990. Epi & Inf. 105. 355-61

Microbial community profiling

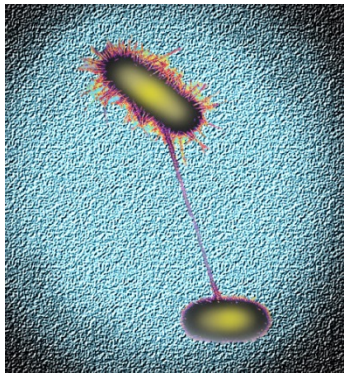
- Culture-independent method targeting the 16S rRNA barcoding gene: present in all bacteria
- Sequences are clustering at various taxonomic levels to provide functional clues
- Rumen microbiota in weaned animal dominated by specific microbial phyla
- Species cut off at 97% similarity operational taxonomic unit (OTU) level
- *E. coli* within the phylum Proteobacteria
- 16S rRNA gene not sufficiently discriminatory for determining ***within species variation***



Jami *et al.*, 2013. ISME, 7:1069-79.

E. coli barcode targets

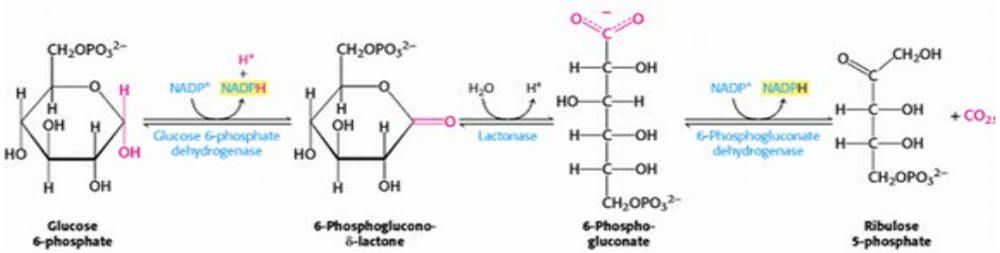
- Focus on hot-spots for recombination/horizontal gene transfer
- O antigen biosynthesis gene clusters prone to recombination
  - 184 recognised *E. coli* serogroups based on antigenic variability
- Representative O antigen biosynthesis gene clusters sequenced[1] and serogroup-specific PCRs[2] developed for isolate identification



[1]Iguchi et al., 2015. DNA Res. 22. 101-7
[2]Iguchi et al., 2015. J Clin Microbiol. 53. 2427-32

gnd – 6-phosphogluconate dehydrogenase

- Housekeeping gene often associated with O antigen biosynthesis gene cluster in Enterobacteriaceae
- Third enzyme reaction of pentose phosphate pathway

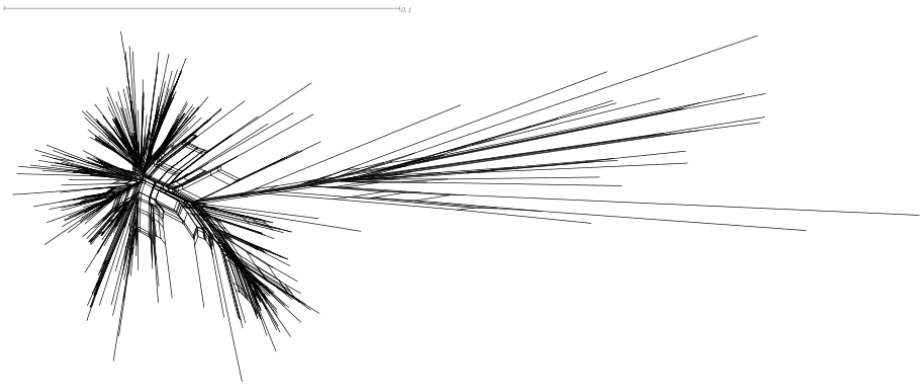


- Described as passive hitch-hiker[1] with existing O antigen biosynthesis gene cluster variants
- Variability noted in prior work through MLEE[2], RFLP[3], sequencing[1,4]

[1]Nelson & Selander, 1994. PNAS. 91. 10227-31
[2]Dykhuizen & Green, 1991. J Bact. 173. 7257-68
[2]Selander & Levin, 1980. Science 210. 545-7
[4]Gilmour et al., 2007. J Med Micro. 56. 620-8

gnd sequence analysis

- Alignment made of >1000 *E. coli gnd* DNA sequences
- Single base SNPs noted between sequences
- Degenerate PCR primers designed for *gnd* amplicon sequencing (conventional and MiSeq)
- Reference database created including 300 unique *gnd* sequence types (gSTs)
  - Covers all 184 serogroups and 35 untypable or rough strains



Detailed study of *E. coli* diversity

- Samples from animals trial to assess role of bifidobacteria on calf health
- Treatment group orally dosed daily with 2 bifidobacterial strains
- RAMS and faecal samples taken from calves (n=23) at 17-18 days of age

