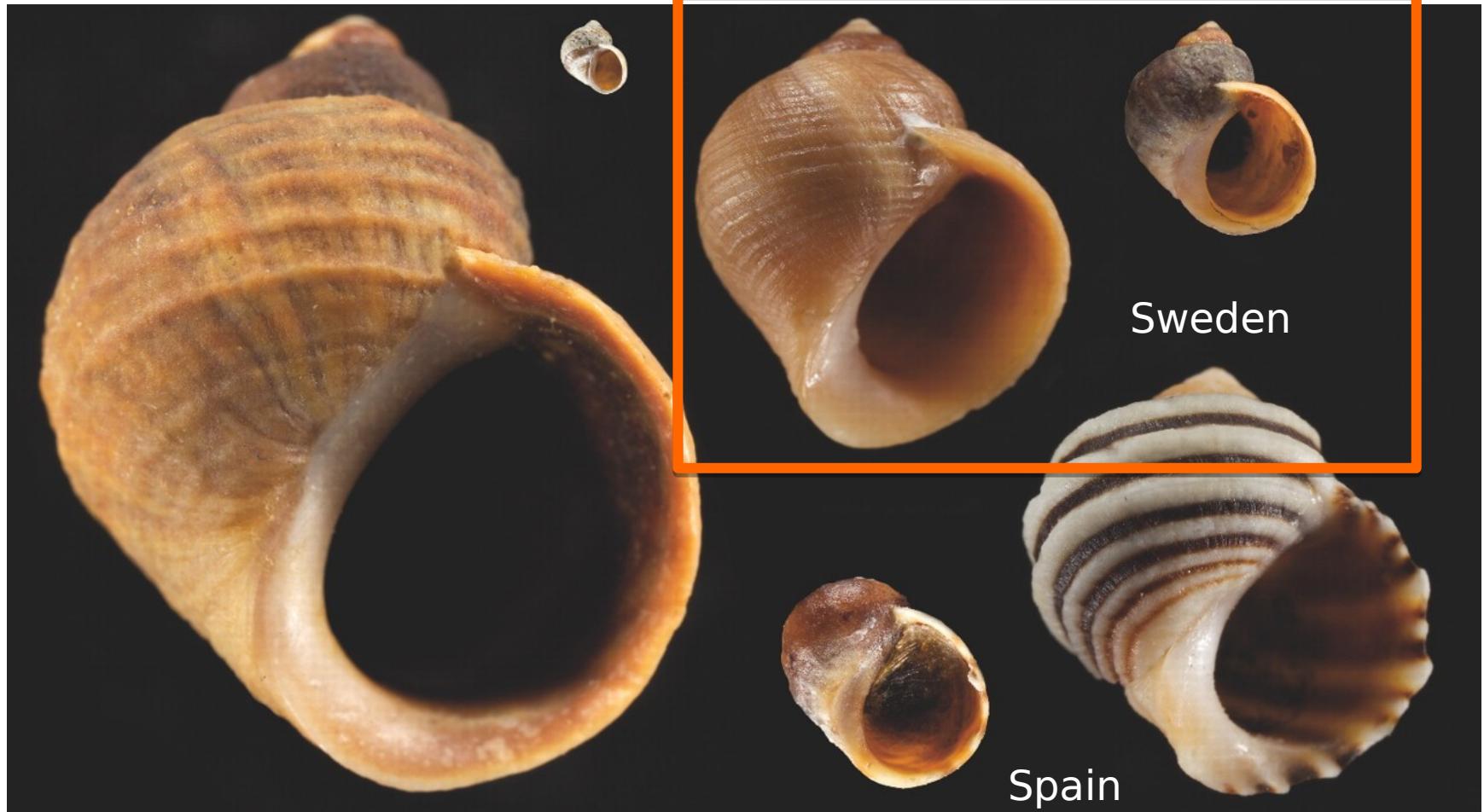


# RAD sequencing reveals parallel and non-parallel genomic divergence in Swedish *Littorina saxatilis* populations

Mark Ravinet<sup>1, 2</sup>, Marina Panova<sup>2</sup>, Roger Butlin<sup>2, 3</sup>, Kerstin Johannesson<sup>2</sup>, Carl André  
National Institute of Genetics, Japan  
University of Gothenburg, Sweden  
University of Sheffield, UK  
[mravinet@nig.ac.jp](mailto:mravinet@nig.ac.jp)

# *Littorina saxatilis* L.



Johannesson et al (2010) Phil Trans Roy S

# Swedish ecotypes

- Divergence between cliffs and boulder habitat – Western Sweden



## WAVE ECOTYPE (AKA EXPOSED)



- Cliff habitat
- Wave action
- Smaller
- Thinner shell
- Bold

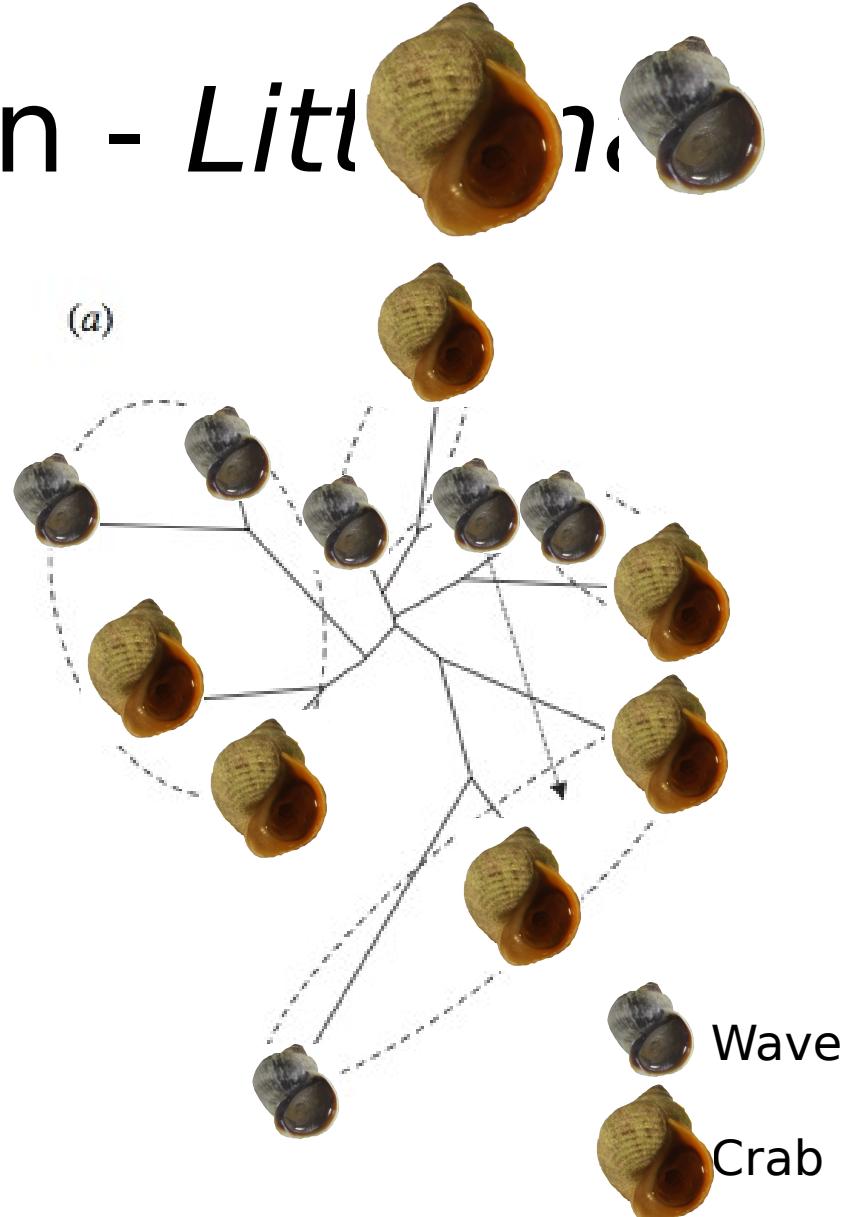
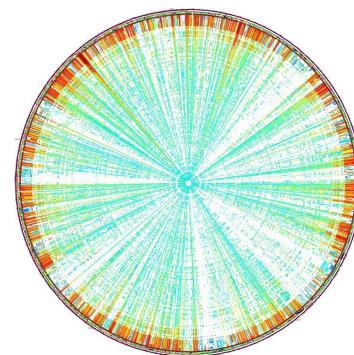
## CRAB ECOTYPE (AKA SHELTERED)



