Evolution and niche adaptation of a multi-host pathogen, *Campylobacter jejuni*, New Zealand

Anja Friedrich

[m]EpiLab, Hopkirk Research Institute, Allan Wilson Centre

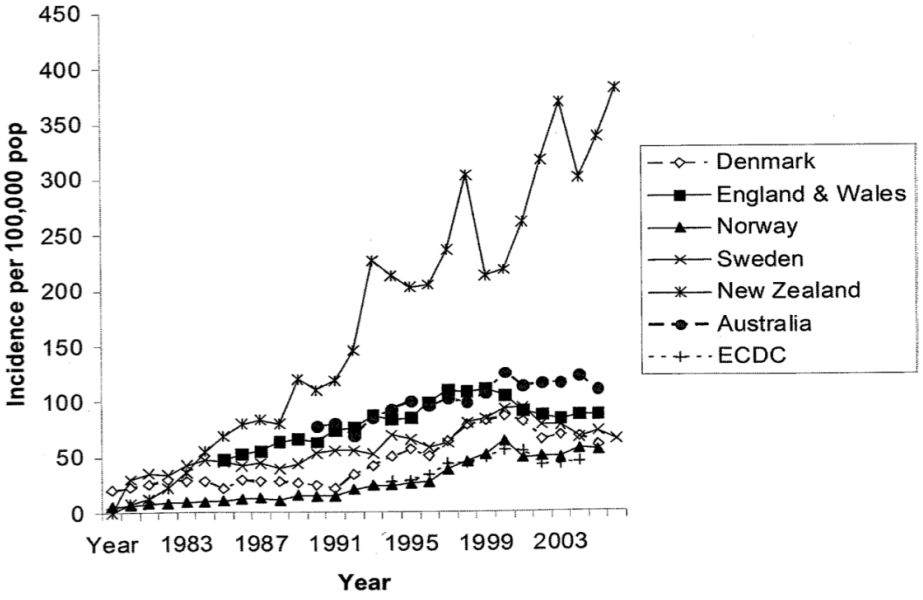
DEPE workshop

10th/11th June 2012

Collaborating Centre

mEpiLab.massey.ac.nz

Why are we doing this?



- most common bacterial cause of human gastroenteritis worldwide
- ability to infect multiple hosts via food and the environment

- New Zealand has a unique geographic location
- questions remain about the genetic basis and ecology of host specificity and niche adaptation

Sears A, Baker MG, Wilson N, Marshall J, Muellner P, Campbell DM, et al. Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. Emerg Infect Dis

Study overview

Study 1: Population structure, evolution and host association of *Campylobacter jejuni* in New Zealand

Study 2: Use of phenotypic microarrays to identify phenotypic characteristics of a range of host associated related genotypes

Study 3: Seasonality of a common *Campylobacter jejuni* genotype which is associated with human infections

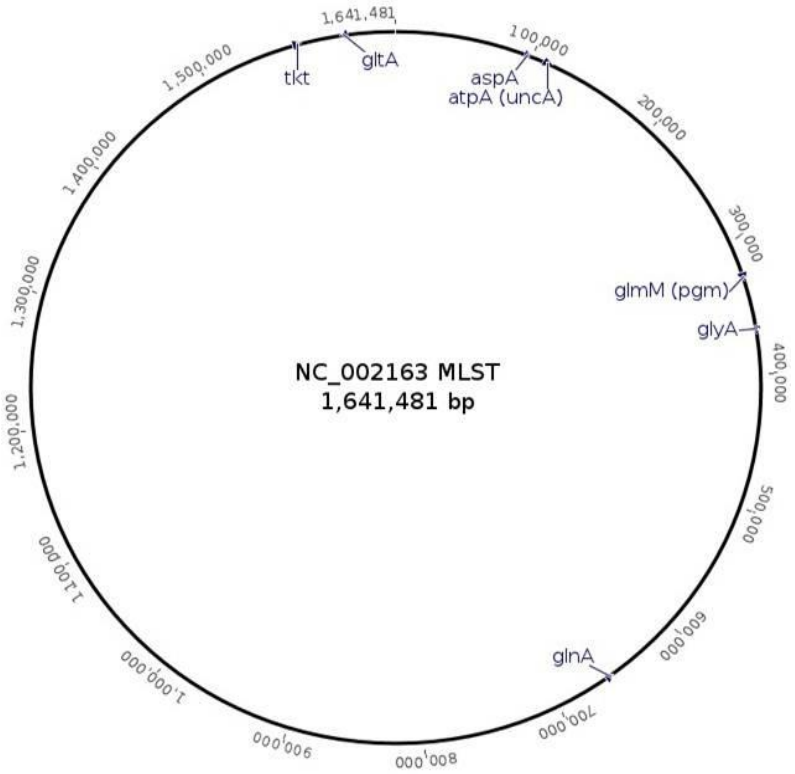
Study 4: Evidence of *in vivo* HGT between two examined isolate lineages

Population structure, evolution and host association of *Campylobacter jejuni* in New Zealand

- *Campylobacter* is able to colonize multiple hosts, and therefore a number of different niches
- the understanding of the relationship between host niche and the lineage structure is important for attempts to reduce the disease
- examining niche segregation and genetic admixture within the New Zealand *Campylobacter* strains
- test of two hypothesis: first, the existence of lineages with low levels of horizontal gene transfer, and their relation to niche specialism and second that *C. jejuni* sequence types is a predictor for host origin

MLST

- nucleotide sequence stretches of 400 to 600 bp from 7 housekeeping genes
- each internal gene fragment is characterized by a unique number
- isolate is characterized by seven numbers building an allelic profile
- ST is defined by allelic profile
- CCs are cluster of closely related STs



Source Type	Sample Type	*aspA*	*gln A*	*glt A*	*gly A*	*pgm*	*tkt*	*unc A*	ST	CC
Chicken		9	25	2	10	22	3	6	52	52
Environment	Duck faeces	2	1	21	3	2	1	5	53	21
Environment	Starling faeces	2	1	21	3	2	1	5	53	21
Chicken		2	1	21	3	2	1	5	53	21

Dataset

- Dataset: ~ 4200 *Campylobacter jejuni* isolates typed by MLST
- animal populations represented by the isolates:
  farmed: chicken, ruminants (sheep, cattle), turkey, farmed ducks
  non-farmed: water birds (wild ducks, geese), rails, starlings, gulls
- environmental and human isolates have been excluded from this study (non amplifying host)

- Software used: BAPS , ClonalFrame, MEGA5

References:

BAPS: J. Corander and P. Marttinen. Bayesian identification of admixture events using Multilocus molecular markers. Molecular ecology, 15(10):2833-2843, 2006

ClonalFrame: X. Didelot and D. Falush. Inference of bacterial microevolution using multilocus sequence data. Genetics, 175(3):1251-1266, 2007

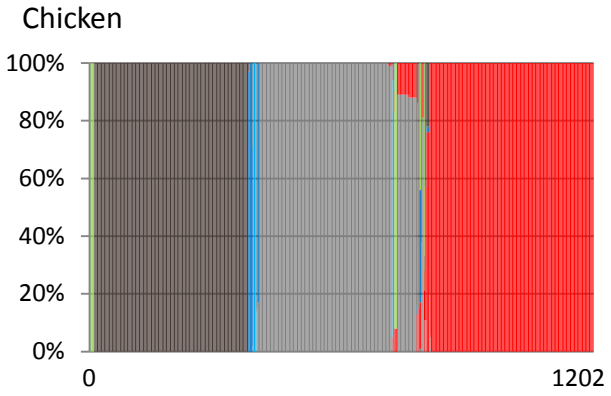
Niche/host associated clusters defined by BAPS

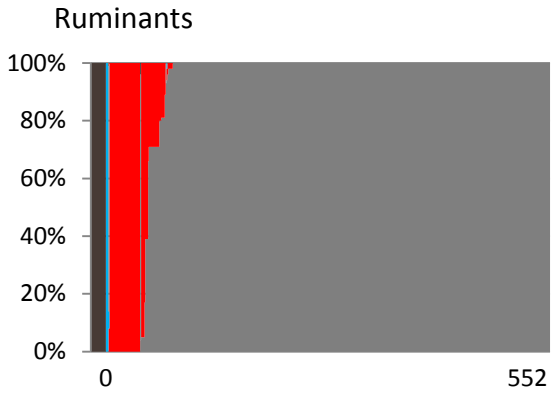
Proportions of isolates (%) in each BAPS- defined cluster which were collected from nine sources

