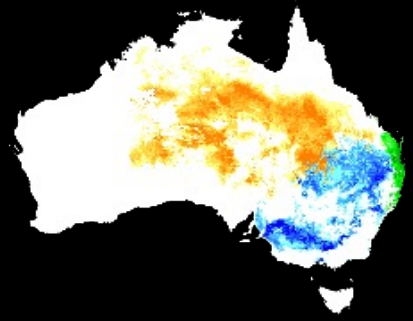
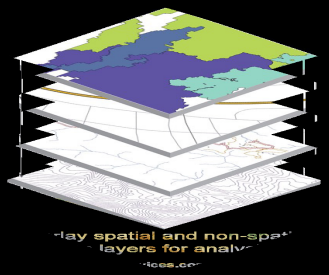




Natural Selection

Luciano Beheregaray, Jonathan (Yuma) Sandoval-Castillo and Chris Brauer

Flinders University

Kioloa, March 2024

Molecular Ecology Lab (MELFU)

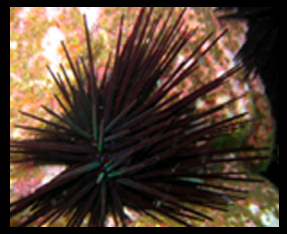
FLINDERS UNIVERSITY

Molecular Ecology Lab

www.molecularecology.flinders.edu.au

We integrate environmental and genomic datasets to address pure and applied questions in evolution, ecology and conservation management

A/Prof. Luciana Möller













FLINDERS UNIVERSITY

Molecular Ecology Lab

www.molecularecology.flinders.edu.au

Dr Jonathan Sandoval-Castillo (Yuma)

Dr Chris Brauer

A few simple (but relevant) distinctions* and observations

Adaptation:
A feature of an organism that serves a purpose in a given environment and is the result of natural selection *in the past*

Natural selection in progress:
Gene frequencies *changing now* because how different individuals interact with their environment

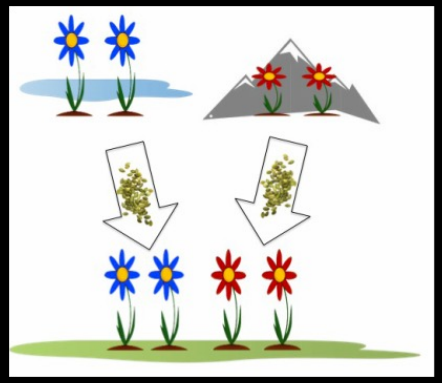
"Selection has nothing to do with what is adequate, or not, for continued survival"

* Modified from:
George C Williams (1966) *Adaptation and Natural Selection*

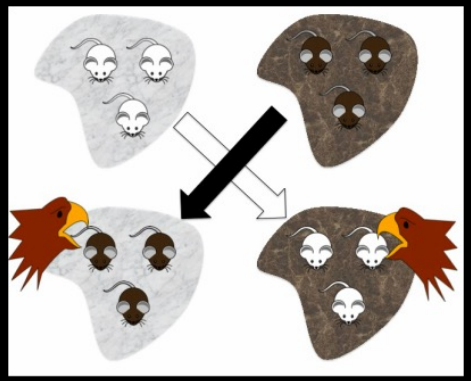
Study (& career?) goals: Explaining adaptations *versus* identifying the signal of selection in progress

Importance of natural history: Understanding spatial non-random mating (i.e. population structure), demographic history and ecology of the study species is key to study natural selection in wild populations

‘Traditional’ ways to study adaptation and natural selection
(without being able to identify its genetic basis)

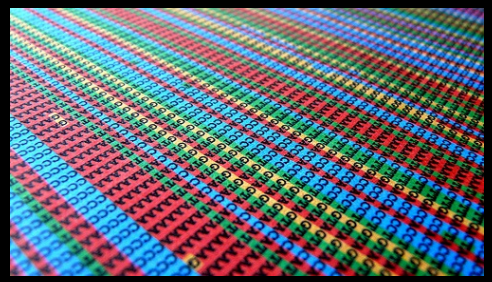


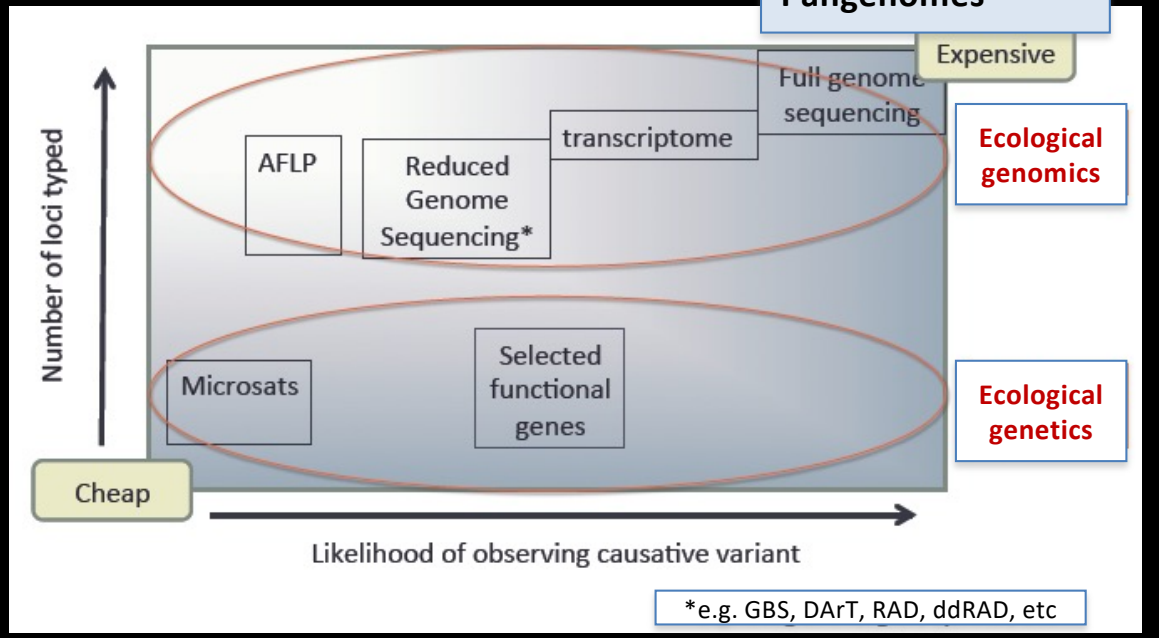
Common garden experiments



Reciprocal transplant experiments

The Era of Genomics





Modified from: Sork (2012) Mol Ecol Symposium, Ottawa

Is reduced-representation sequencing enough?

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 142–152 doi: 10.1111/1755-0998.12635

NEWS AND VIEWS

OPINION

Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation

DAVID B. LOWRY,*† SEAN HOBAN,‡§ JOANNA L. KELLEY,¶ KATIE E. LOTTERHOS,**
LAURA K. REED,†† MICHAEL F. ANTOLIN‡‡ and ANDREW STORFER¶

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 362–365 doi: 10.1111/1755-0998.12669

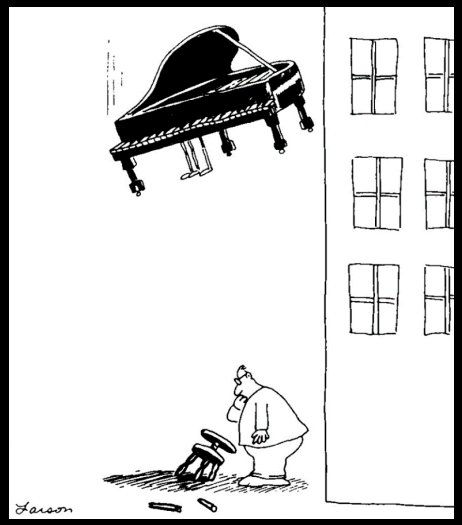
COMMENT

Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations

JULIAN M. CATCHEN,* PAUL A. HOHENLOHE,† LOUIS BERNATCHEZ,‡ W. CHRIS FUNK,§
KIMBERLY R. ANDREWS¶ and FRED W. ALLENDORF**

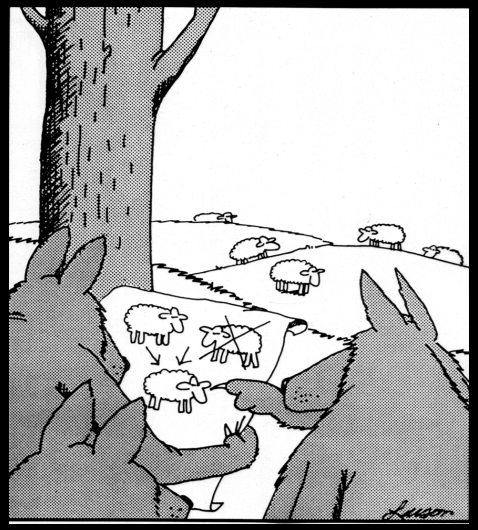
Results should be presented in the context of the experimental characteristics known about the system (including extent of LD) and study conclusions tempered based on their power to detect effects

Random events



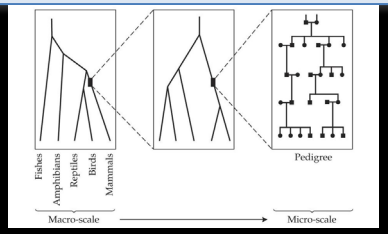
EVOLUTION

Natural selection (non-random)



STANDING GENOMIC VARIATION IN POPULATIONS

NEUTRAL VARIATION
- genome-wide signal
- no impact on fitness

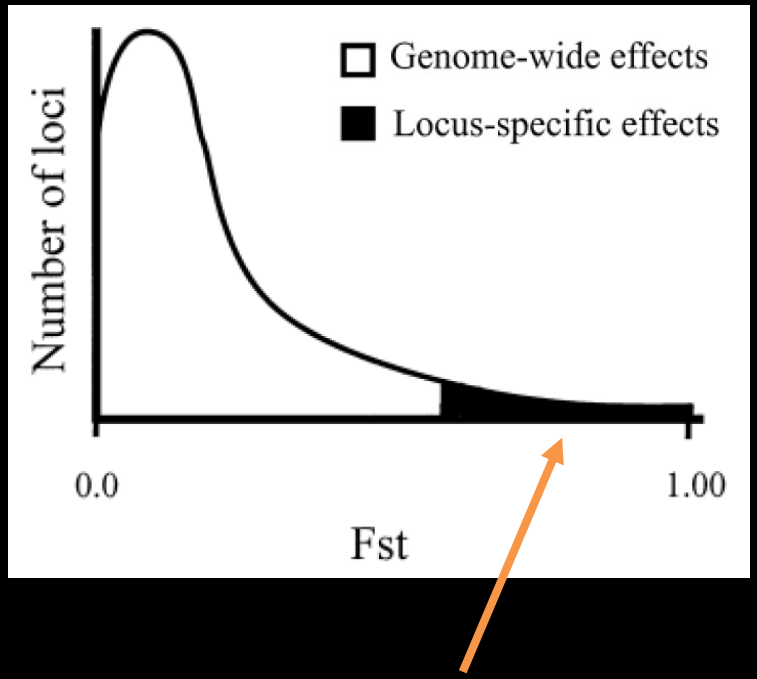


ADAPTIVE VARIATION
- no genome wide signal
- impacts on fitness

Phenotypic plasticity (this can become adaptive)

Identifying natural selection: Outlier tests

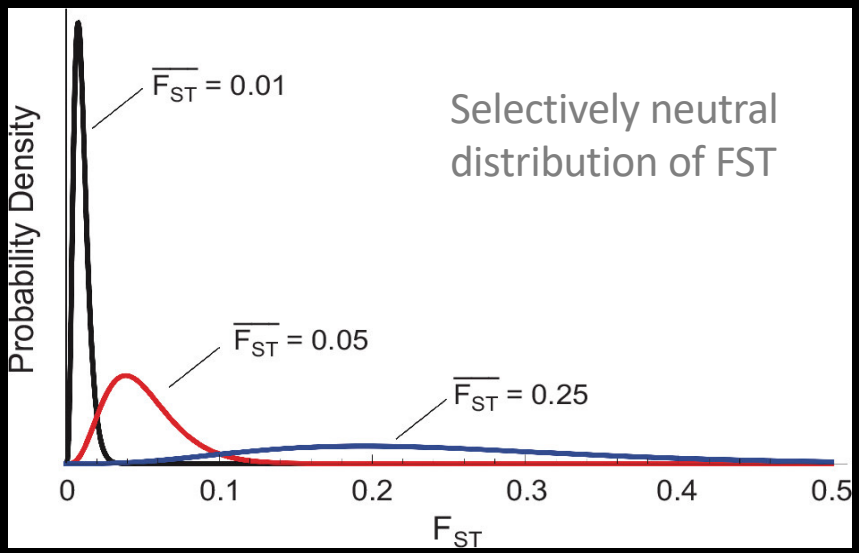
- Frequencies of many alleles often vary across populations due to genome-wide demographic effects = 'neutral' SNPs
- Frequencies of some alleles might co-vary with fitness traits or with environmental variation due to natural selection = 'adaptive SNPs'



