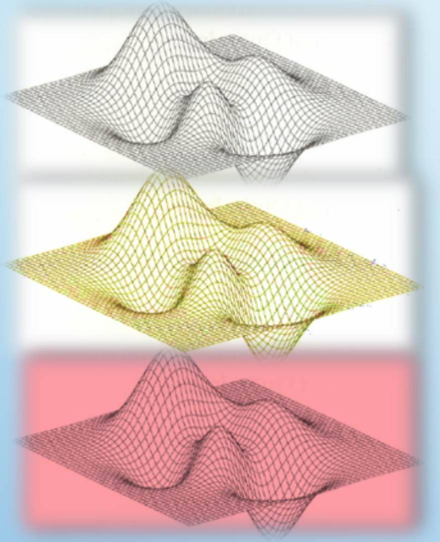
DETECTING SELECTION AND LOCAL ADAPTATION

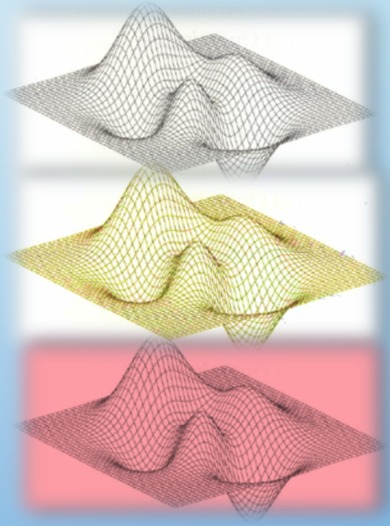


Rachael Dudaniec
Macquarie University
Sydney, Australia

Workshop on Population and Speciation Genomics, Cesky Krumlov, 2020

PART 1 –THEORY AND METHODS

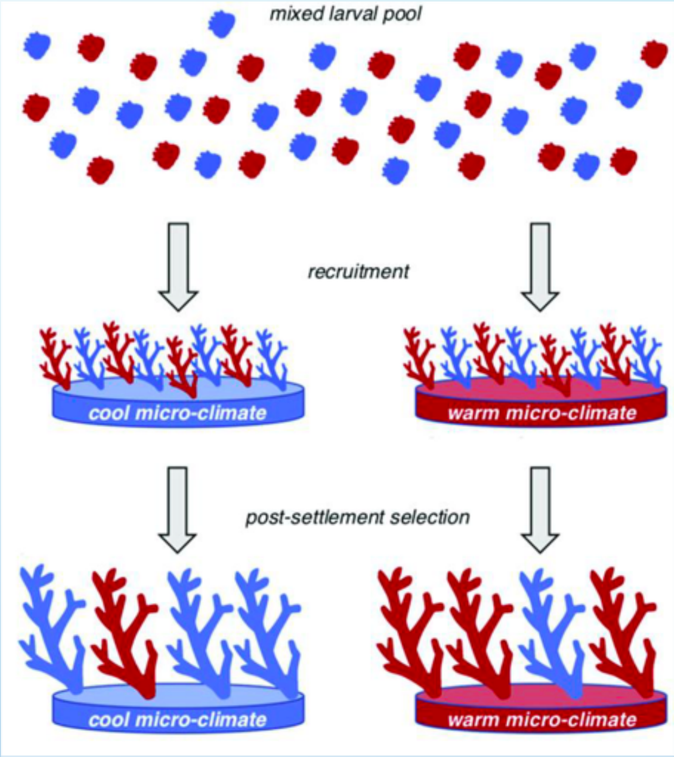
- Local adaptation – what and why?
- Methods of detecting local adaptation
- Differentiation-based analyses (Fst outlier tests)
- Environmental Association Analyses (EAA/GEA/ GxE)
- Parameters and sampling effects



LOCAL ADAPTATION: WHAT AND WHY?

- Local adaptation is the response to **differential selective pressures among populations**, acting on genetically controlled **fitness differences** among individuals.

- Important for the generation and maintenance of biodiversity, species range shifts, the dynamics of species interactions.



Thomas et al. (2018)

LOCAL ADAPTATION: WHAT AND WHY?

- Occurs when selection is **spatially heterogeneous** and strong, relative to other evolutionary forces
- **Genetic drift** reduces additive genetic variance/causes random fixation of genotypes and reduces local adaptation
- **Gene flow** is generally thought to inhibit local adaptation (i.e. via ‘gene swamping’)? But not always…

Adaptive genes may be maintained at intermediate gene flow in temporally variable environments

- selection needs variability to act on

WHY IS LOCAL ADAPTATION IMPORTANT?

Adaptive variation contributes towards mediating the ***vulnerability*** *of a species to extinction*

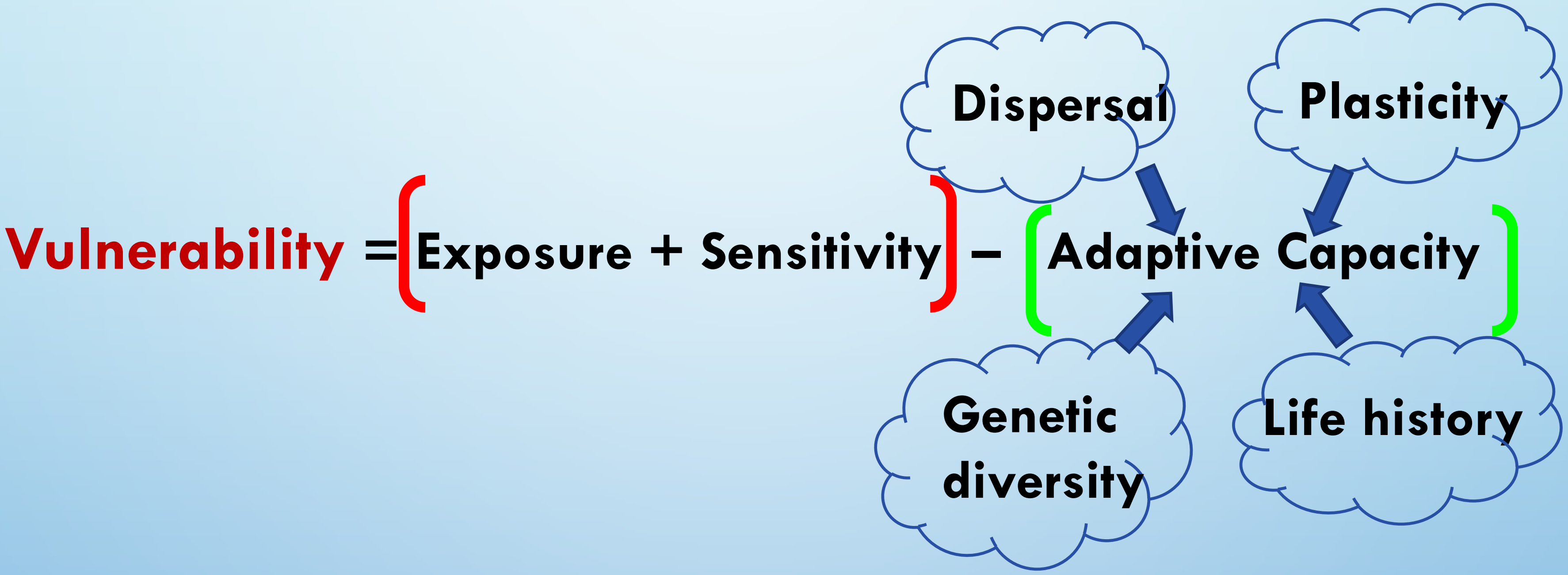
A species' ***vulnerability*** *is a function of its* ***sensitivity*** *and* ***exposure*** *to fitness stressors, and is mediated by the* ***adaptive potential*** *of the species (both ecological and evolutionary), and the* ***capacity for conservation management***
(Williams et al. 2008)

WHY IS LOCAL ADAPTATION IMPORTANT?

$$\text{Vulnerability} = \left[\text{Exposure} + \text{Sensitivity}\right] - \left[\text{Adaptive Capacity}\right]$$

Williams et al. 2008

WHY IS LOCAL ADAPTATION IMPORTANT?



Vulnerability = [Exposure + Sensitivity] – [Adaptive Capacity]

Dispersal

Plasticity

Genetic diversity

Life history

Williams et al. 2008

TYPES OF RESEARCH QUESTIONS

Study-system specific

- Will species X adapt during range expansion/ invasion?
- Are hybrids of X+ X more locally adapted than parents or visa versa?
- Does declining species X have ‘enough’ adaptive variation to survive climate change?

TYPES OF RESEARCH QUESTIONS

Study-system specific

- Will species X adapt during range expansion/ invasion?
- Are hybrids of X+X more locally adapted than parents or visa versa?
- Does declining species X have ‘enough’ adaptive variation to survive climate change?

Theoretical

- Does high gene flow limit local adaptation?
- Does local adaptation act in parallel across species or environments?
- Are there common patterns of local adaptation across species with respect to demography, traits, or evolutionary history?

KNOWLEDGE OF ADAPTIVE CAPACITY CAN INFLUENCE SPECIES' MANAGEMENT

PNAS

Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections

Orly Razgour[a,b,1], Brenna Forester[c], John B. Taggart[d], Michaël Bekaert[d], Javier Juste[e], Carlos Ibáñez[e], Sébastien J. Puechmaille[f,g,h], Roberto Novella-Fernandez[a], Antton Alberdi[i], and Stéphanie Manel[j]

[a]Biological Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom; [b]School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdo
Stirling FK9 4LA, Uni
des Sciences de l'Evo

Evolutionary Applications

Open Access

Evolutionary approaches to environmental, biomedical and socio-economic issues

REVIEW AND SYNTHESES | Open Access |

Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation

Sarah P. Flanagan, Brenna R. Forester, Emily K. Latch, Sally N. Aitken, Sean Hoban

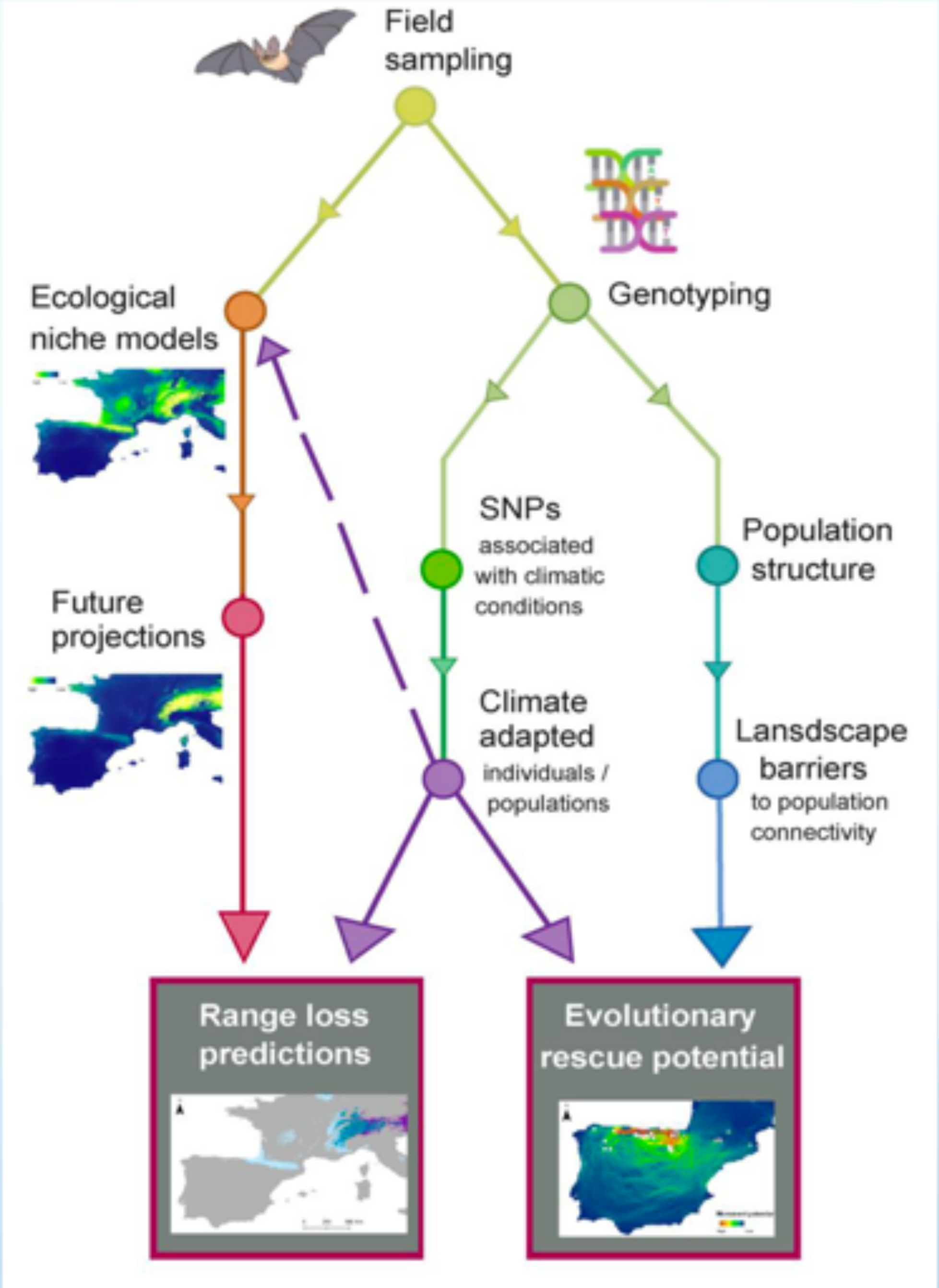
Conservation Biology

Essay | Full Access |

Adaptive introgression as a resource for management and genetic conservation in a changing climate