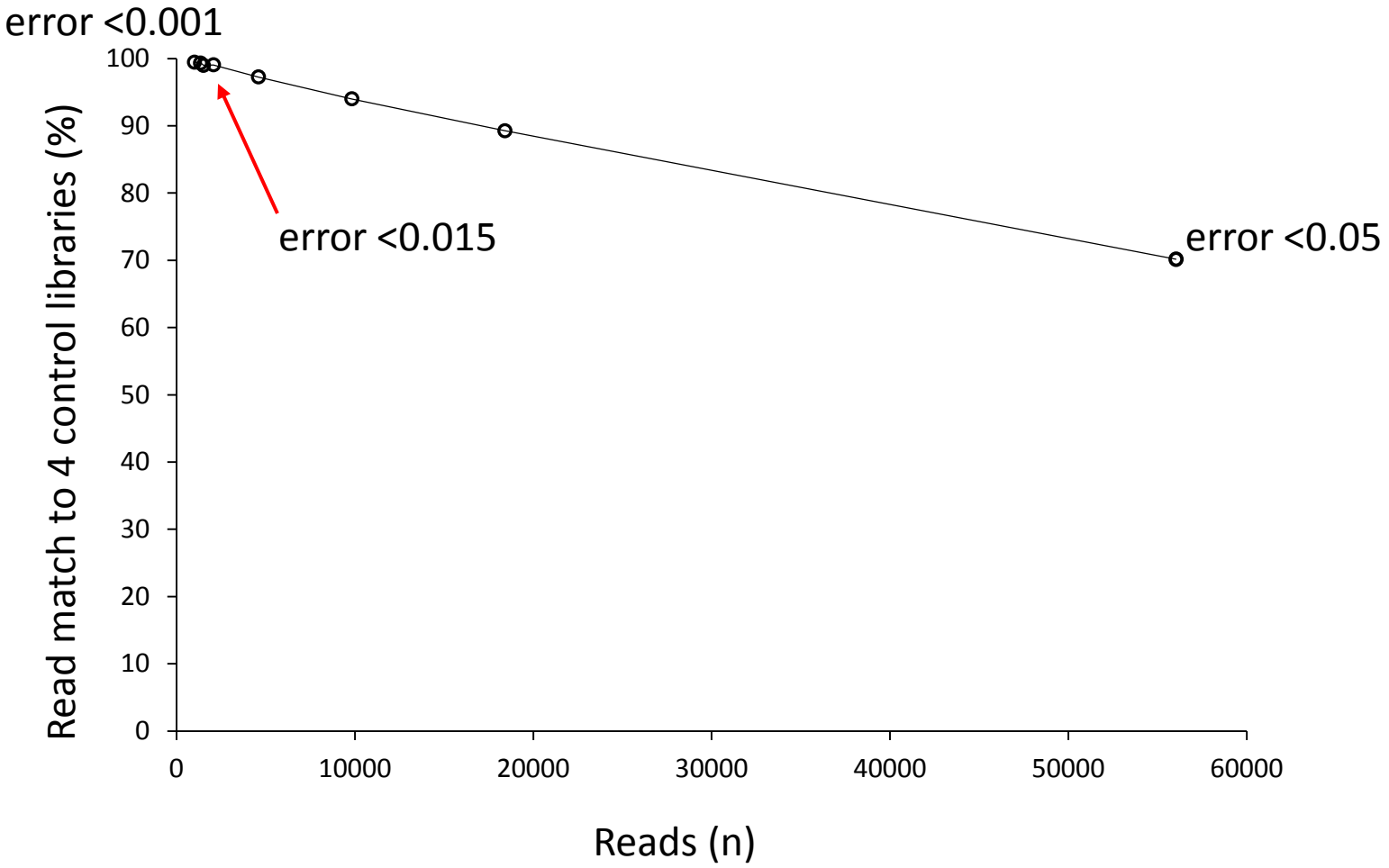
  - Faeces (23)
  - mTSB pre-enrichment (23)
  - mTSB post-enrichment – boiled lysate (23)
  - mTSB post-enrichment – Roche kit (23)
  - ***Defined synthetic control libraries (4)***

- Barcoded *gnd* amplicons generated from DNA extracts
- MiSeq (2 x 250bp PE); reads (>150bp) assembled using SolexaQA^{++}[1]

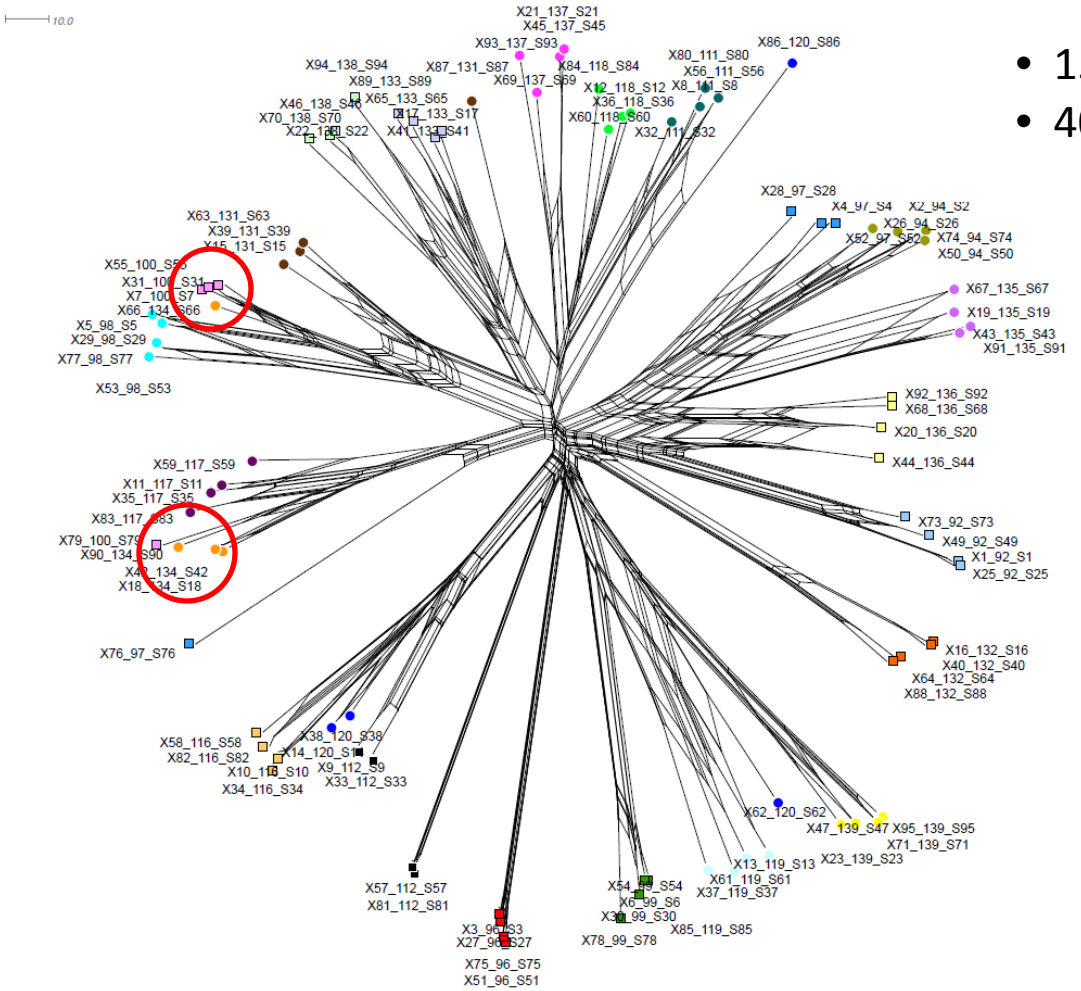
RAMS – recto-anal mucosal swabs

[1]Cox et al., 2010. BMC Bioinformatics. 11. 485

Defined control libraries – impact of read error on sequence match



NN of *E. coli* community diversity (gST original proportions)