Loci with exceptionally high Fst values are 'outlier loci' and *might* represent adaptive loci

Black *et al.* (2001) *Annu. Rev. Entomol.*

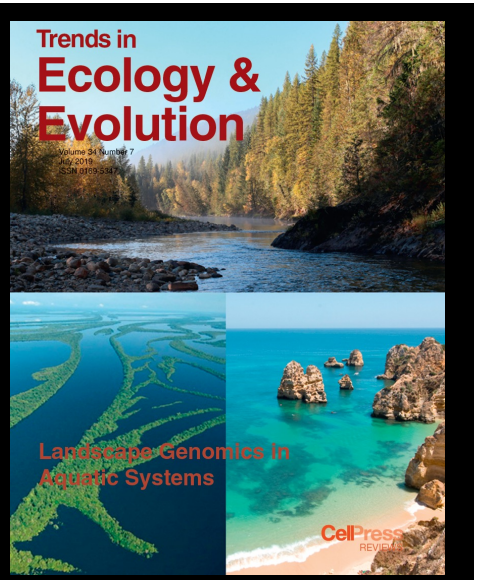
Pitfall: difficult test for highly differentiated populations



Hoban *et al.* (2016) *American Naturalist*

Landscape Genomics:

The study of genome-wide neutral and adaptive variation of ecologically diverse populations sampled across heterogeneous landscapes



Grummer *et al.* (2019)

Review

Ten years of landscape genetics

Stéphanie Manel[1,2] and Rolf Holderegger[3,4]

[1] Laboratoire LPED "Population Environnement, Développement" - UMR151 - Aix Marseille University - IRD Marseille, France
[2] UMR Botanique et BioinforMatique de l'Architecture des Plantes, CIRAD, TA A51/PS2 34398 Montpellier Cedex 5, France
[3] WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[4] Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, CH-8092 Zürich, Switzerland

Volume 19, Issue 17

SPECIAL ISSUE ON LANDSCAPE GENETICS

Pages: 3489-3835

September 2010

Tansley review

Genomic variation across landscapes: insights and applications

Jason G. Bragg[1], Megan A. Supple[1], Rose L. Andrew[1,2] and Justin O. Borevitz[1,3]

ECOLOGY LETTERS

Ecology Letters, (2015) **18**: 1–16 doi: 10.1111/ele.12376

IDEA AND PERSPECTIVE

Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation

Received: 18 March 2017 | Revised: 16 March 2018 | Accepted: 19 March 2018

DOI: 10.1111/mec.14584

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations

Brenna R. Forester[1] | Jesse R. Lasky[2] | Helene H. Wagner[3] | Dean L. Urban[1]

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 4348–4370 doi: 10.1111/mec.13322

INVITED REVIEWS AND SYNTHESES

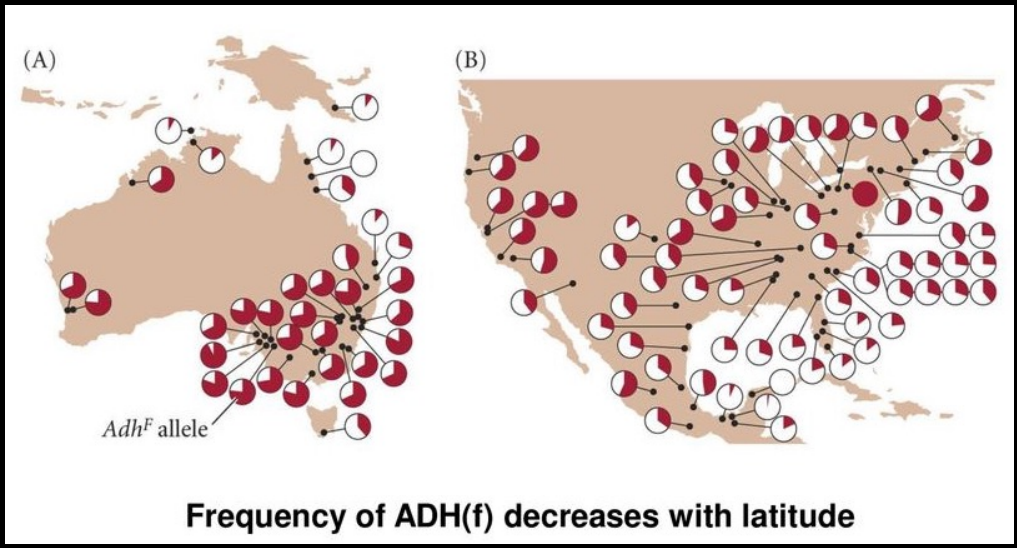
A practical guide to environmental association analysis in landscape genomics

Identifying natural selection: Genotype-Environment Association (GEA) tests

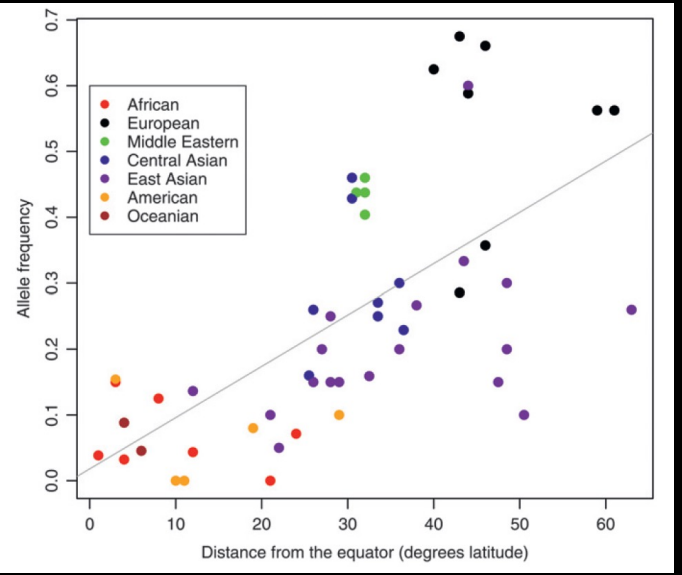
- Strong divergent selection between geographic locations >> local adaptation / differences in average phenotype >> differences in allele frequencies can be maintained in the face of gene flow

GEA tests: Identify candidate adaptive SNPs by testing for direct associations between variation in allele frequencies of each SNP and environmental predictors over the landscape

Pitfall:
non-independence of populations
(shared histories and environments)



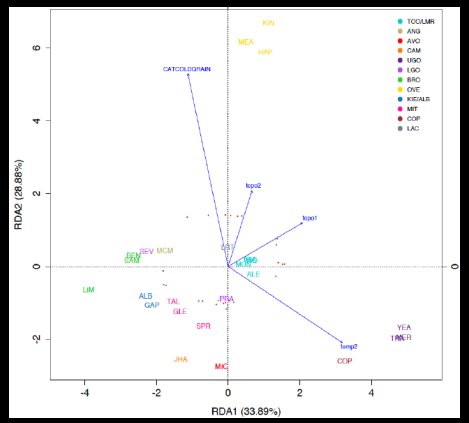
Oakeshott *et al.* (1982) *Evolution*



Thompson *et al.* (2004) *Am. J. Hum. Genet*

Identifying natural selection:
Genotype-Environment Association (GEA) tests

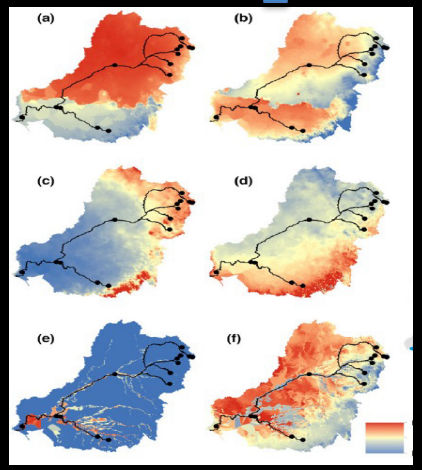
Redundancy analysis (RDA) models linear combinations of the explanatory variables (environmental predictors) that explain linear combinations of the response variables (SNPs) >> identify covarying SNPs associated with the multivariate predictors



Received: 29 June 2021 | Accepted: 26 August 2021
DOI: 10.1111/2041-210X.13722
Methods in Ecology and Evolution
REVIEW ARTICLE

Redundancy analysis: A Swiss Army Knife for landscape genomics

Thibaut Capblancq[1] | Brenna R. Forester[2]



RDA

- best balance between low false positive and high true positive rates
- robust across a range of demographic and spatial scenarios, sampling designs and types of selection
- good power to detect polygenic adaptation

Forester *et al.* (2018) *Molecular Ecology*



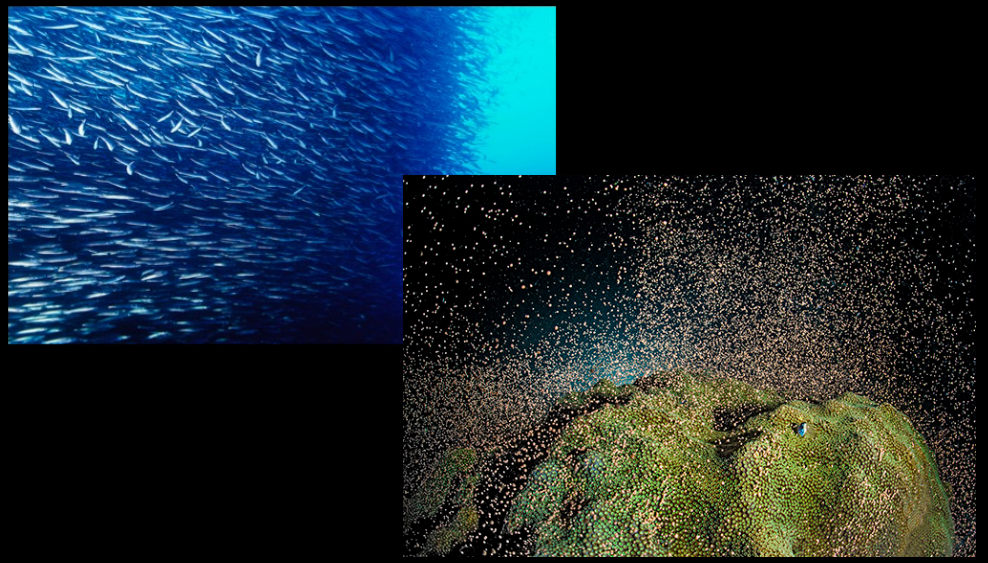


Mini (super mini 😊) case studies about Natural Selection from our lab

- Adaptation with gene flow
- Social biology and ecology
- Population history
- Eco-evo simulations
- Ecological niche models and genomic vulnerability
- GWAS of a disease phenotype