- Boulder habitat
- Crab predation
- Larger
- Thicker shell
- Less bold

# Parallel evolution - Littorina

- Repeated, independent evolution of ecotypes in the face of parallel selection
- Are the same genes involved in parallel evolution on this small scale?
- How parallel is this parallelism?



Panova et al (2006); Makinen et al (2008); Johannesson et al (2010)

# Rationale

- Top-down approach using high throughput sequencing to identify:
  - Loci experiencing positive parallel selection between ecotypes
  - Loci strongly associated with ecotypes across islands
  - Loci strongly associated with phenotypic variation within ecotypes



Three islands on Swedish coast – two ecotypes present on each

Kosterhavets  
Nationalpark  
Kosterhavets  
nationalpark

Sydkoster



Jutholmen



Saltö

A

Ramsö

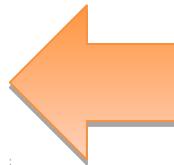


# Study design

- 144 individuals – Ramsö, Saltö, Jutholmen



Geometric  
morphometrics and  
phenotyping



24 crab    24 wave  
48 per population



Restriction-site  
associated DNA (RAD)  
sequencing

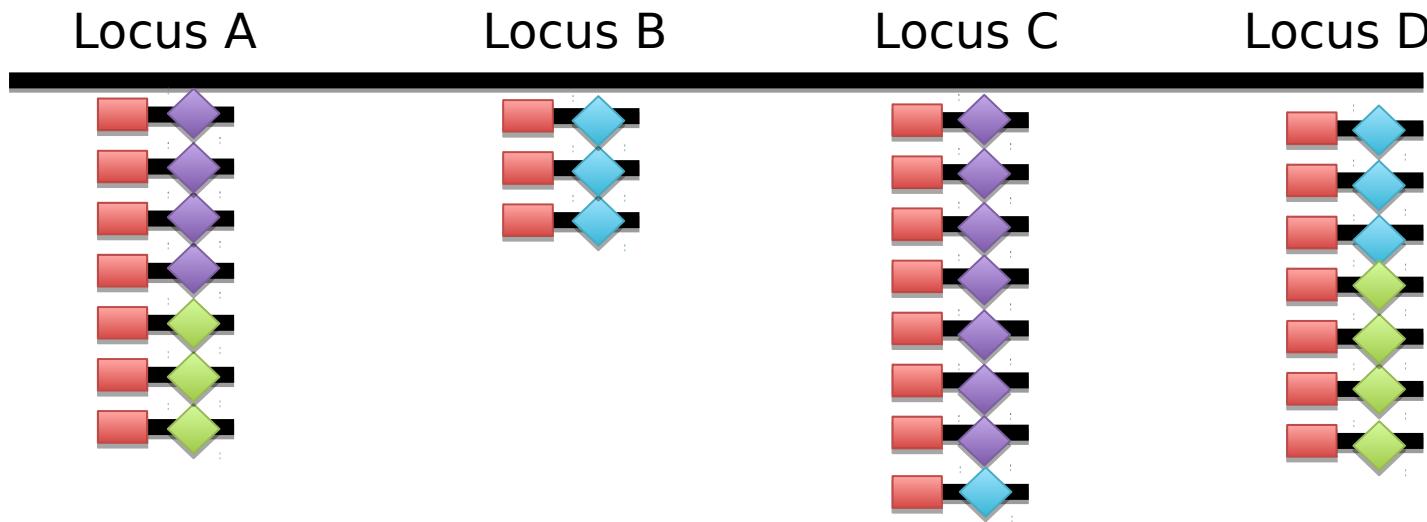
# RAD-sequencing in brief

- Single digest RADs – Baird et al (2008)
  - simple example with two individuals



# RAD loci: reference genome

- Align reads to a reference genome
- Identify RAD loci and call SNPs
- Simple!



# RAD *de novo* assembly

- *de novo* assembly is more complicated
- Must match identical sequences



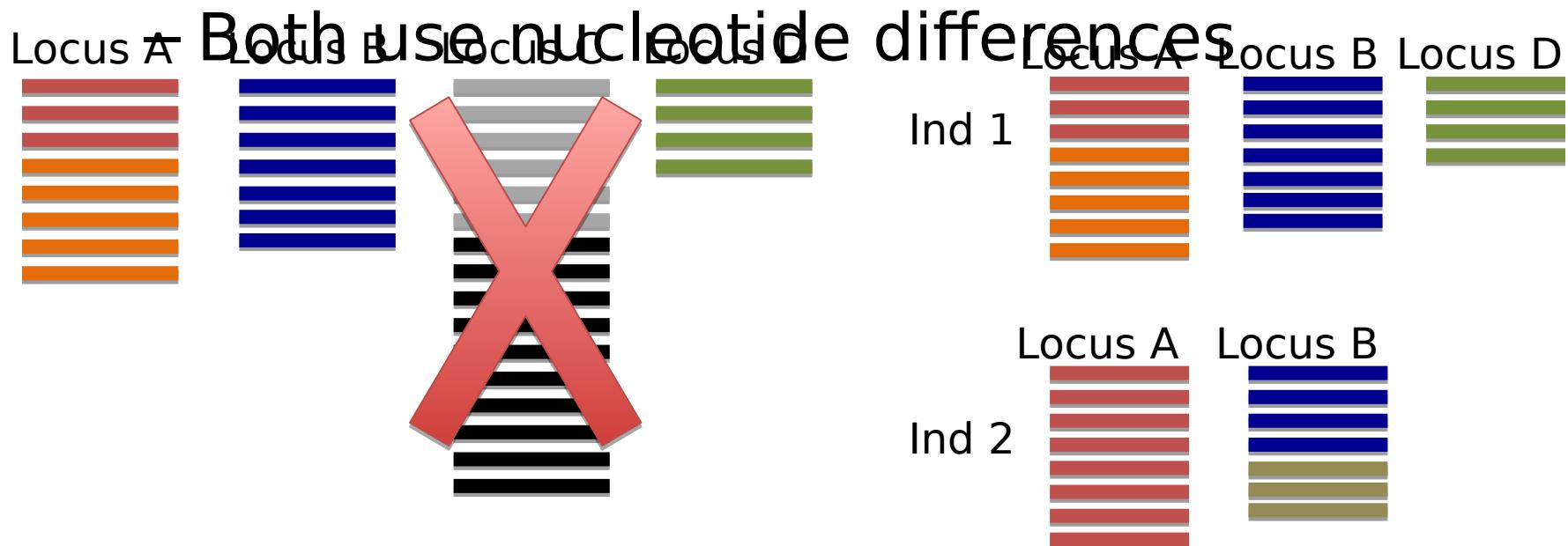
# RAD *de novo* assembly

- *de novo* assembly is more complicated
  - Must match identical sequences into stacks
  - Essentially alleles...



