

# Fishing for heritability in the Salmon Gill Microbiome

**Why statisticians should get out into the field**

Elle Saber  
Australian National University

**“Let’s not just improve our model but take  
our model and convince someone to go  
out and sample”**

**Cheng Soon Ong**

**“Then ask if you can come along”**

**Elle Saber**

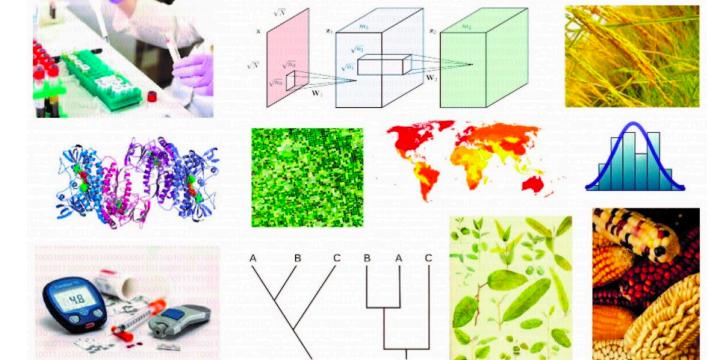
# Acknowledgements

## Supervisory team:

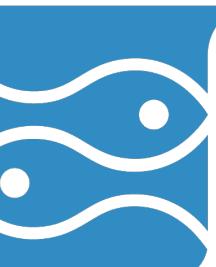
Professor Eric Stone (ANU), Dr Klara Verbyla (Center for Aquaculture Technology),  
Dr Gen Nowak (ANU)

### CSIRO

Joel Slinger  
Peter Kube  
James Wynn  
Andrew Bisset  
Brad Evans  
Richard Taylor  
Scott Cooper  
Nigel Valentine  
Dr Ben Maynard  
Dr Megan Rigby



ANU Biological Data Science Institute



THE CENTER FOR  
AQUACULTURE  
TECHNOLOGIES

### SALTAS

Aaron Wingrove  
Damien Nieuwsteeg  
Ree Oliver  
Jack Brown

### TASSAL

David Cameron  
Rob Newham  
Joana Ferreira Mendez

### Huon Aquaculture

Adrian Steenholdt  
Jarod Wells  
Dr. Jasmine Knowles  
Andrew Bourke  
Emmett Gurr  
Damien Nieuwsteeg



# Motivation

# Background

## Salmon Breeding Program

- North American Atlantic Salmon originating from wild stocks in Nova Scotia in the 1960s
- Atlantic Salmon Breeding Program operated by Salmon Enterprises of Tasmania Pty Ltd (SALTAS).
- Salmon under study are a sample of the marine grow-out salmon used in the genomic selection program.



Received: 14 June 2021 | Revised: 9 September 2021 | Accepted: 10 September 2021

DOI: 10.1111/eva.13304

SPECIAL ISSUE ORIGINAL ARTICLE

Evolutionary Applications Open Access WILEY

### Commercial implementation of genomic selection in Tasmanian Atlantic salmon: Scheme evolution and validation

Klara L. Verbyla<sup>1</sup> | Peter D. Kube<sup>2</sup> | Bradley S. Evans<sup>3</sup>

<sup>1</sup>CSIRO, Black Mountain, Canberra, ACT, Australia

<sup>2</sup>CSIRO, Hobart, Tas., Australia

<sup>3</sup>Tassal Operations, Hobart, Tas., Australia

#### Correspondence

Klara L. Verbyla, Center for Aquaculture Technology, San Diego, CA, USA.  
Email: [kverbyla@auquatechcenter.com](mailto:kverbyla@auquatechcenter.com)

#### Abstract

Genomic information was included for the first time in the prediction of breeding values for Atlantic salmon within the Australian Salmon Enterprises of Tasmania Pty Ltd selective breeding program in 2016. The process to realize genomic selection in the breeding program began in 2014 with the scheme finalized and fully implemented for the first time in 2018. The high potential of within family selection to accelerate genetic gain, something not possible using the traditional pedigree-based approach,

# Motivation

## Fish gills are important

- Gill health drives salmon performance, yet gill diseases like amoebic gill disease (AGD) remain major challenges.
- The gill microbiome may influence health, but we know far less about it compared to gastrointestinal microbiomes.
- Do host genetics shape the gill microbiome or does it simply reflect environmental exposure?