Seascape genomics of highly dispersive broadcast spawners

Environmental heterogeneity drives local adaptation despite high connectivity

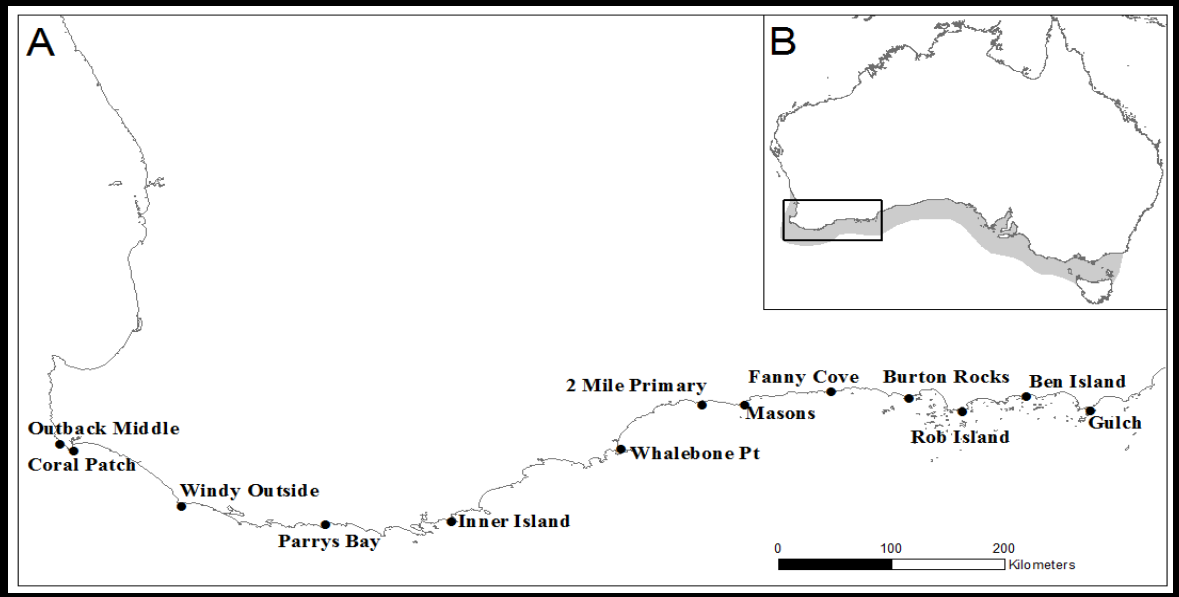


Haliotis roei

13 sites

n = 371

10,803 SNPs

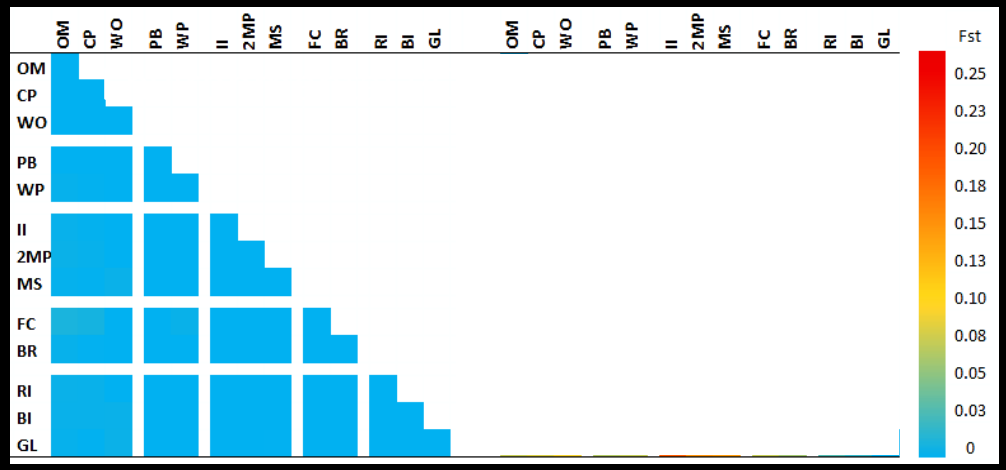


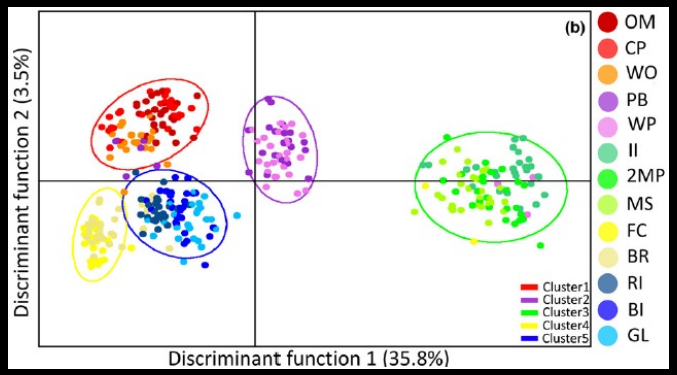
Sandoval-Castillo *et al.* (2018) *Molecular Ecology*

Adaptive divergence with high gene flow

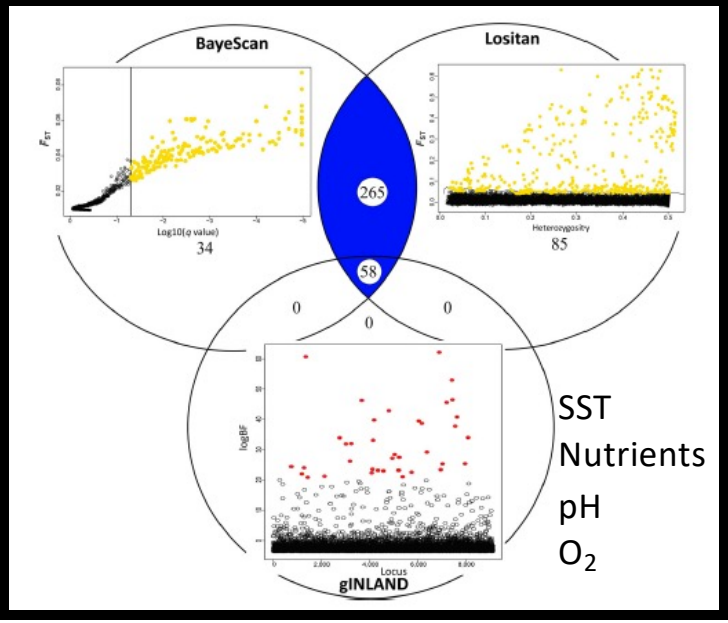
All SNPs

Candidate adaptive SNPs

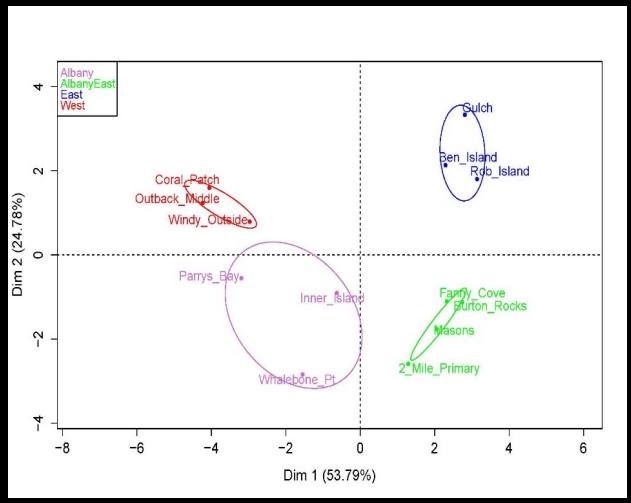




Tests for selection

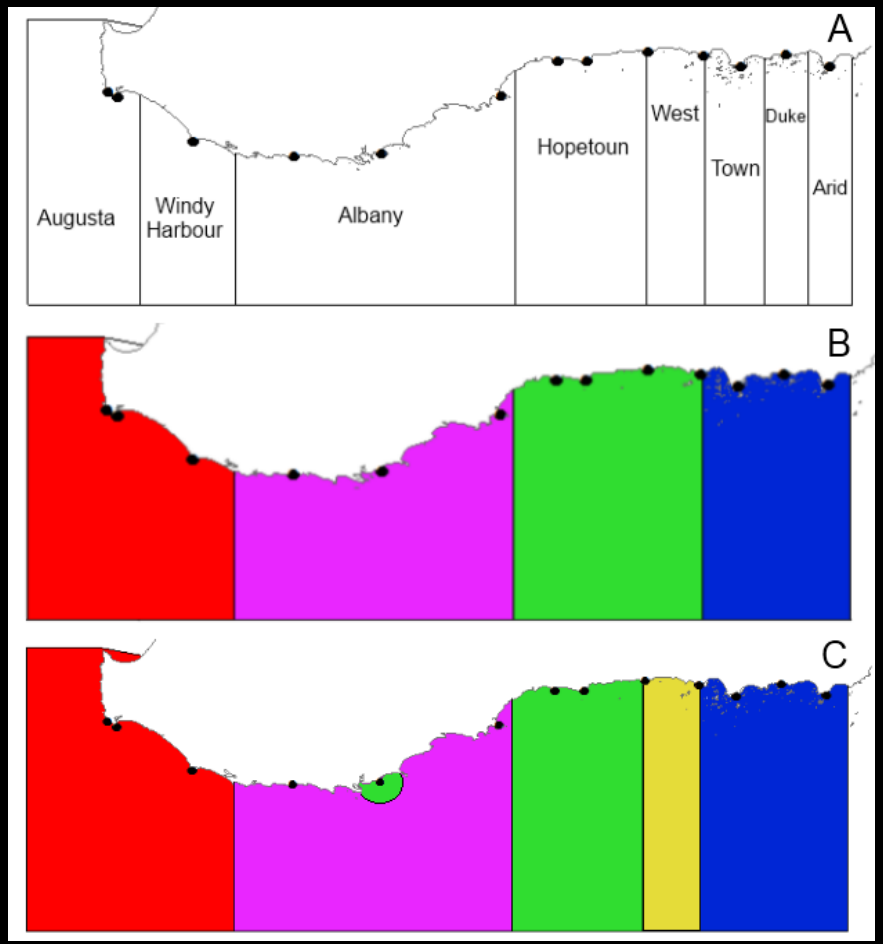


’Adaptive’ populations >> environmental PCAs



Sandoval-Castillo *et al.* (2018) *Molecular Ecology*

Adaptive divergence with high gene flow



8 fisheries management regions

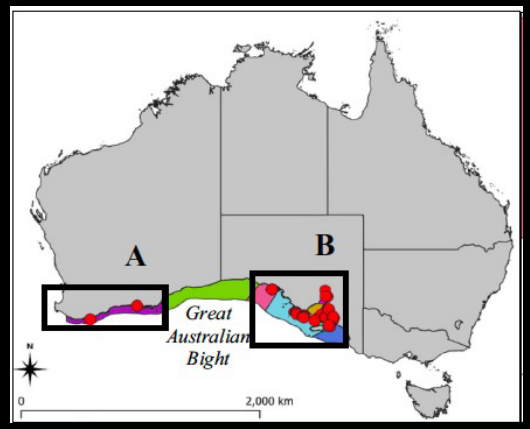
4 environmental regions

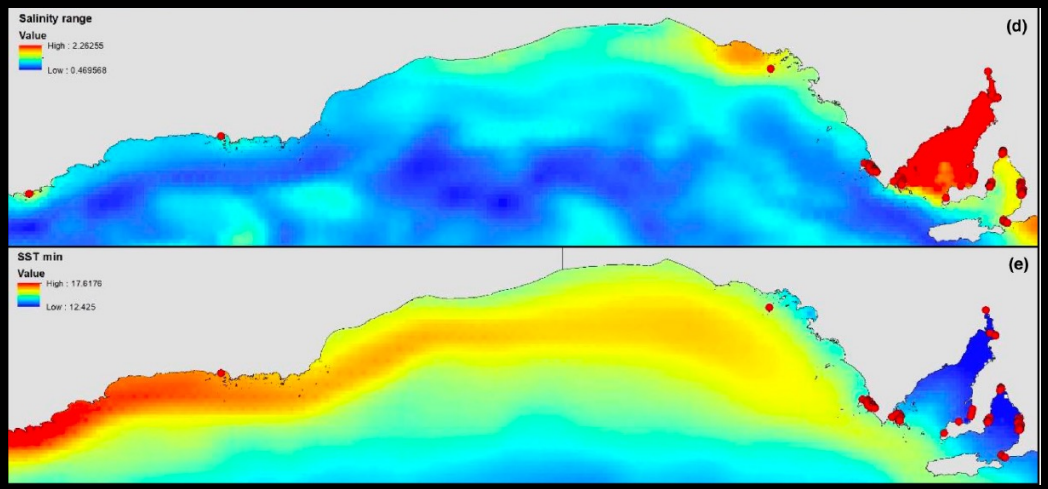
5 adaptive clusters for broodstock collection and breeding

Sandoval-Castillo *et al.* (2018) *Molecular Ecology*

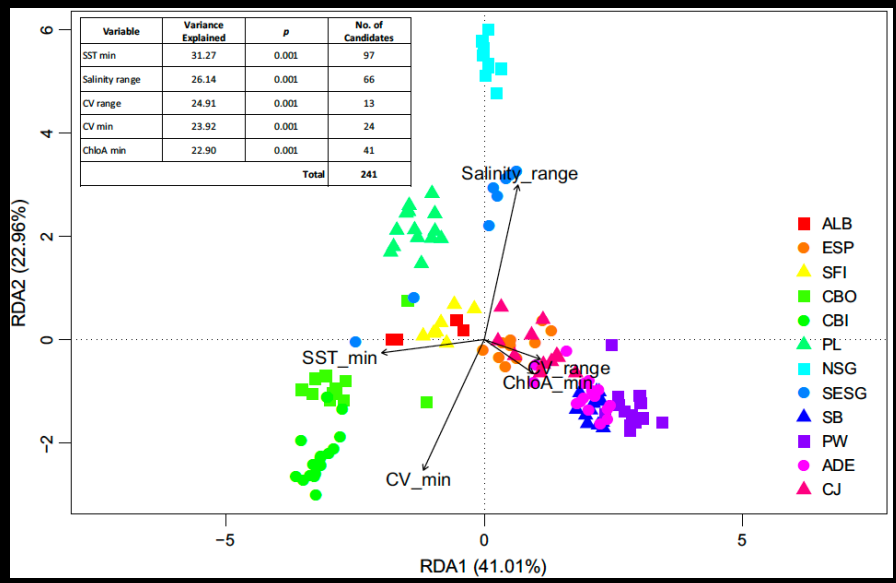
Social biology and ecology

12 sites
n = 139
8,104 SNPs





RDA (5 retained env variables)