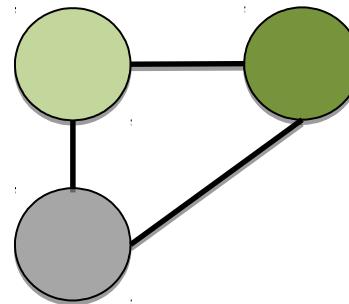
# RAD *de novo* assembly

- *de novo* assembly is more complicated
  - Alleles matched *within* individuals
  - Loci matched *between* individuals



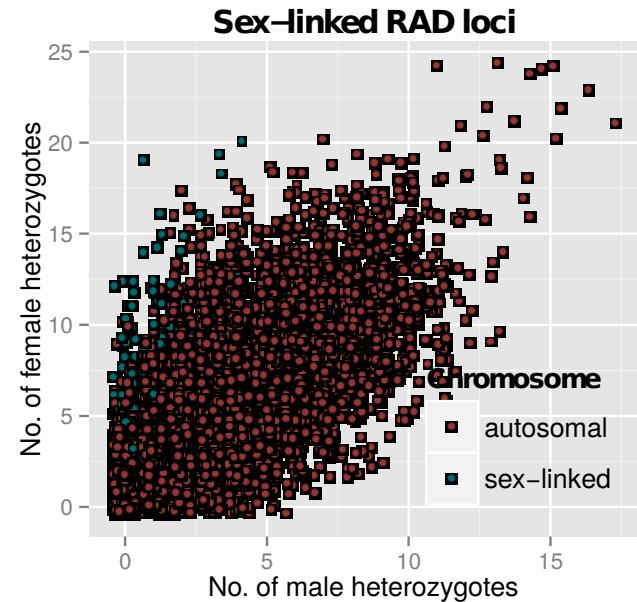
# Some RAD issues

- High nucleotide diversity can make *de novo* assembly difficult
- Always a balance:
  - too low – alleles called as separate loci
  - too high – loci are merged



# De novo assembly and filtering

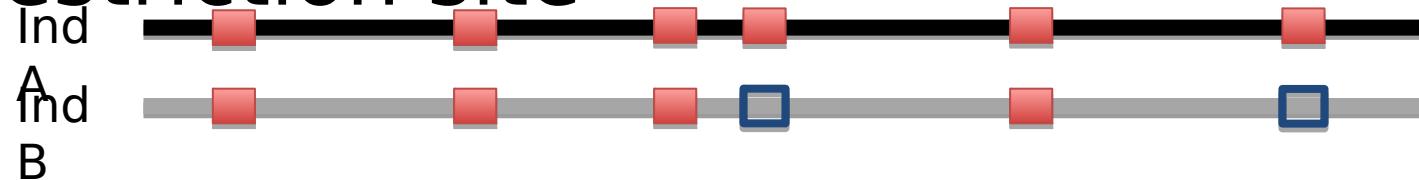
- 8 nucleotide differences within and between individuals
- 103 919 loci across all 144 individuals
  - Filtering – must be polymorphic, must occur in >50 % individuals on an island
- ~9400 loci per island – but further filtering required...



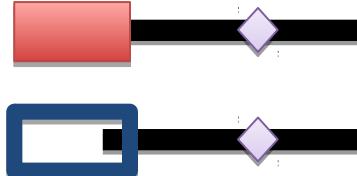
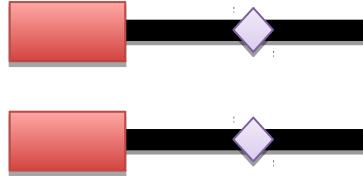
Likelihood-based test to remove loci with sex linkage

# Null alleles

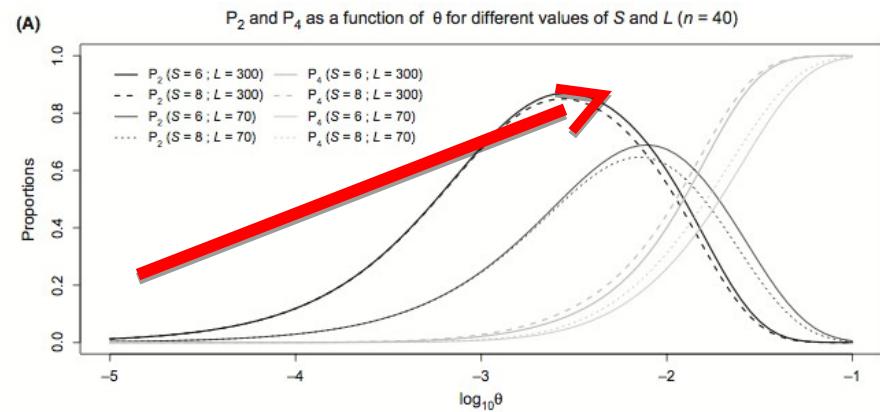
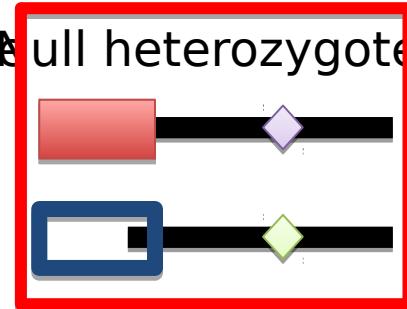
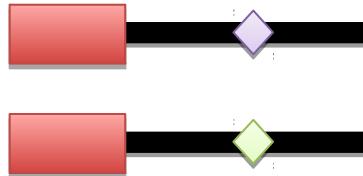
- Caused by mutations disrupting restriction site