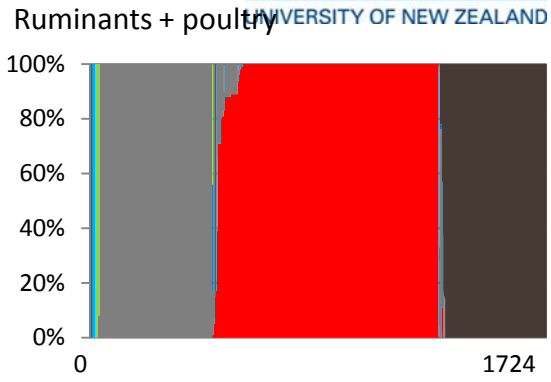
Source	1(47)	2(54)	3(532)	4(106)	5(55)	6(52)	7(144)	8(422)	9(85)	10(20)	11(24)	12(81)	13(65)	14(107)	15(56)	16(33)	17(92)
farmed	**0.03**	**0.03**	**0.29**	**0.06**	**0.03**	**0.03**	**0.08**	**0.23**	**0.05**	**0.00**	**0.01**	**0.04**	**0.01**	**0.04**	**0.01**	**0.02**	**0.05**
chicken	0.00	0.04	0.26	0.09	0.03	0.00	0.11	0.31	0.06	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.01
farmed ducks	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.92	0.00	0.00	0.00
sheep	0.11	0.01	0.33	0.00	0.04	0.16	0.02	0.04	0.04	0.00	0.01	0.13	0.00	0.00	0.00	0.00	0.12
turkey	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00
cattle	0.07	0.01	0.48	0.00	0.04	0.03	0.04	0.03	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.01	0.16
non-farmed	**0.01**	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.00**	**0.10**	**0.00**	**0.17**	**0.09**	**0.03**	**0.03**	**0.25**	**0.30**	**0.00**	**0.00**
ducks & geese	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.11	0.00	0.02	0.01	0.00	0.48	0.28	0.07	0.00	0.00
gulls	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.05	0.19	0.00	0.67	0.00	0.00
rails	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
starlings	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.23	0.05	0.12	0.09	0.05	0.00	0.00
Dominant CC	403	52	21	48	48, 206	21	48	45	692, 1034	1304, U/A	177, 677	42	1034, U/A	354	1332, 1275, U/A	257	61

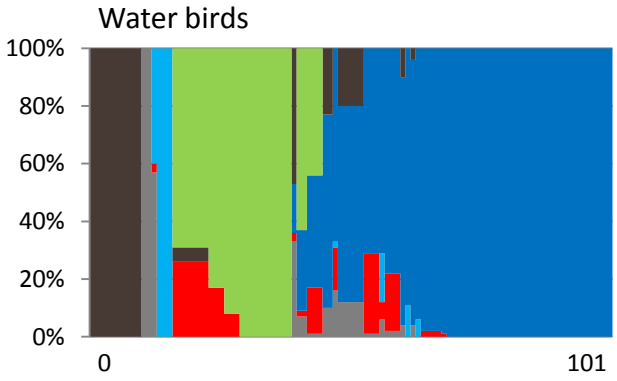
The number of isolates in each cluster is given in brackets next to the cluster number.
The weighted sum of isolates (%) from farmed and non- farmed subcategories is given in bold.