Pratt *et al.* (2022) *Molecular Ecology*

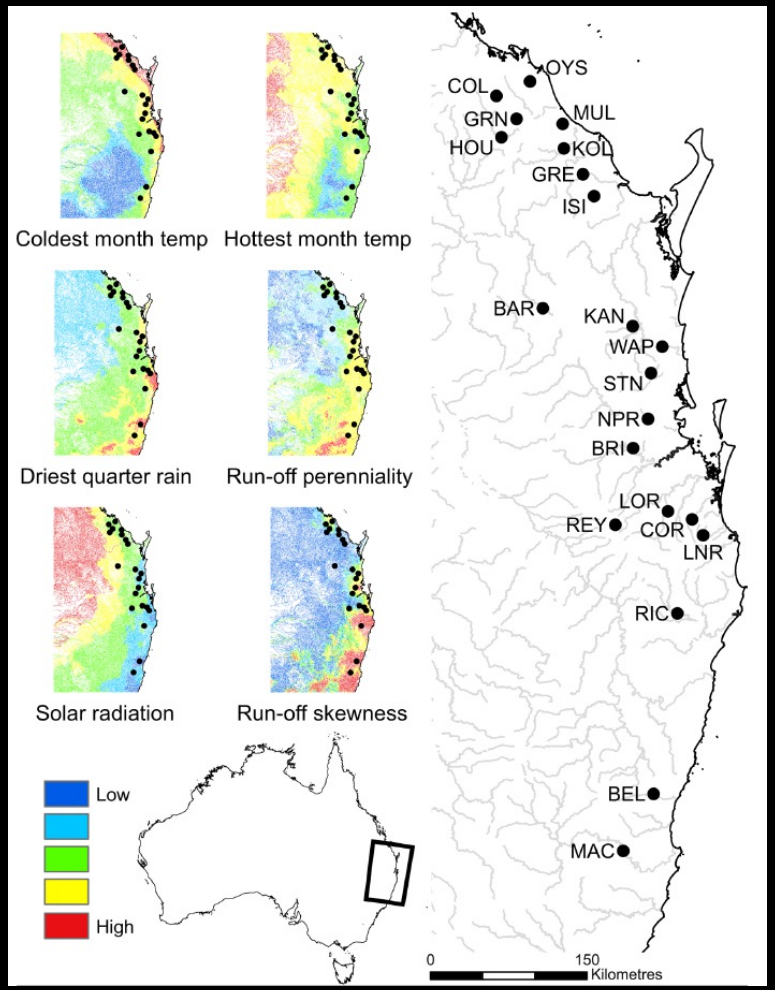
Genome Biology and Evolution (2023) GBE

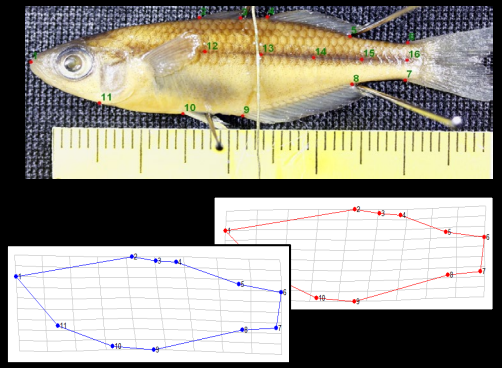
Genomic Divergence and the Evolution of Ecotypes in Bottlenose Dolphins (Genus *Tursiops*)

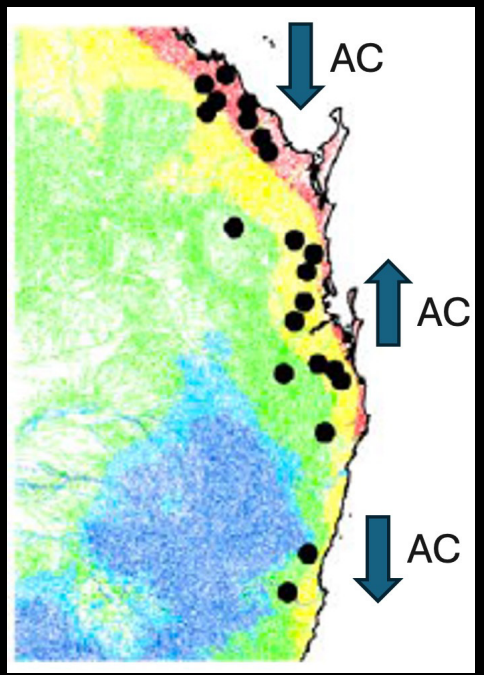
Eleanor A.L. Pratt[1,2], Luciano B. Beheregaray [1], Pedro Fruet[3,4,5], Gabriela Tezanos-Pinto[6], Kerstin Bilgmann[7], Nikki Zanardo[1,2,8], Fernando Diaz-Aguirre [1,2], Eduardo R. Secchi[3,4], Thales R.O. Freitas[9], and Luciana M. Möller [1,2,*]

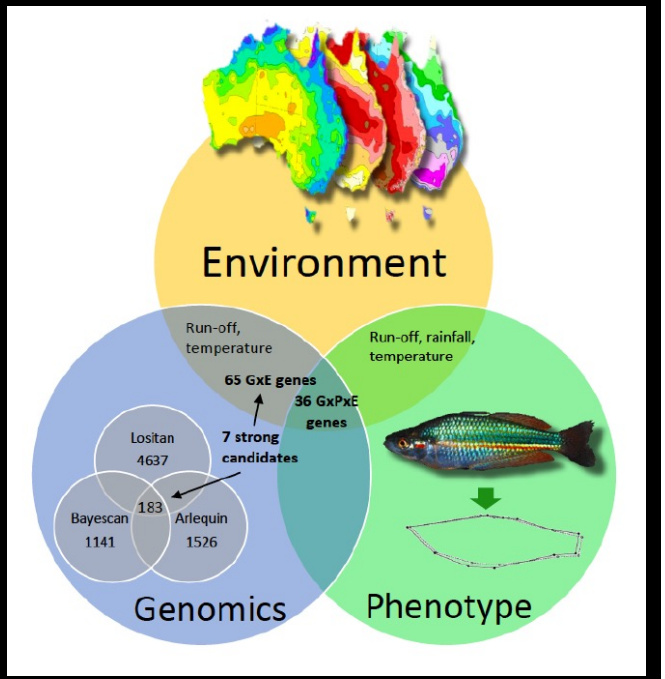
Latitudinal gradient and Population History

21 sites
n = 420
17,047 SNPs





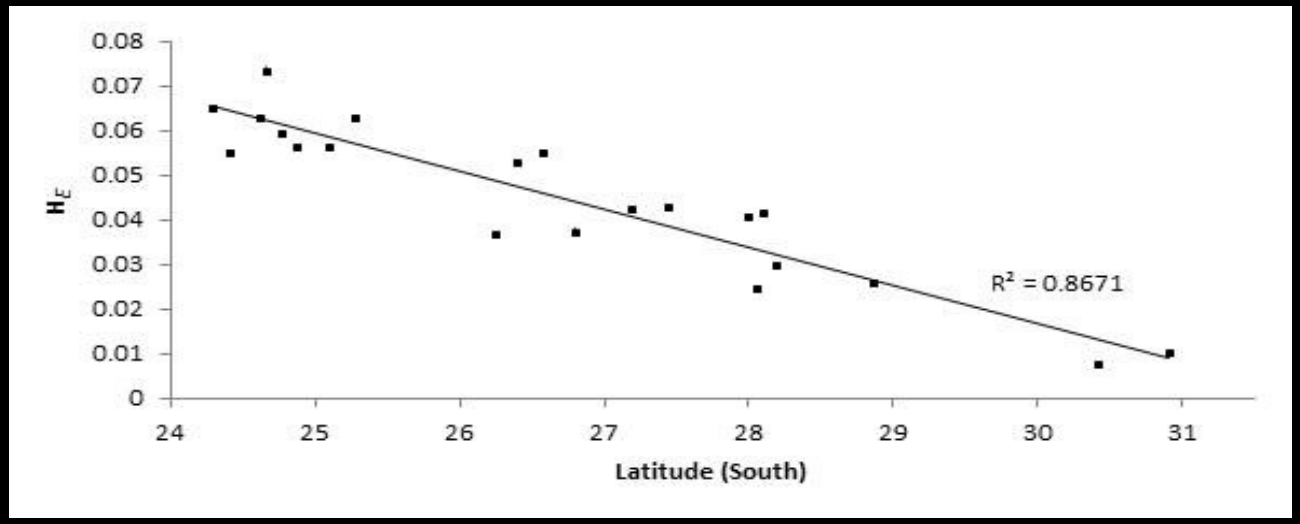




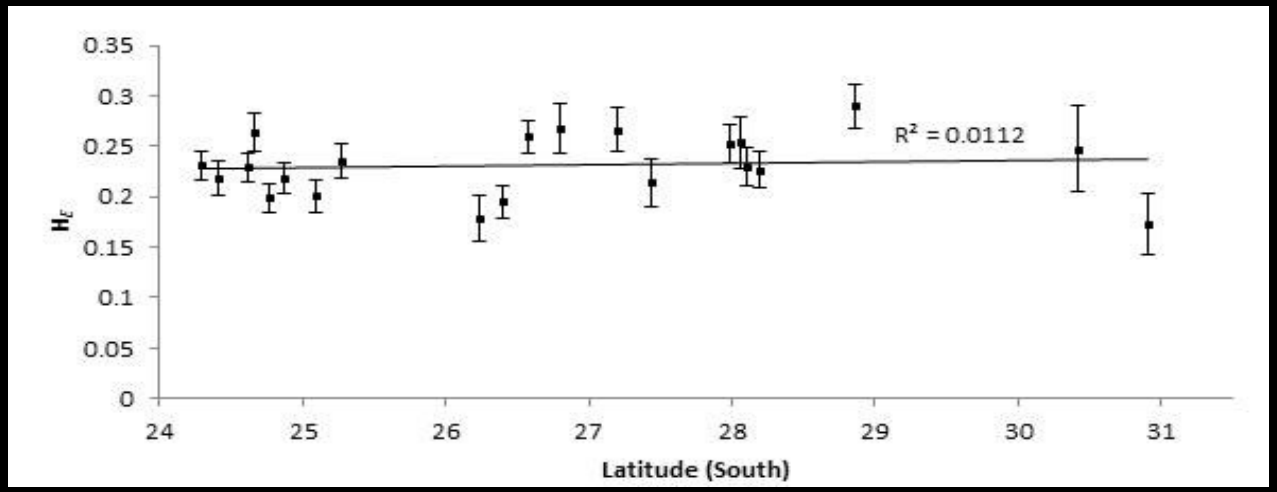
Smith *et al.* (2020) *Molecular Ecology*

Latitudinal gradient and Population History

All SNPs: strong cline in genetic diversity (population size - Pleistocene history)

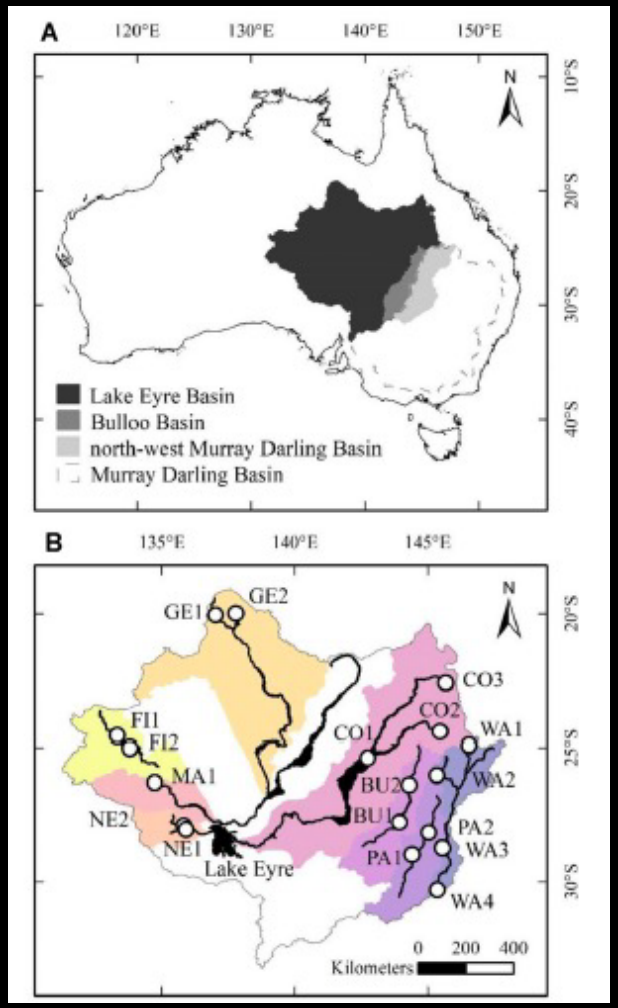


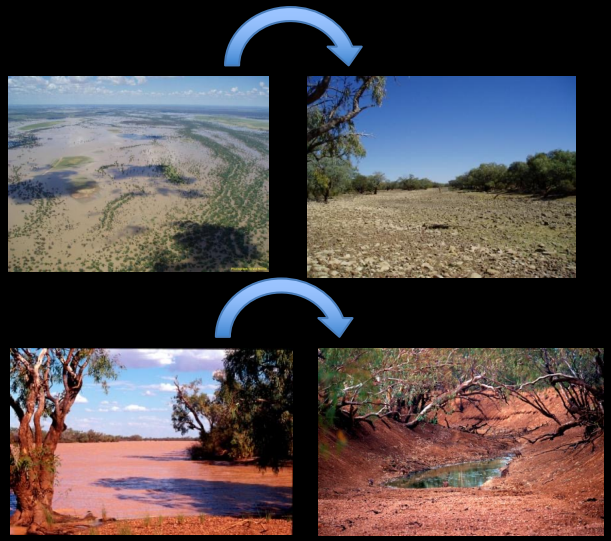
Adaptive SNP dataset: out with the cline! Evidence for balancing selection