Images from: Taylor, Richard., et al. (2016). Gill score guide-amoebic gill disease (AGD) management training document. *Tassal Operations Pty: Hobart, Australia.*

# Is the Salmon Gill Microbiome Heritable?

# What is heritability?

Proportion of variance explained by (additive genetics)

The total variance of a trait can be partitioned into the genotypic variance and the environmental variance\*:

$$V_P = V_G + V_E$$

The genotypic variance can be further partitioned into additive, dominance and interaction (epistatic) effects

$$V_P = V_A + V_D + V_I + V_E$$

Broad sense heritability:

$$H^2 = \frac{V_G}{V_P}$$

Mostly of interest to plant breeders  
See Fanti's talk on Friday!

Narrow sense heritability:

$$h^2 = \frac{V_A}{V_P}$$

of interest to me

# Sampling Design

*The road to hell is paved with  
good intentions*

# Microbiome Data

- Gills swabs from a *sample* of 800 salmon each from **two** separate cohorts at separate locations.
- Sequencer-based measurement of bacterial content of the gill (16S analysis)
- Microbiome data is a set of reads (counts) against an OTU (operational taxonomical unit)



# Design considerations

## The Plan

- Sampling designed to maximise genetic variation (randomised within families)
- Water samples from the sedation baths were planned as a background reference
- Sampling had to coincide with commercial operations (an existing gill measure)

## The Implementation

- Sampling design mostly achieved with some exceptions
- Recording of baths failed due to technical issues so it was not possible to match water samples to bath numbers or individuals to bath numbers

# The Model

## Estimated using ASReml-r

BLUP Model treating microbiome phenotype

$$\text{Model: } y = Xb + Z_n u_n + Z_g u_g + e$$

$y$  is a vector of taxa abundance, in this case aggregated to *genus and transformed to a log proportion (362 distinct genera)*