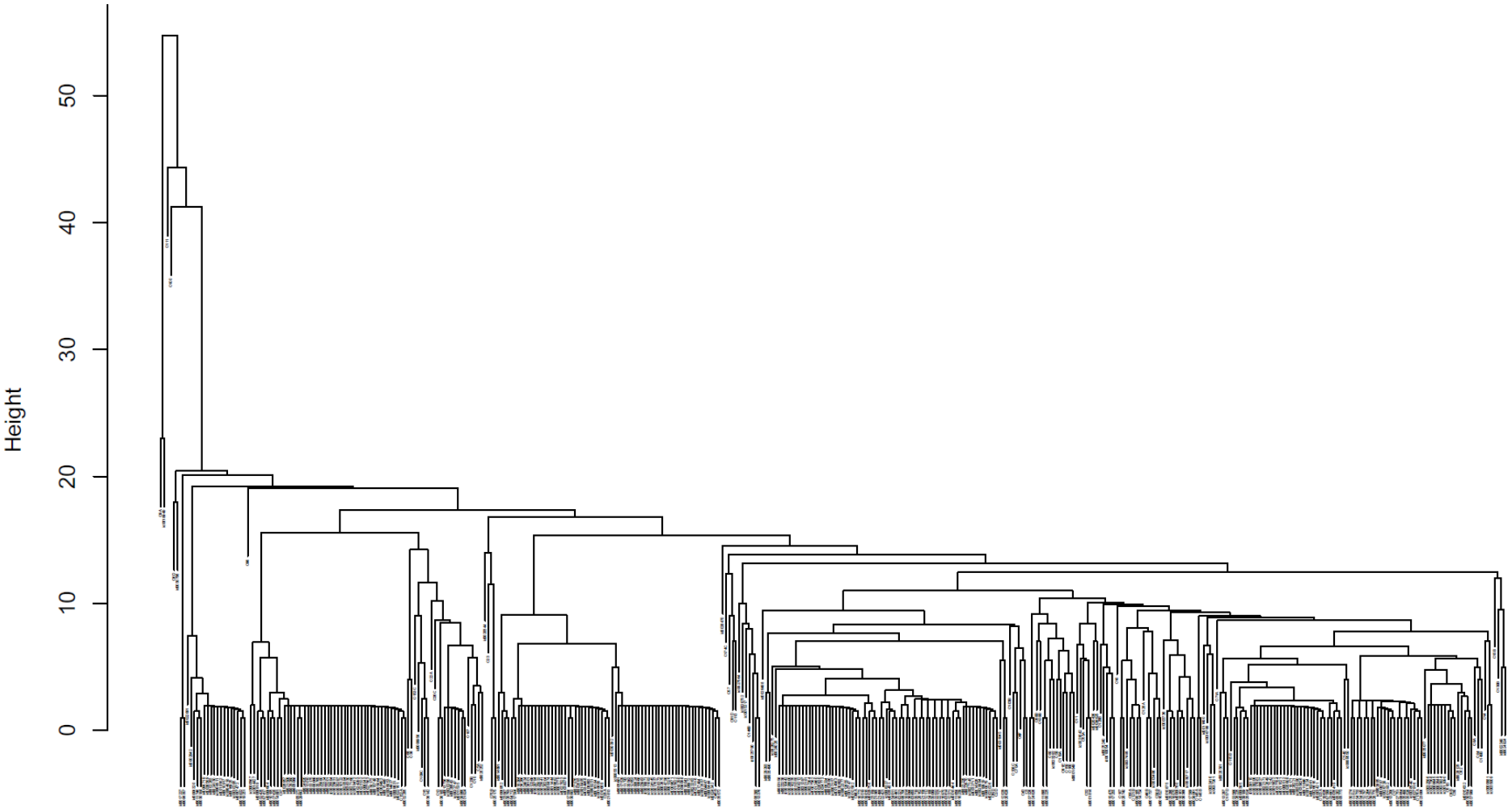


- 15thou dataset: 191649 reads
- 403 *gnd* sequence types (>10 reads)