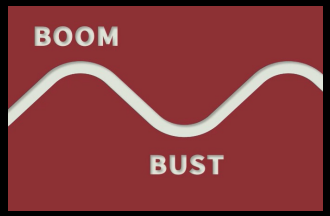


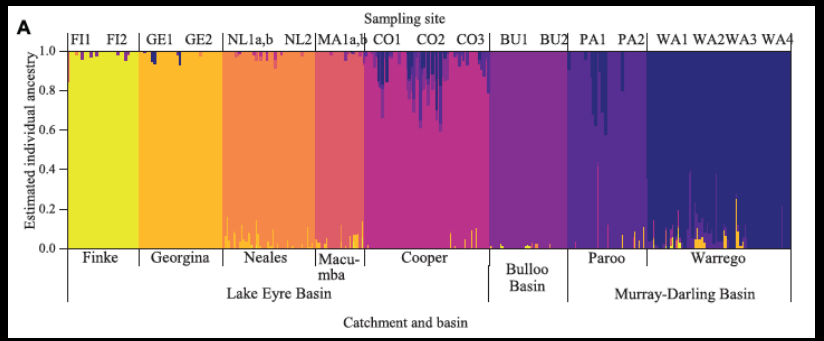
Smith *et al.* (2020) *Molecular Ecology*

Eco-evo simulations to model population dynamics









18 sites
n = 351
21,294 SNPs

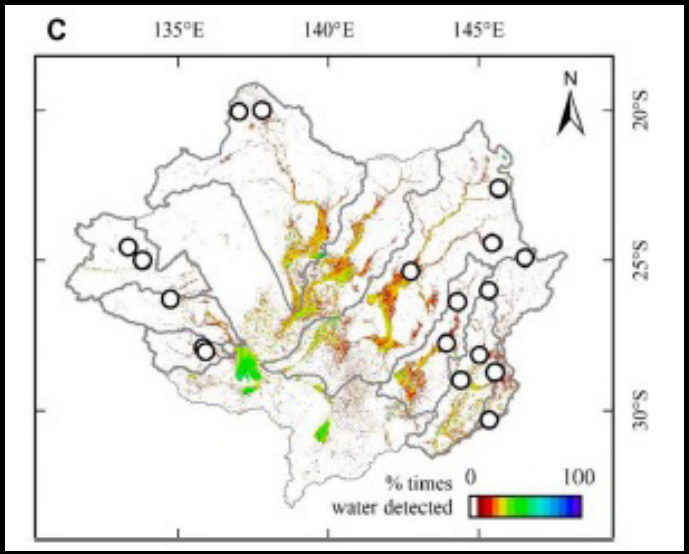
Attard *et al.* (2022) *Evolution*

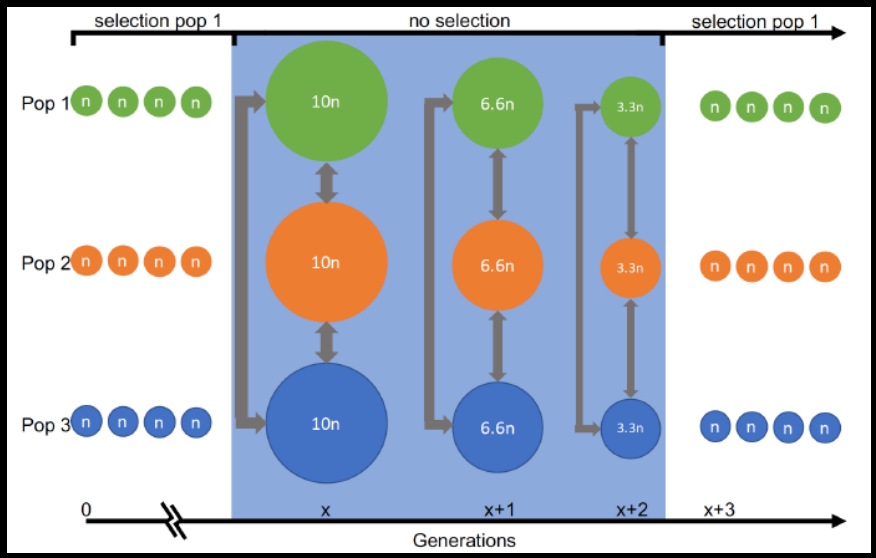
Eco-evo simulations to model population dynamics

GEAs using '*Water Observations from Space*' (detected over 25 years)

+

Eco-evo simulations (SLiM 3.1) to model boom-bust dynamics

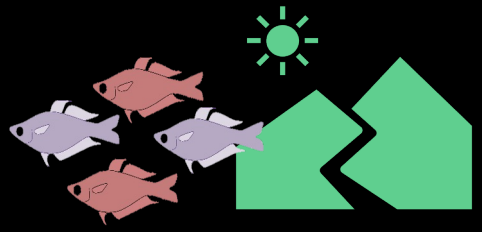


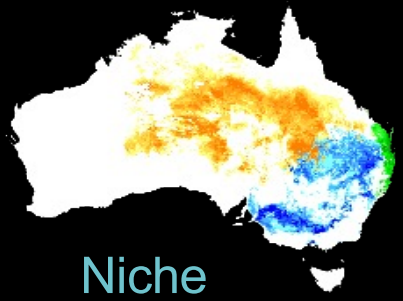


Adaptive diversity associated with aridity can be maintained in very small populations and integrates with rare events of connectivity to allow persistence in the desert

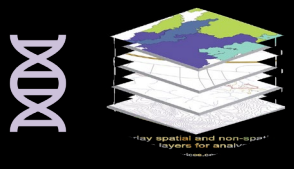
Attard *et al.* (2022) *Evolution*

Ecological niche models and genomic vulnerability





Niche distribution models (present and future)



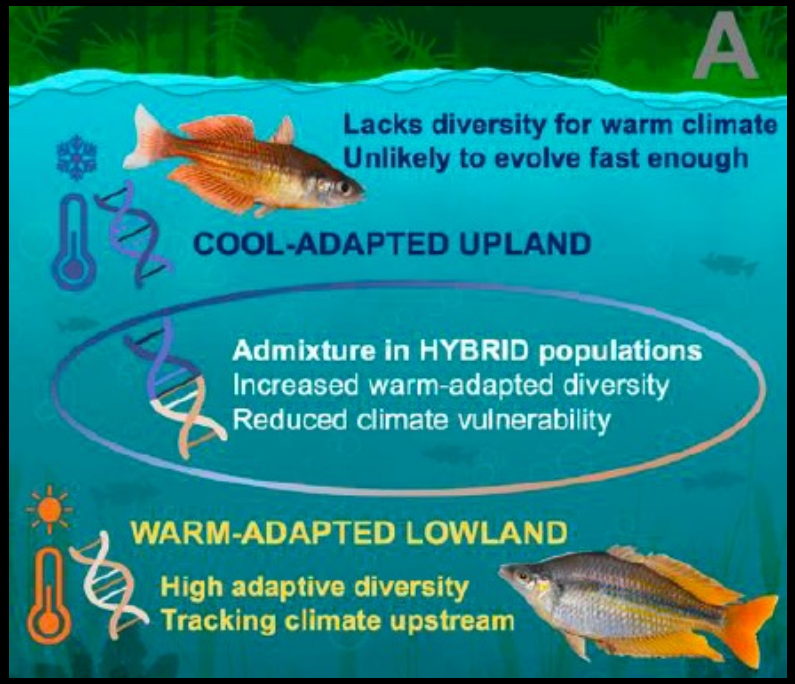
GEAs

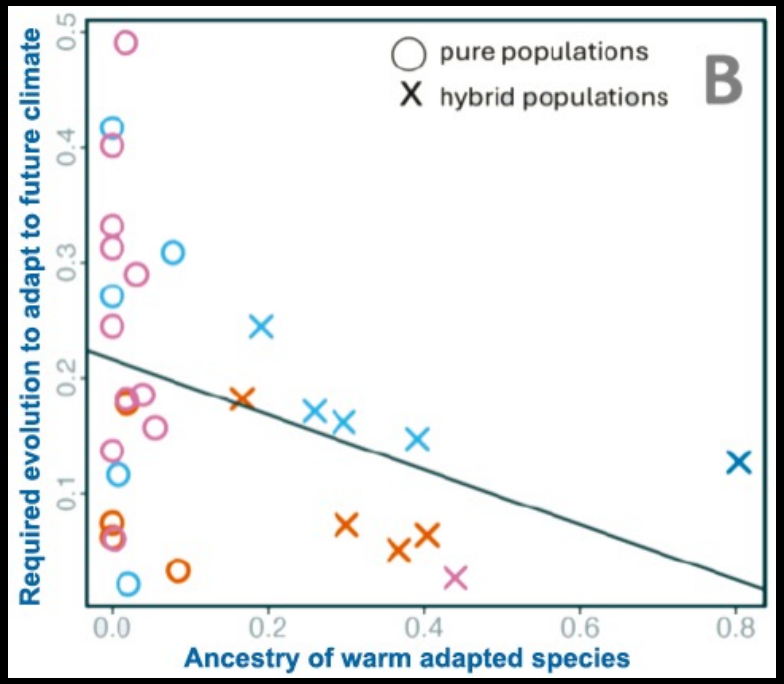
Amount of evolution needed to maintain present adaptation in future climates

Genomic vulnerability (Genomic offset)

Populations with high GV are more vulnerable to CC

Ecological niche models and genomic vulnerability





Natural hybridisation reduces vulnerability to climate change

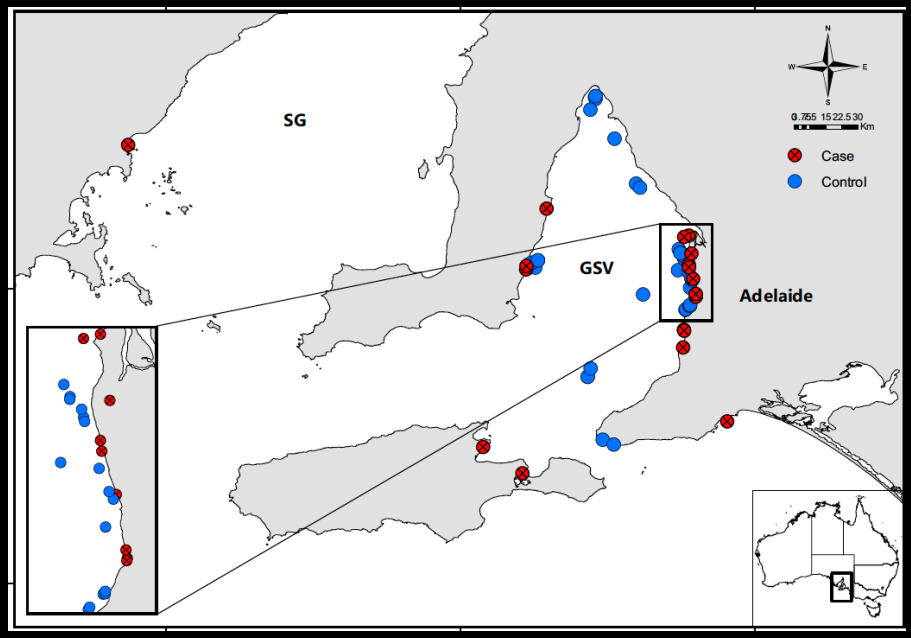
Brauer *et al.* (2023) *Nature Climate Change*

Clarivate Highly Cited Paper Web of Science™

Genome-wide association study (GWAS) of a disease phenotype

CeMv (cetacean morbillivirus)

Controls *vs* *Cases*



10.2 million SNPs
Multiple stringent association tests (age, sex and relatedness)

DOI: 10.1111/mec.15873

SPECIAL ISSUE: WHOLE-GENOME SEQUENCING IN MOLECULAR ECOLOGY

MOLECULAR ECOLOGY WILEY

Whole genomes reveal multiple candidate genes and pathways involved in the immune response of dolphins to a highly infectious virus

Kimberley C. Batley[1,2] | Jonathan Sandoval-Castillo[1] | Catherine M. Kemper[3] | Nikki Zanardo[1,2] | Ikuko Tomo[3] | Luciano B. Beheregaray[1] | Luciana M. Möller[1,2]