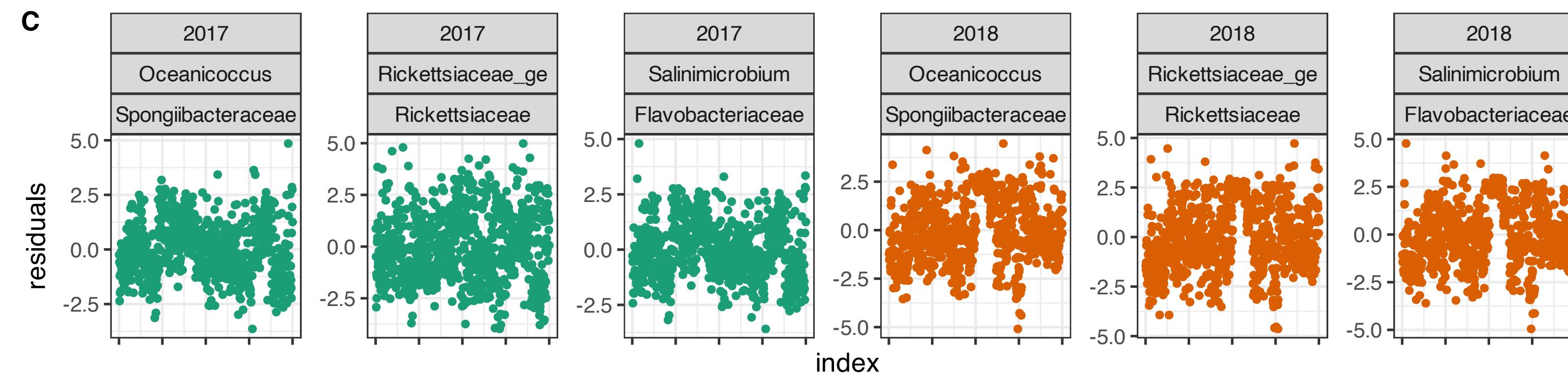
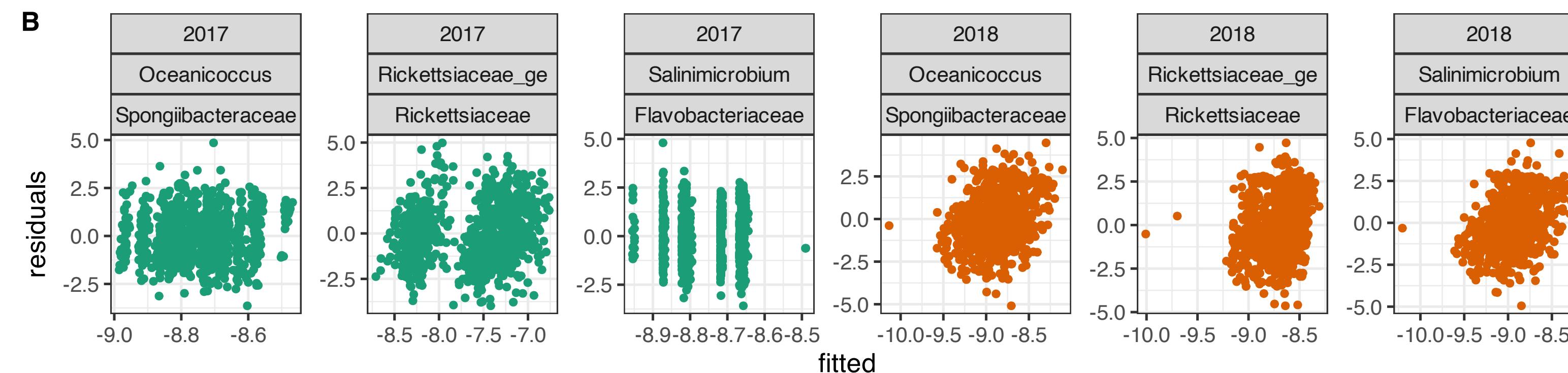
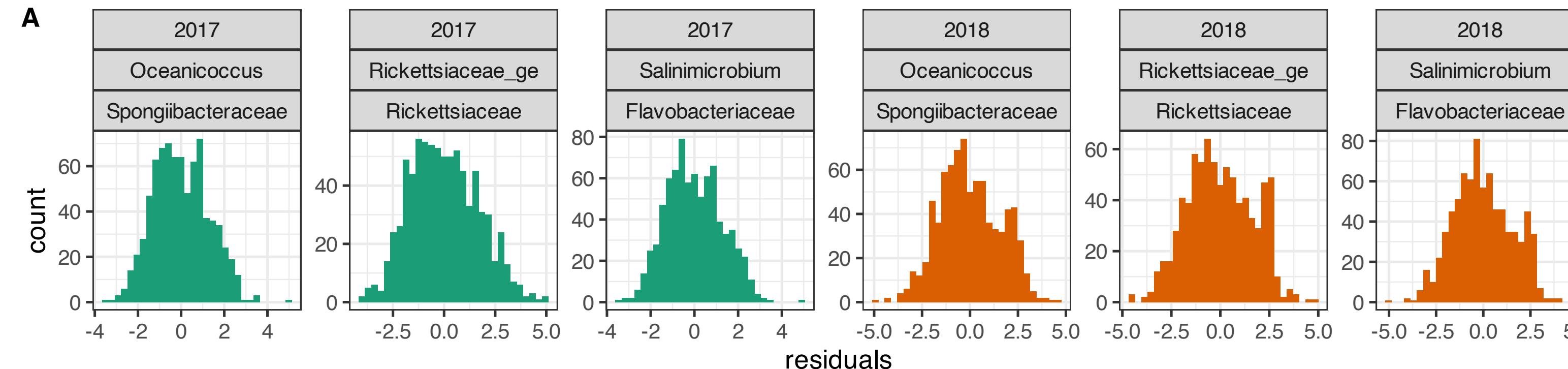
$b$  is the vector of fixed effects and  $X$  the corresponding design matrix

$Z_n u_n$  captures the non-genetic random effects (Initially just the day of sampling)

$Z_g u_g$  captures the genetic effects, it's assumed  $u_g \sim N(0, A)$  where  $A$  is the numerator relationship matrix calculated from the pedigree.