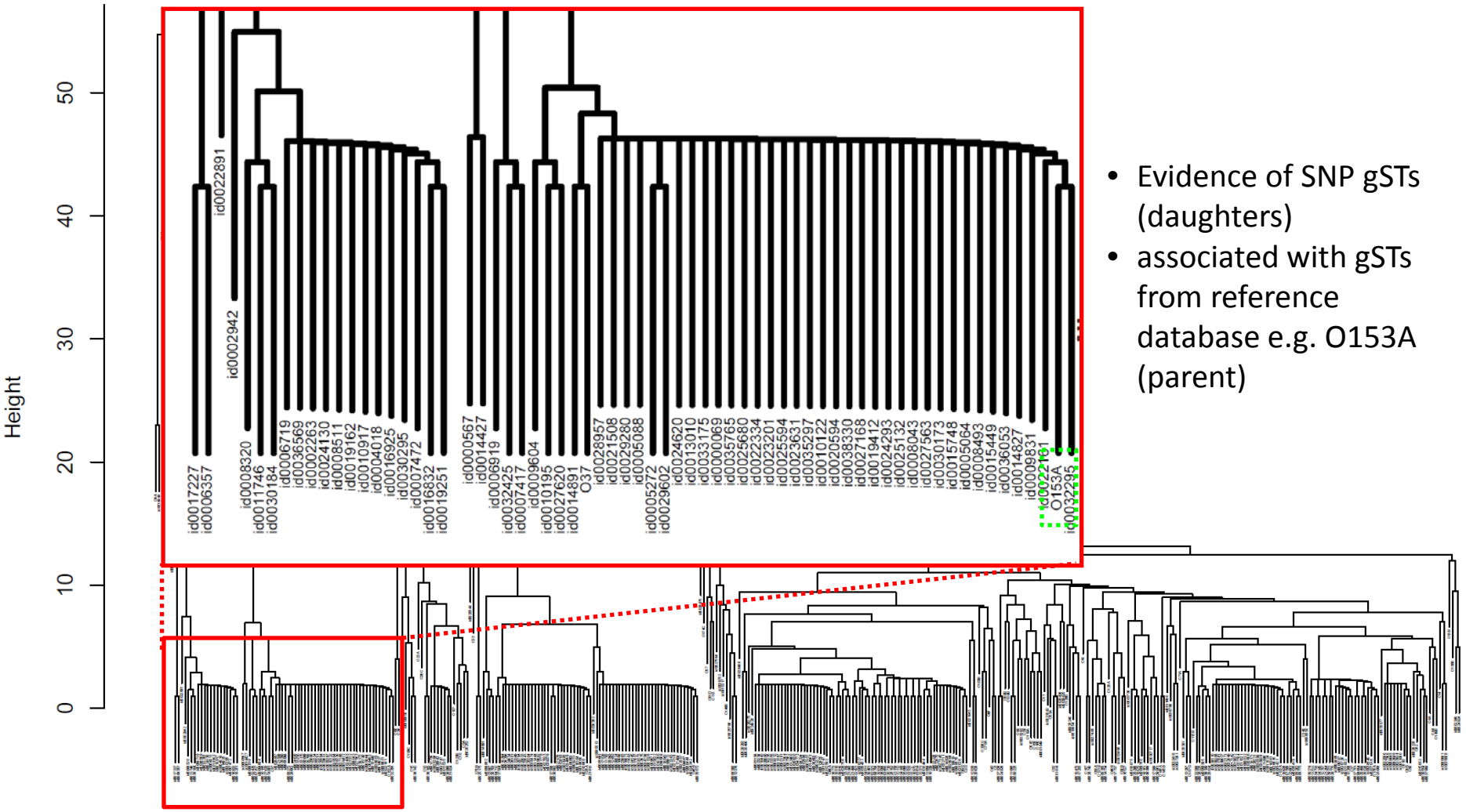
agresearch

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Clustering of 403 gST sequences