Hands-on section

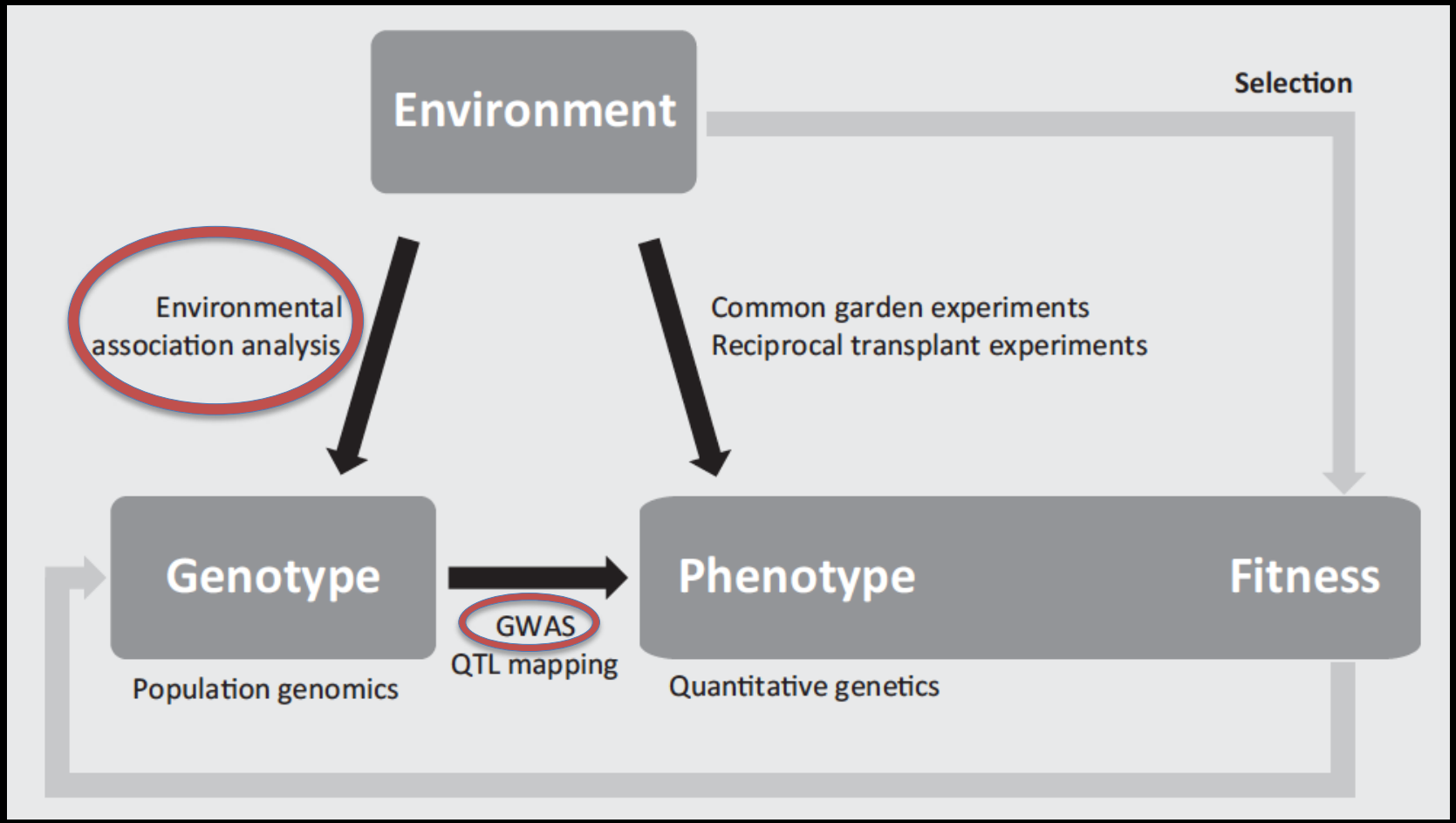
Chris Brauer

Genotype-Environment Association (GEA)

Jonathan Sandoval-Castillo
(Yuma)

Genome-Wide Association Study (GWAS)

Detecting signal of natural selection, environmental factors and genes involved in local adaptation



source of variation that can be quantified

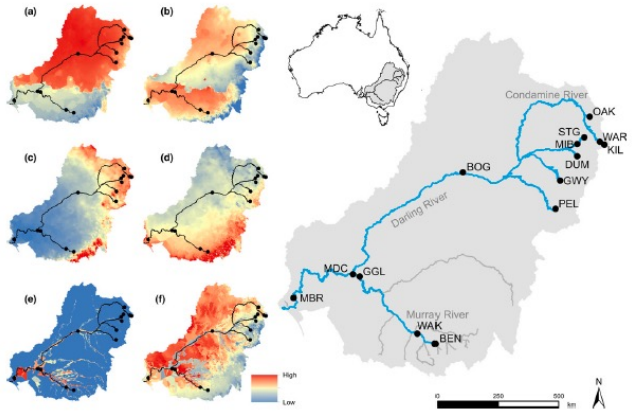
direction of evolution between cause and effect

how selection acts on the different levels

Rellstab *et al.* (2015) *Molecular Ecology*

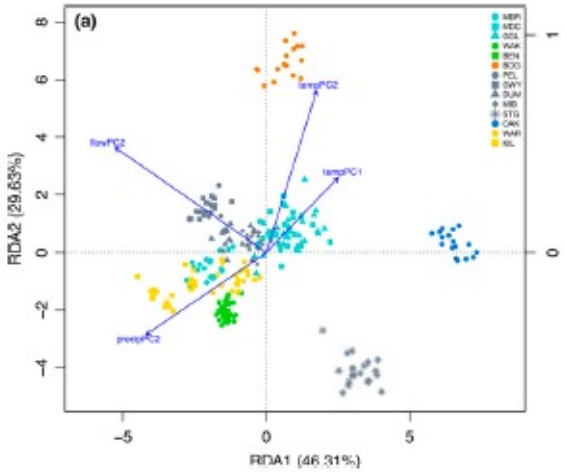
Genotype–environment association (GEA)

Chris Brauer, Flinders University



How does the environment shape genetic variation?

GEA methods test for signals of adaptive divergence associated with environmental variation



ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system

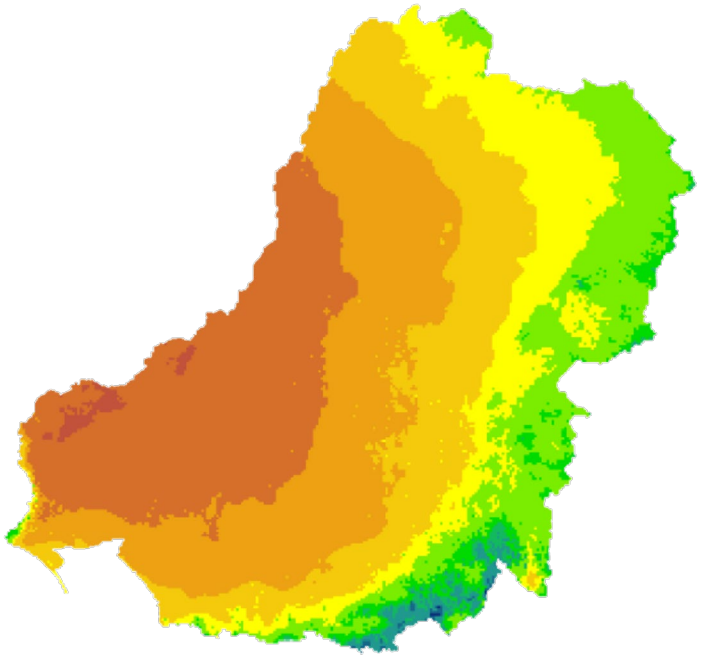
Chris J. Brauer[1] | Peter J. Unmack[2] | Steve Smith[1,3] | Louis Bernatchez[4] | Luciano B. Beheregaray[1]

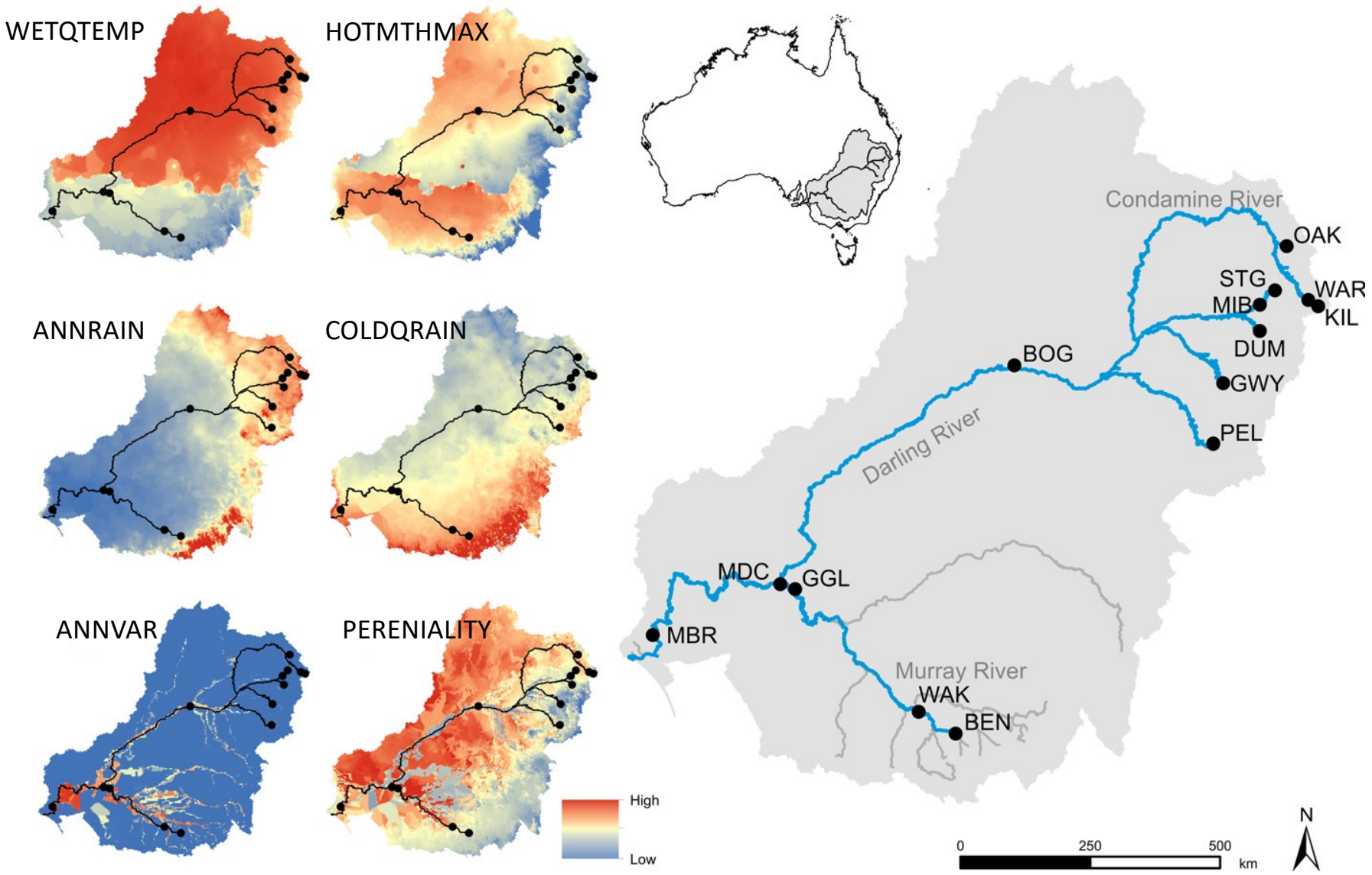
Murray river rainbowfish
(*Melanotaenia fluviatilis*)

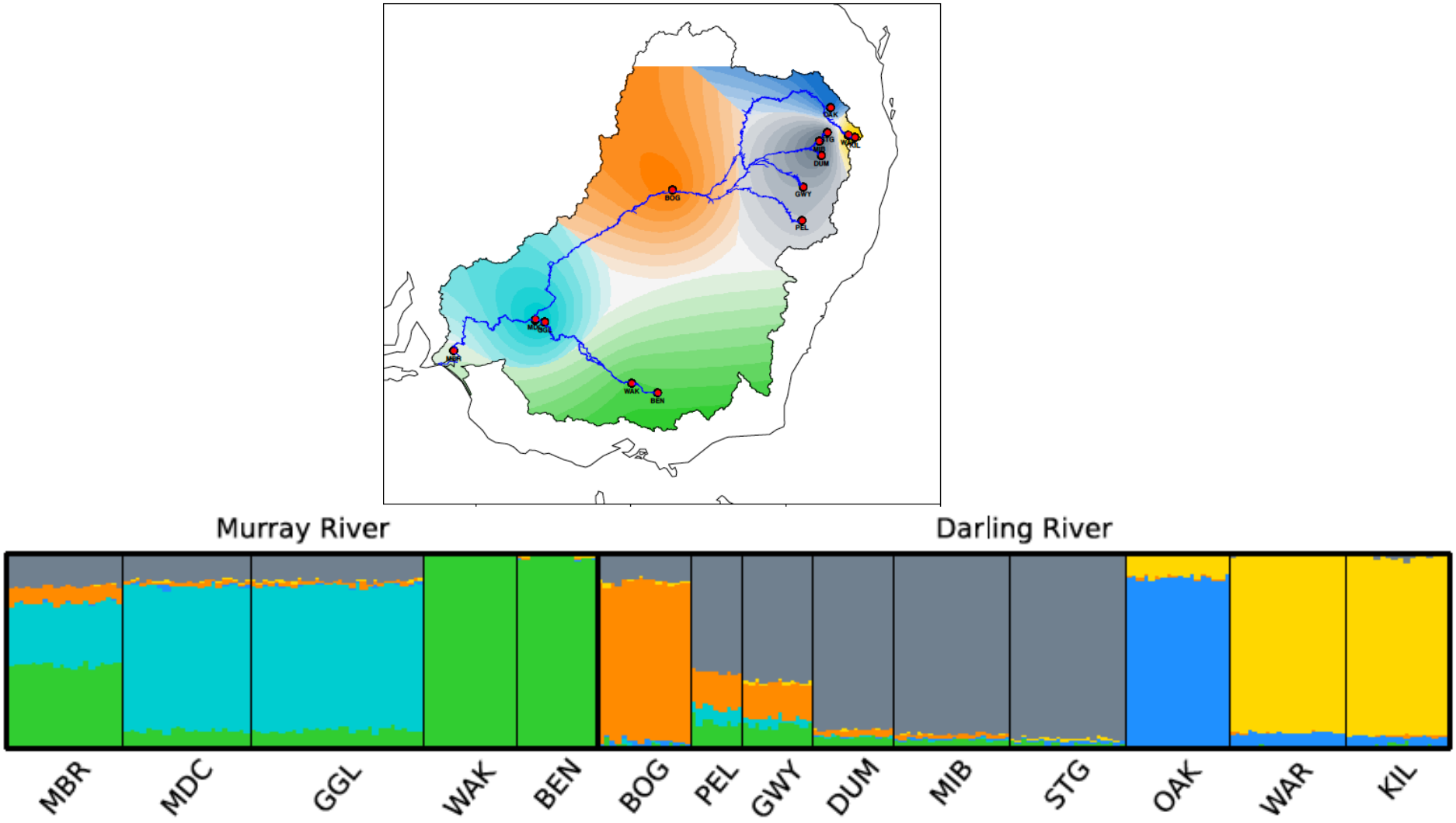
Generalist species
Some capacity for dispersal