Normal homozygote Null homozygote



Normal heterozygote Null heterozygote

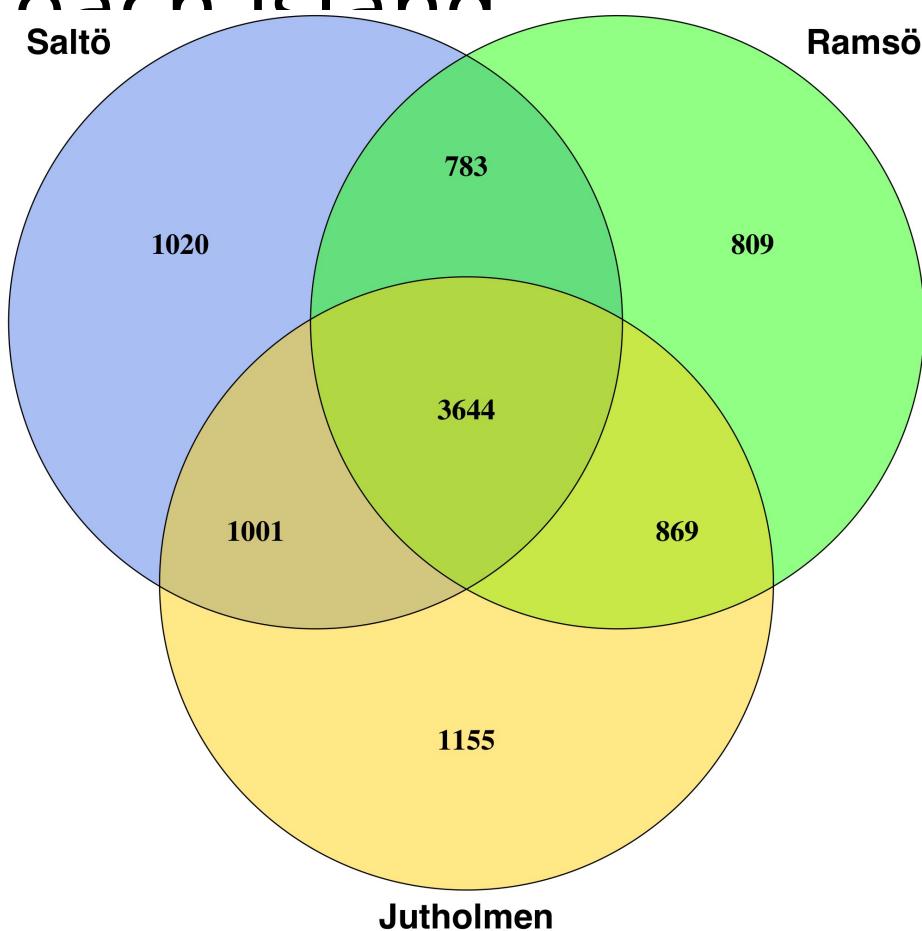


$$\Theta (4N_E\mu)$$

Called as homozygotes causing homozygosity excess Gautier et al (2013) Mol Ecol

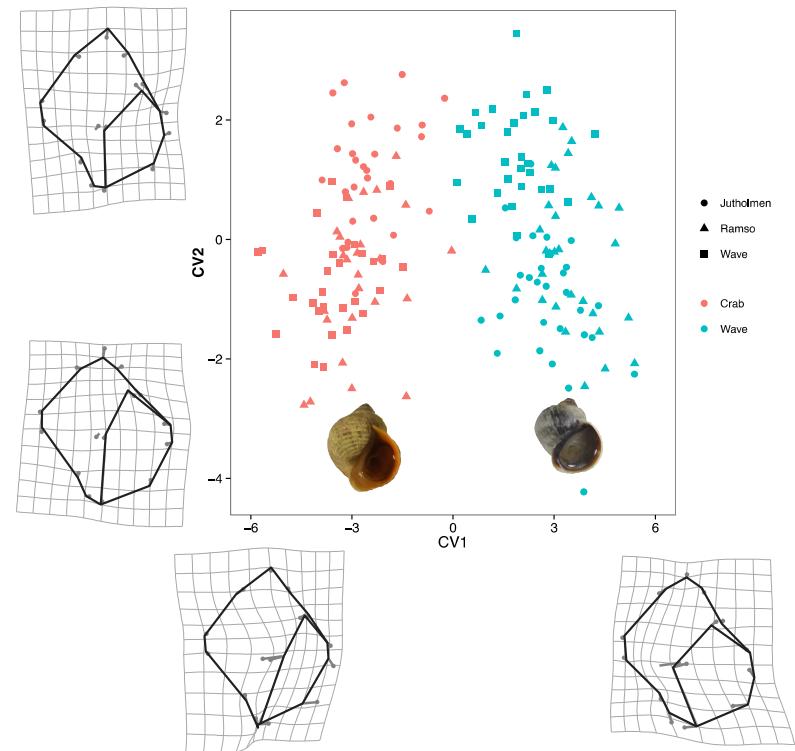
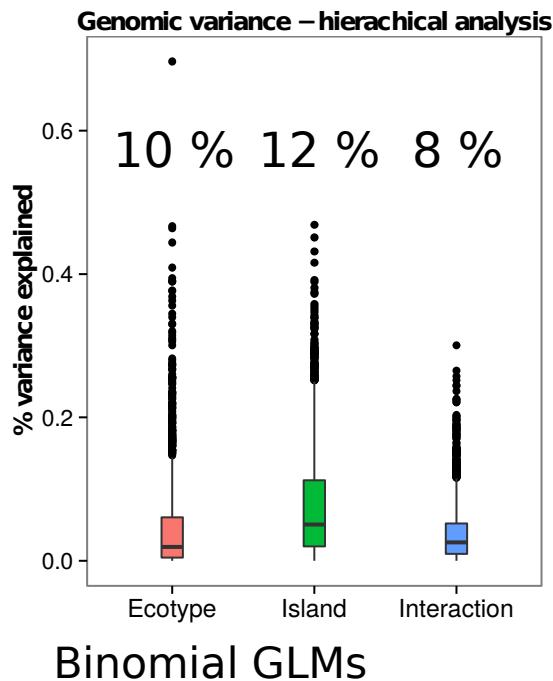
# Strict filtering

- Three different datasets, one for each island
  - 9281 loci
  - ~1/3 shared between all three islands
  - ~1/3 are unique to each island



# Quantifying parallelism

- Strong parallel evolution of shell morphology amongst ecotypes
- Genomic parallelism is low



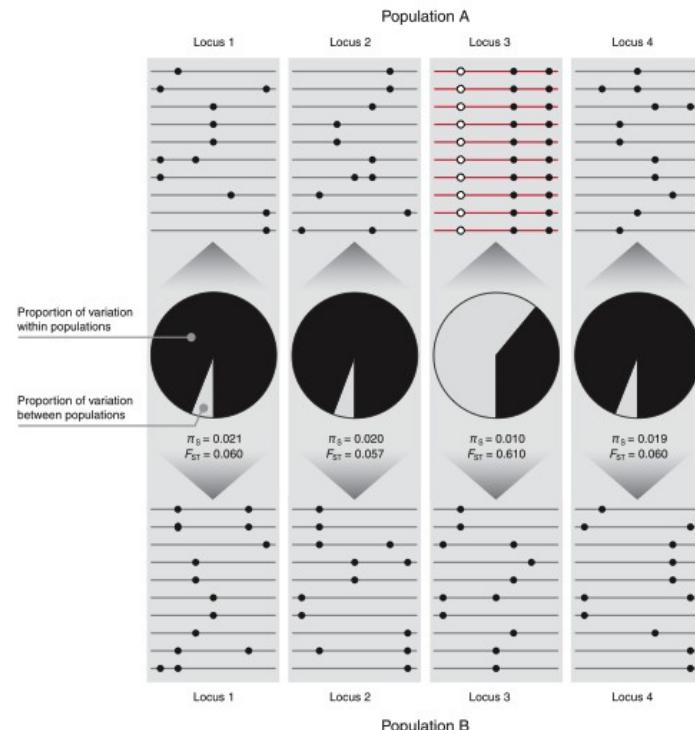
Factor	PVE
Ecotype	90%
Island	37%
Interaction	39%