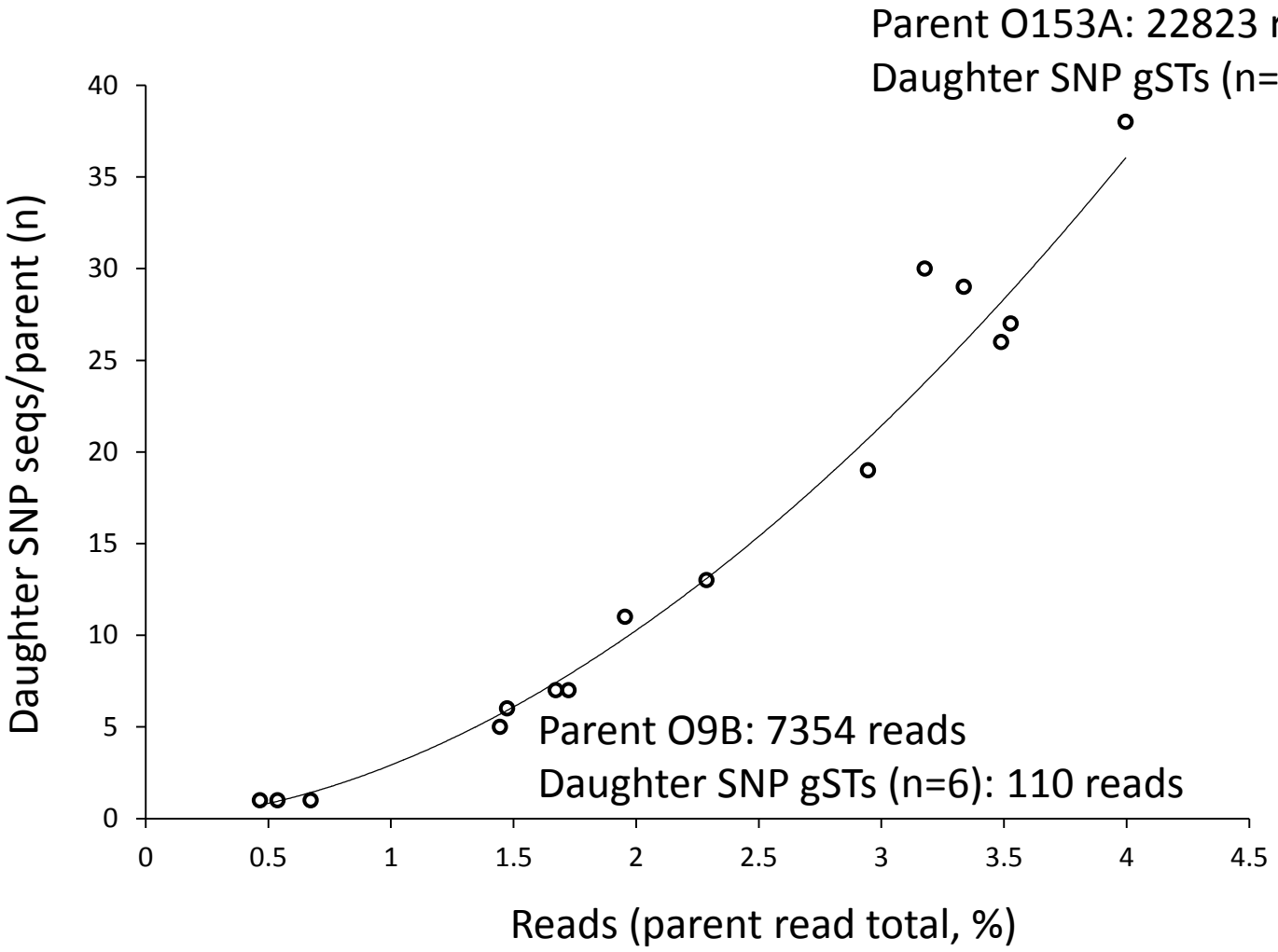


Clustering of 403 gST sequences



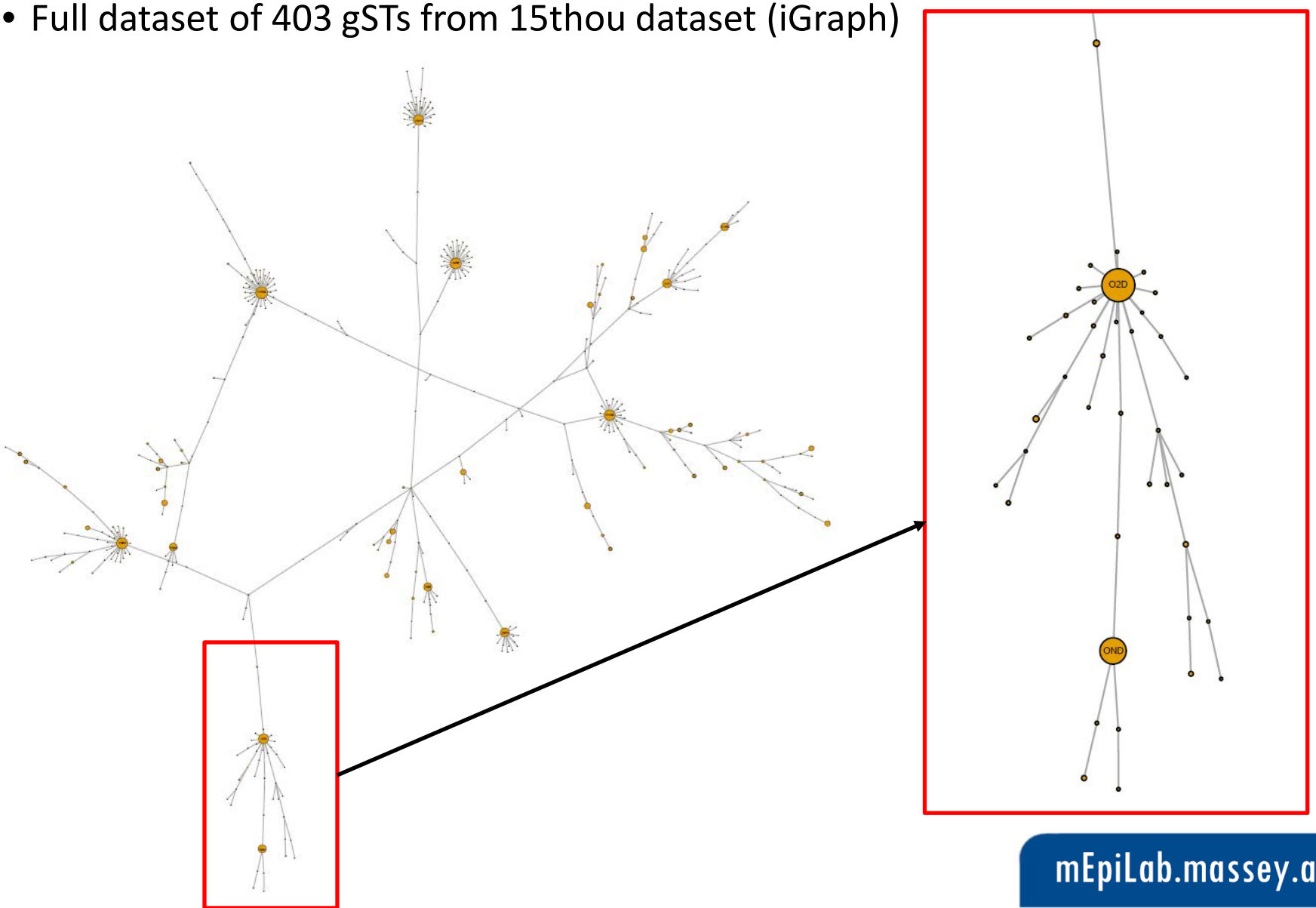
- Evidence of SNP gSTs (daughters)
- associated with gSTs from reference database e.g. O153A (parent)

Parent/daughter SNP sequences and read numbers (15thou dataset)



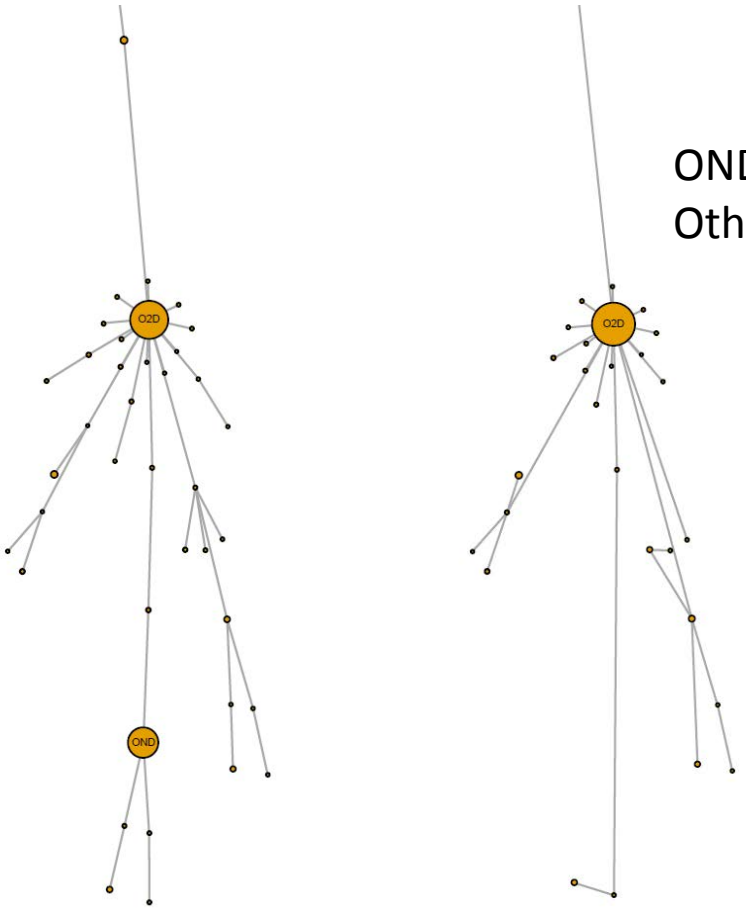
Full data analysis

- Full dataset of 403 gSTs from 15thou dataset (iGraph)



Cluster analysis (CD-HIT)

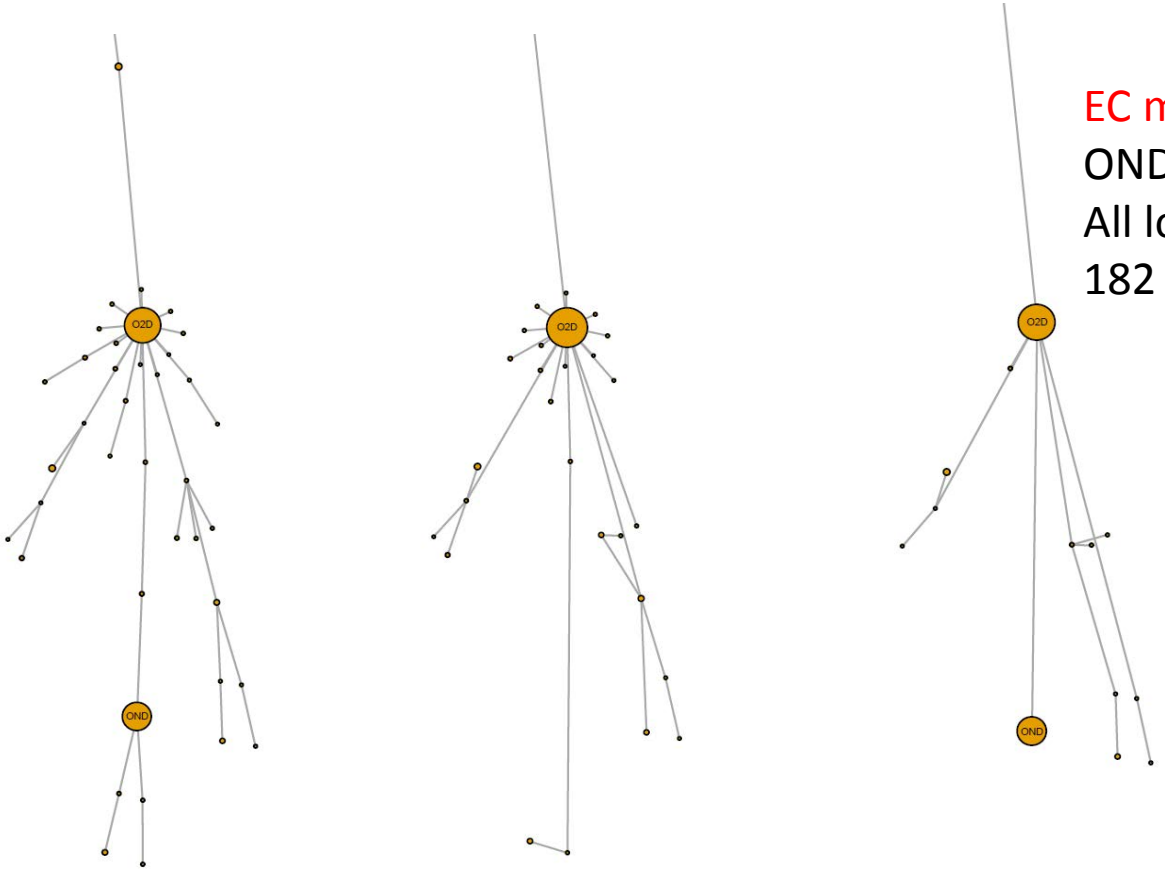
- Cluster analysis of 403 gSTs from 15thou dataset (CD-HIT)
- 218 gSTs (148 novel gSTs) at 99.6% seq identity level
- 11 gSTs from reference database merged with other gSTs: differ by single base SNPs