Jill A. Hamilton, Joshua M. Miller

INTEGRATING ADAPTIVE GENETIC VARIATION



1. Collect field and genomic data

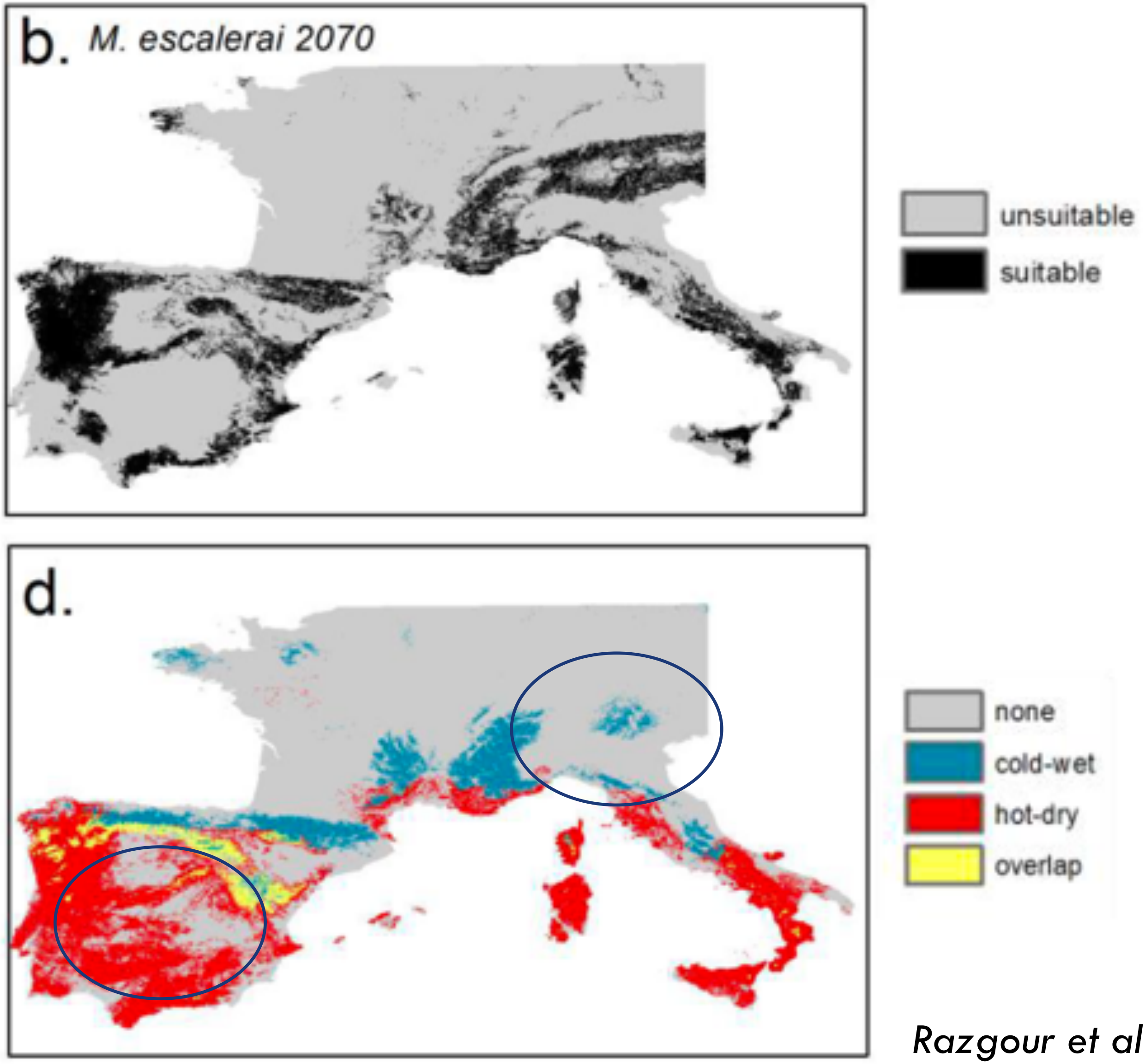
2. Use niche models to assess vulnerability and future ranges

3. Use SNP candidates to project species ranges and evolutionary potential

4. Complement with neutral landscape genetic connectivity

Razgour et al. 2019

SDM
informed by
adaptive
genetic
variation



Razgour et al. 2019

Future losses were overestimated when adaptive variability was NOT accounted for

LOCAL ADAPTATION: HOW TO MEASURE?

Population genomics/
Fst outlier tests?

Quantitative Genetics?

Fitness experiments?

?

Environmental Association Analysis (EAA)?

Genome Wide Association Studies?

Common Gardens/
Reciprocal transplants?

LOCAL ADAPTATION: HOW TO MEASURE?

Population genomics/
Fst outlier tests?

Quantitative Genetics?

Fitness experiments?

There is no consensus on the best way to measure it!

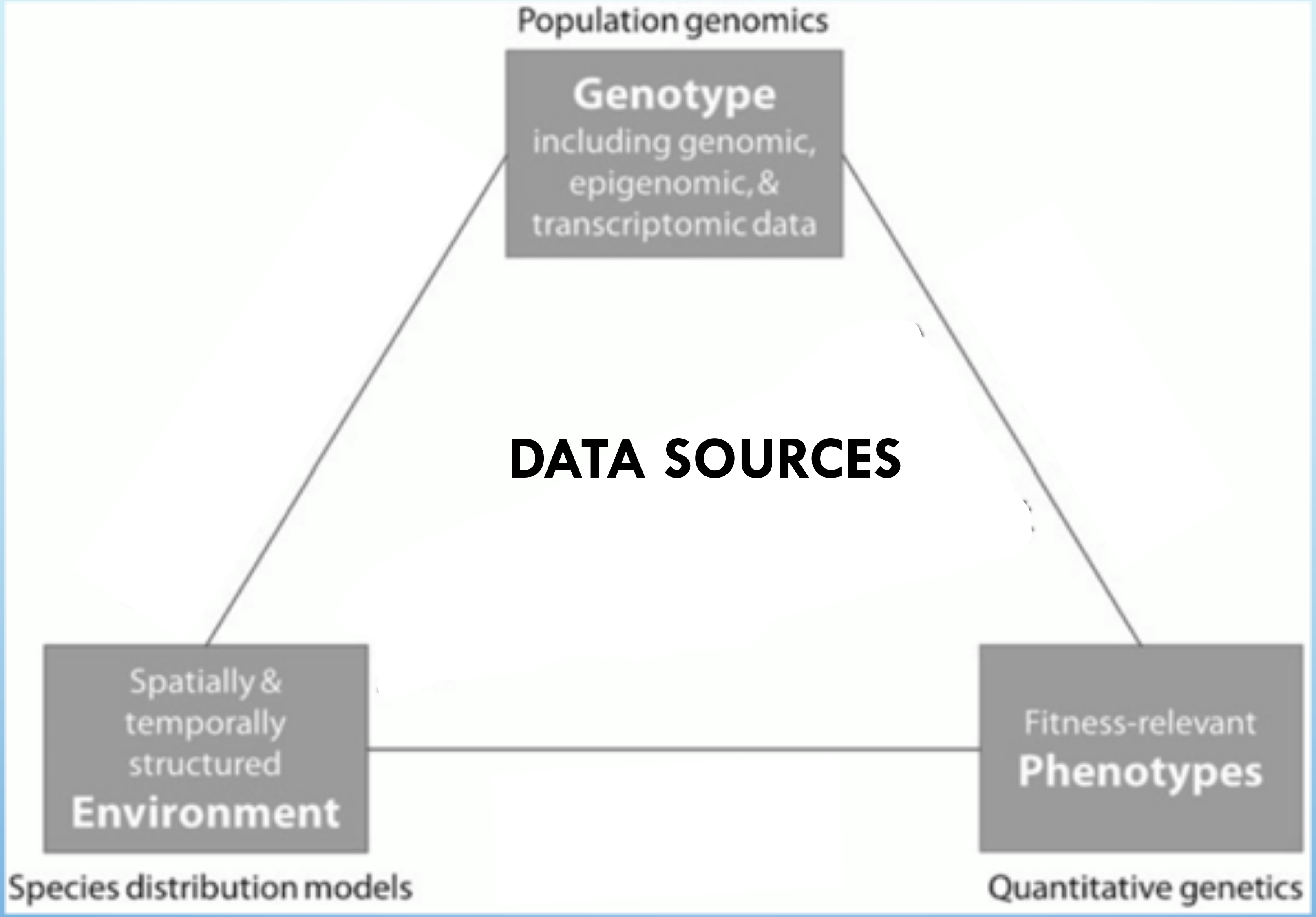
An appropriate combination is recommended☺

Environmental Association Analysis (EAA)?

Genome Wide Association Studies?

Common Gardens/
Reciprocal transplants?

HOW ARE WE DETECTING LOCAL ADAPTATION?



Population genomics

Genotype
including genomic, epigenomic, & transcriptomic data

DATA SOURCES

Spatially & temporally structured
Environment

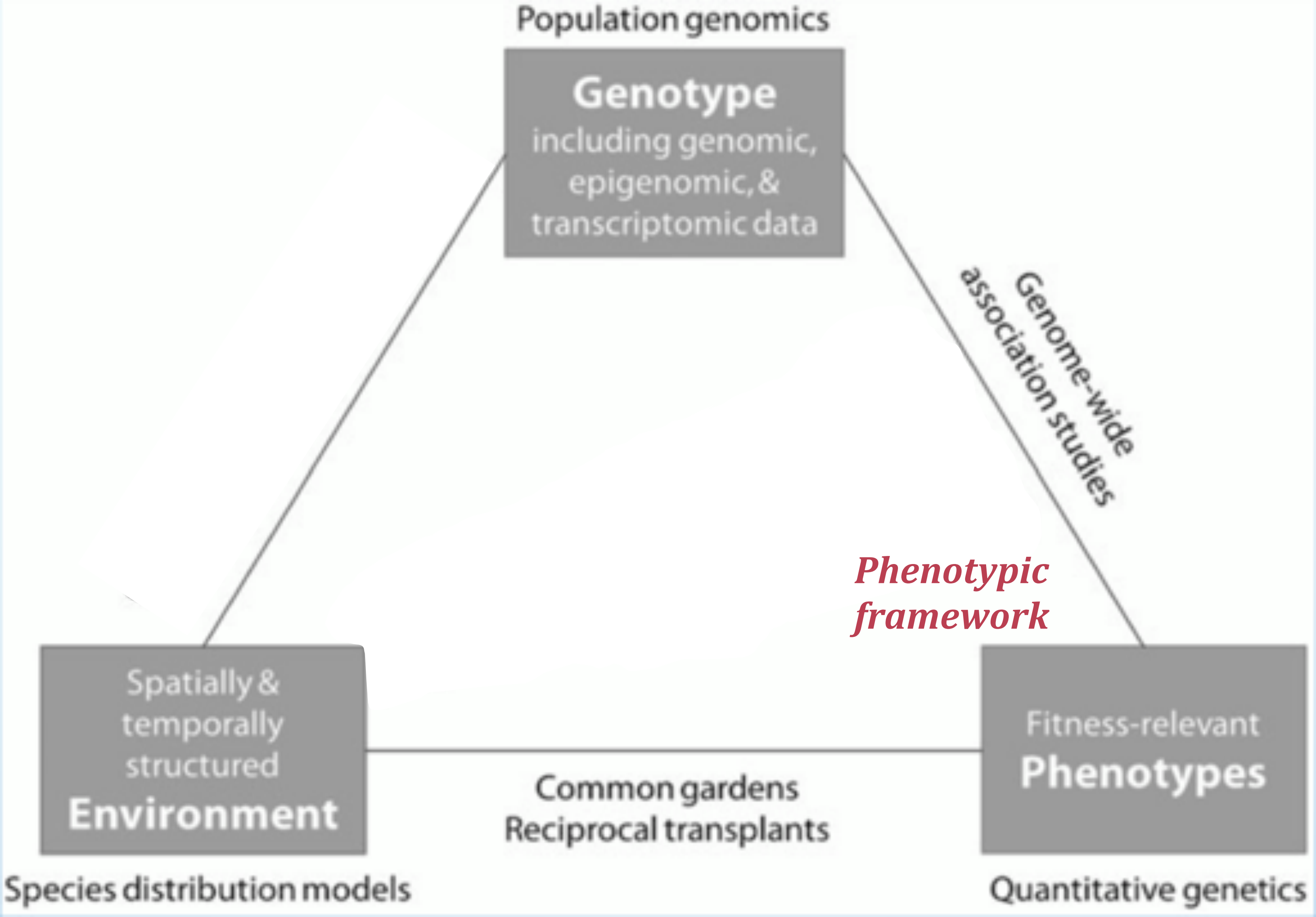
Species distribution models

Fitness-relevant
Phenotypes

Quantitative genetics

Modified from Forester et al. (2018)

HOW ARE WE DETECTING LOCAL ADAPTATION?