MANCOVA and PVE estimation

# Detecting outliers using $F_{ST}$

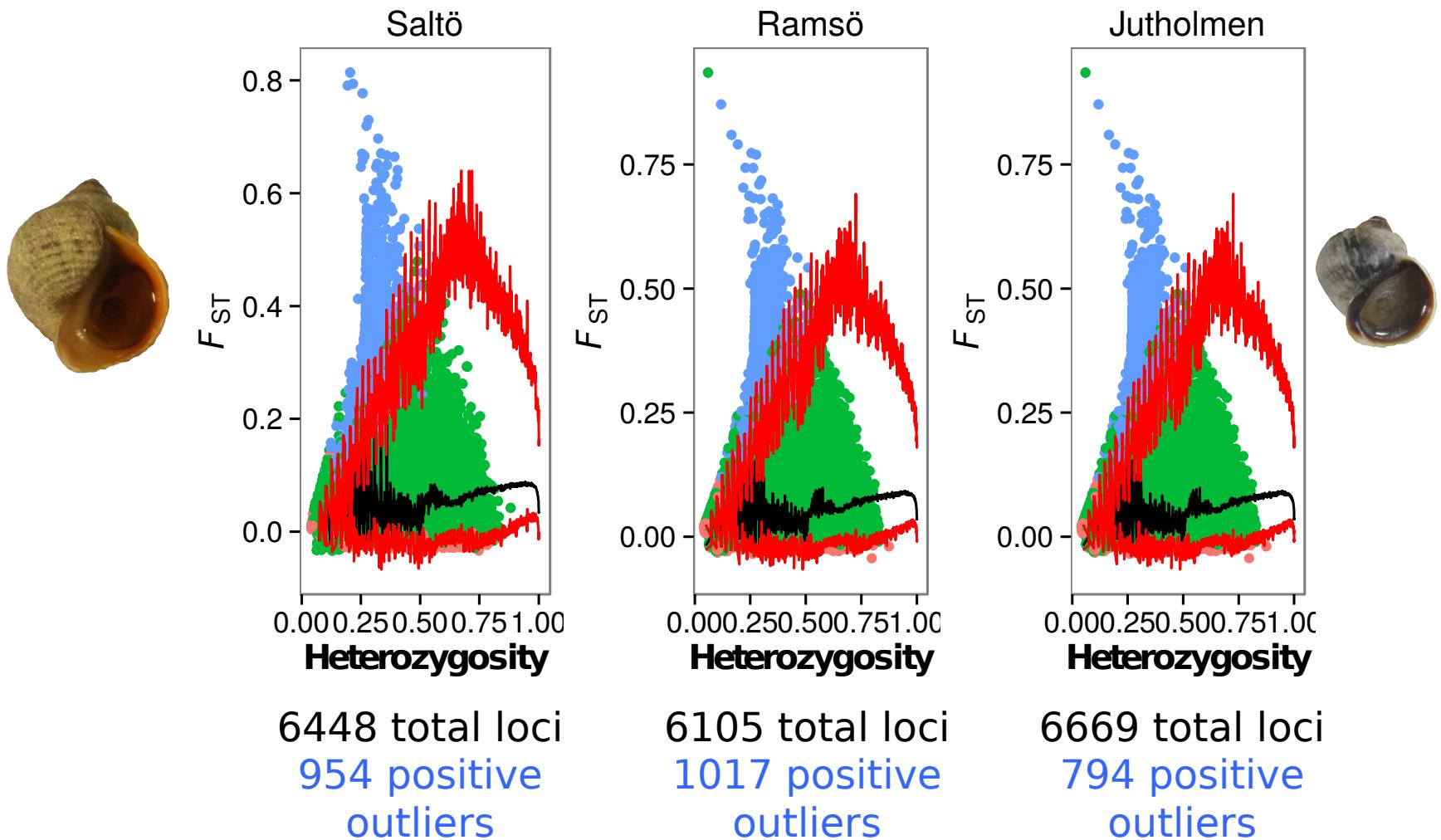
- Reduction of within-population diversity relative to increase in between-population diversity
- fdist - coalescent simulations to estimate expected neutral  $F_{ST}$  distribution
- Loci outside 99% CI are potentially under selection