Most southern rainbowfish
Less abundant in south
Steep climatic gradient

Local adaptation?
249 individuals, 17503 SNPs




WETQTEMP
HOTMTHMAX
ANNRAIN
COLDQRAIN
ANNVAR
PERENIALITY
High
Low
Condamine River
OAK
STG
WAR
MIB
KIL
DUM
BOG
GWY
PEL
Darling River
MDC
GGL
MBR
Murray River
WAK
BEN
0
250
500
km
N

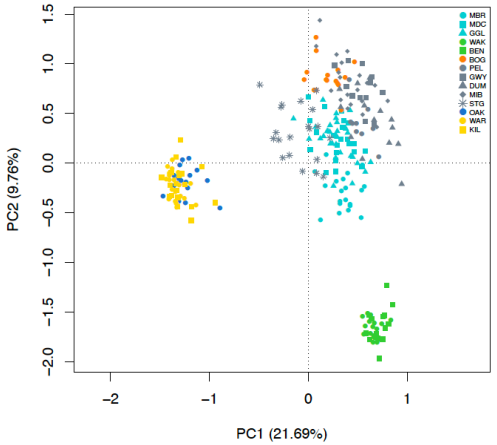

OAK
STG
MIB
WAR
KIL
DUM
GWY
BOG
PEL
MDC
GGL
MBR
WAK
BEN
Murray River
Darling River
MBR
MDC
GGL
WAK
BEN
BOG
PEL
GWY
DUM
MIB
STG
OAK
WAR
KIL

Multivariate ordination

Two major types

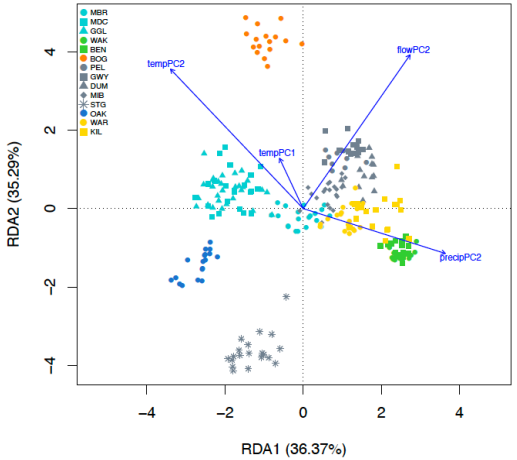
Unconstrained

Identifies axes that account for variation in the data e.g. PCA, CA



Constrained

Combines unconstrained ordination with regression analysis, with the axes of variation being constrained to linear combinations of a set of predictor variables e.g. CCA, RDA



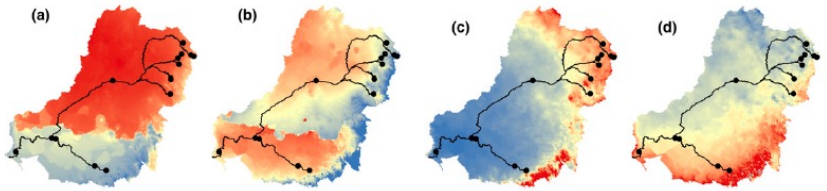
Redundancy analysis

Y ~ x

Genotypes ~ environmental variables

	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6	SNP_7	SNP_8	SNP_9	SNP_10
MBR_10	1	1	2	2	2	2	1	1	2	1
MBR_11	1	1	2	2	2	2	0	2	2	0
MBR_12	1	2	2	2	2	1	0	2	2	1
MBR_14	0	2	2	2	2	2	1	1	2	2
MBR_15	0	1	2	2	2	2	1	1	2	2
MBR_16	0	2	2	2	2	2	0	2	2	2
MBR_17	0	0	2	2	2	2	2	2	2	2
MBR_18	0	1	2	2	2	1	1	2	2	0
MBR_19	1	2	2	2	2	1	0	2	2	2
MBR_1	0	1	2	2	1	2	0	2	2	2

~



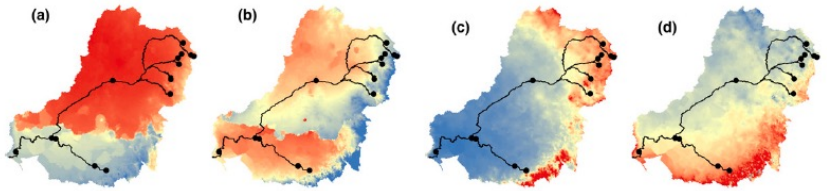
Partial redundancy analysis

Y ~ x + Condition(z)

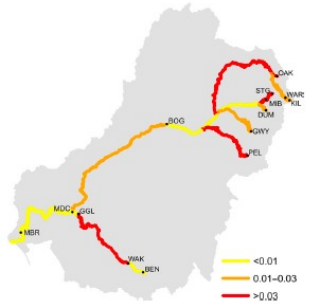
Genotypes ~ environmental variables + Condition(space)

	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6	SNP_7	SNP_8	SNP_9	SNP_10
MBR_10	1	1	2	2	2	2	1	1	2	1
MBR_11	1	1	2	2	2	2	0	2	2	0
MBR_12	1	2	2	2	2	1	0	2	2	1
MBR_14	0	2	2	2	2	2	1	1	2	2
MBR_15	0	1	2	2	2	2	1	1	2	2
MBR_16	0	2	2	2	2	2	0	2	2	2
MBR_17	0	0	2	2	2	2	2	2	2	2
MBR_18	0	1	2	2	2	1	1	2	2	0
MBR_19	1	2	2	2	2	1	0	2	2	2
MBR_1	0	1	2	2	1	2	0	2	2	2

~



+ Condition(



)

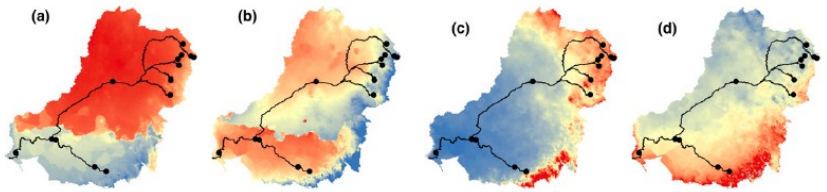
Partial redundancy analysis

Y ~ x + Condition(z)

Genotypes ~ environmental variables + Condition(space)

	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6	SNP_7	SNP_8	SNP_9	SNP_10
MBR_10	1	1	2	2	2	2	1	1	2	1
MBR_11	1	1	2	2	2	2	0	2	2	0
MBR_12	1	2	2	2	2	1	0	2	2	1
MBR_14	0	2	2	2	2	2	1	1	2	2
MBR_15	0	1	2	2	2	2	1	1	2	2
MBR_16	0	2	2	2	2	2	0	2	2	2
MBR_17	0	0	2	2	2	2	2	2	2	2
MBR_18	0	1	2	2	2	1	1	2	2	0
MBR_19	1	2	2	2	2	1	0	2	2	2
MBR_1	0	1	2	2	1	2	0	2	2	2

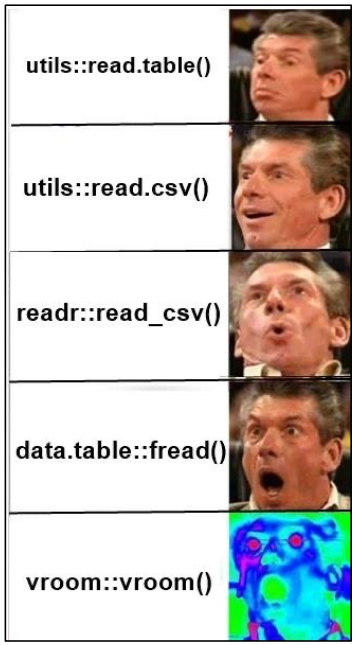
~

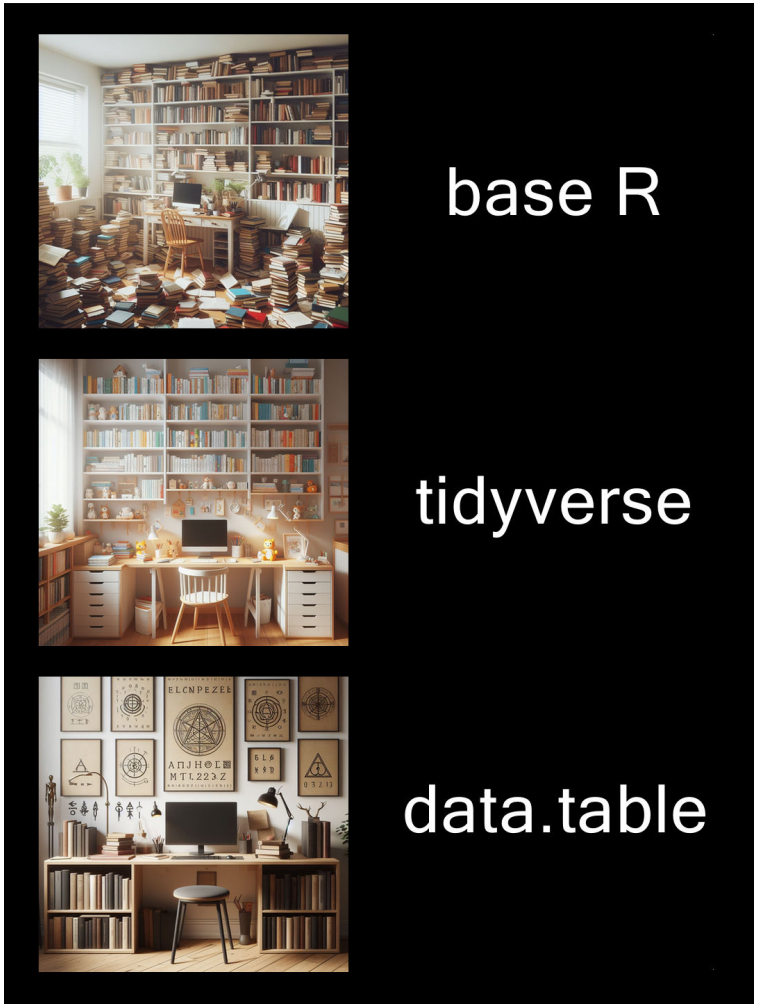


+ Condition(XY)

Coding in R: the following examples are how I do it, **not** how you should do it…