Population genomics

Genotype
including genomic,
epigenomic, &
transcriptomic data

Genome-wide
association studies

Phenotypic framework

Spatially &
temporally
structured
Environment

Common gardens
Reciprocal transplants

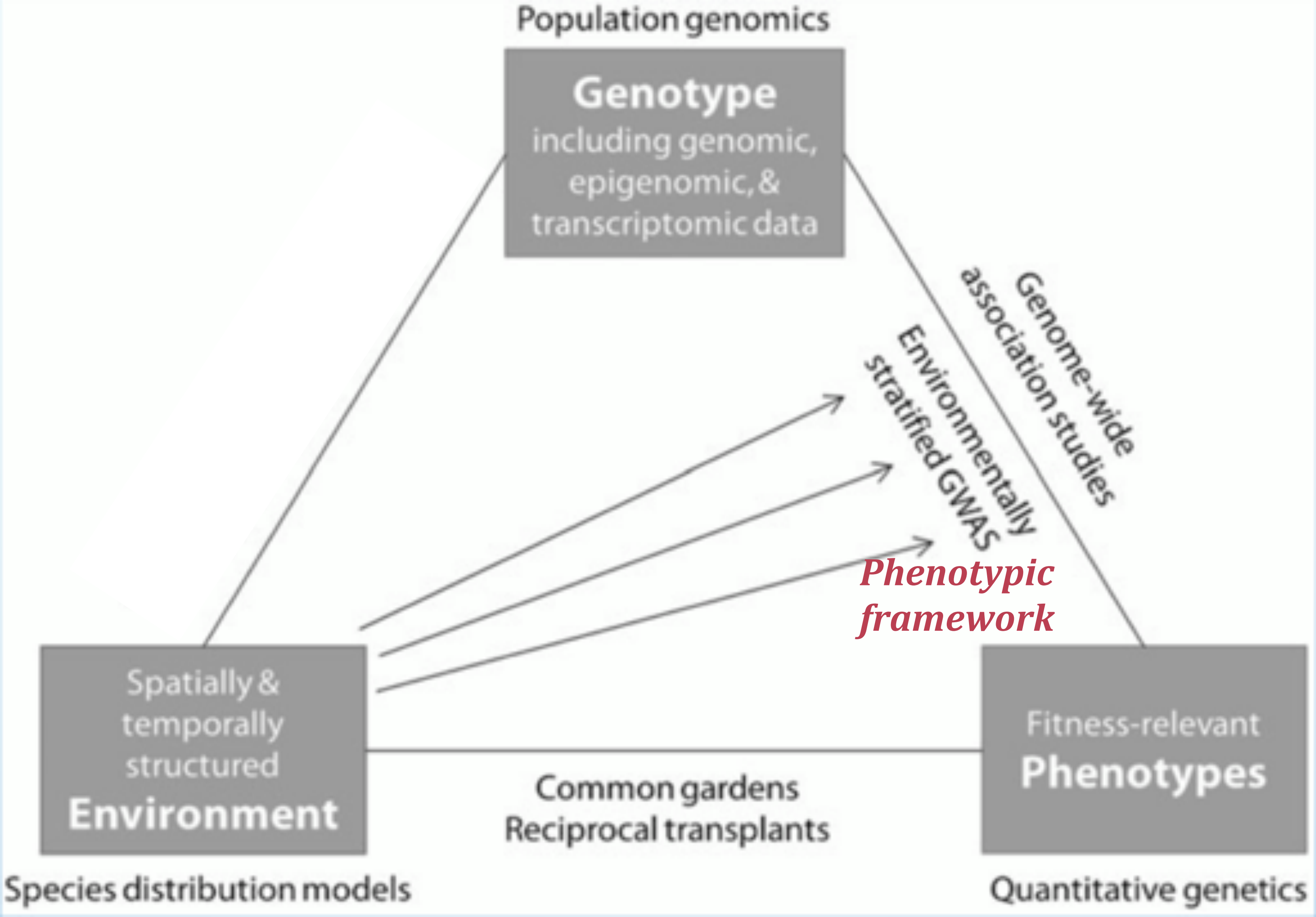
Fitness-relevant
Phenotypes

Species distribution models

Quantitative genetics

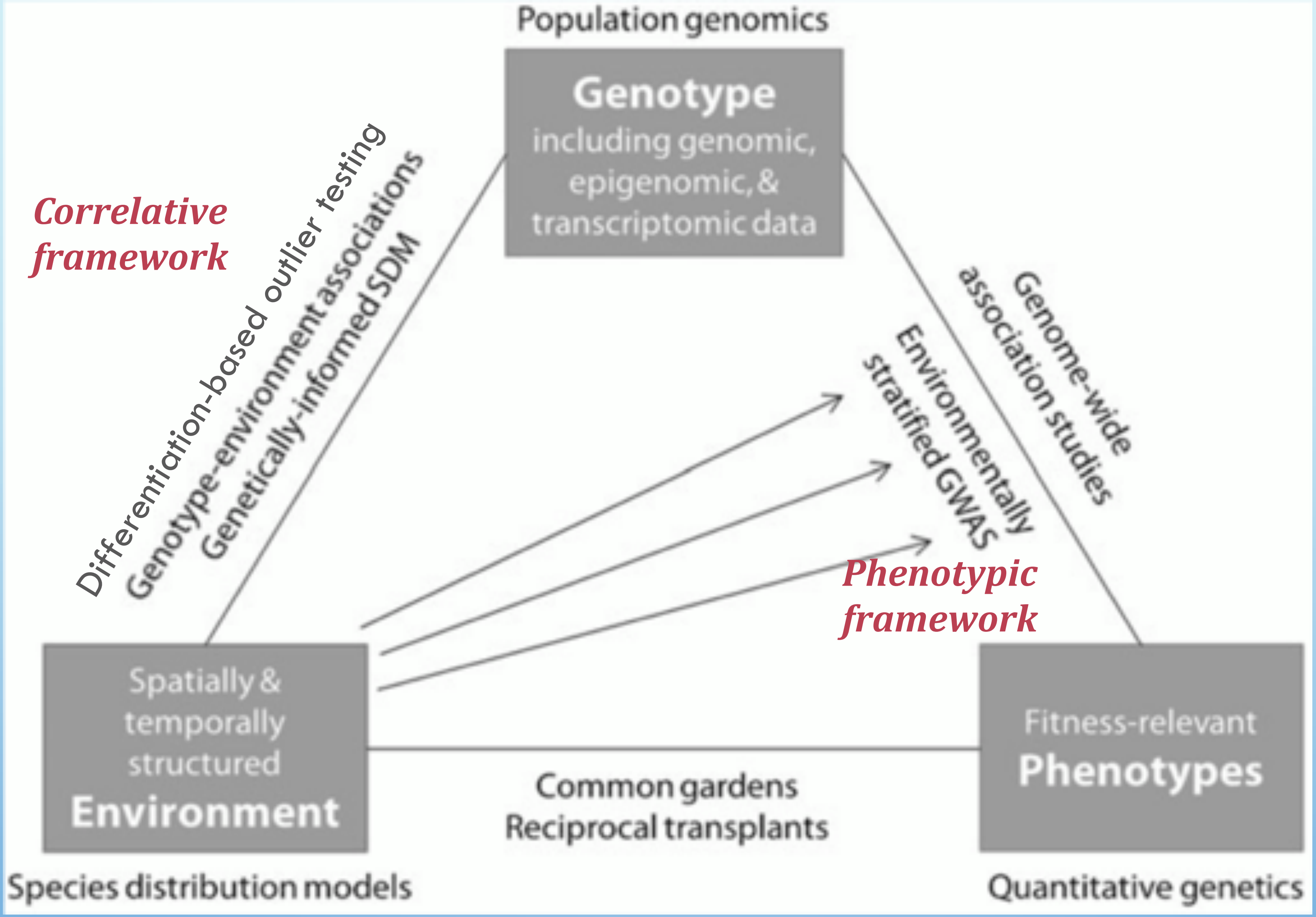
Modified from Forester et al. (2018)

HOW ARE WE DETECTING LOCAL ADAPTATION?



Modified from Forester et al. (2018)

HOW ARE WE DETECTING LOCAL ADAPTATION?



Population genomics

Genotype
including genomic, epigenomic, & transcriptomic data

Correlative framework

Differentiation-based outlier testing
Genotype-environment associations
Genetically-informed SDM

Genome-wide association studies

Environmentally stratified GWAS

Phenotypic framework

Spatially & temporally structured
Environment

Species distribution models

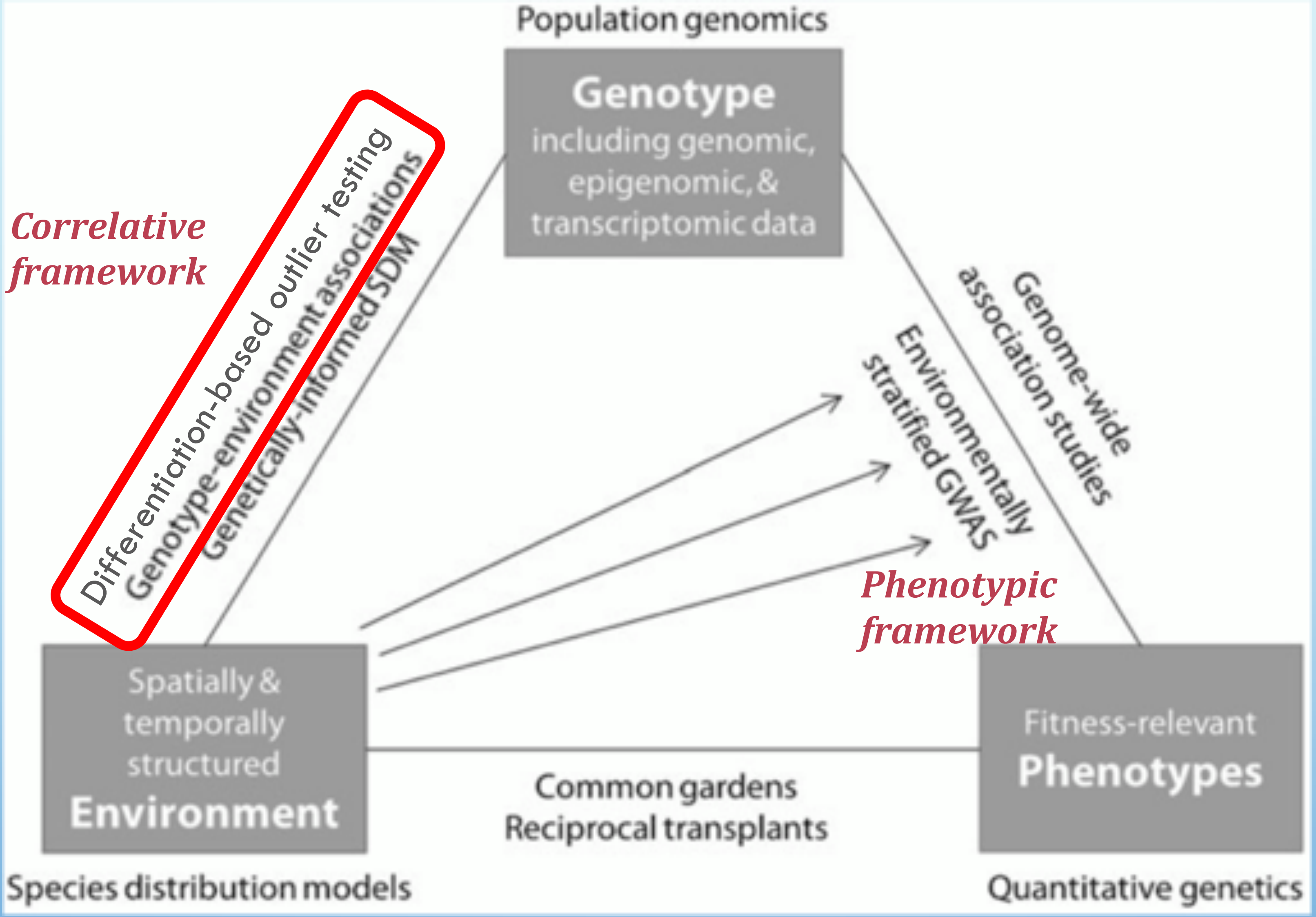
Fitness-relevant
Phenotypes

Quantitative genetics

Common gardens
Reciprocal transplants

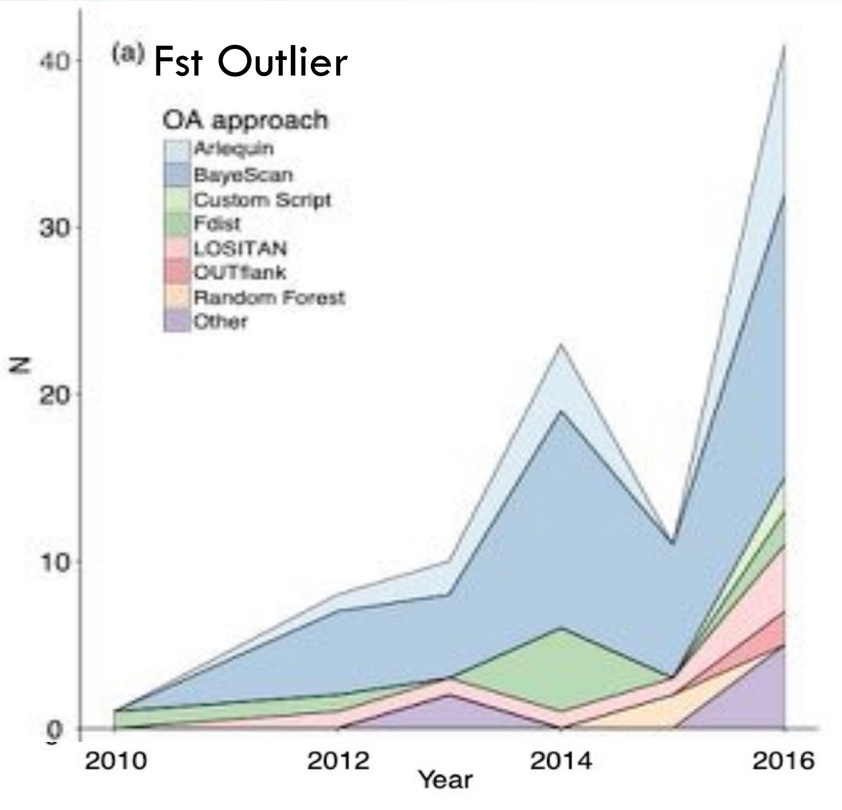
Modified from Forester et al. (2018)