OND (6780 reads) SNP gST merged into O2D
Other SNPs gSTs remain

Error correction model analysis

- 15thou dataset modelled to identify genuine gSTs
- True gSTs were identified by including specified prior error based on parent/daughter read abundance
- 221 gSTs removed where abundance made up of at least 50% read error



EC model

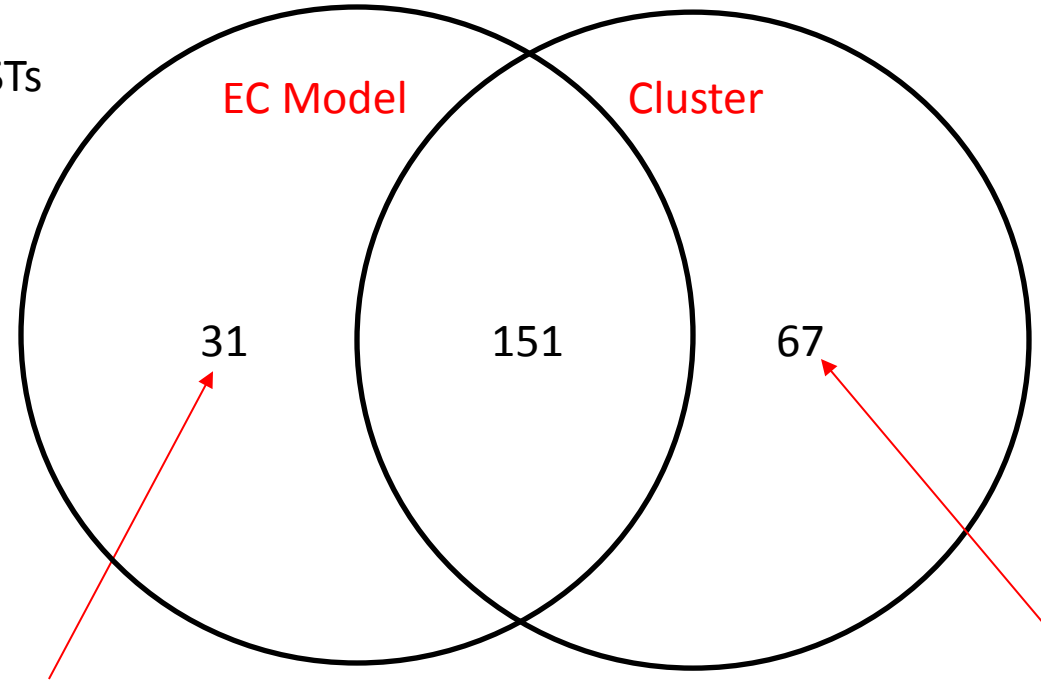
OND (6780 reads) retained

All low abundance SNPs removed

182 gSTs remain (59% novel gSTs)

Comparison of methods (403 gSTs)