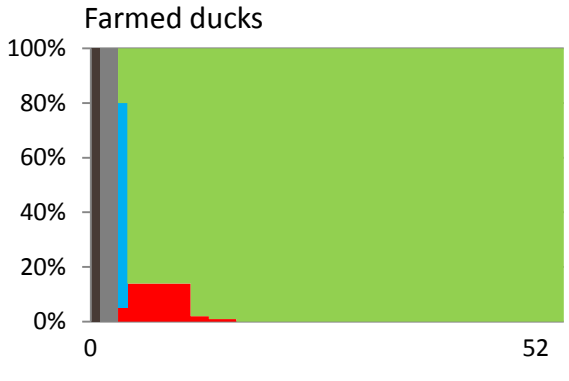
Admixture plots based on BAPS analysis

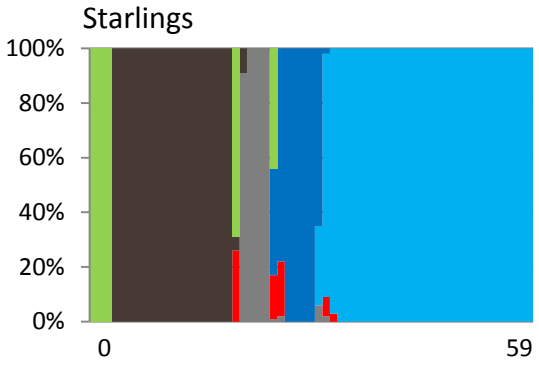




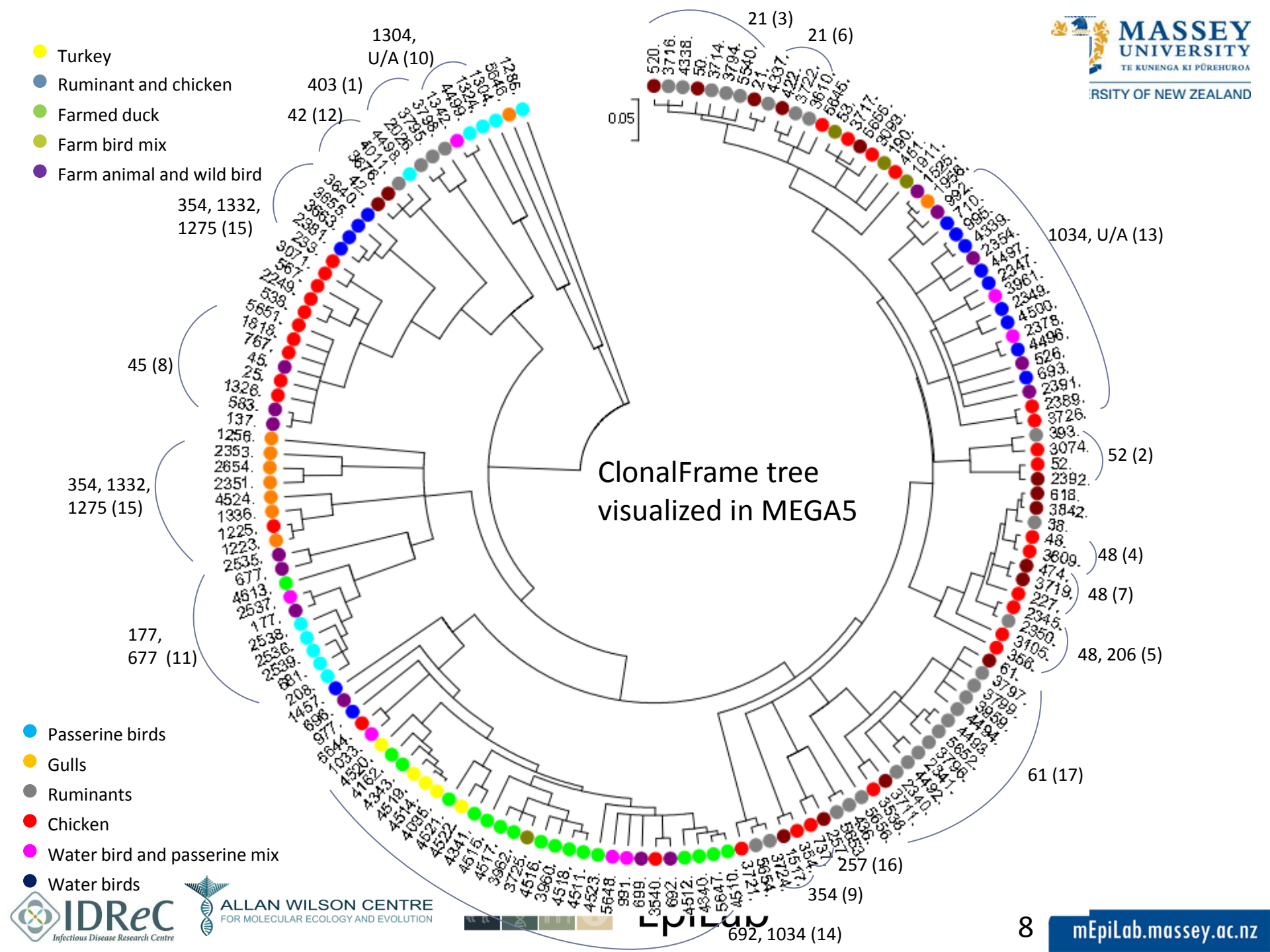




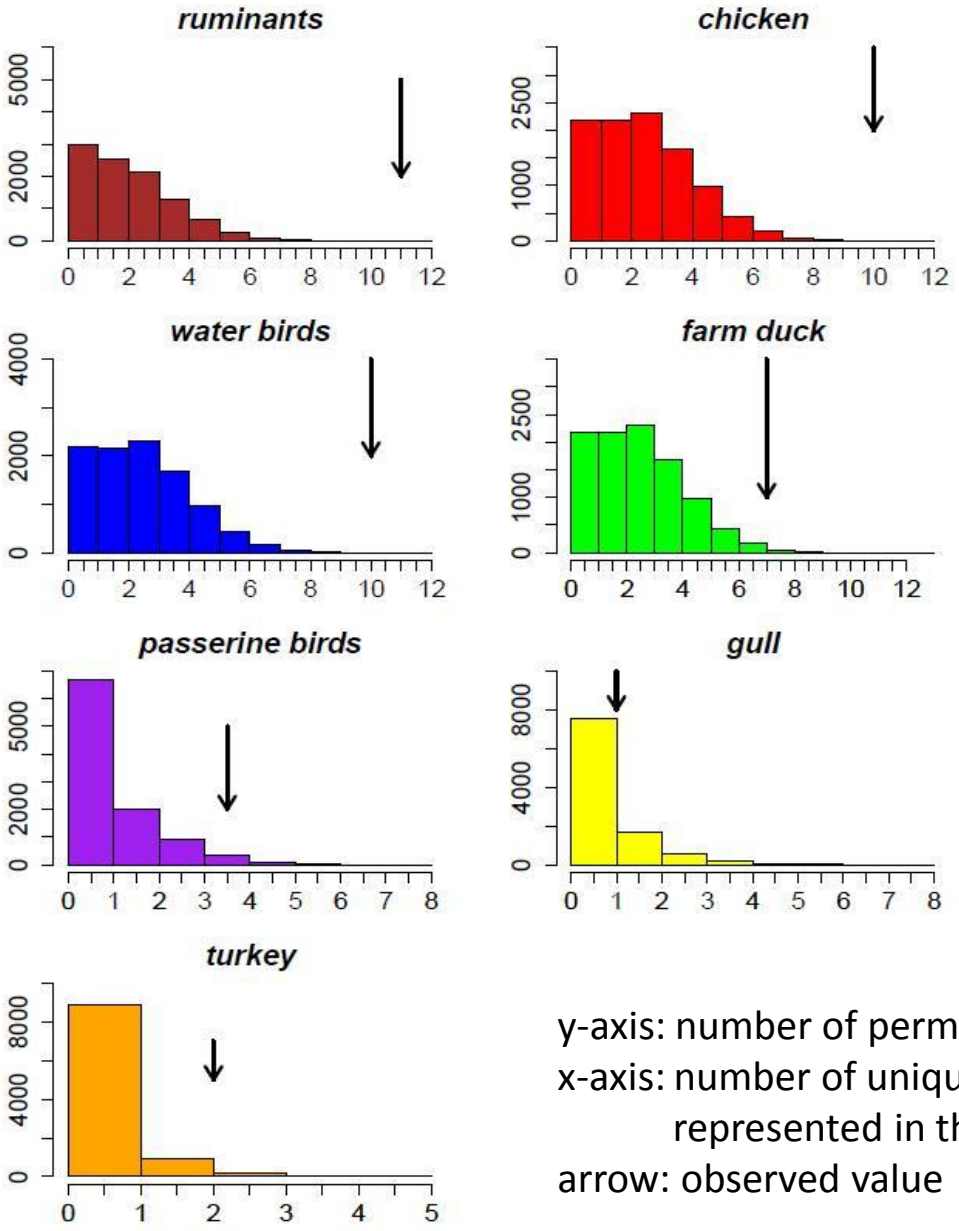


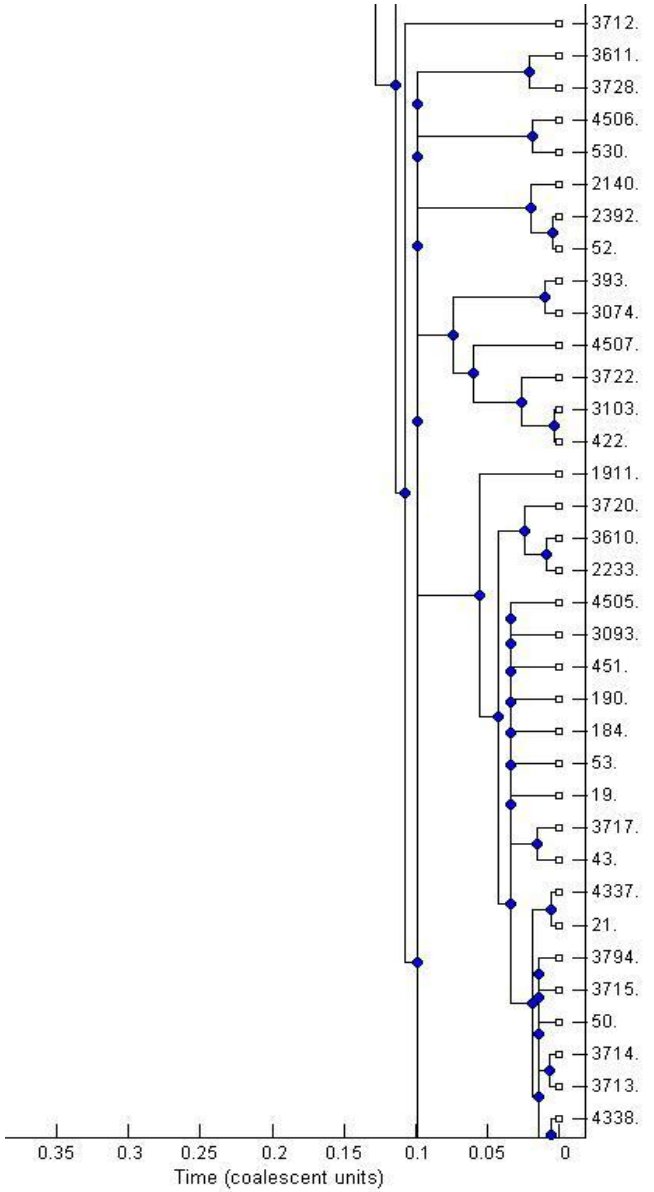


Cluster 1, 3, 5, 6, 12, 17	ruminants	403, 21, 48, 206, 42
Cluster 13 and 15	water birds	1034, 1332, 1275, U/A
Cluster 14	farmed ducks	692, 1034
Cluster 2, 4, 7, 9, 16	chicken	52, 48, 354, 257
Cluster 10 and 11	starlings	1304, 177, 677, U/A
Cluster 8	rum + poultry	45

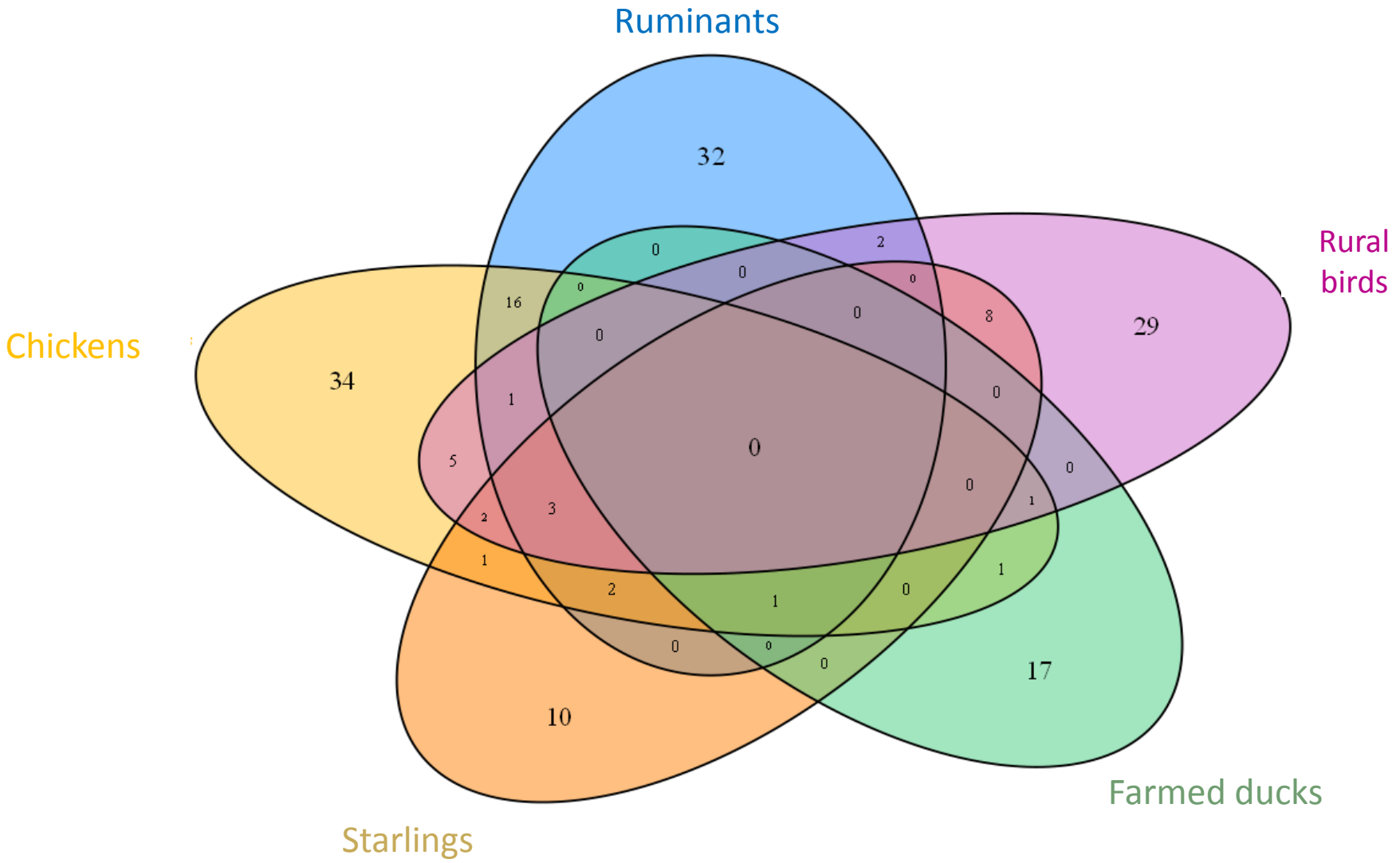

Turkey
Ruminant and chicken
Farmed duck
Farm bird mix
Farm animal and wild bird
1304, U/A (10)
403 (1)
42 (12)
354, 1332, 1275 (15)
45 (8)
354, 1332, 1275 (15)
177, 677 (11)
21 (3)
21 (6)
0.05
1034, U/A (13)
52 (2)
48 (4)
48 (7)
48, 206 (5)
61 (17)
257 (16)
354 (9)
692, 1034 (14)
ClonalFrame tree visualized in MEGA5
Passerine birds
Gulls
Ruminants
Chicken
Water bird and passerine mix
Water birds