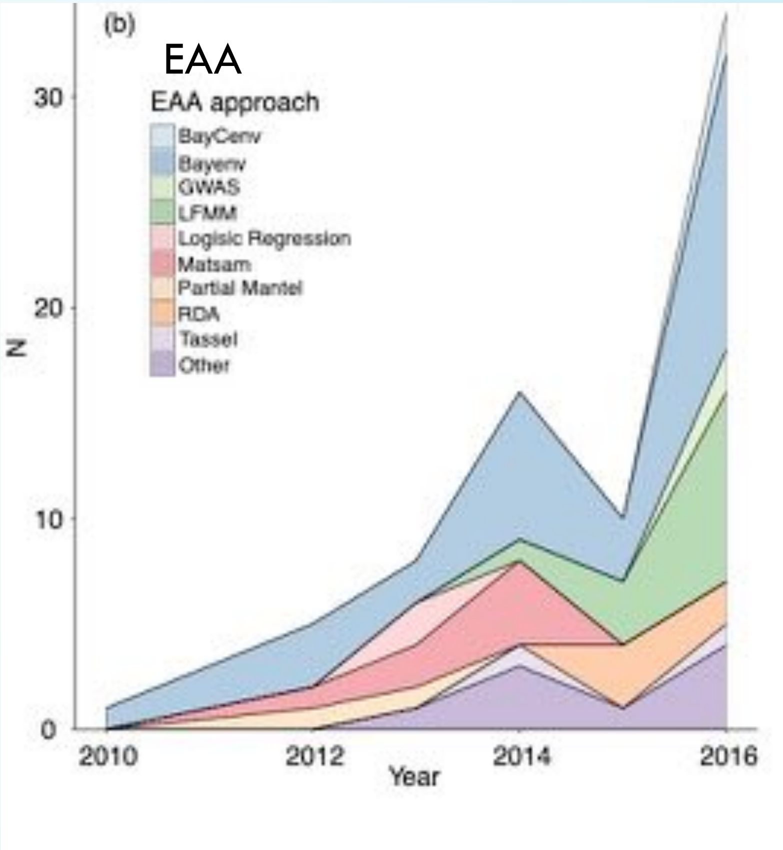
HOW ARE WE DETECTING LOCAL ADAPTATION?



Modified from Forester et al. (2018)

USE OF FST OUTLIER TESTS AND EAA IS INCREASING





- **N = 66 studies (2010-2016) applying EAA or Fst outlier tests**

INVITED REVIEWS AND SYNTHESES

WILEY MOLECULAR ECOLOGY

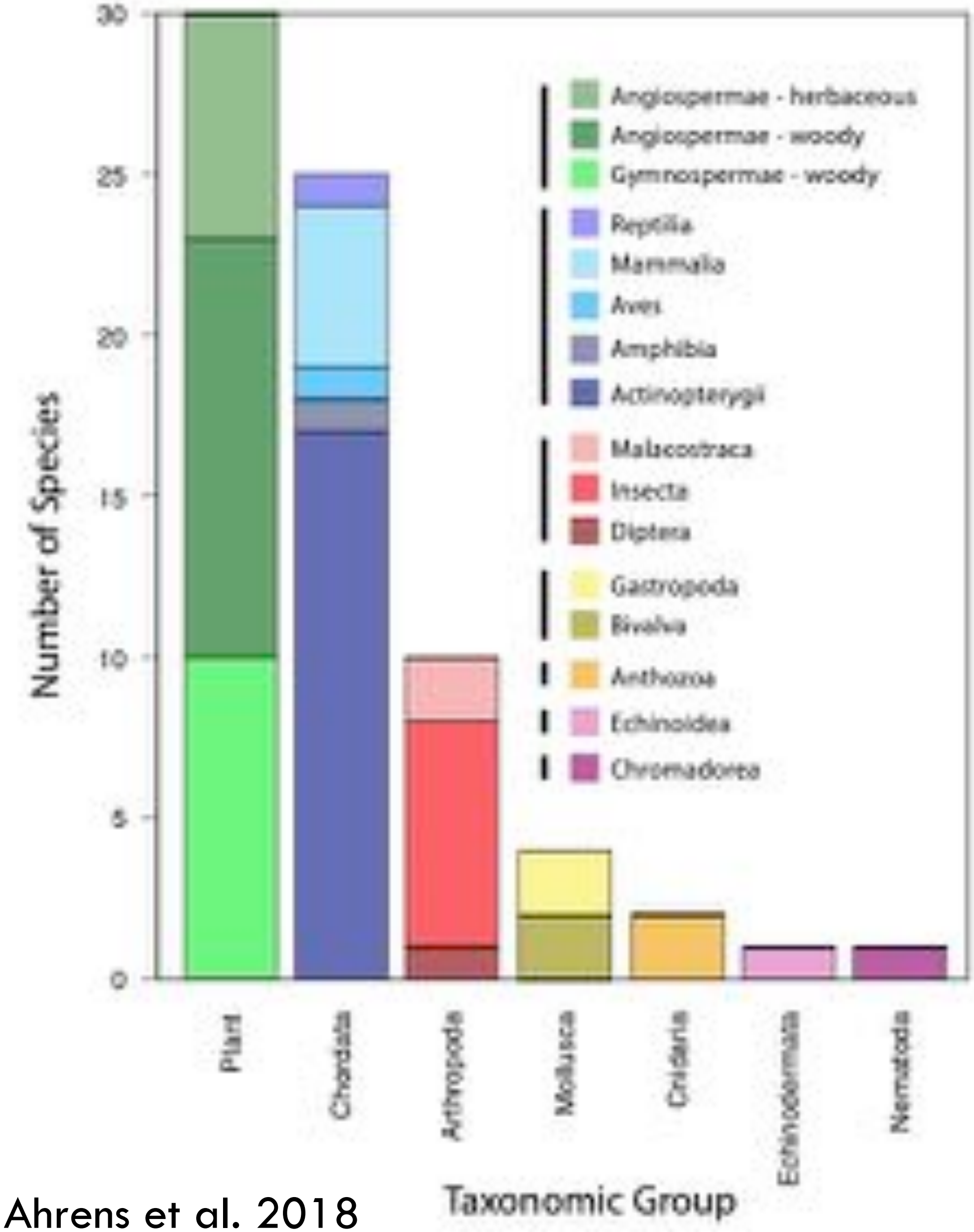
The search for loci under selection: trends, biases and progress

Collin W. Ahrens[1] | Paul D. Rymer[1] | Adam Stow[2] | Jason Bragg[3] | Shannon Dillon[4] | Kate D. L. Umbers[1,5] | Rachael Y. Dudaniec[2]

[1]Hawkesbury Institute for the Environment

Ahrens et al. 2018

MANY TAXONOMIC GROUPS REPRESENTED



Ahrens et al. 2018

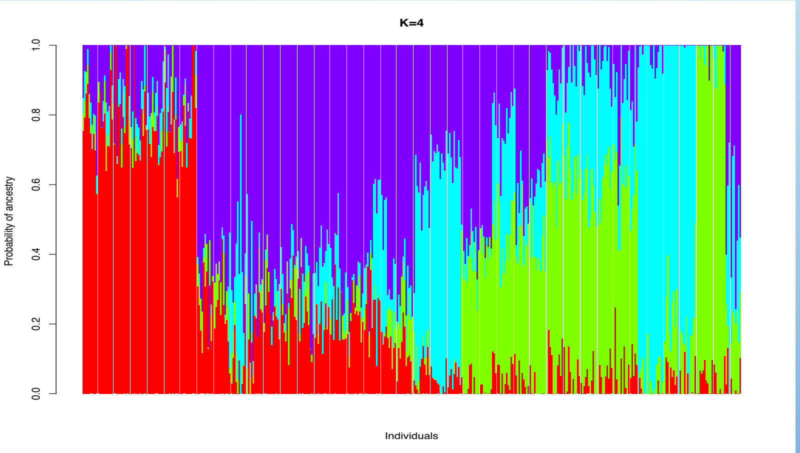
- N = 66 studies (2010-2016) applying EAA or Fst outlier tests

- Mostly trees, plants, then fish, mammals and insects

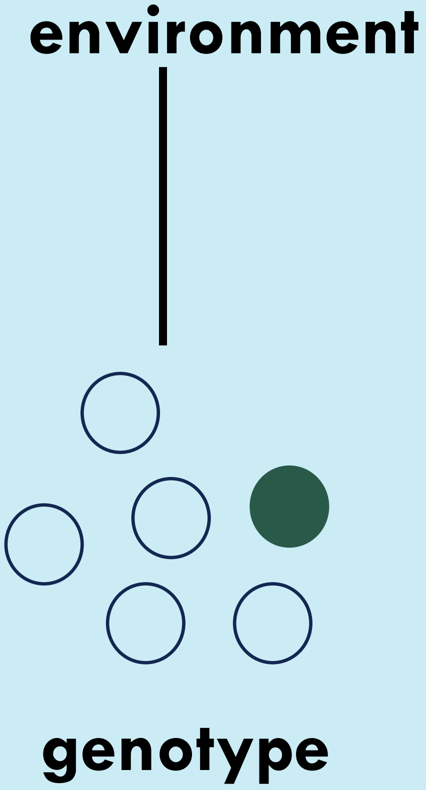
THE ISSUE OF GENETIC STRUCTURE

When pairwise Fst is not identical between populations we can get an excess of false positives

- Genetic drift in small, patchy populations = false positives
- When environment tracks genetic structure = false negatives
- Introgression, hybridization = false positives