154 read error gSTs in both models



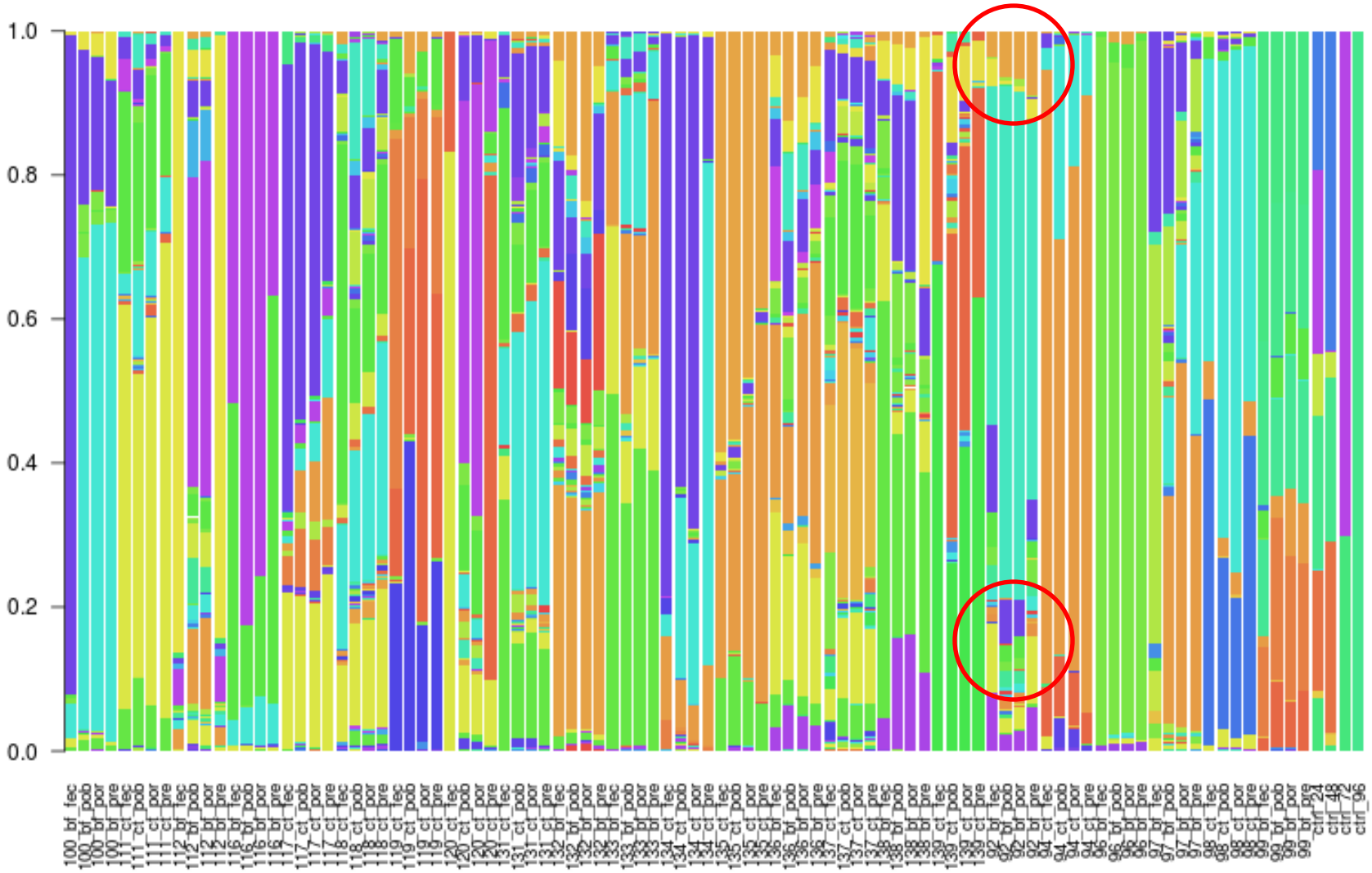
Less likely to be generated by error
Low abundance gSTs
High abundance gSTs

Included in error
Low abundance (novel gSTs) gSTs clustered with high abundance gST e.g. O2D and OND

agresearch

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Relative proportion of gSTs (n=182) across libraries (EC model)