Statistical comparison based on evolutionary clades and host associated STs

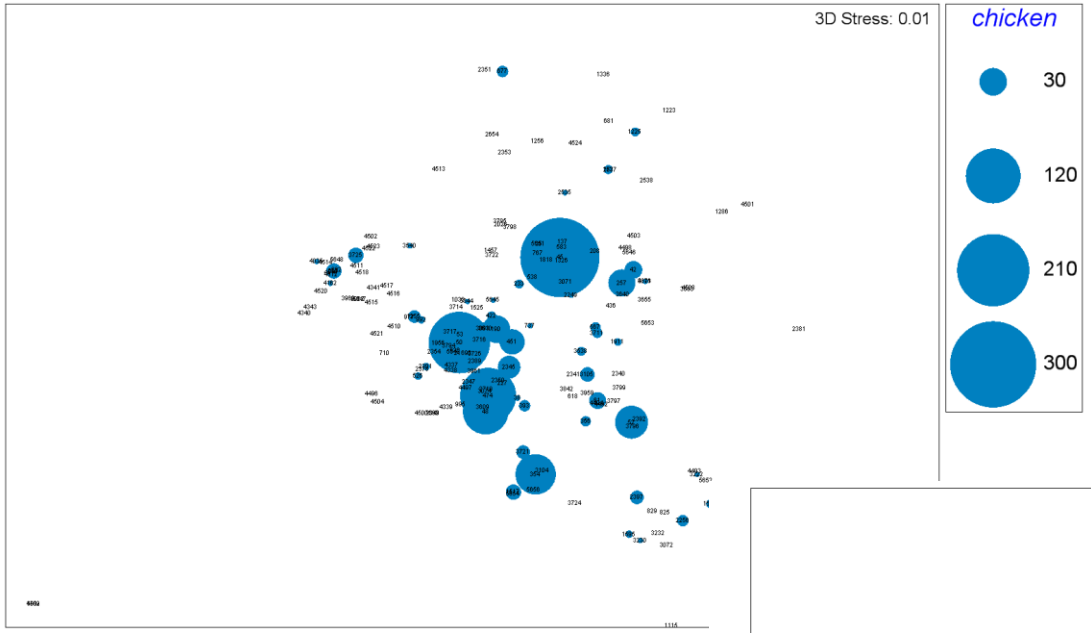


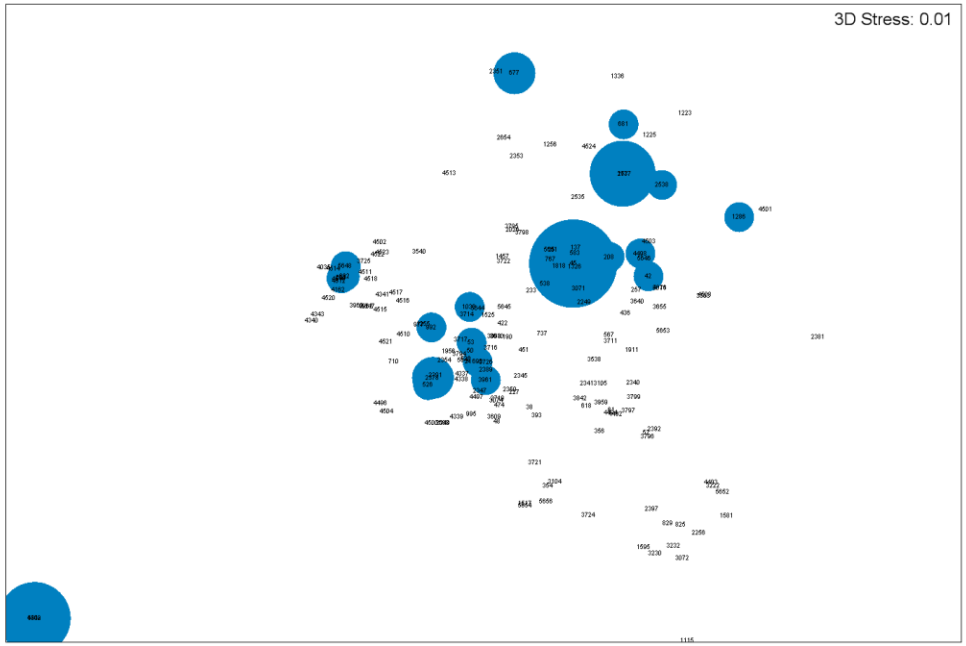


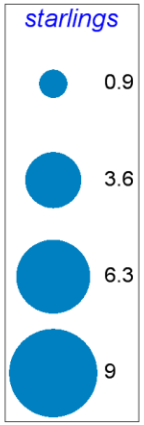
Venn diagram of host associated STs in *C.jejuni*



Comparison of population structure of agriculture and wild species







Clarke, KR (1993). Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117-143.

Discussion

- findings show host specific lineages in New Zealand and strains with low/ high levels of horizontal gene transfer
- *C. jejuni* genotype is a predictor for host or niche (farmed?) origin
- results are consistent with previous research where distinct *Campylobacter* lineages were associated with different host species
- dominance of genetically similar or identical 'farm-type' *C.jejuni*, present in organisms as distantly related as cattle and chickens, may suggest that selection for these types transcends host species in the farmed environment
- manipulation of the host niche may have contributed to genetically distant domestic animals (chickens and cattle) sharing similar genotypes

Study overview

Study 1: Population structure, evolution and host association of *Campylobacter jejuni* in New Zealand

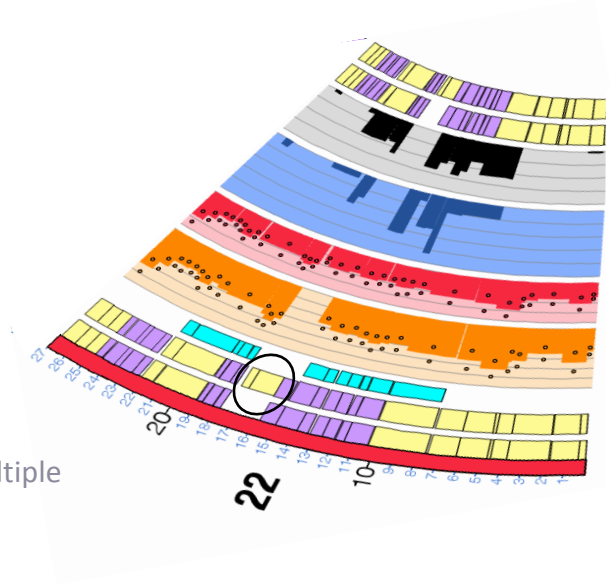
Study 2: Use of phenotypic microarrays to identify phenotypic characteristics of closely related strains

Study 3: Seasonality of a common *Campylobacter jejuni* genotype which is associated with human infections

Study 4: Evidence of *in- vivo* HGT between two examined isolate lineages

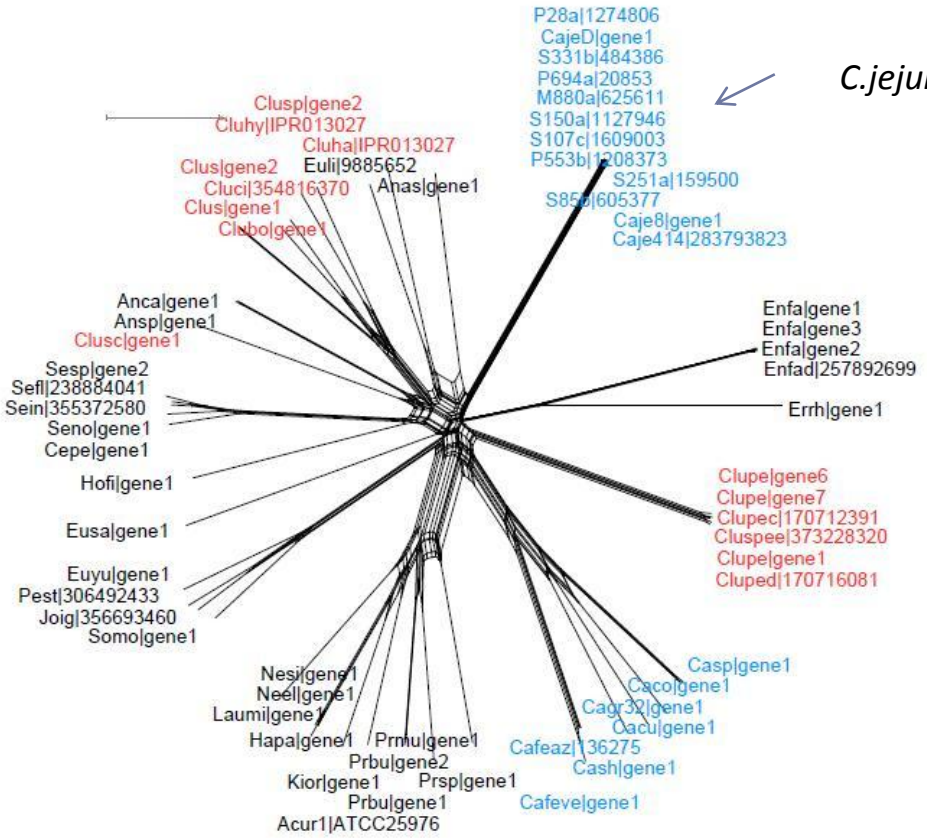
Background

- ST-474, rarely found anywhere outside New Zealand, accounts for quarter of notified human *Campylobacter* cases
- insertion between ORFs Cj1069-Cj1070 with >99% identity to *ykg*C (pyridine nucleotidedisulfide - oxidoreductase protein)
- rare in ST-42 (4/21) and ST-474 (6/47), common ST-61
- association of *ykg*C with ruminants or ruminant faeces may indicate a specific niche for this ST-474 variant
- *ykg*C insertion associated with particular phenotype? Distinct differences?
- PM system will be described later