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2}$$



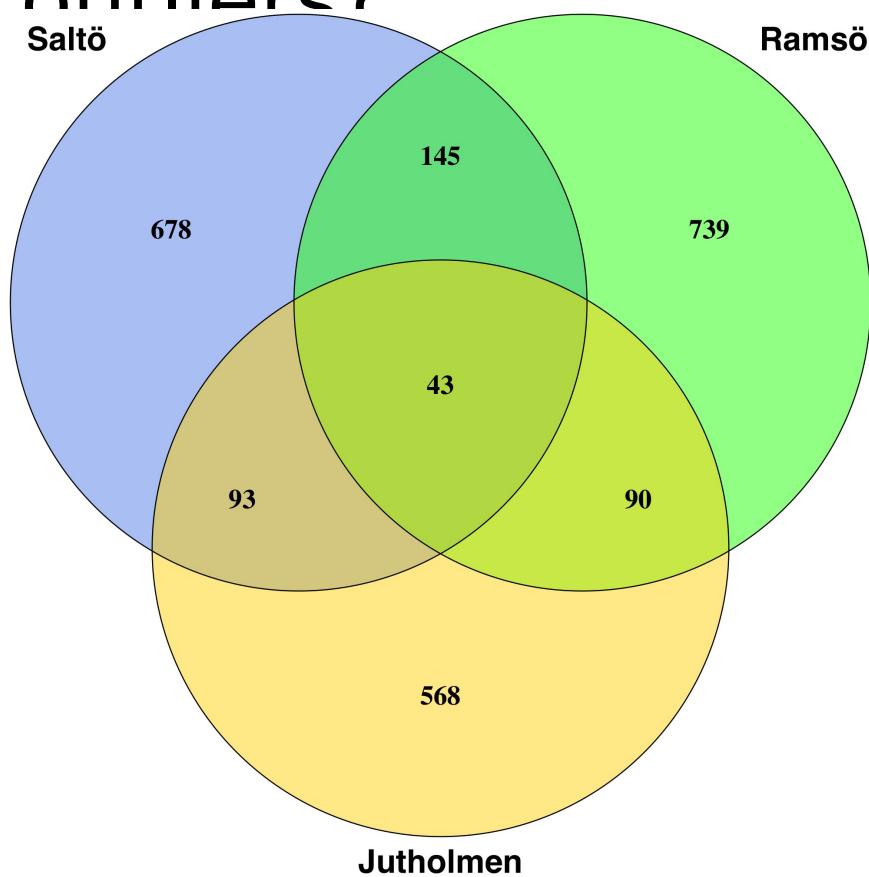
# Identifying selection

- Fdist outlier analysis – 5% FDR



# Shared outliers

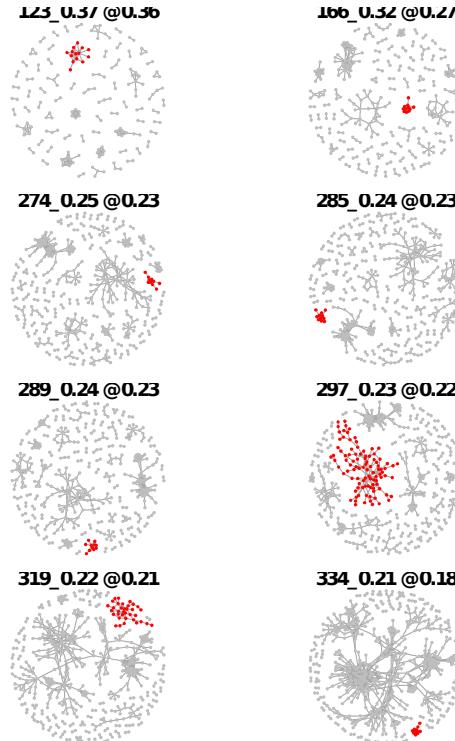
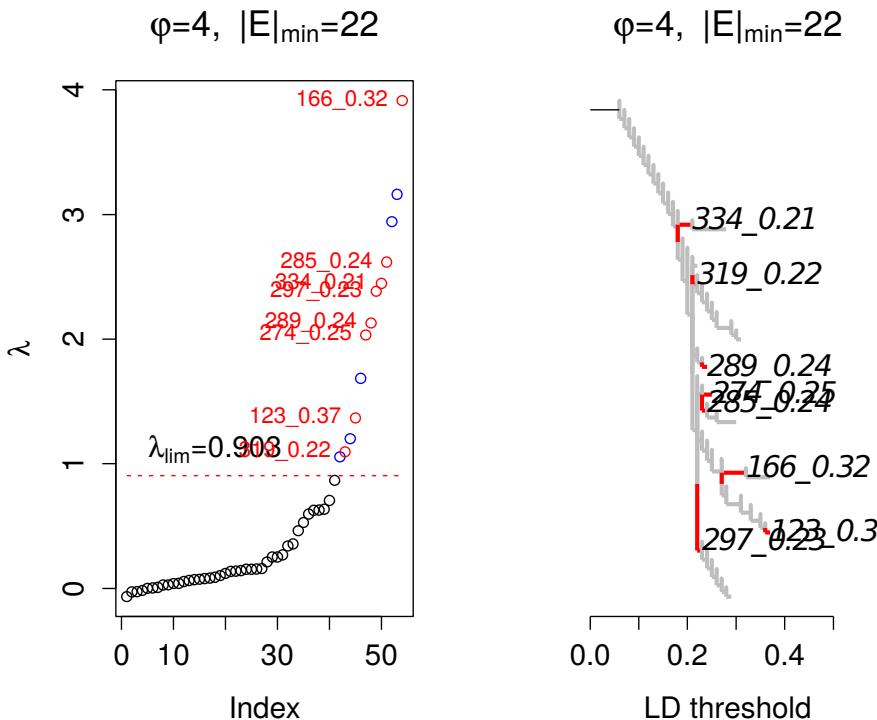
- What proportion of loci are shared outliers?



- 0.5 % shared between all three islands
- ~1% shared by two islands
- ~7% are unique outliers

# Other outliers

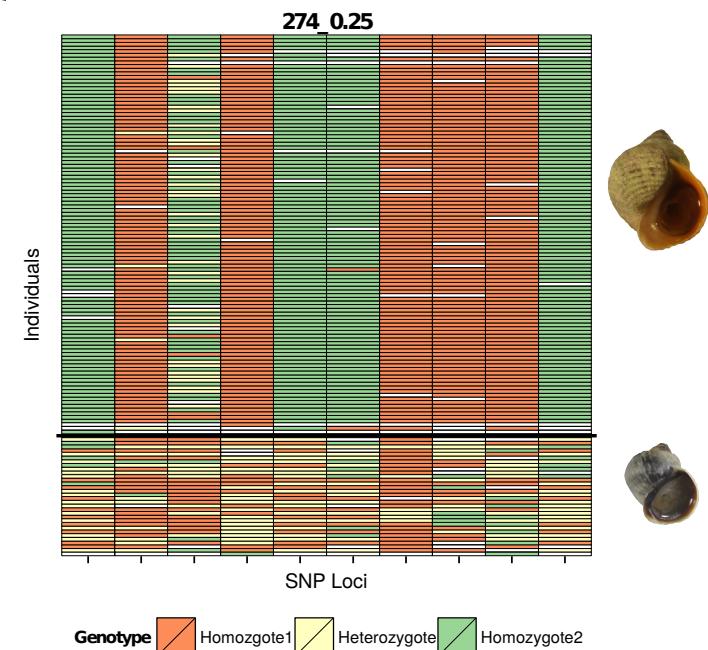
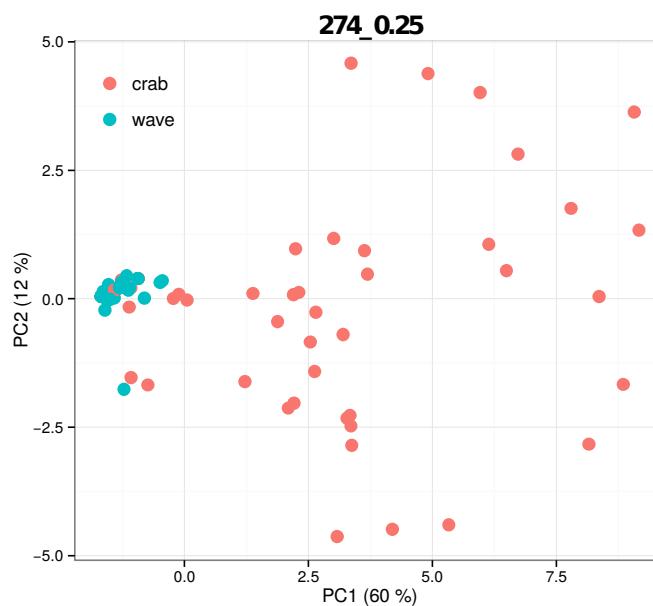
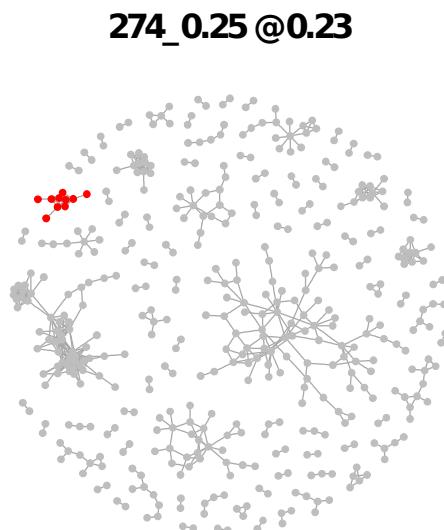
- Strong linkage disequilibrium between loci may indicate genomic inversions, local adaptation or population structure
- LDna – network analysis of linkage disequilibrium