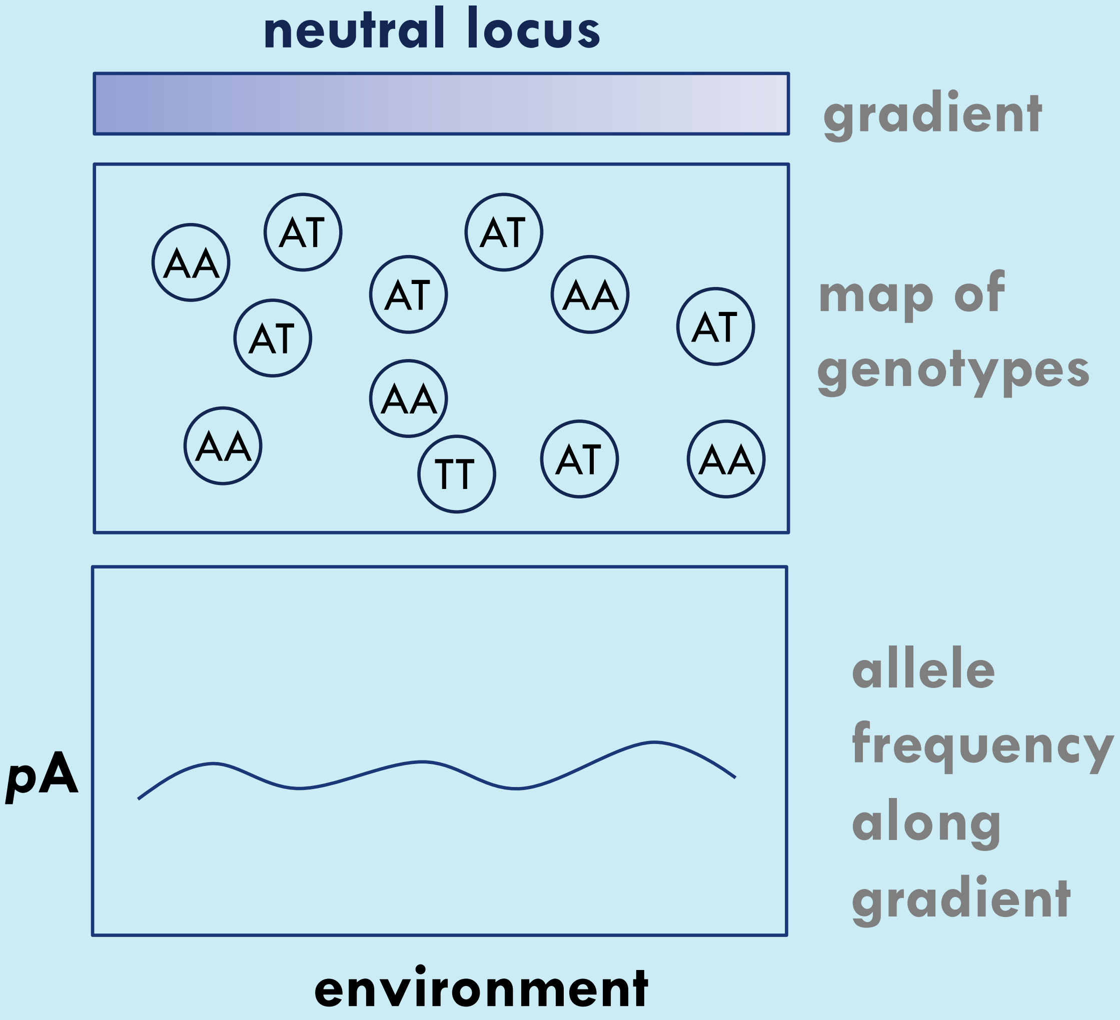


THE ISSUE OF GENETIC STRUCTURE



environment

genotype



neutral locus

gradient

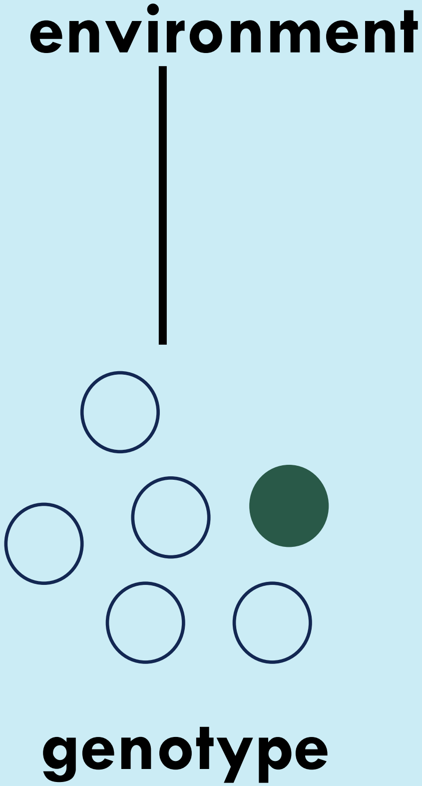
map of genotypes

allele frequency along gradient

pA

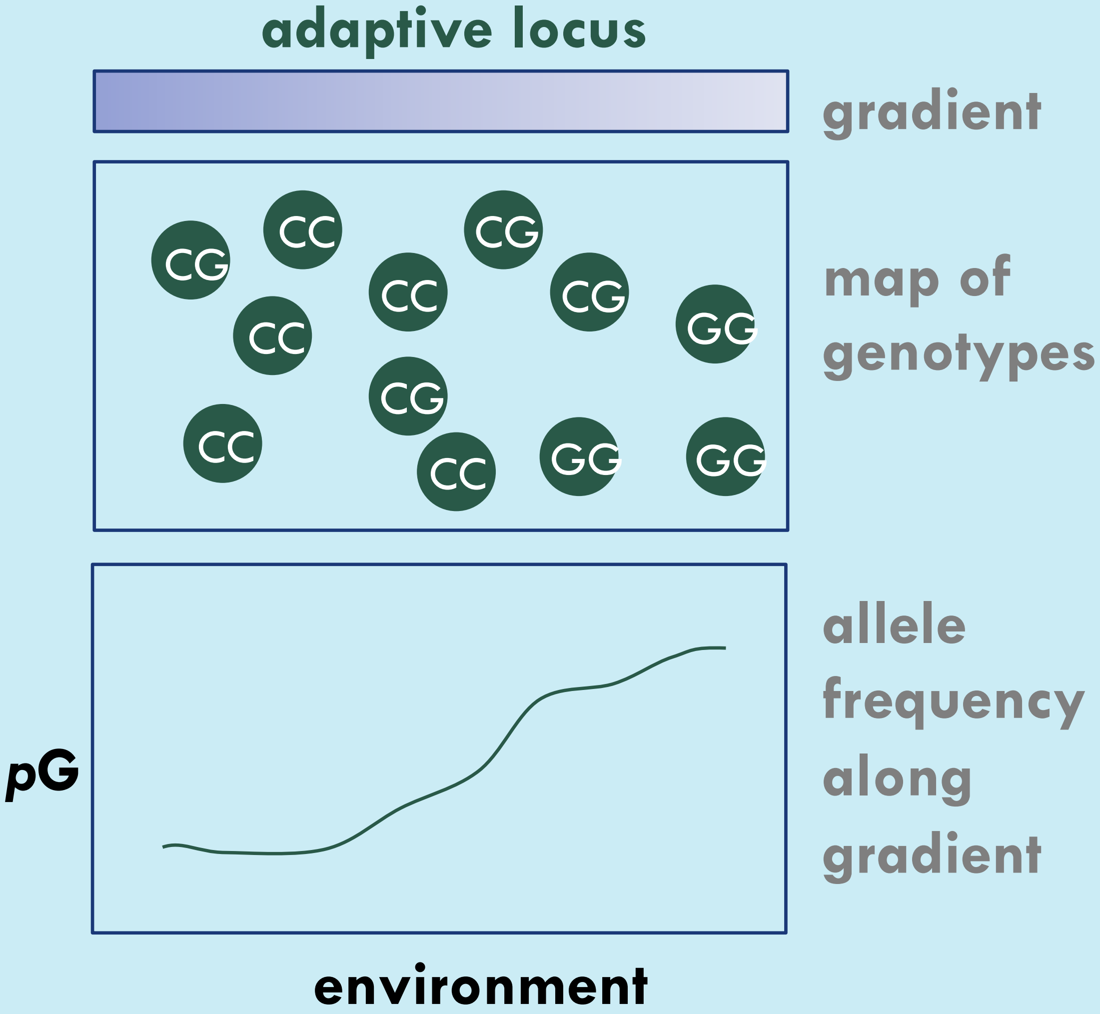
environment

THE ISSUE OF GENETIC STRUCTURE



environment

genotype



adaptive locus

gradient

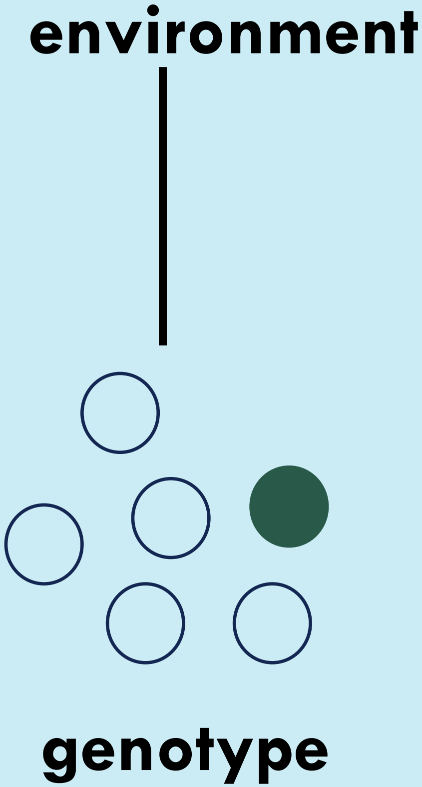
map of genotypes

allele frequency along gradient

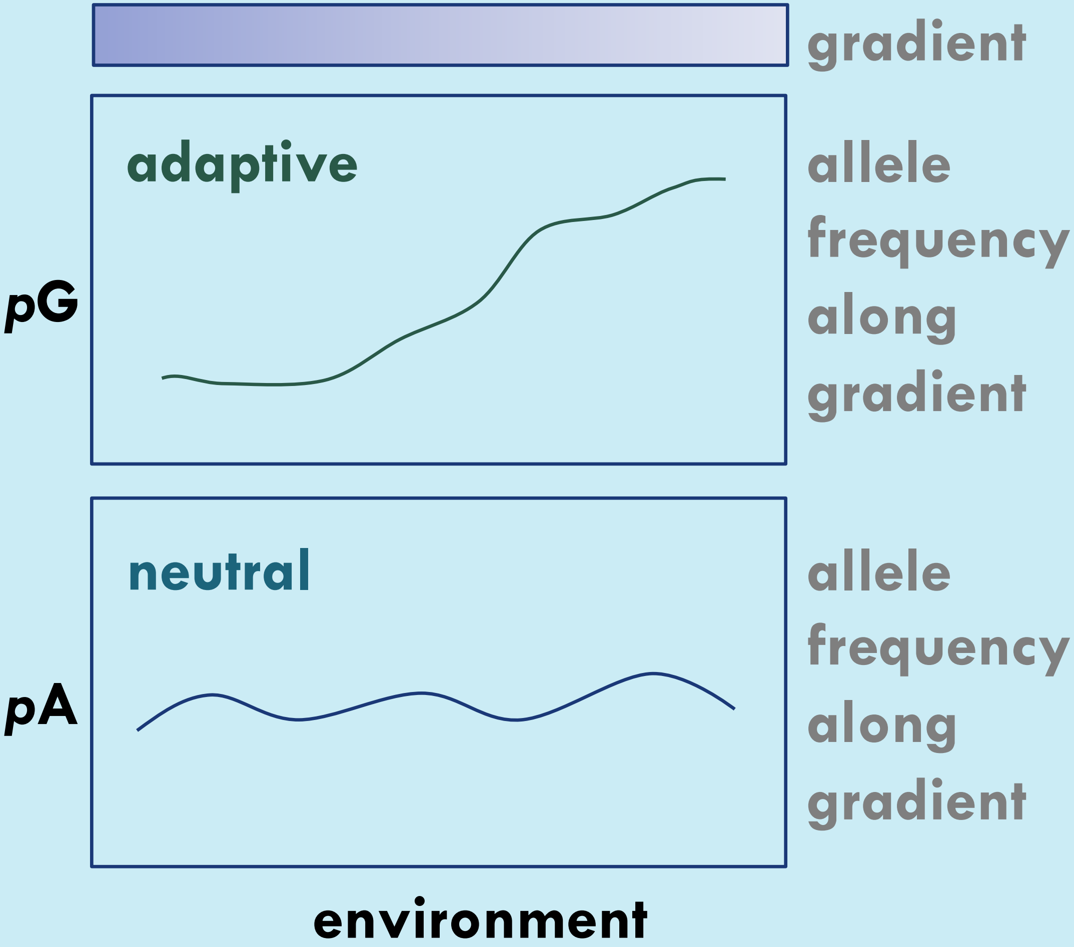
pG

environment

THE ISSUE OF GENETIC STRUCTURE


environment

genotype


gradient

adaptive

pG

allele frequency along gradient

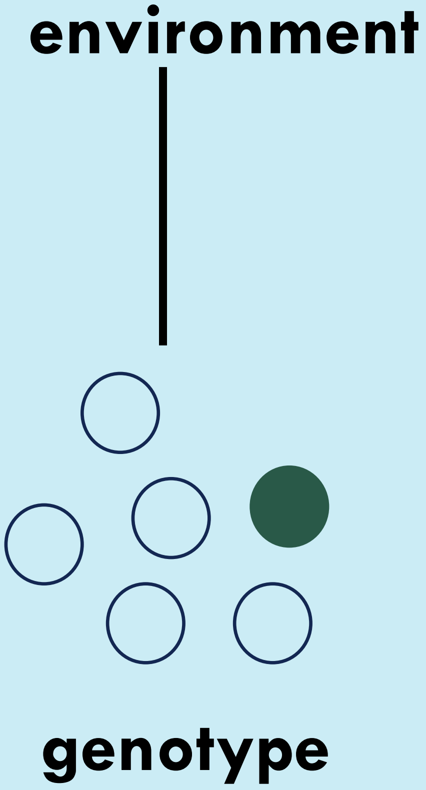
neutral

pA

allele frequency along gradient

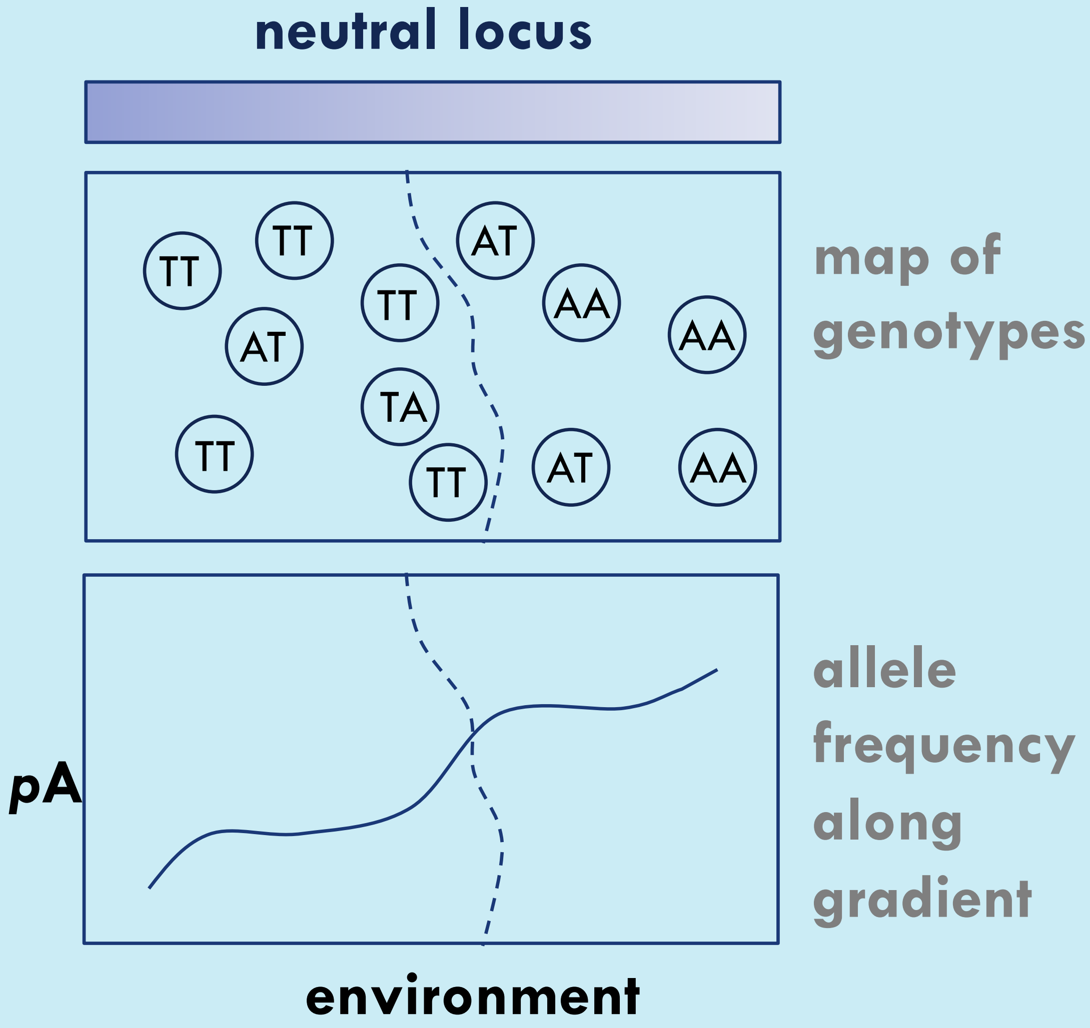
environment

THE ISSUE OF GENETIC STRUCTURE



environment

genotype



neutral locus

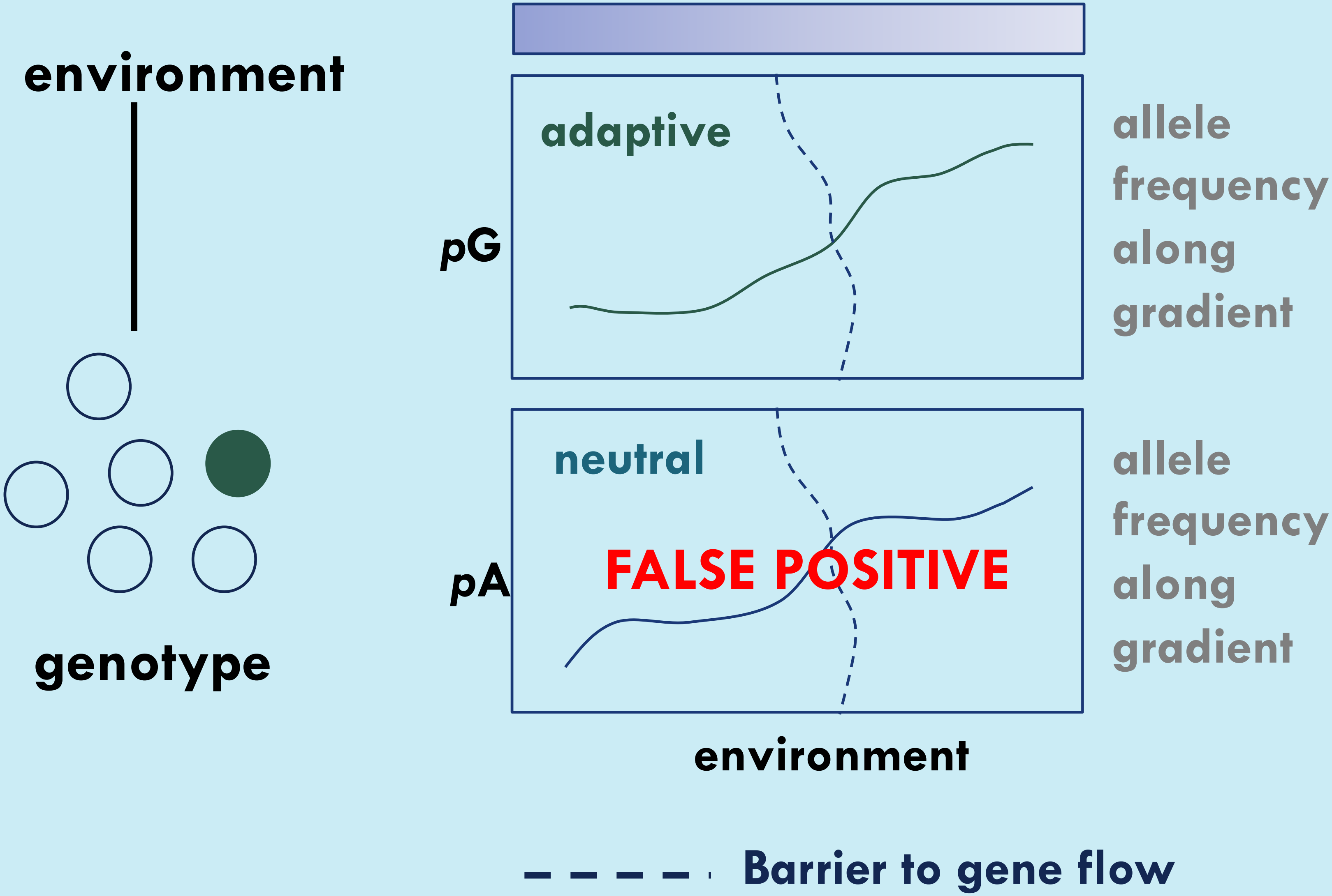
map of genotypes

allele frequency along gradient