# The bad news

There is a pattern in the residuals when you order by sample day and time



# The Model

## Estimated using ASReml-r

BLUP Model treating microbiome phenotype

$$\text{Model: } y = Xb + Z_n u_n + Z_g u_g + e$$

$y$  is a vector of taxa abundance, in this case aggregated to *genus and transformed to a log proportion (362 distinct genera)*

$b$  is the vector of fixed effects and  $X$  the corresponding design matrix

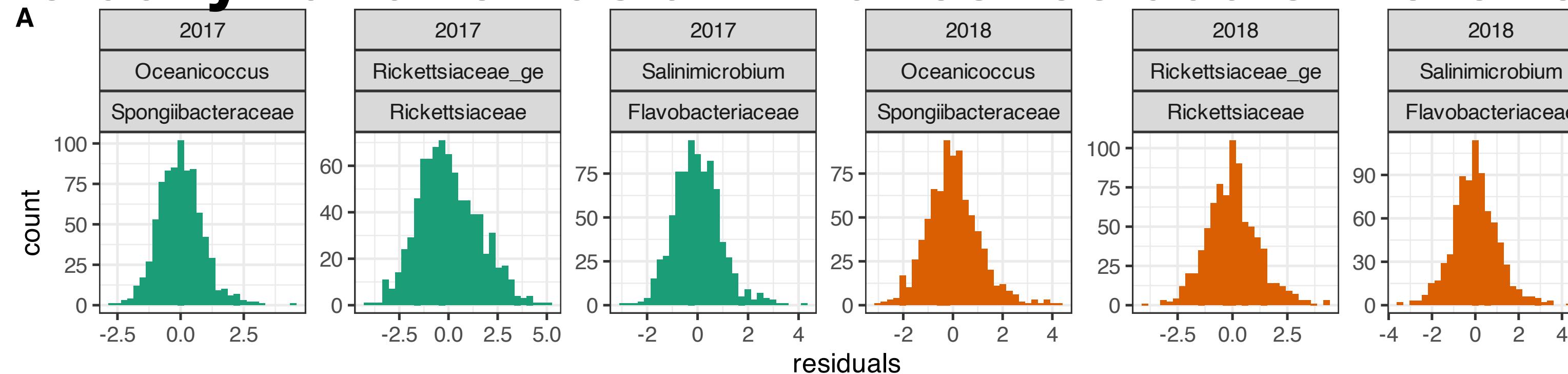
$Z_n u_n$  captures the non-genetic random effects (Updated to **day:ar1(time)**)

$Z_g u_g$  captures the genetic effects, it's assumed  $u_g \sim N(0, A)$  where  $A$  is the numerator relationship matrix calculated from the pedigree.

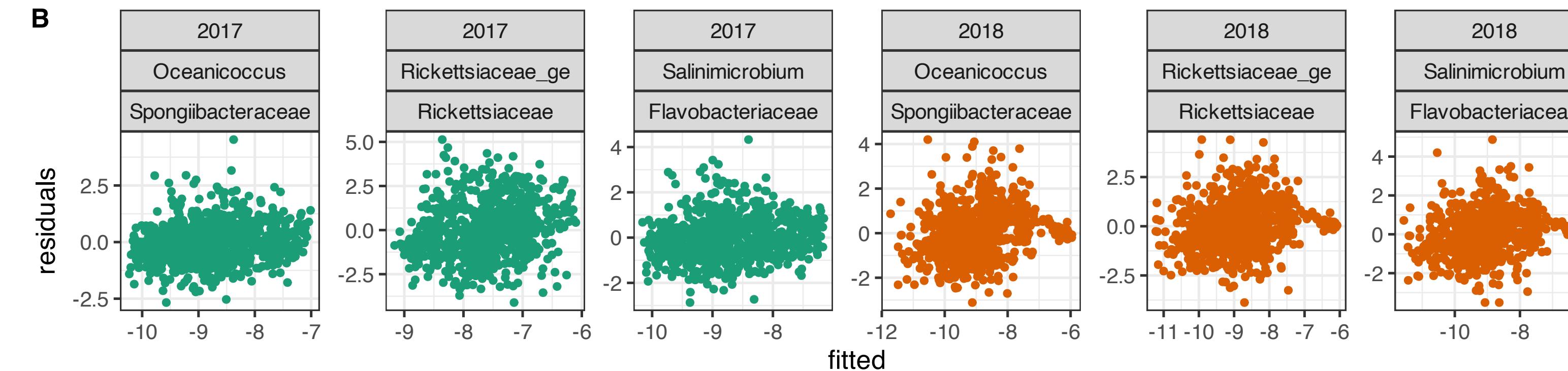
# Better?

