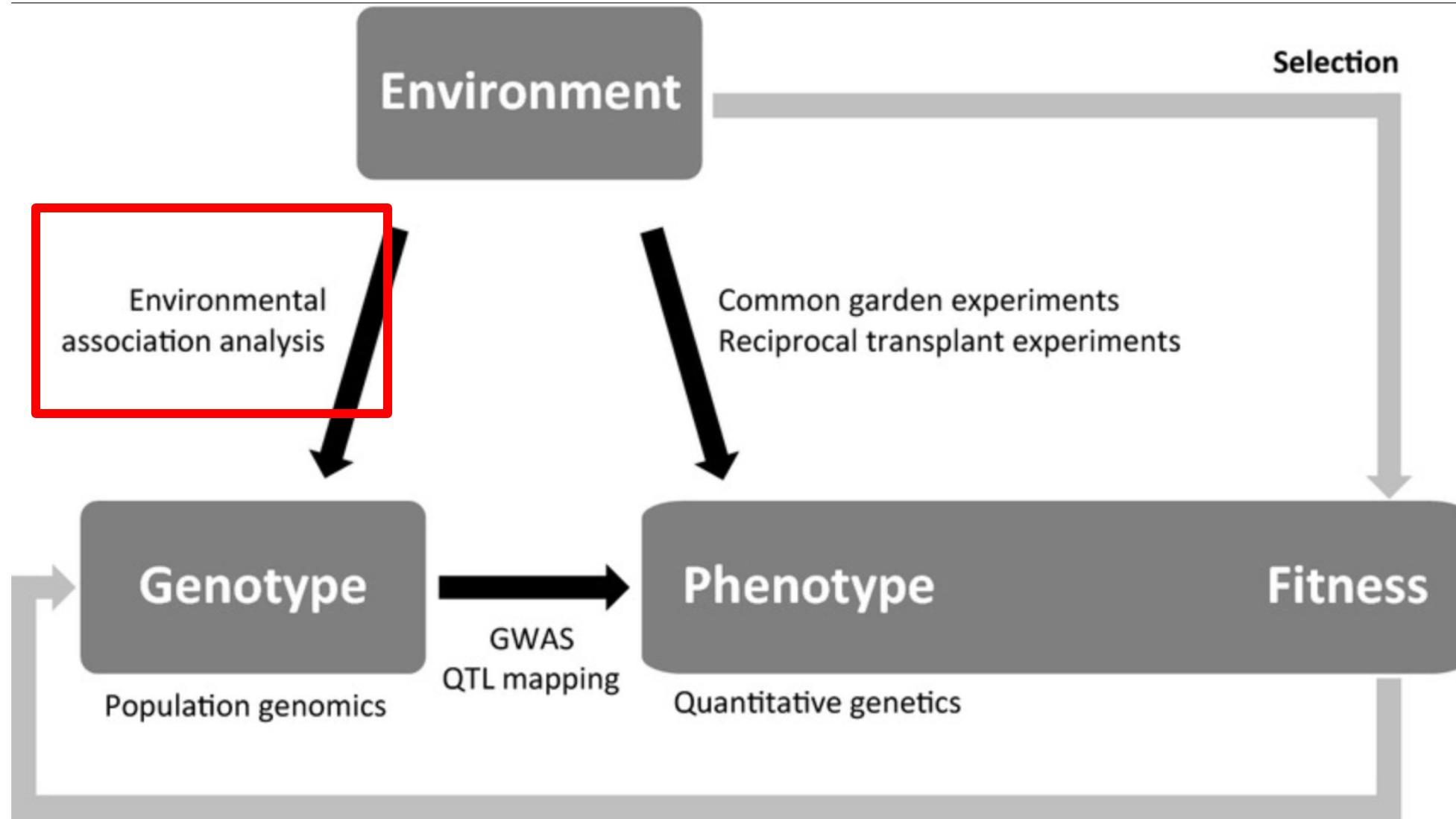


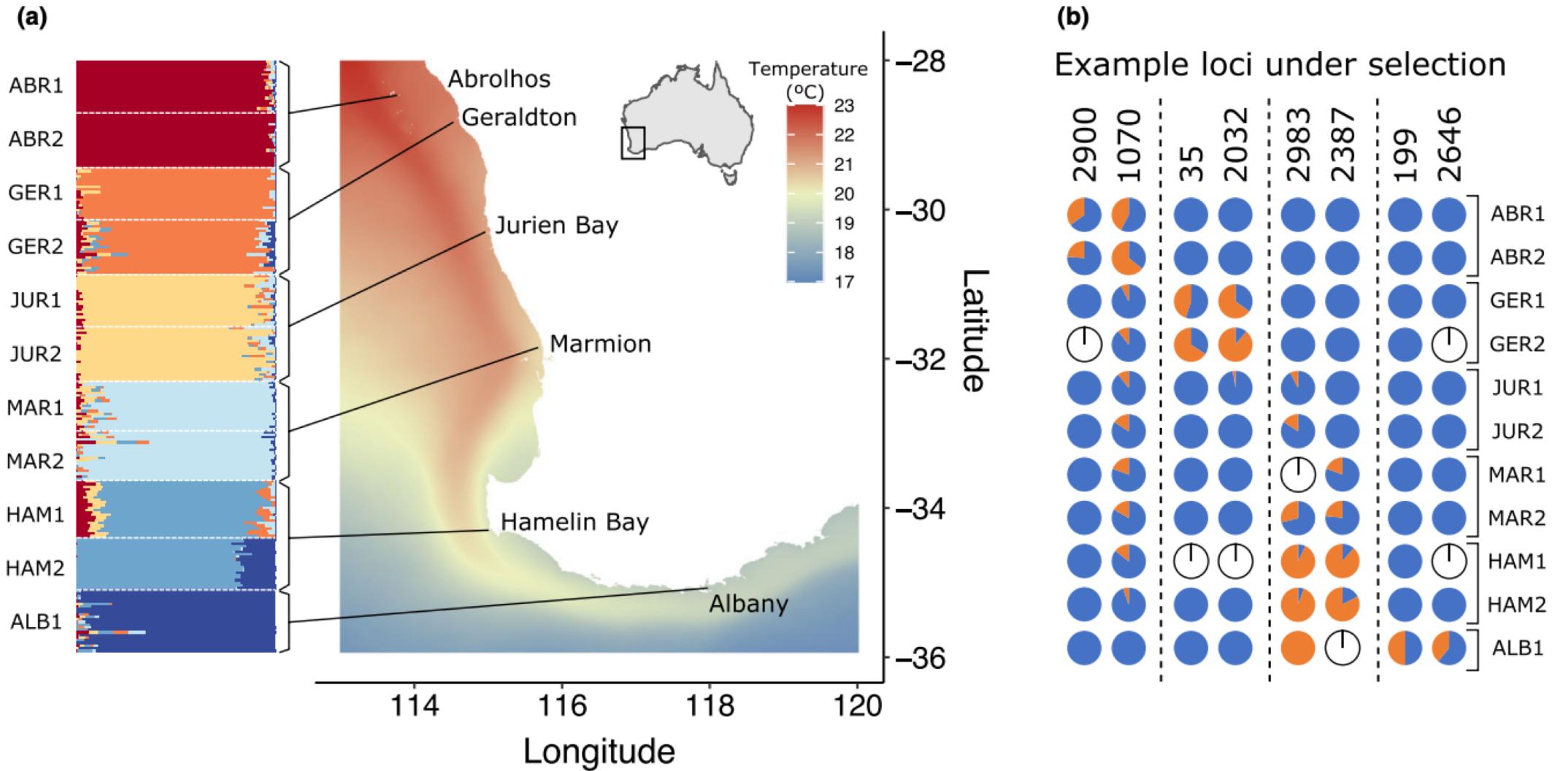
Genome-Environment Associations (GEAs)

Marine Genomics S22
Week 10 – June 2nd 2022





What SNPs are associated with the environment?



Method		SNPs	Mean temperature	Temperature range	Diffuse attenuation
F_{ST}	BAYESCAN	179 (164)	—	—	—
GEA	Ifmm	220 (197)	56	59	105
	RDA	155 (117)	39	103	13

Note: The number of candidate SNPs uniquely identified by each method is shown in parentheses.

For Ifmm and RDA the number of loci is broken down into the three environmental variables tested.