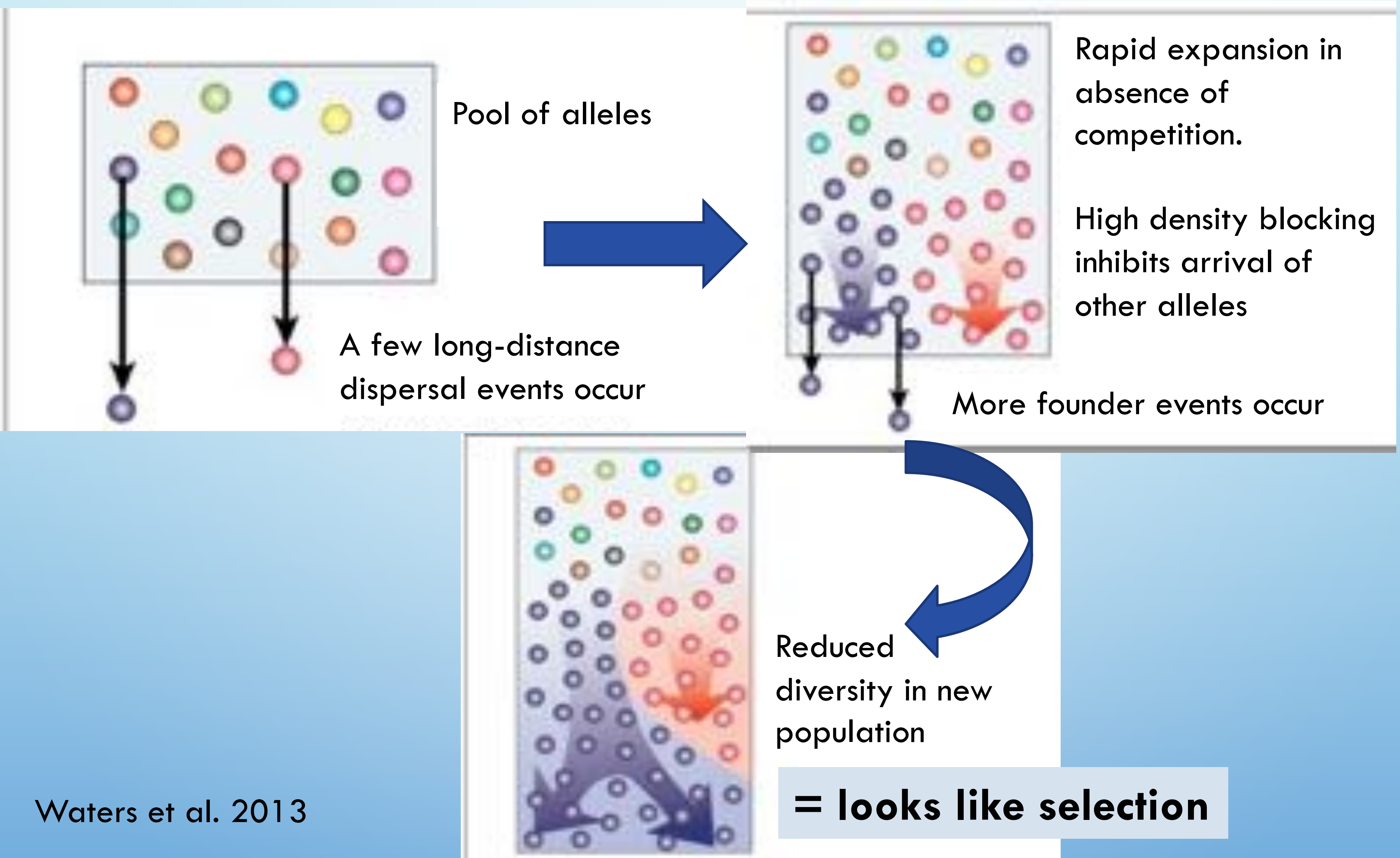
environment

- - - - **Barrier to gene flow**

THE ISSUE OF GENETIC STRUCTURE



EFFECTS OF RANGE EXPANSION: ALLELE SURFING/ FALSE POSITIVES



Pool of alleles

A few long-distance dispersal events occur

Rapid expansion in absence of competition.

High density blocking inhibits arrival of other alleles

More founder events occur

Reduced diversity in new population

= looks like selection

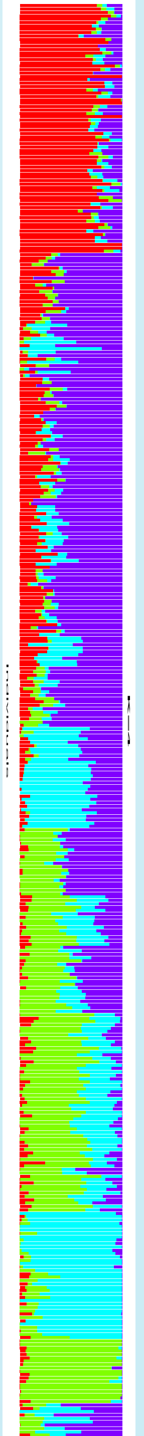
Waters et al. 2013

HOW DO WE ACCOUNT FOR GENETIC STRUCTURE?

- Genetic clustering (LFMM, OutFlank)
- Kinship matrix (BayeEnv)
- Spatial predictors (spatial autocorrelation, GWR)
- PCA (RDA, pcadapt- principal coordinates)
- Generalised Dissimilarity Modelling
- Simulation of neutral data under demographic model

  (Harris & Munshi-South 2016)

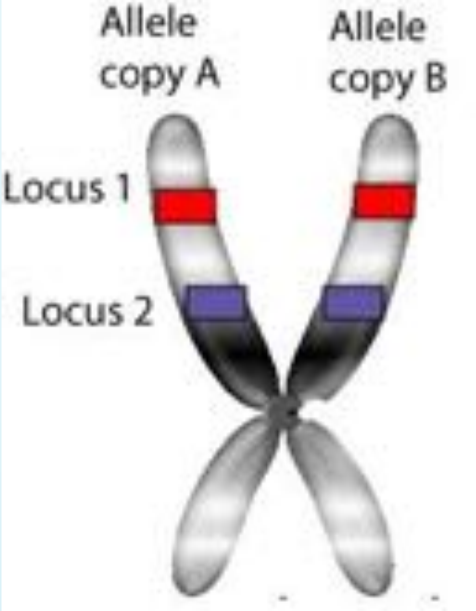
Accounting for neutral population structure can reduce false positive rates…but also reduce power.