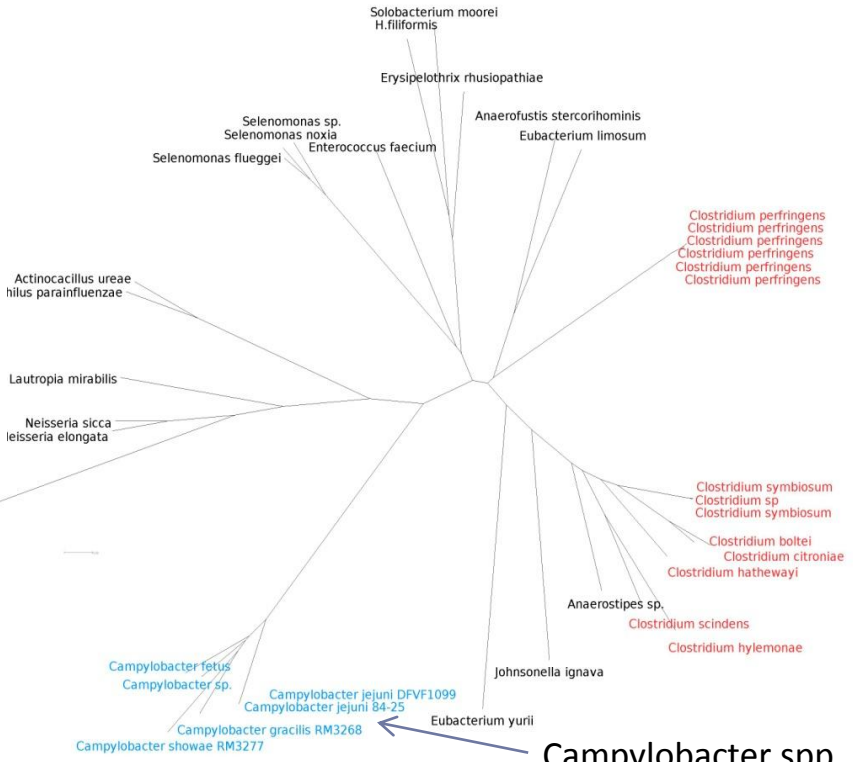
Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, et al. (2011) Whole-Genome Comparison of Two *Campylobacter jejuni* Isolates of the Same Sequence Type Reveals Multiple Loci of Different Ancestral Lineage. PLoS ONE 6(11): e27121. doi:10.1371/journal.pone.0027121

*ykg*C phylogeny network vs. 16S phylogeny tree



C.jejuni

*ykg*C network



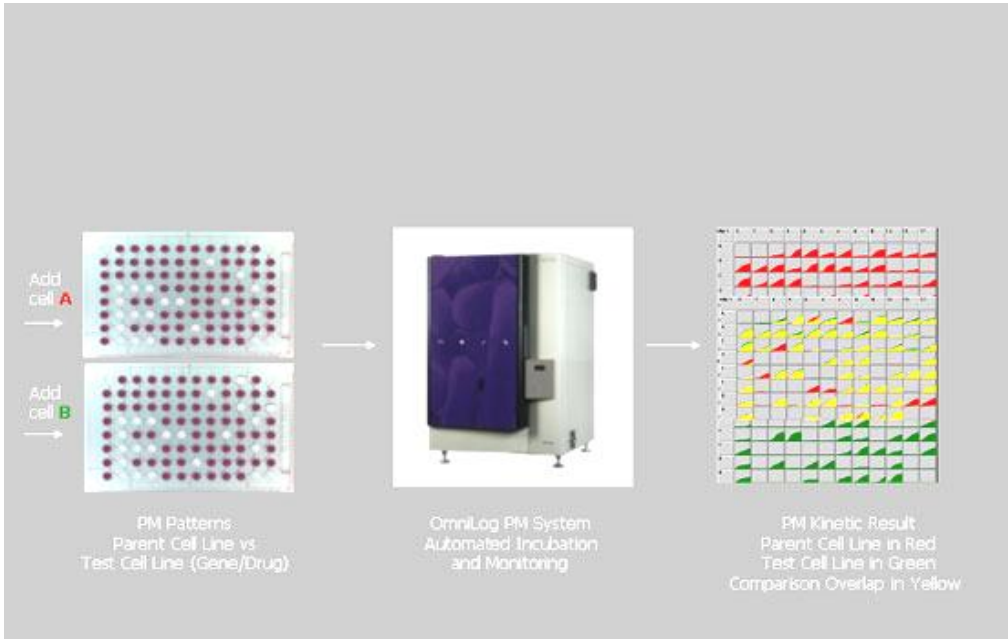
16S tree

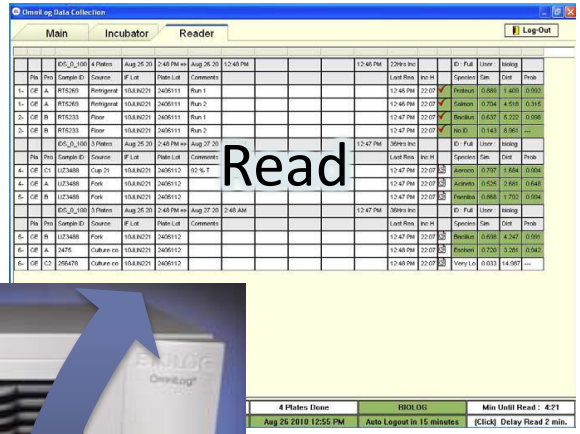
other Campylobacter spp.

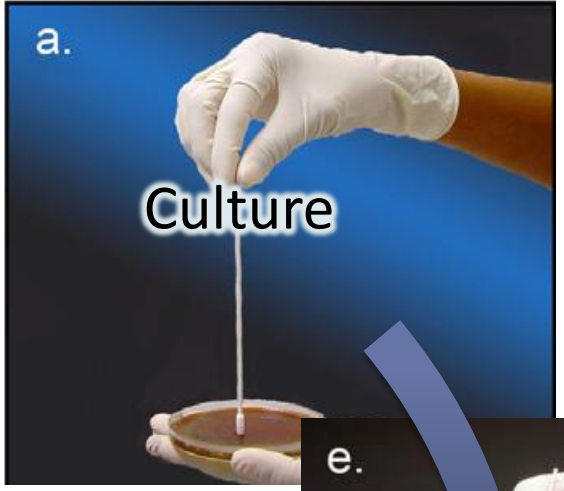
Campylobacter spp.

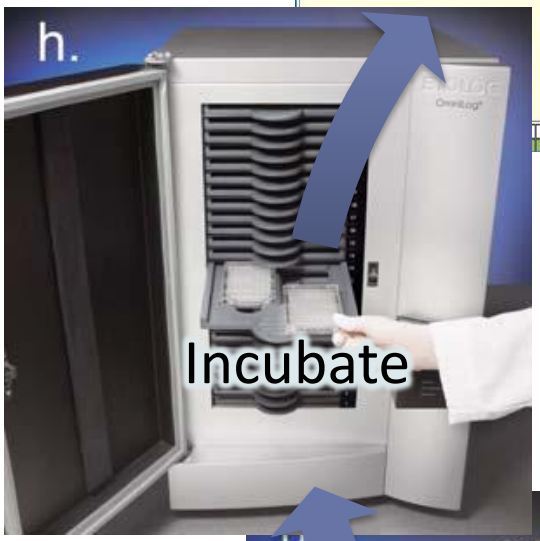
the Omnilog system

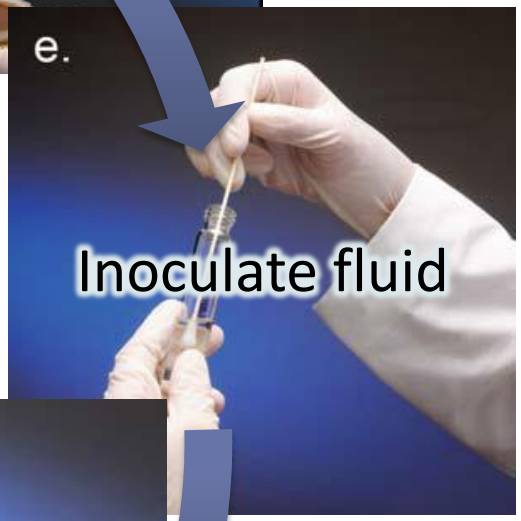
- high throughput phenotypic microarray (PM) system
- up to 20 96-well plates
- each well of the array (96) is designed to test the ability of the bacteria to utilise different sources of nutrient (e.g. carbon, nitrogen, sulphur,...) or to test for varying effects of osmotic, pH environment...
- plates are incubated for up to 96h
- respiration leads to reduction of tetrazolium dye
- intensity of the colour is recorded every 15 min by a CCD camera











MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

IDReC Infectious Disease Research Centre

ALLAN WILSON CENTRE
FOR MOLECULAR ECOLOGY AND EVOLUTION

mEpiLab

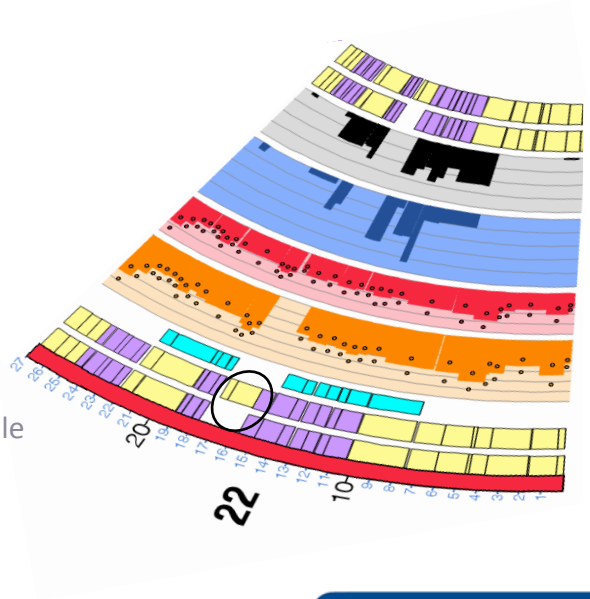
mEpiLab.massey.ac.nz

Overview of isolates used in the study

Host	Sequence Type	ykgC +	ykgC -	C. coli insert +	C. coli insert -	strain ID
Human	474	√				H22082
Poultry	474		√			P110b
Poultry	474	√				P694a
Ruminant	474	√				S168b
Human	474		√			73020
Ruminant	474		√			S330a
Ruminant	2026			√		S22b
Human	2026			√		28548
Human	42	√				H180
Human	42		√			H550
Ruminant	42	√				M602b
Ruminant	42		√			S355b
Ruminant	42		√			S263b
Human	61	√				H450b
Ruminant	61	√				S276b

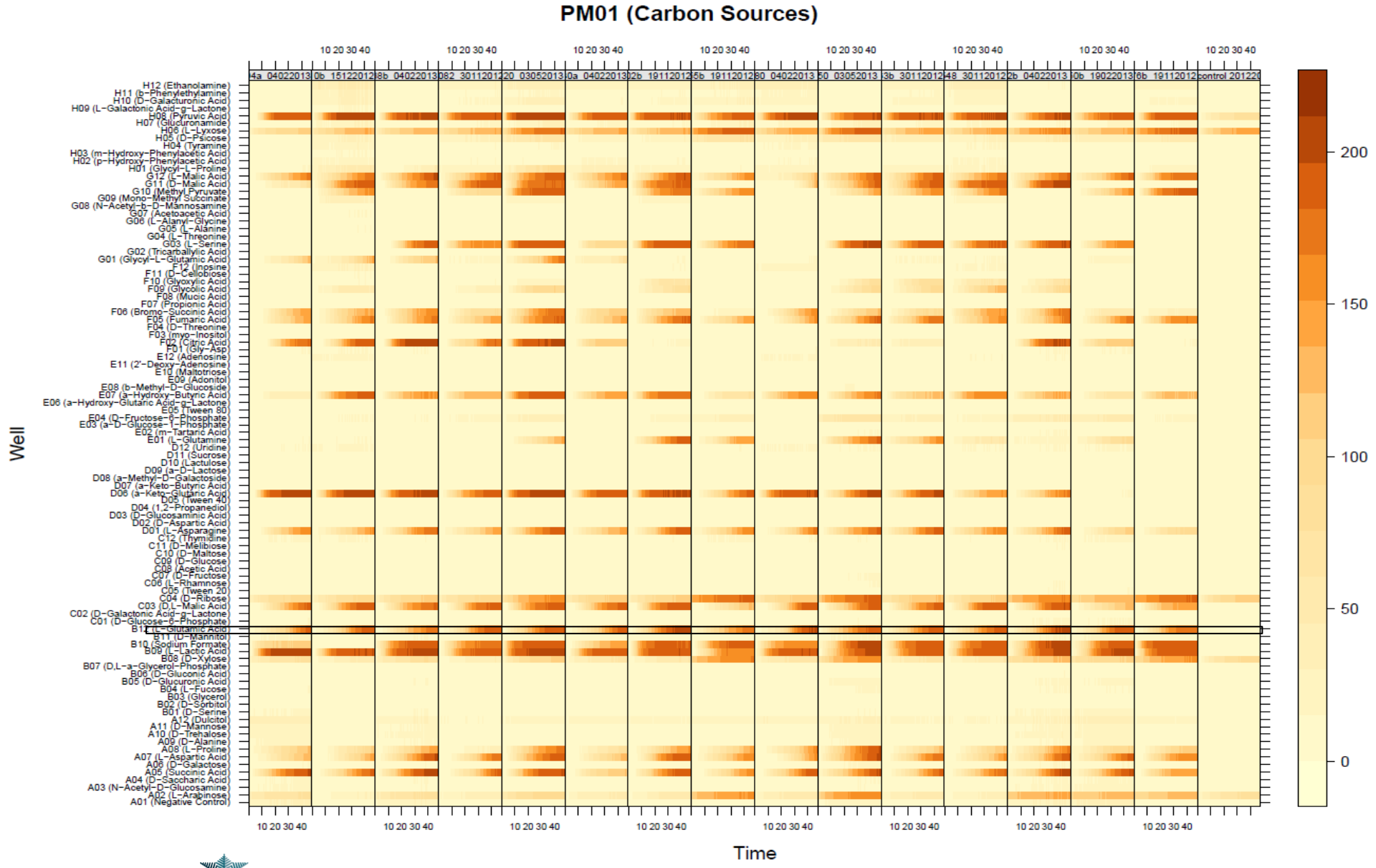
- Ruminants and humans body temperature:

  ~ 38°C, poultry body temperature: ~ 42°C

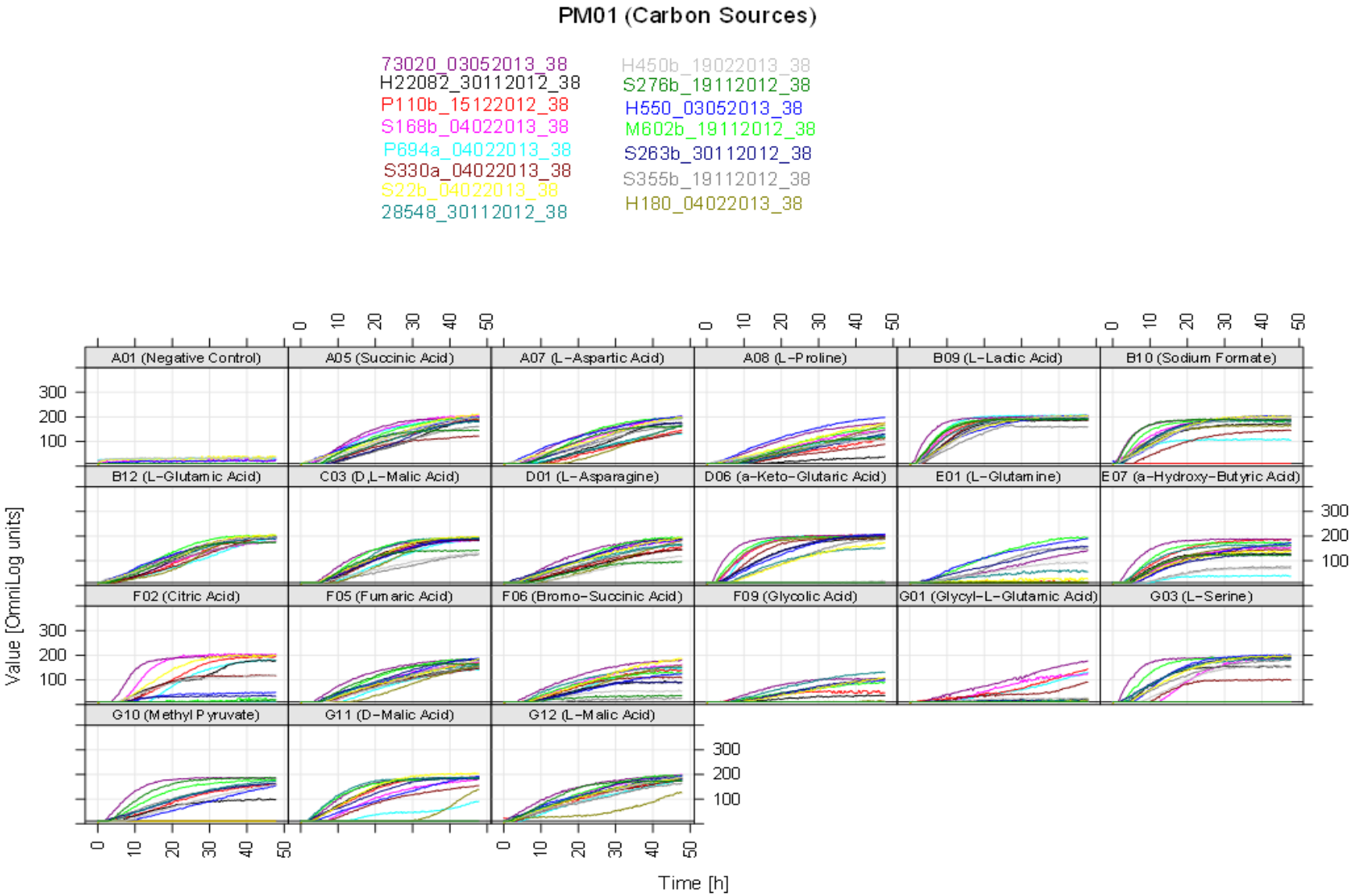


Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, et al. (2011) Whole-Genome Comparison of Two *Campylobacter jejuni* Isolates of the Same Sequence Type Reveals Multiple Loci of Different Ancestral Lineage. PLoS ONE 6(11): e27121. doi:10.1371/journal.pone.0027121