Blocks of loci associated with each other

# Ecotype clusters

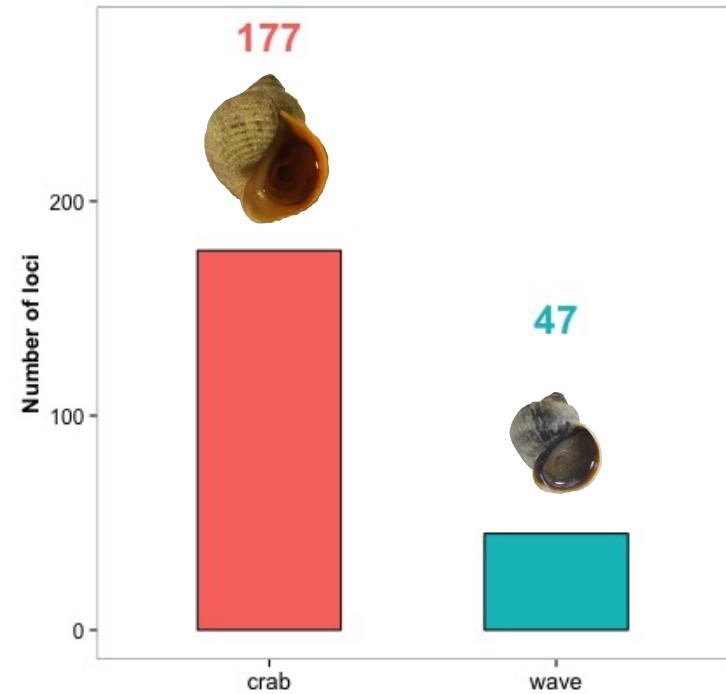
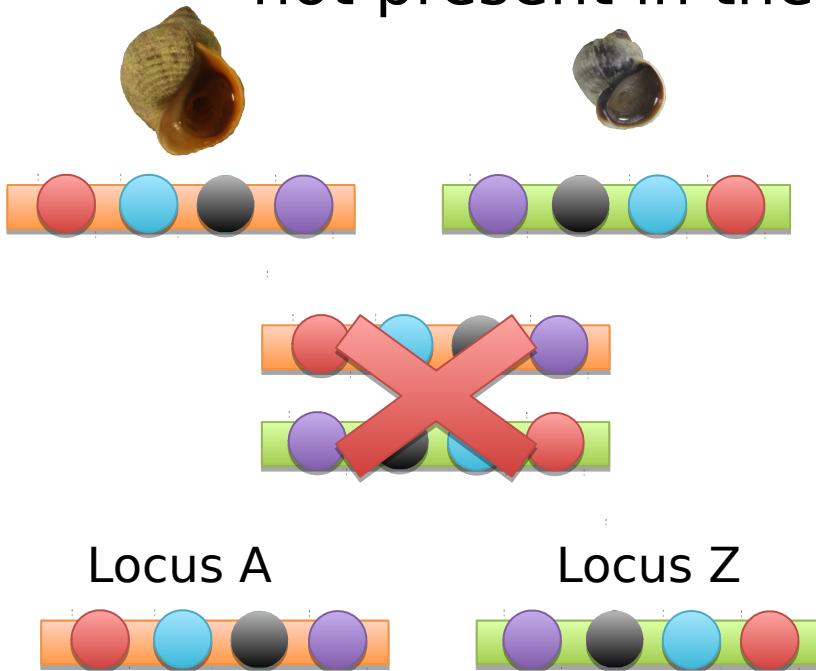
- 274\_0.25 with 10 loci contributes strongly to ecotype difference



- Includes two outliers shared between all three islands and five which are outliers on at least one island

# Other outliers

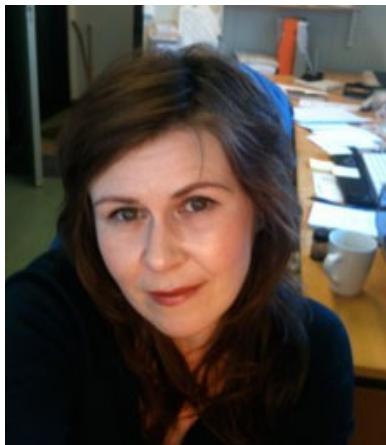
- Ecotype specific loci
  - Highly divergent (i.e. ancient) alleles may not merge in a *de novo* assembly
  - Screened dataset for loci fixed in one ecotype but not present in the other



# Summary

- De novo RAD assembly can be problematic for *Littorina saxatilis* and null alleles make things worse!
- Our top-down approach has identified **loci** under parallel selection, **loci** associated with ecotype and loci associated with phenotype
- The next step is to try use these loci to identify the genes that may be involved in ecotype formation

# Acknowledgements



Marina  
Panova

Valuable assistance:

- Anja Westram
- Kevin Keenan
- Petri Kemppainen

Carl André

Kerstin  
Johanness  
on

Roger Butlin



# Wave-type colour phenotypes

- Association tests using binomial GLMs.
- Loci above 99<sup>th</sup> quantile of PVE.



17 loci



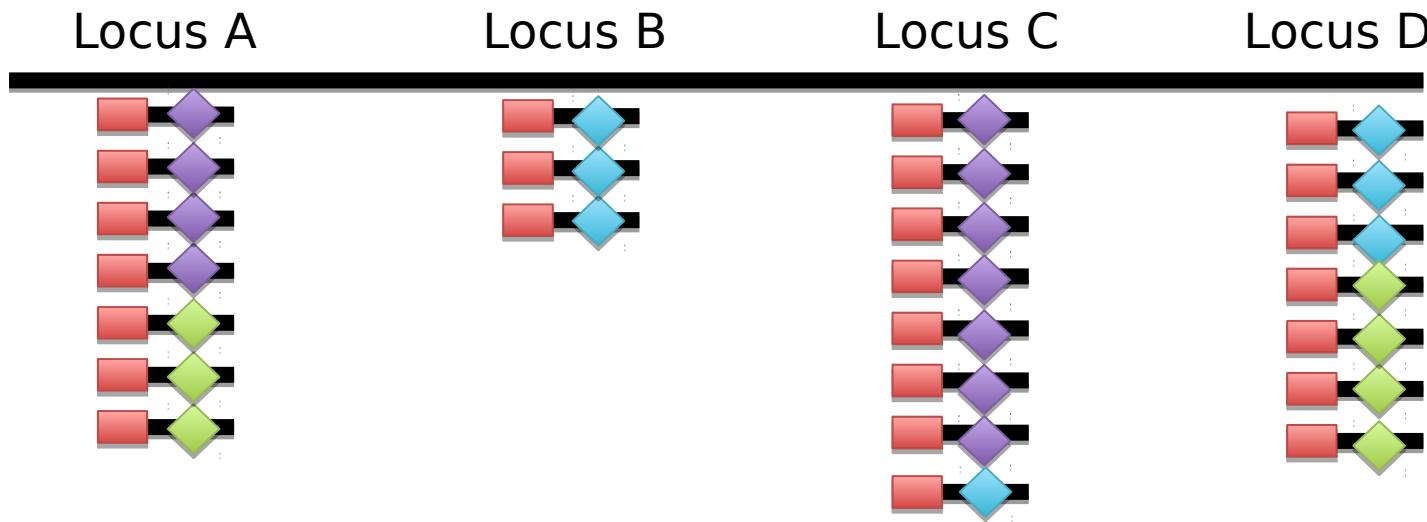
12 loci



17 loci

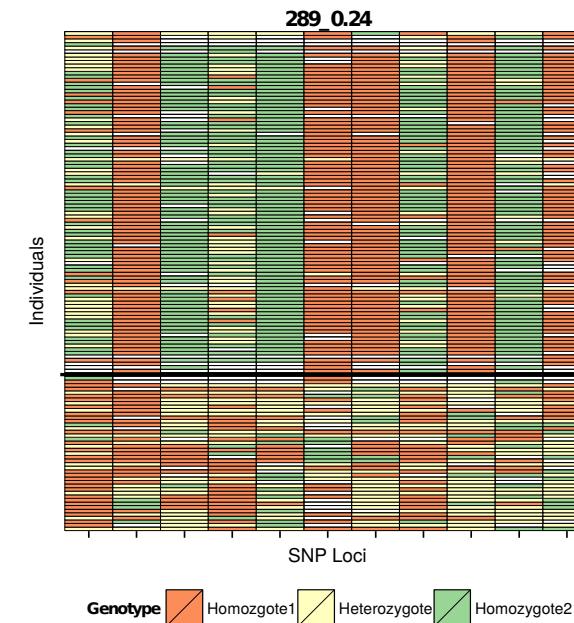
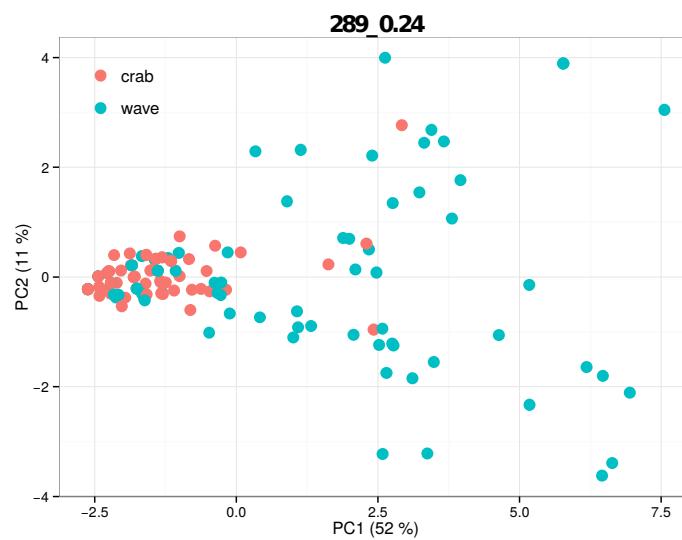
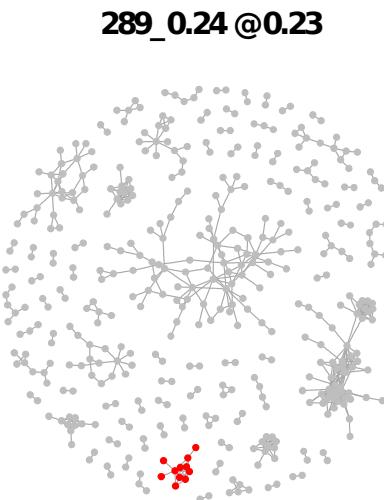
# RAD loci: reference genome