## Modelling the daily variation as ar1 makes residuals more respectable

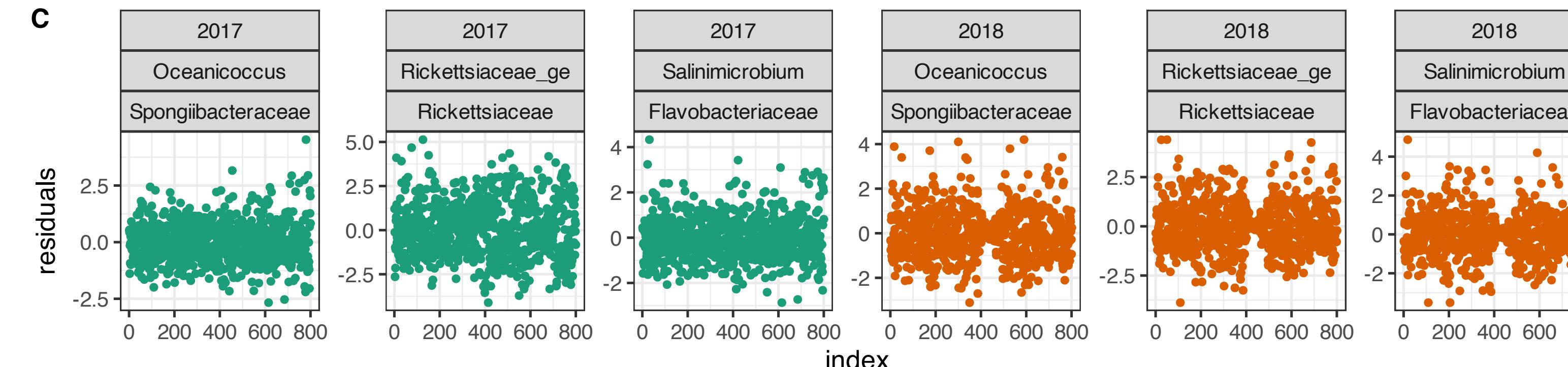
A



B



C

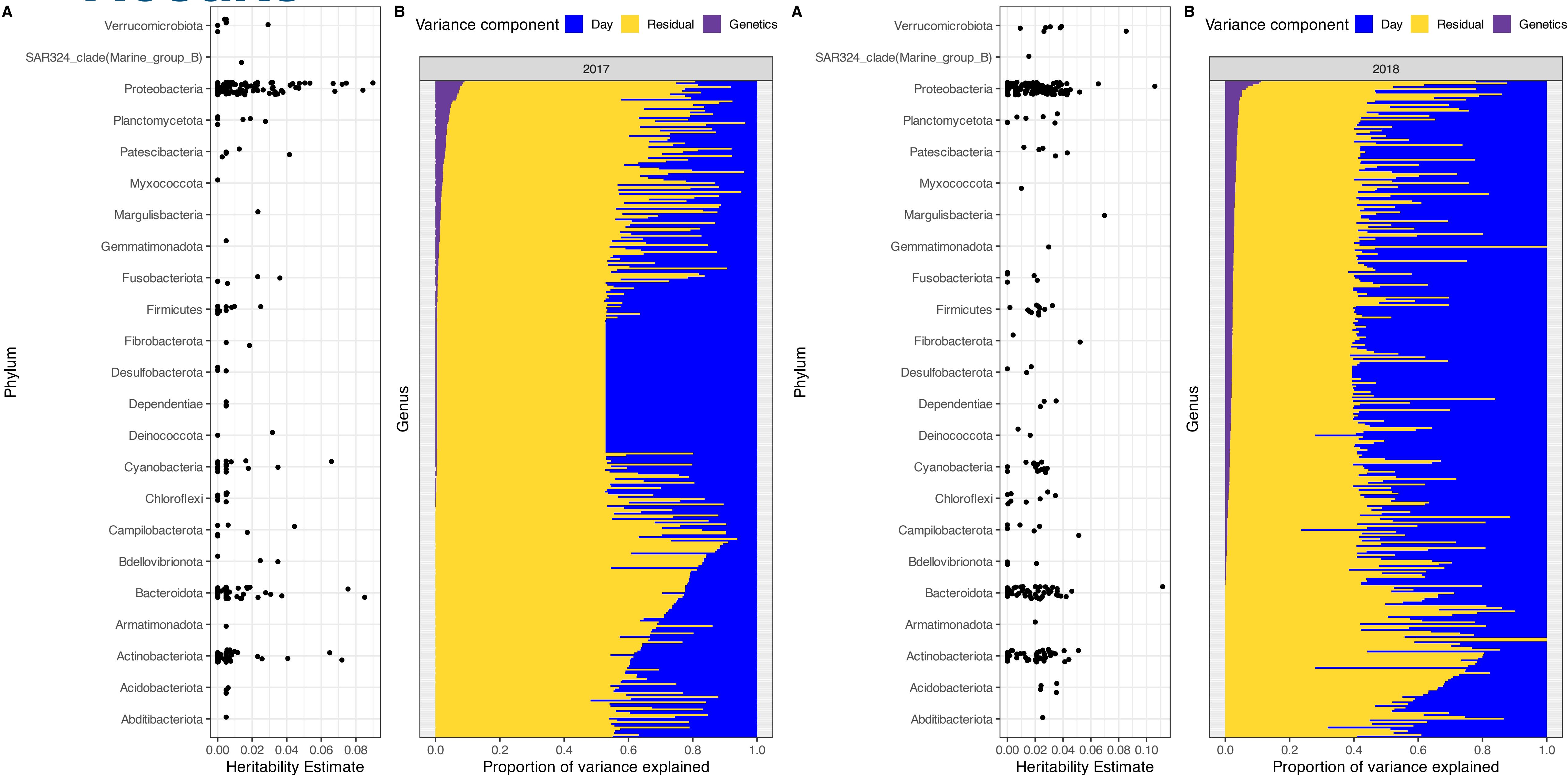


# Results

# Is the Salmon Gill Microbiome Heritable?

maybe

# Results



# Are these taxa interesting?

The most heritable taxa fall into three broad groups:

- **Disease / stress-associated** (e.g. *Lewinella*, *Cyclobacteriaceae*, *Thiothrix*, *Chryseobacterium*, *Bacteroidia*, *Microtrichales*, *Arcobacteraceae*, *Acrophormium*), which elsewhere are linked to gill disease, lesions, or stressed/diseased marine species.
- **Nutrient-processing taxa** (e.g. *Rhodobacteraceae*: *Octadecabacter*, *Sulfitobacter*, *Sedimentitalea*; *Lentisphaera*, *Margulisbacteria*, *Rhizobiaceae*, *SAR11 Clade Ia*) involved in nitrogen or sulphur cycling and degradation of organic matter at the gill surface.
- **Environmental background taxa** likely reflecting water column communities, algae associations and biofilms.

This one is particularly interesting  
implicated in salmonid Rickettsial Septicaemia

# Conclusions

# Is the Salmon Gill Microbiome Heritable?

Maybe

But the environment matters

more

**“Hell is paved with good meanings,  
but heaven is full of good works”**

Design

~~Hell~~ is paved with good meanings,  
but ~~heaven~~ is full of good works”

analysis