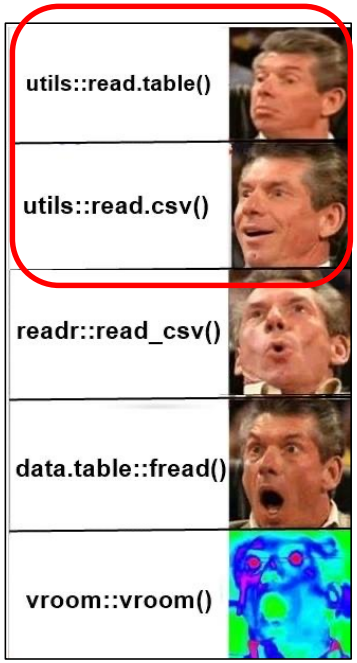
Use whatever works…

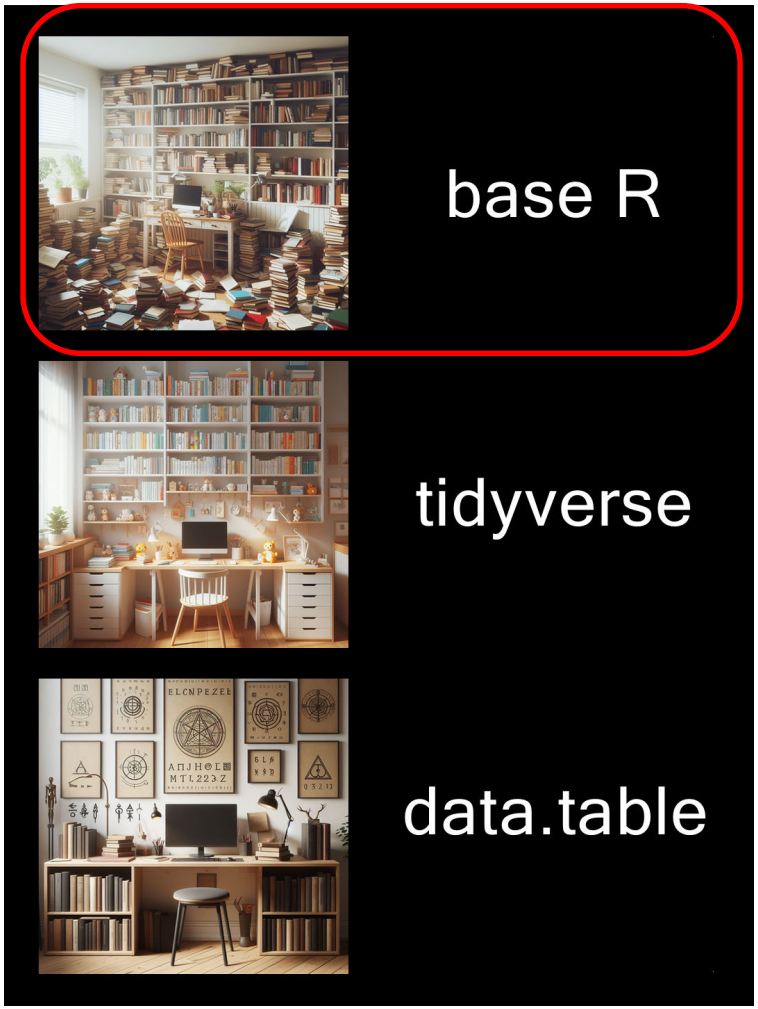




Coding in R: the following examples are how I do it, **not** how you should do it…

Use whatever works…
For me that (mostly) looks like this



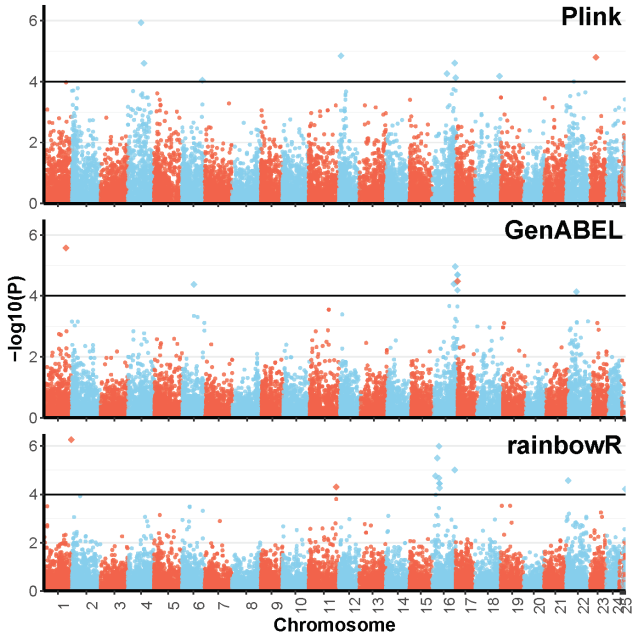


Genome-wide association study (GWAS)

Jonathan Sandoval Flinders University, MELFU

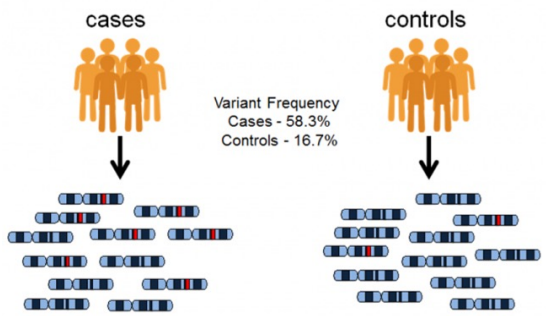
GWAS tests for association between genotypes and phenotypic traits, including diseases, physical characteristics, behavioural traits, and responses to environmental stimuli.

They are important tools to understand biological mechanisms underlying phenotypes, and ultimately to predicting phenotypes from genomic information.



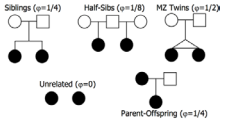
Genome-wide association study (GWAS)

GWAS involves genotyping a large number of genetic markers across the genome several individuals.



Covariate factors that should be considered:

Population structure
Relatedness
Sex bias
Age bias

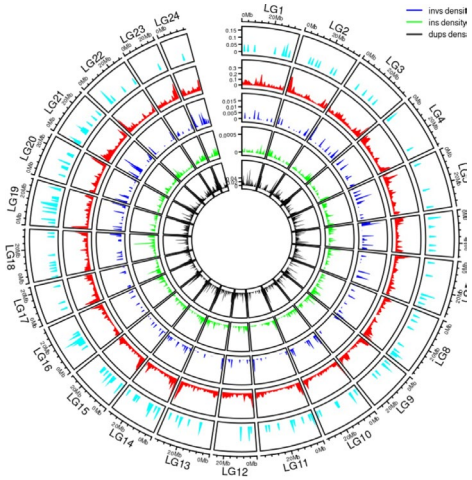


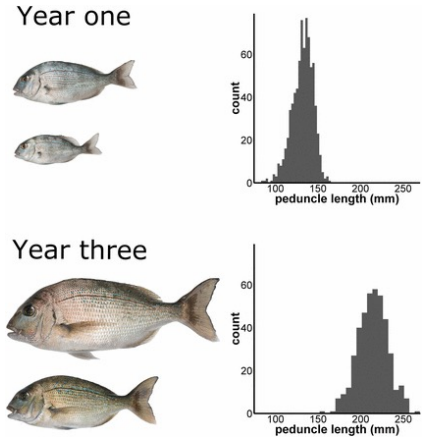
OXFORD G3 Genes | Genomes | Genetics

G3, 2022, **12(3)**, jkac015
https://doi.org/10.1093/g3journal/jkac015
Advance Access Publication Date: 20 January 2022
Genomic Prediction

Genomic prediction of growth in a commercially, recreationally, and culturally important marine resource, the Australian snapper (*Chrysophrys auratus*)

Jonathan Sandoval-Castillo [1], Luciano B. Beheregaray,[1] and Maren Wellenreuther [2,3,*]





G3 Genes|Genomes|Genetics, Volume 9, Issue 4, 2019, Pages 1027–1035

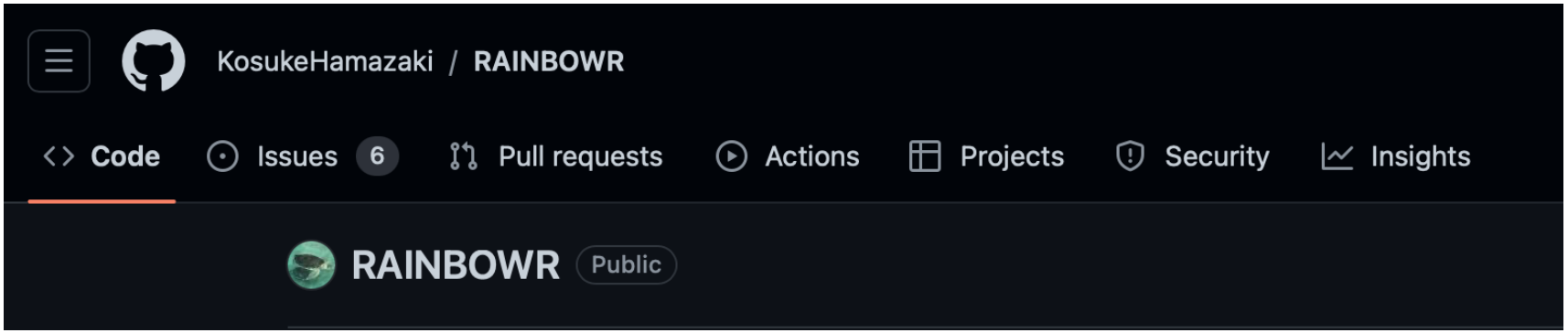
180 Small
183 Large

ddRAD
17,490 high-quality SNPs

Tune Machine Learning Algorithms in R (random forest case study)

by **Jason Brownlee** on July 31, 2020 in **R Machine Learning**

158



RandomForest

Supervised learning machine algorithm consists of a collection of regression or classification (decision) trees.

It is a versatile and powerful tool:

GEA, GWAS, Genetic offset / ”vulnerability”

Benefits

Lower risk of overfitting
Flexibility (diversity of data and hypotheses)
Easy to determine important features (genes and environmental variables)

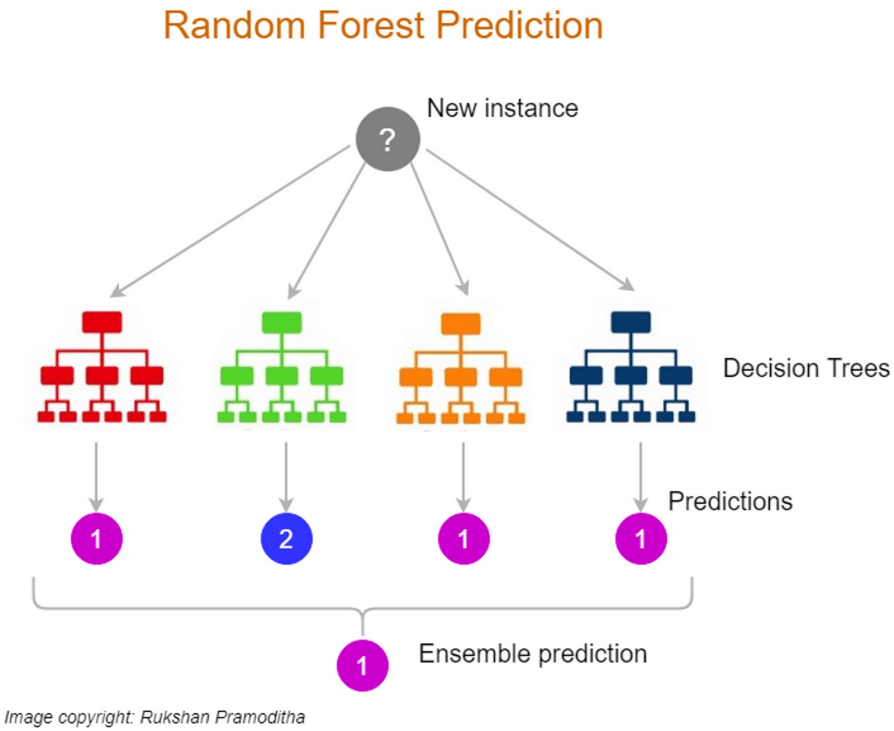


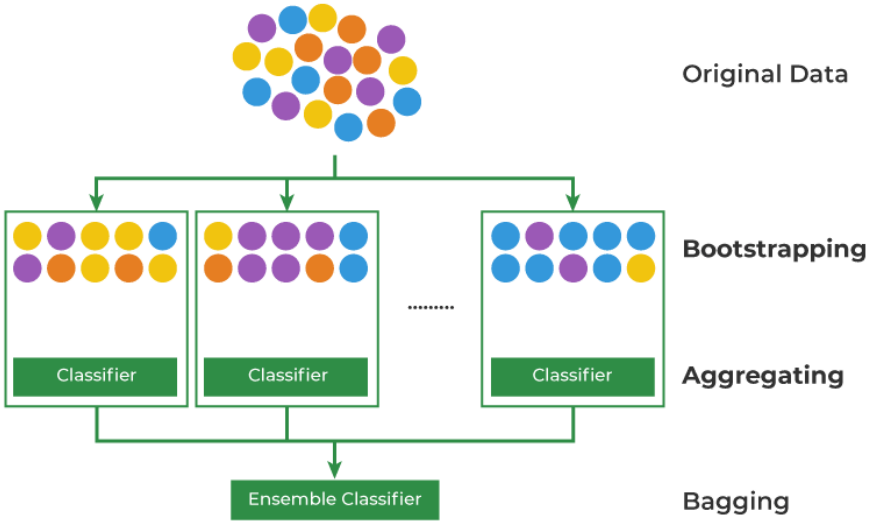
Image copyright: Rukshan Pramoditha

RandomForest

Bagging (bootstrap aggregation)

Training data set replacement 70-80%

Test (Validation) data set 30-20%



https://www.geeksforgeeks.org/

Two main components of the Random Forest algorithm (and other decision tree variants) are selection of features and the tree structure:

Hyperparameters

n_estimators
max_depth
min_samples_split
max_leaf_nodes

number of tress (ntree)
number of features to try (mtry)

↑ performance, accuracy

ntree = 100-500
mtry = total/3 or sqrt(total)

↑ computation resources

RAINBOWR

Linear / logistic regression
Additive, epistatic

Haplotype-based analysis
(without prior information)

Can control for several cofactors
- Relatedness (matric of covariance)
- Pop structure (Q matric from structure)
- Covariate factors, sex, age etc (phonotype table