DEALING WITH LINKAGE DISEQUILIBRIUM

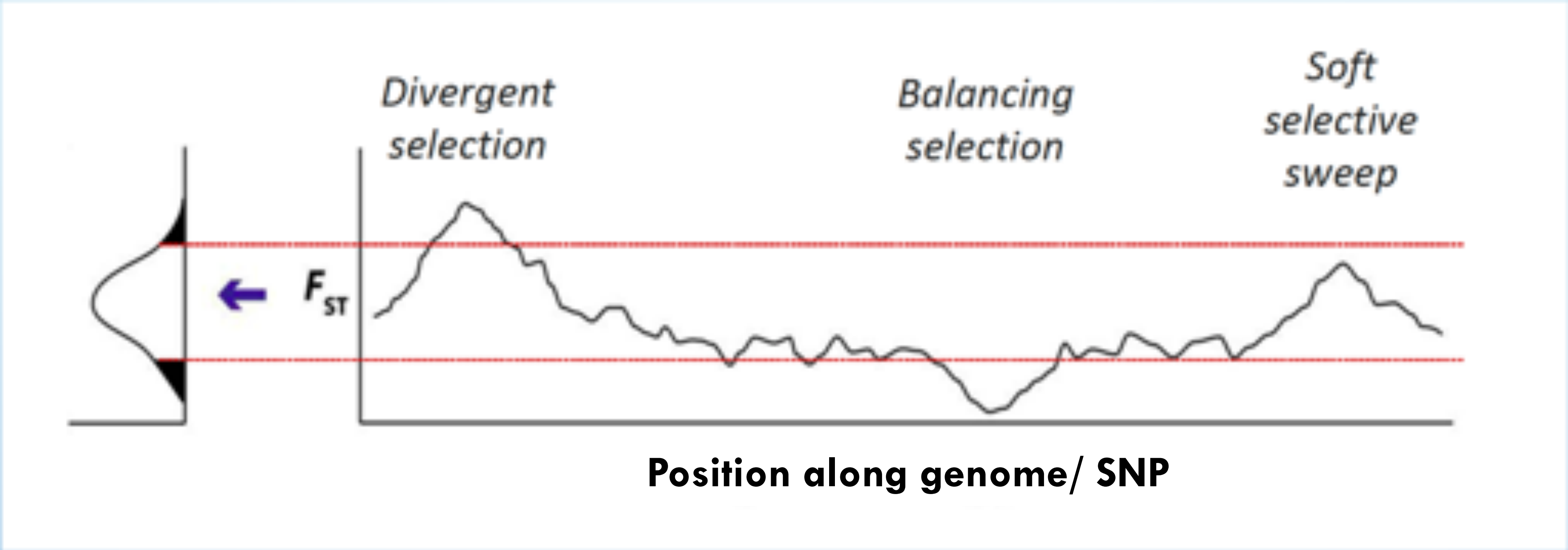
The nonrandom association of alleles at two or more loci

- LD can inform collinearity between SNPs, or create inflated numbers of correlated outliers
- You may filter SNP data sets based on LD summary statistics, and/or using a single SNP per RAD tag
- Use unlinked dataset for calculating neutral population structure to avoid bias



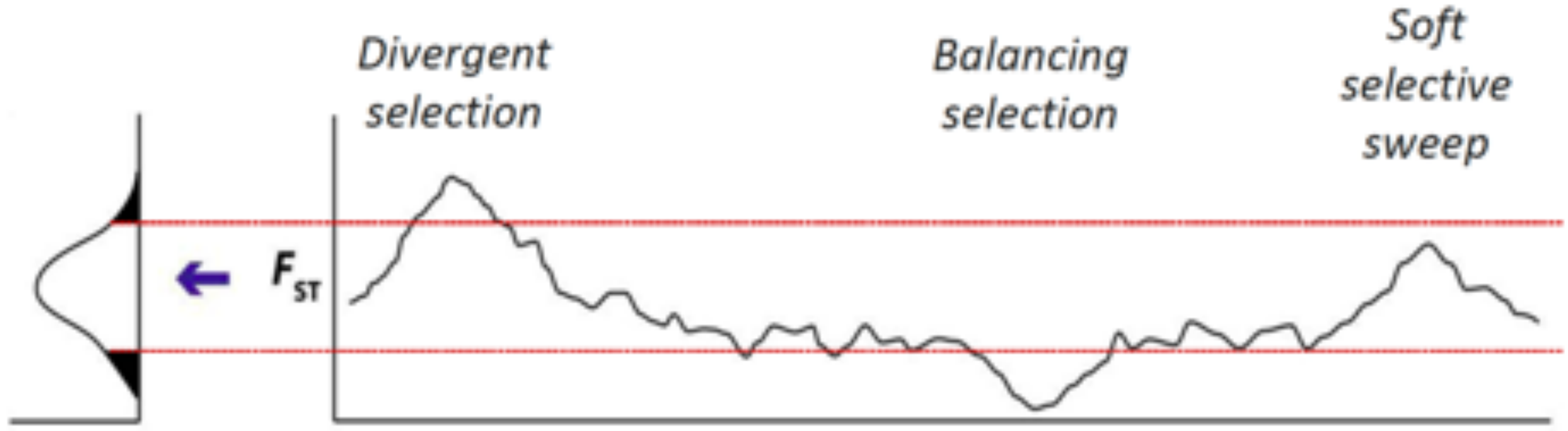
DIFFERENTIATON-BASED (F_{ST}) OUTLIER TESTS

- Identifies loci with higher or lower Fst than expected from the Fst distribution expected under neutrality – usually based on X^2 approximations of Fst
- Fst values on the tail ends are identified as outliers
- No environmental or geographic data needed – but need hypotheses!



DIFFERENTIATON-BASED (F_{ST}) OUTLIER TESTS

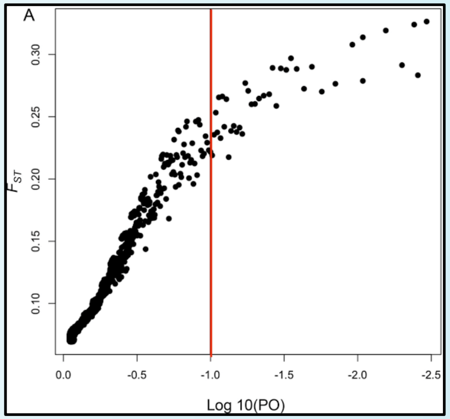
- Good when environmental effects are unknown, or data deficient
- Risk missing important environmental drivers of selection
- Best for detecting large effect loci (less sensitive to small effect)
- Best for divergent, positive selection – negative less clear
- Works well under IBD (Lotterhos & Whitlock 2015)



Position along genome/ SNP

DIFFERENTIATON-BASED (F_{ST}) OUTLIER TESTS

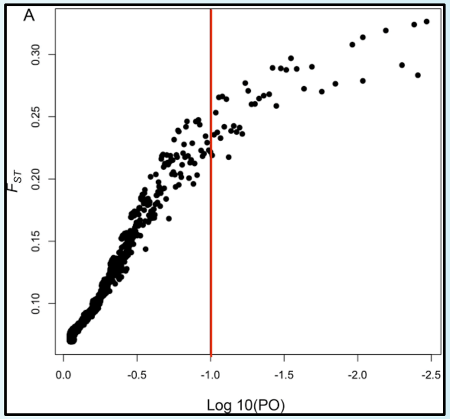
- **Bayescan** – Bayesian method to estimate the relative probability that each locus is under selection (Foll and Gaggiotti 2008) – *high false positive rate*



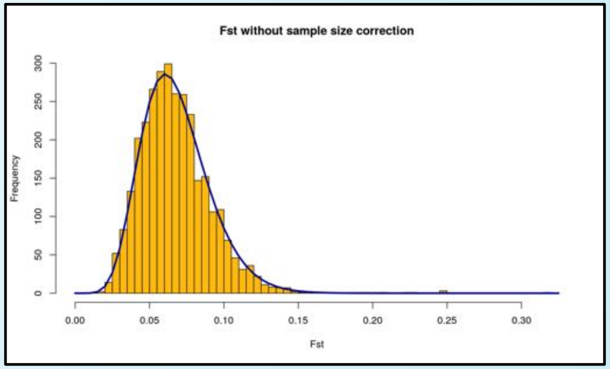
See Hoban et al. 2016 for more

DIFFERENTIATON-BASED (F_{ST}) OUTLIER TESTS

- **Bayescan** – Bayesian method to estimate the relative probability that each locus is under selection (Foll and Gaggiotti 2008) – *high false positive rate*



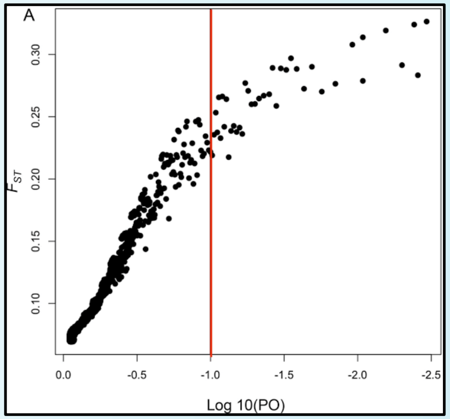
- **OutFLANK**– calculates null Fst distribution to separate diversifying SNPs with high Fst from neutral SNPs (Whitlock & Lotterhos 2015) - *conservative*



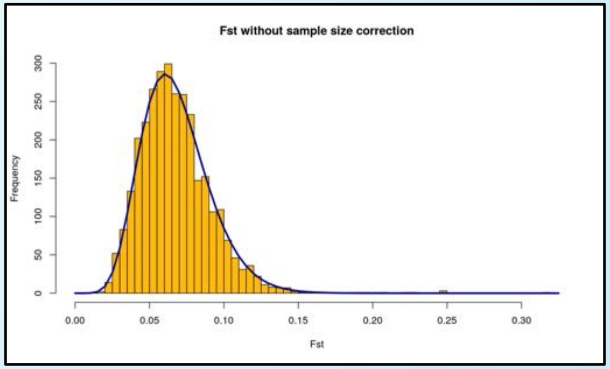
See Hoban et al. 2016 for more

DIFFERENTIATON-BASED (F_{ST}) OUTLIER TESTS

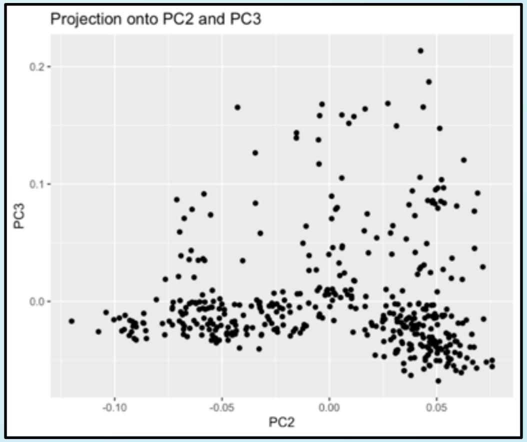
- **Bayescan** – Bayesian method to estimate the relative probability that each locus is under selection (Foll and Gaggiotti 2008) – *high false positive rate*



- **OutFLANK**– calculates null Fst distribution to separate diversifying SNPs with high Fst from neutral SNPs (Whitlock & Lotterhos 2015) - *conservative*



- **pcadapt** –principal components analysis. Tests correlation between genetic variation and principal components. (Luu et al. 2017)- *performs well with admixture*