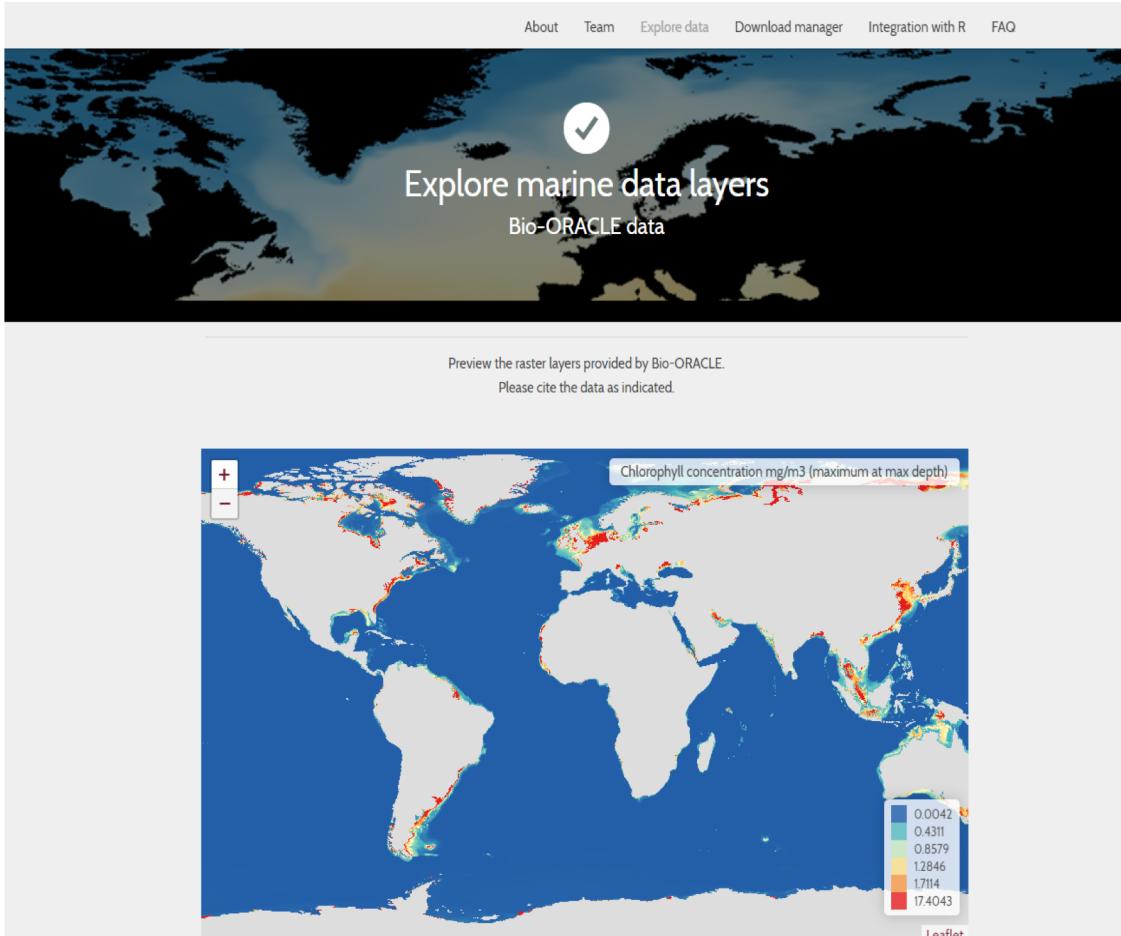
TABLE 4 *E. radiata* gene models containing candidate SNPs that result in missense mutations

gene	Variant	AA change	RNAseq support	BLASTX hit
evm.TU.scf7180001285637.2	658G>A	Ala220Thr	No	similar to ankyrin 2,3/unc44 (CBJ31919.1)
evm.TU.scf7180001305554.1	397C>T	Arg133Cys	No	NA
evm.TU.scf7180001329928.1	199G>T	Asp67Tyr	No	NA
evm.TU.scf7180001289356.2	1484C>A	Thr495Asn	No	Endo-1,3-beta-glucanase, family GH81 (CBJ48932.1)
evm.TU.scf7180001289906.1	1670C>G	Ala557Gly	No	hypothetical protein Esi_0012_0173 (CBN74090.1)
evm.TU.scf7180001287055.3	881T>G	Val294Gly	Yes	Similarity to ribonuclease D (CBN78705.1)
evm.TU.scf7180001286527.1	925A>T	Ser309Cys	Yes	transmembrane protein, putative (CBJ27369.1)
evm.TU.scf7180001298253.1	195T>A	Asp65Glu	No	NA