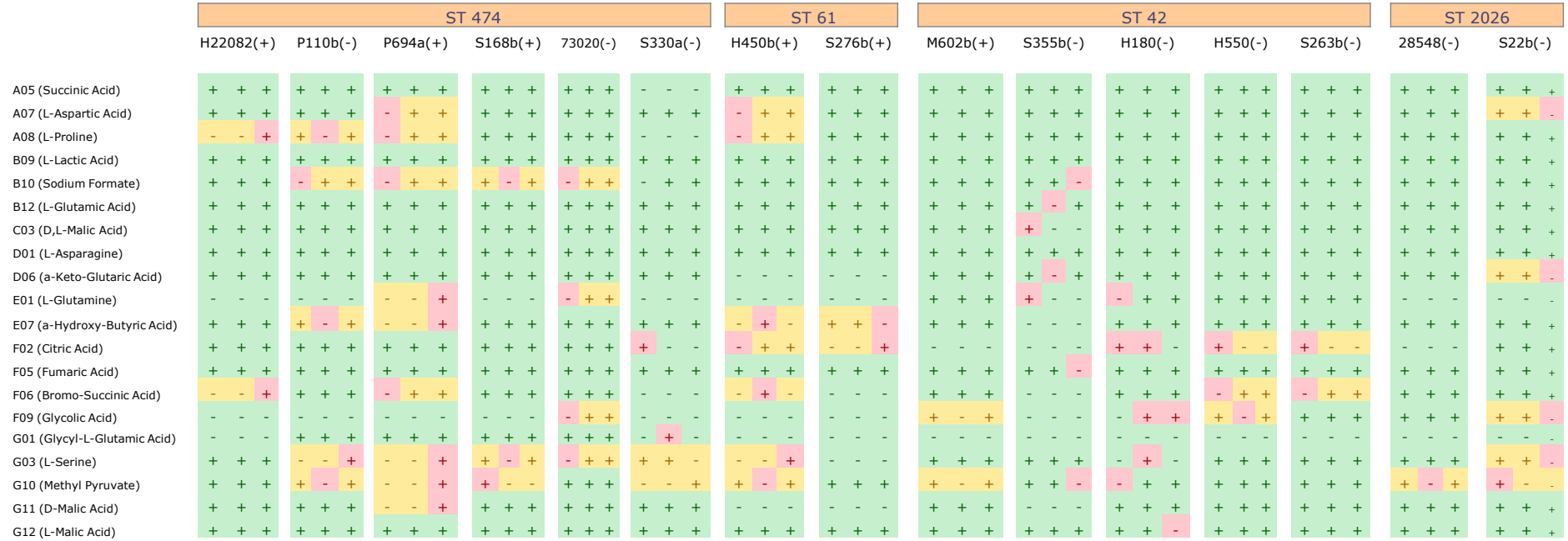
Level plot across all isolates and assays



XY- plot on selected assays



Utilisation across replicates



Key:

- consistent across all 3 replica
- 2 out of 3
- poor repeat

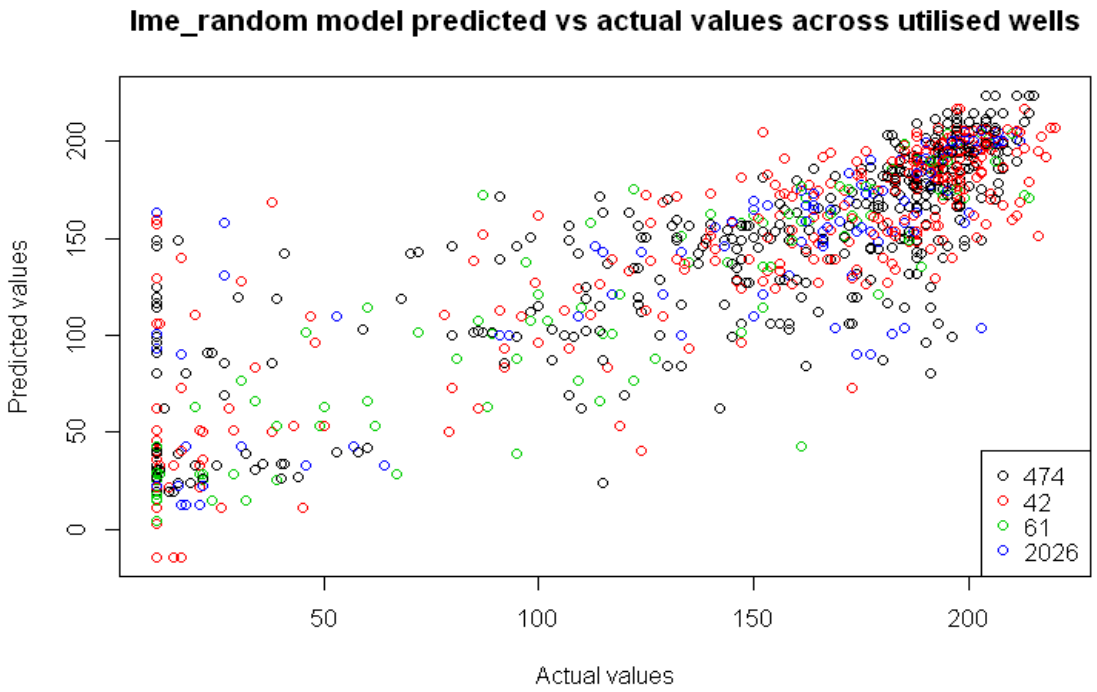
Statistical based analysis...

- Relationship between phenotype and genotype?
  - Phenotype – carbon source assay
  - Genotype – *ykg*C (presence/absence), MLST type
  - Random effect – isolate (replicates)
  - Outcome variable: A-value (maximum height of the curve)

- Possible approaches:

  - Generalised linear models (without random effect)
  - Linear mixed effects models (with stratification and random variables)
  - REEM trees (Regression Trees with Random Effects for Longitudinal Data)
  - Permanova

Comparison of the different regression models



lme_random <- lme(value ~ variable/(MLST + ykgC) , random=~1|isolate, data=data_long)

AIC value 8561.595
logLik: -4178.797

glm: AIC value 9284.6

lmer: AIC value 8562
logLik: -4179

REEMtree: logLik: - 4570.27

Permanova results

Factors

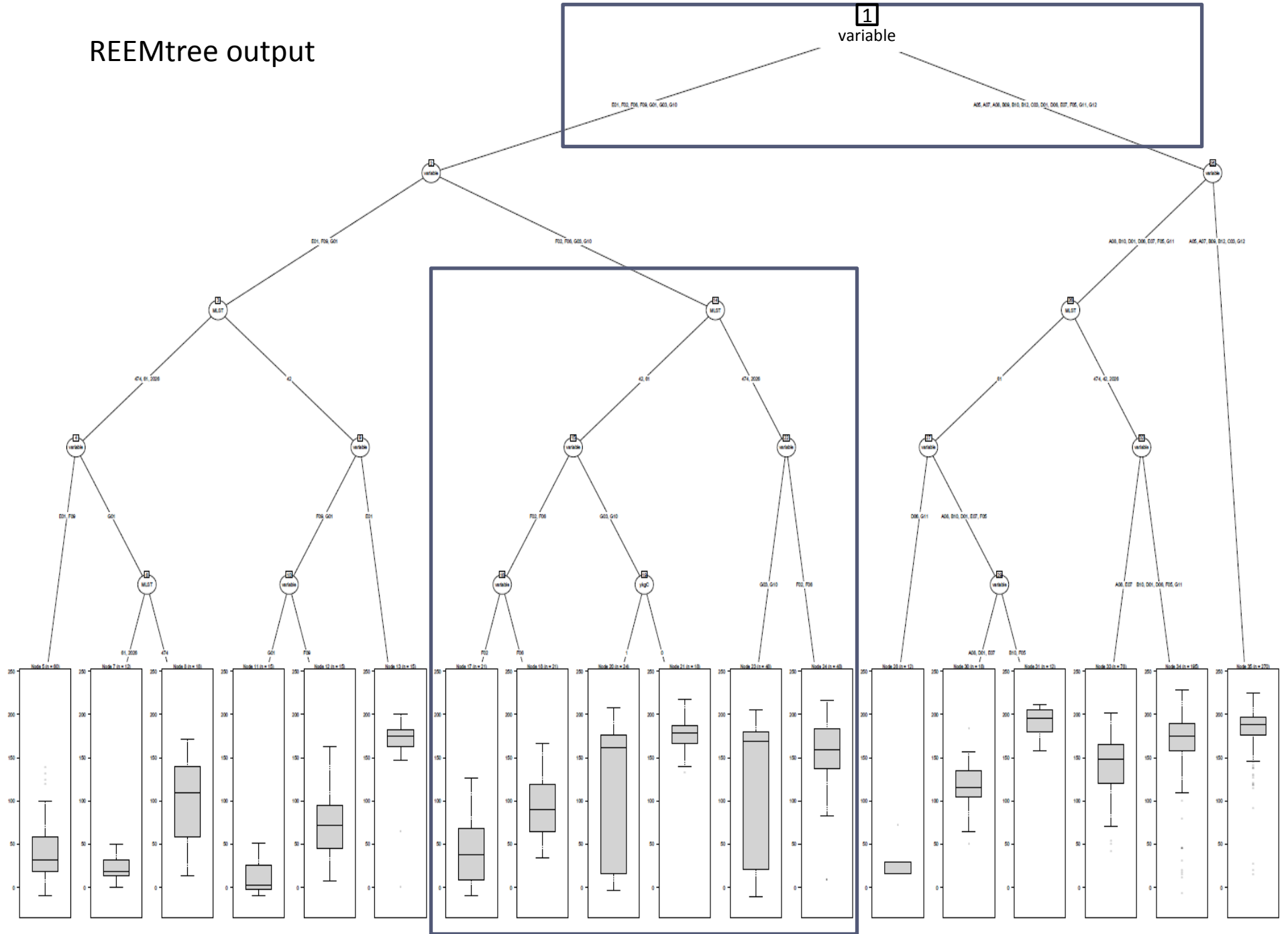
	Name	Abbrev.	Type	Levels
•	MLST	ML	Fixed	5
•	isolate	is	Random	16
•	*ykg*C	yk	Fixed	2