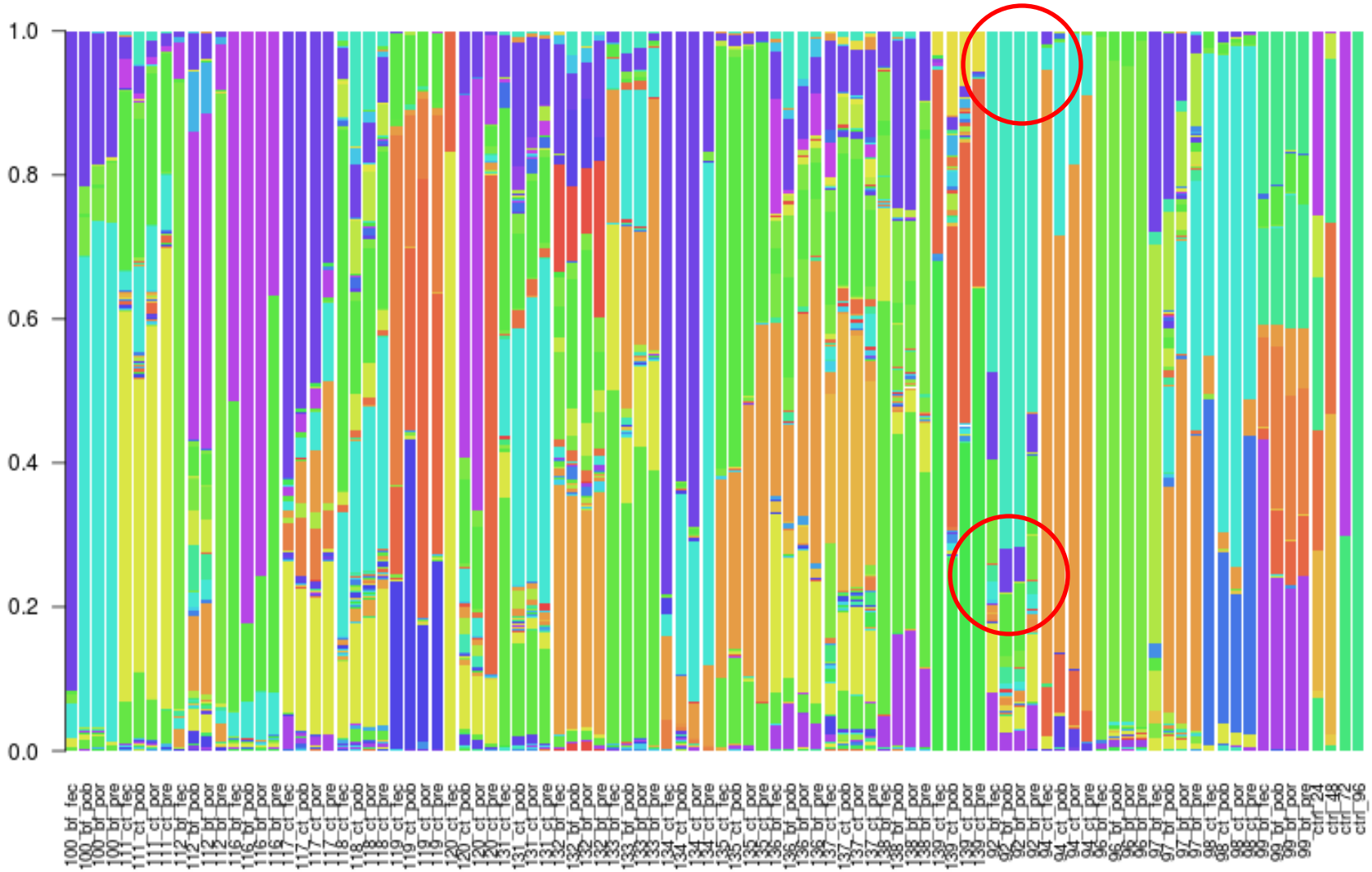


PERMANOVA - 77% variation associated with animal, 3% extraction method

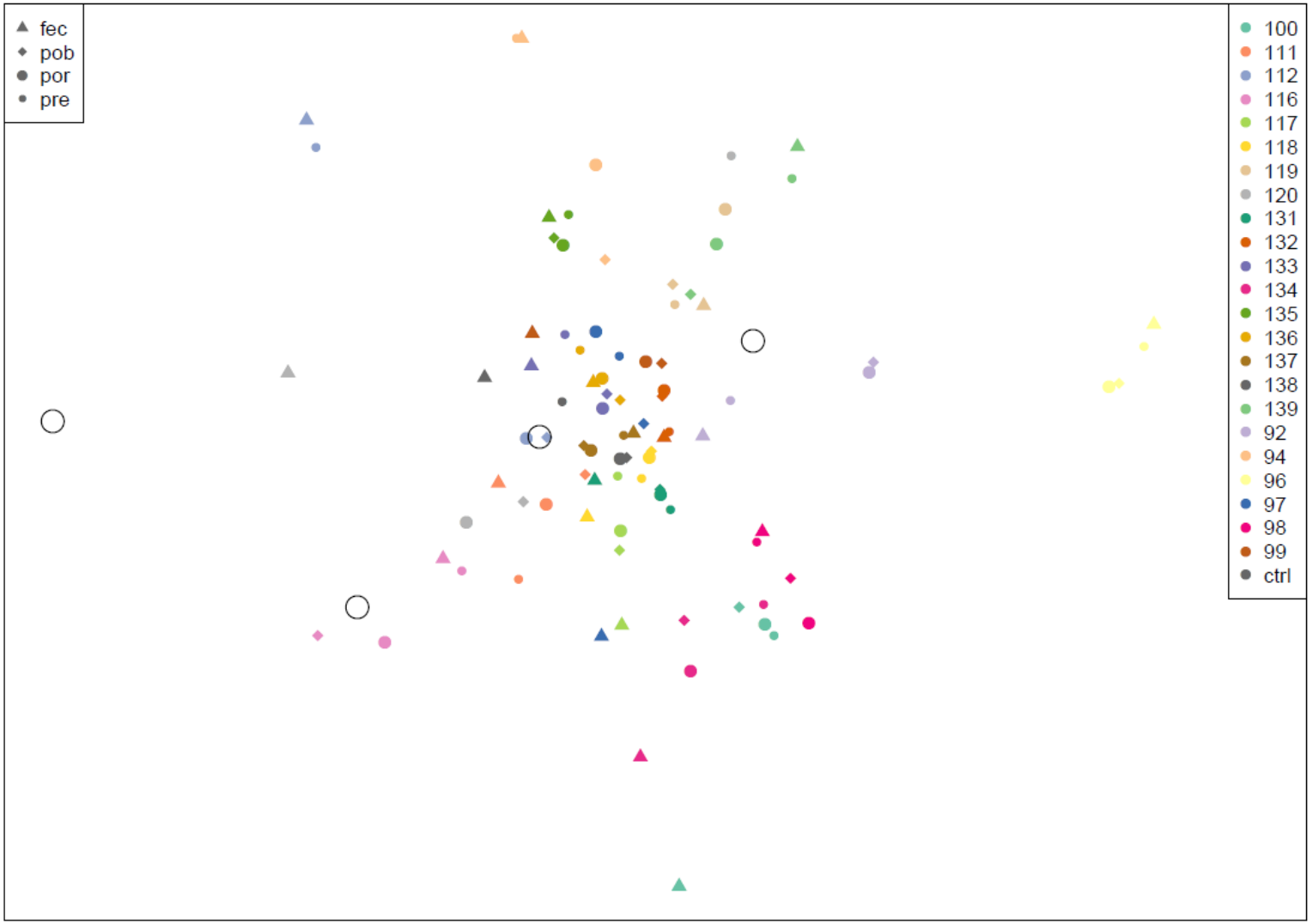
agresearch

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Relative proportion of gSTs (n=218) across libraries (cluster)



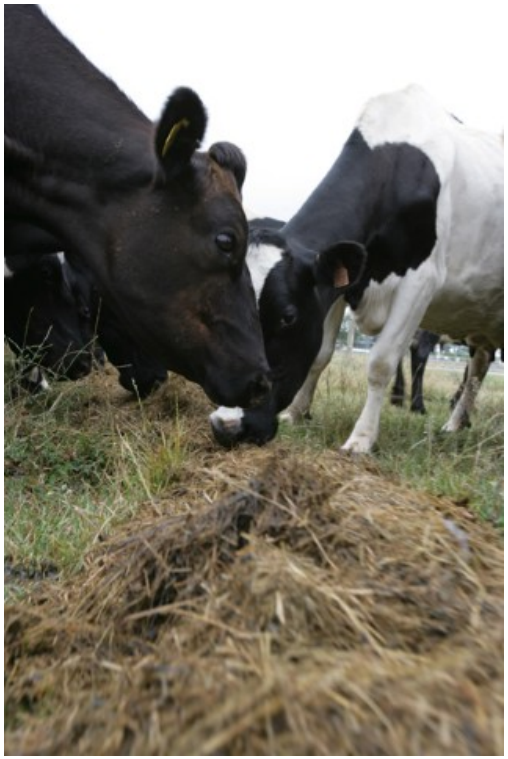
MDS plot of gSTs (relative abundance by library)



Libraries grouped by animal

Conclusions

- *gnd* candidate barcode gene for establishing *E. coli* diversity
- Culture-based methods underestimate *E. coli* community diversity from cattle
  - Up to 104 gST per library (average 35 per animal)
- Many bovine *E. coli* remain to be serogrouped (no gST in reference database
  - 106/107 novel gST best match to *E. coli* gST
- Animal main driver of *E. coli* community diversity in cattle
- No treatment (bifidobacteria) effect
- STEC7 not present in this calf cohort



Future work

Research method to:

- Demonstrate temporal changes in commensal *E. coli* community profile during STEC infection event in cattle
- Identify temporal changes in *E. coli* community structure
  - During maturation of bovine gut – birth to weaning & beyond
  - Before/during/after interventions or stress (disease, antibiotics, calving)
- Examine *E. coli* diversity between species (cattle and sheep) and of contrasting health status
- Shotgun approach to detect industry/clinically important *E. coli*
- Culture-independent approach to assist with targeted *E. coli* isolation
- Preliminary identification of serogroup/serotype (Sanger Sequencing)

Acknowledgements

- Patrick Biggs – MiSeq data analysis & bioinformatics
- Jonathan Marshall – mathematical modelling and statistics
- Rose Collis – PCR, Sanger sequencing
- Angie Reynolds – *gnd* library preparation

- New Zealand Genomics Ltd at Massey University

- Rose Collis was the recipient of an AgResearch Core-funded Summer Studentship (2014-2015 and 2015-2016)

- This work was funded through AgResearch Core (Curiosity) Funding

mEpiLab

OIE Collaborating Centre for Veterinary Epidemiology and Public Health