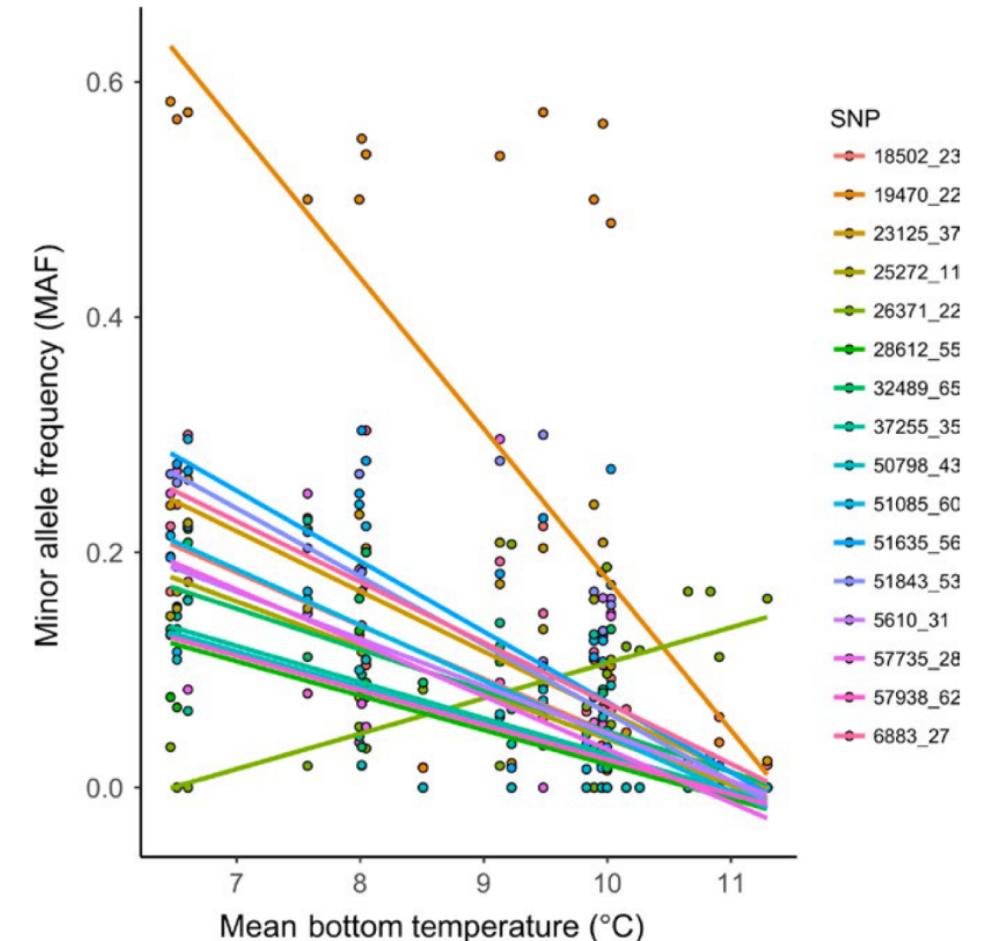
Note: The DNA variant (Variant) and resulting amino acid change (AA change) is shown with the functional annotations based on the best BLASTX hit of the gene model to the NCBI database, GenBank IDs are shown in parentheses. RNA-sequencing support for the gene model is based on SRA data (SRR3709347 and SRR3705662) mapped to the *E. radiata* draft genome. The numbers in 'Variant' and 'AA change' refer to the position of the nucleotide and amino acid respectively.



Environmental data ~ Genomic data



<https://www.bio-oracle.org/explore-data.php>



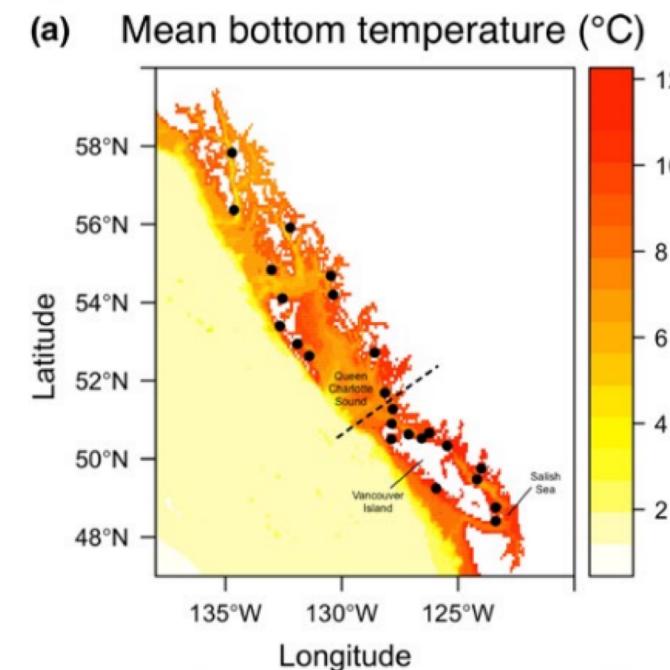
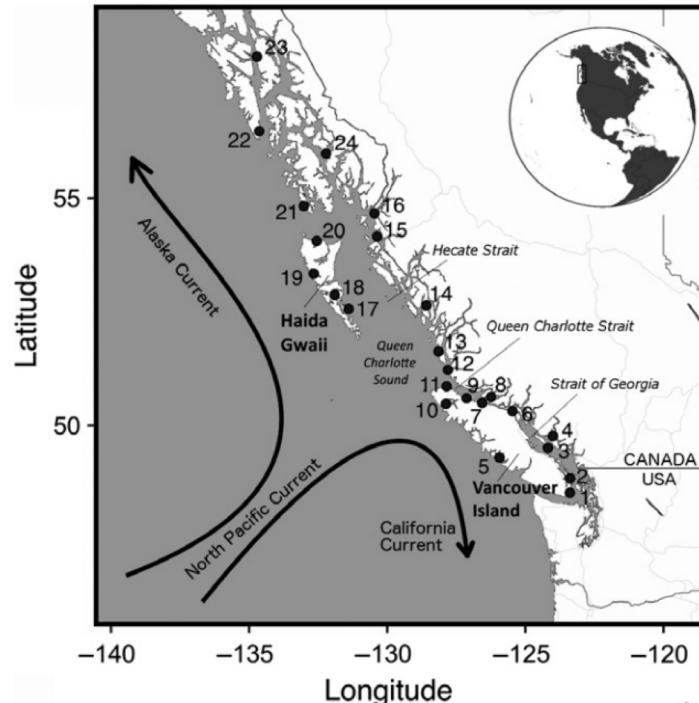
Xuereb et al. (2018)