- Align reads to a reference genome
- Identify RAD loci and call SNPs
- Simple!



# Ecotype clusters

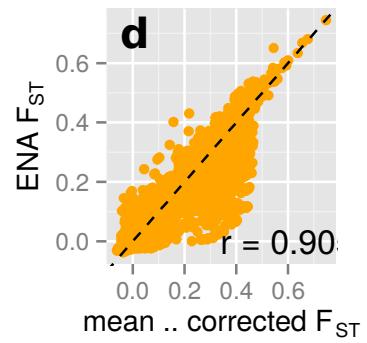
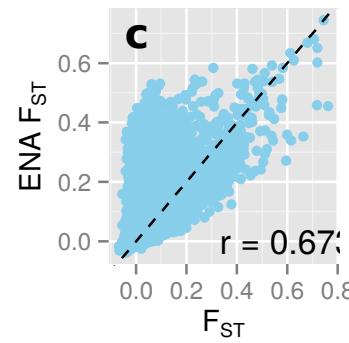
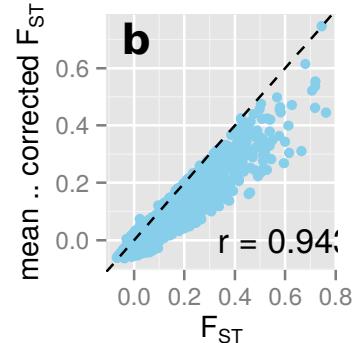
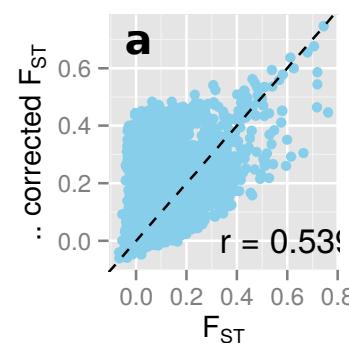
- 289\_0.24 with 11 loci also contributes strongly to ecotype difference



- No shared outliers but 6 of the 11 loci are outliers on at least one island

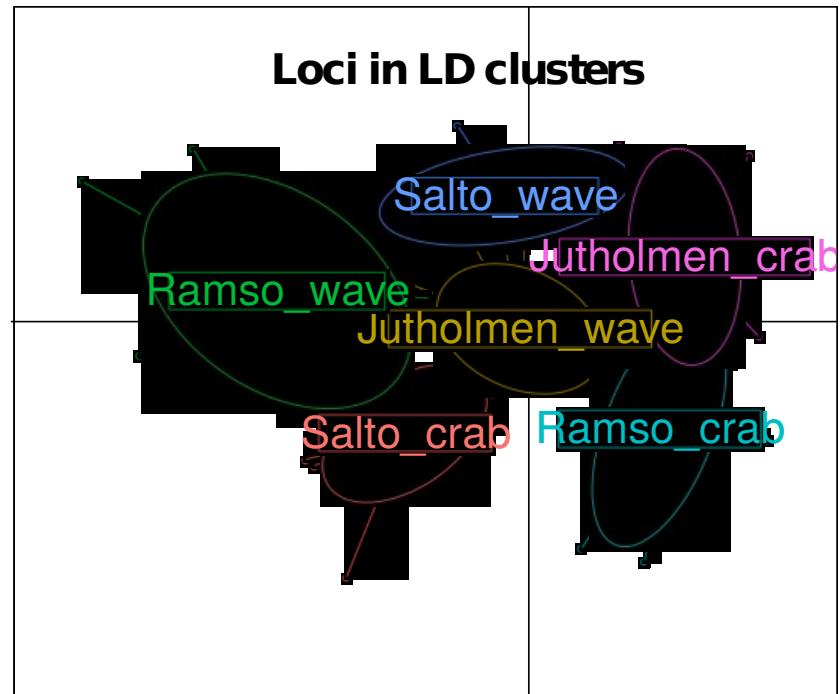
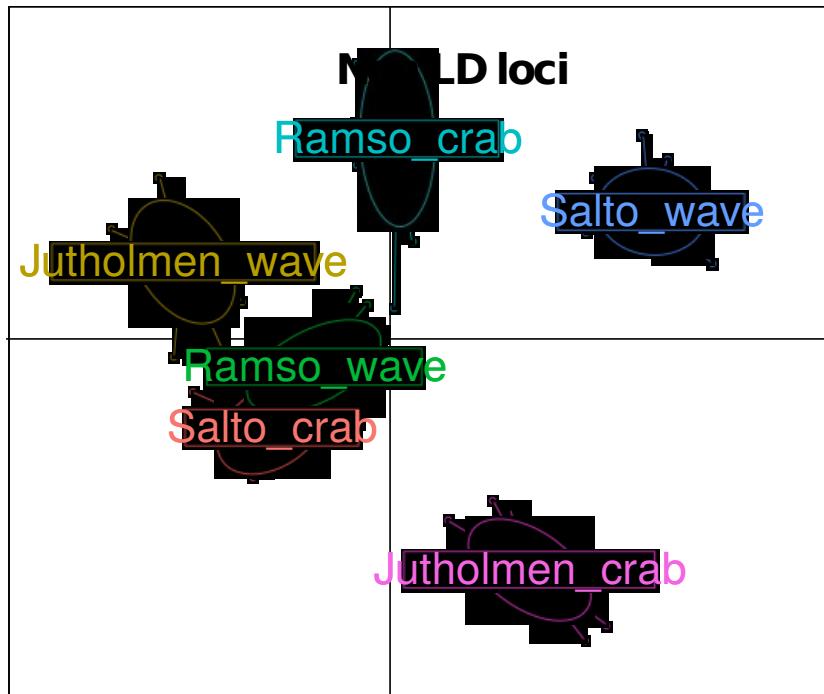
# Dealing with null alleles

- No established method for RADs
- Implemented 2 microsatellite methods
  - HWE to estimate frequency of null alleles
  - Correct observed allele frequencies and downstream estimates - i.e.  $F_{ST}$



# Population structuring

- Population structure not clear for non LD loci
- Strong LD loci show mixture of ecotype and population structure



Discriminant analysis PCA with  $k$  means clustering