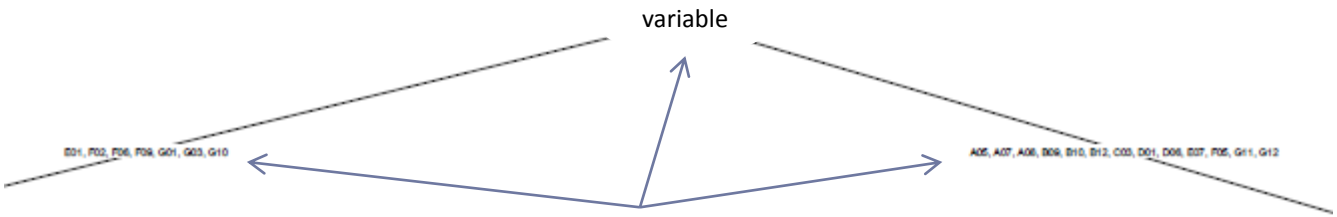
PERMANOVA table of results

	Source	df	SS	MS	Unique Pseudo-F	P(perm)	perms
•	MLST	4	1.584	3.96	13.269	0.0001	9915
•	*ykg*C	1	1.634	1.63	0.5475	0.6832	9951
•	Res	40	1.194	2.98			
•	Total	45	2.835				

McArdle and Anderson 2001, Anderson 2001

REEMtree output

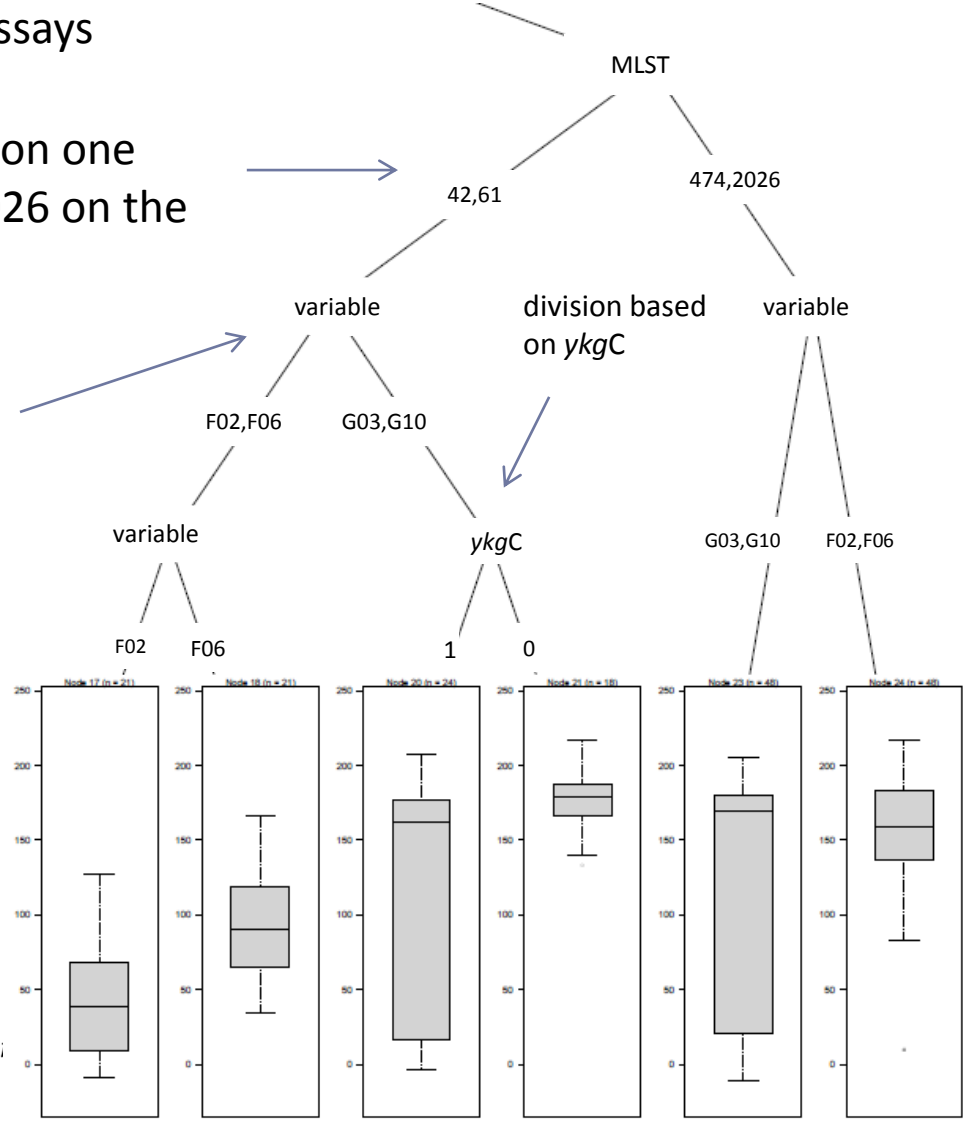




root assigning data to different assays

divided into ST-42 and ST-61 on one branch and ST-474 and ST-2026 on the other branch

division of ST42 and ST61 into different wells



F02: Citric Acid
F06: Bromo- Succinic Acid
G03: L-Serine
G10: Methyl Pyruvate

Findings so far…

- Identified a carbon source for PM4 and PM9 used consistently by examined isolates
- Differences in phenotypic profiles related to MLST type but not to the presence or absence of *ykg*C (in carbon utilisation)

Sequence based analysis

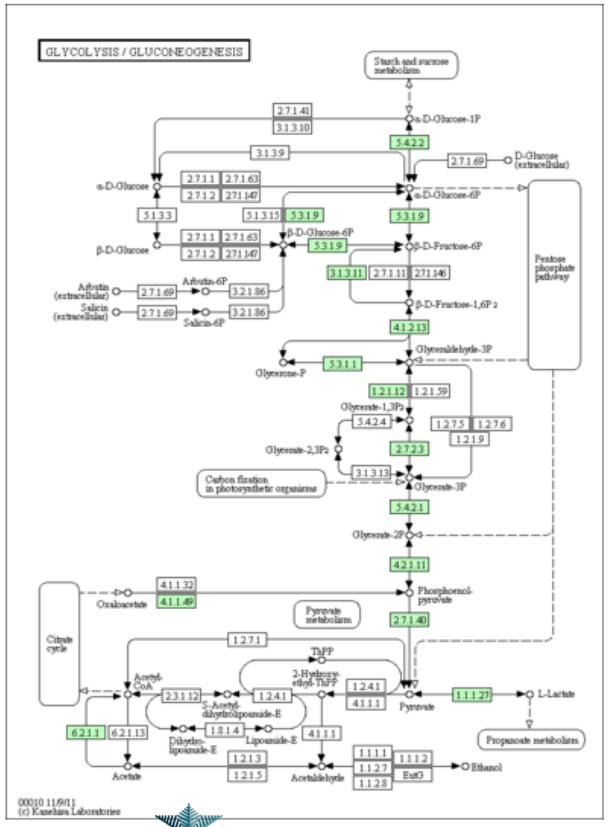
MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

KEGG **Glycolysis / Gluconeogenesis - Campylobacter jejuni RM1221**

[Pathway menu | Organism menu | Pathway entry | Download KGML | Hide description | User data mapping]

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glycerone-P, glyceraldehyde-3P, glycerate-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00307]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glycerone-P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which often corresponds to operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].

Campylobacter jejuni RM1221 | Go | 55%



- Kinetic data output from Omnilog is able to connect directly to KEGG (Kyoto Encyclopaedia of Genes and Genomes) and identify enzymes associated with the carbon source of interest
- Last non-commercial version is from 2011
- Limited identified pathways for *C.jejuni*, other possibilities are to identify enzyme in *E.coli* and find orthologous in *C.jejuni* through Brenda or NCBI

BRENDA
The Comprehensive Enzyme Information System

Next steps…

- Completion of the PM9 study
- Identification of enzymes/ operon complexes related to the carbon sources of interest
- Identification of SNPs and possible relatedness to phenotypic differences
- Analysis of pathways in *Campylobacter jejuni*

Acknowledgments

- Supervisors:
  - Prof Nigel French
  - Prof Paul Rainey
  - Prof Mike Hendy
  - Dr. Patrick Biggs
  - Dr. Anne Midwinter

- Dr. Jonathan Marshall, Dr. Steve Ritchie

- Allan Wilson Centre Doctoral Scholarship