Population structure

If population structure follows environmental gradient, it can confound GEA

Many GEA programs will correct for structure (via different statistical analyses)

Can also run GEAs within each pop if enough sample sites



GEA analysis

Correlation between environmental data and genomic data

LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies
DOI: <https://doi.org/10.1093/molbev/msz008>

Kevin Caye, Basile Jumentier, Johanna Lepeule, Olivier François

Molecular Biology and Evolution, Volume 36, Issue 4, April 2019, Pages 852–860,

<https://doi.org/10.1093/molbev/msz008>

Published: 17 January 2019

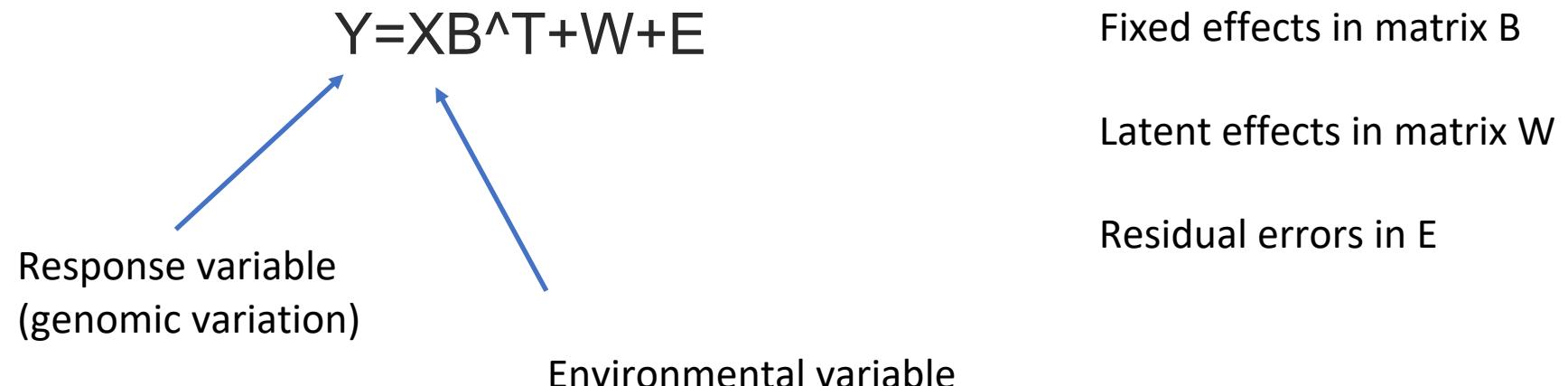
 PDF  Split View  Cite  Permissions  Share ▾

We'll be using the R package lfmm (latent factor mixed model)
Basically, a regression model with fixed and latent effects

$$Y = XB^T + W + E$$

Response variable (genomic variation) Environmental variable

Fixed effects in matrix B
Latent effects in matrix W
Residual errors in E



Data for class today

3966 SNPs from 685
Pacific sea
cucumbers

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*)

Amanda Xuereb¹  | Laura Benestan² | Éric Normandeau² | Rémi M. Daigle¹ |
Janelle M. R. Curtis³ | Louis Bernatchez² | Marie-Josée Fortin¹