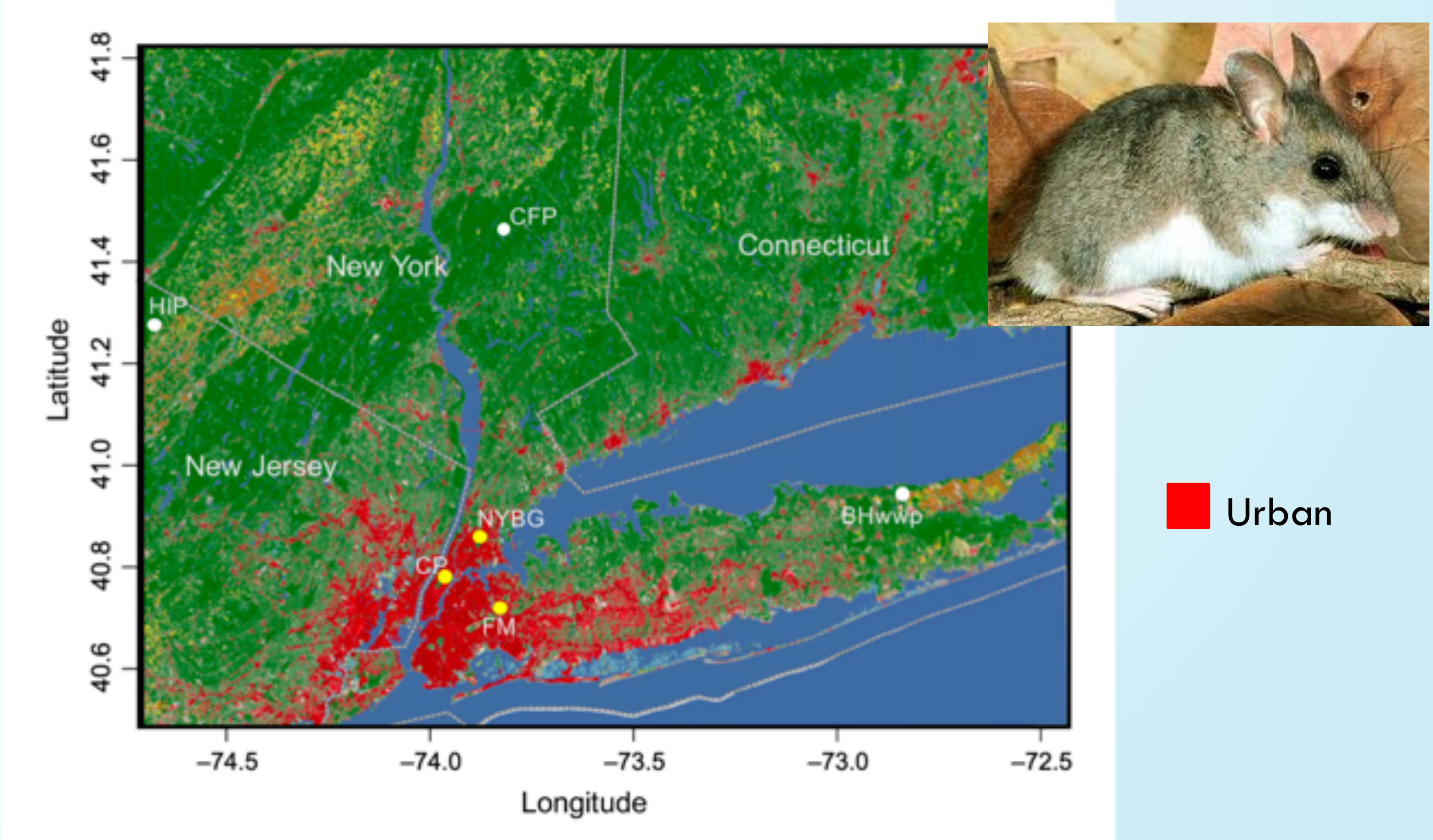
See Hoban et al. 2016 for more

CASE STUDY– FST OUTLIER TESTS

Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*)

Stephen E. Harris[1*] | Jason Munshi-South[2]

2017- Molecular Ecology

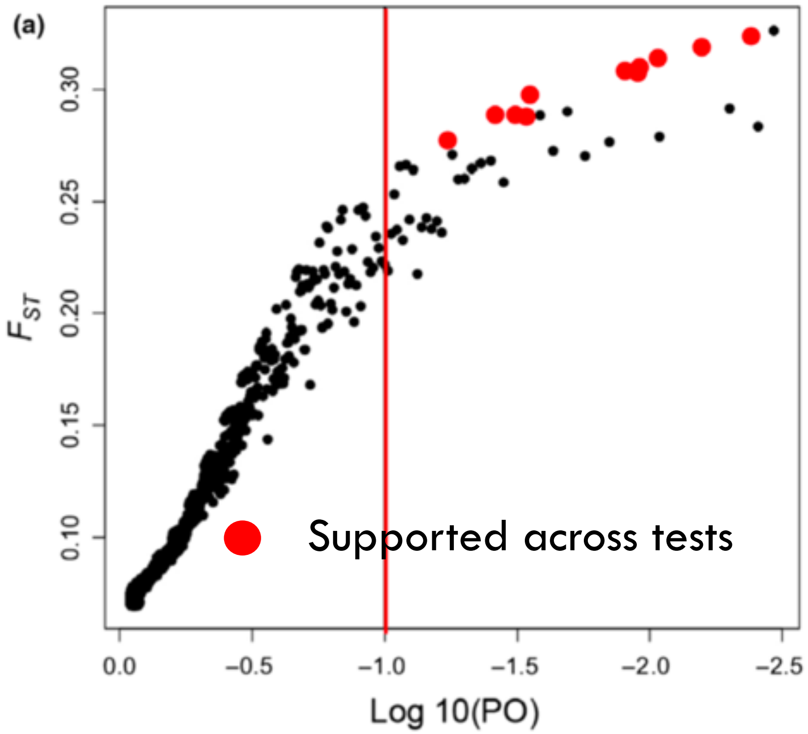


🟥 Urban

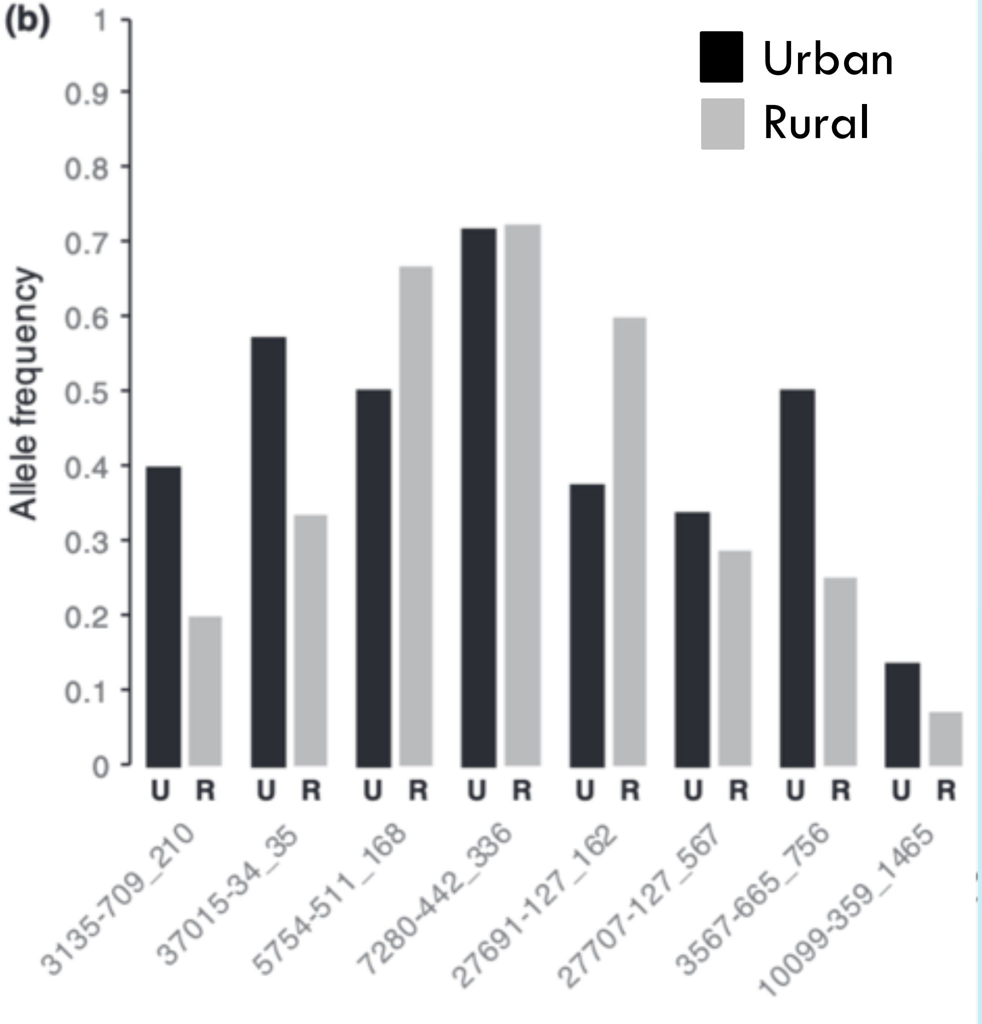
CASE STUDY– FST OUTLIER TESTS

Bayescan

Urban vs Rural



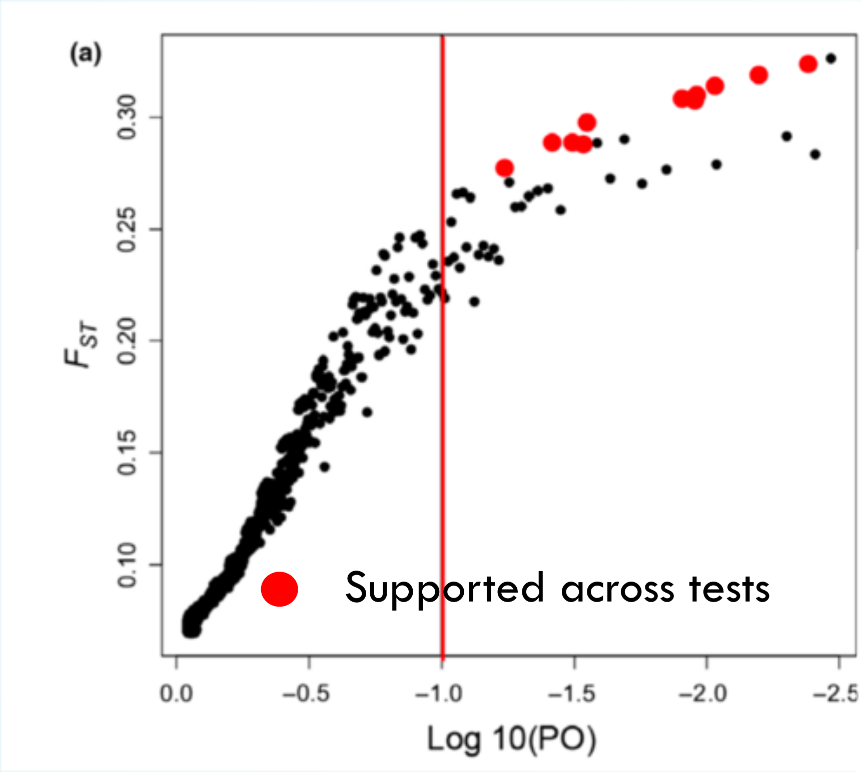
● Supported across tests



■ Urban

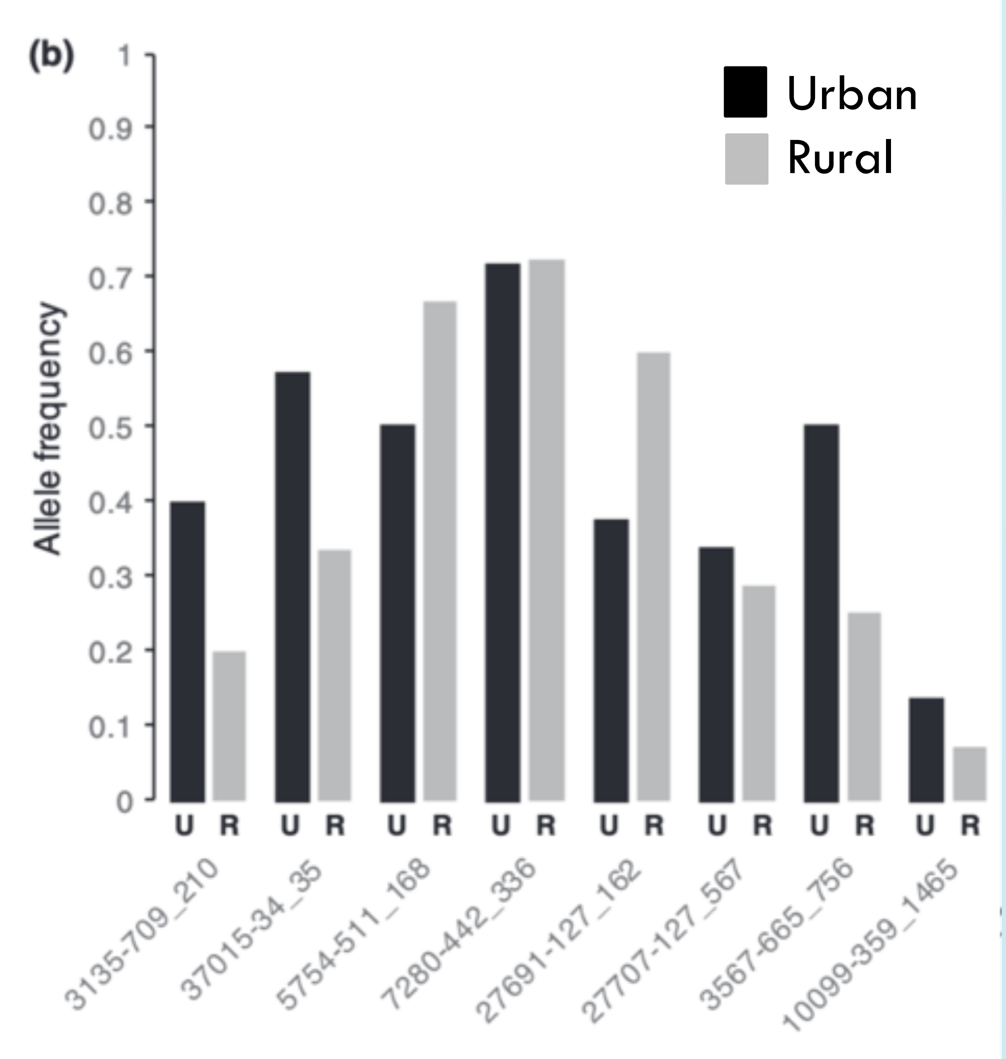
▨ Rural

CASE STUDY– FST OUTLIER TESTS



Bayescan

Urban vs Rural



- Gene annotations associated with dietary metabolism
- Urban vs rural energy budgets, physiological stressors or diets ?

ENVIRONMENTAL ASSOCIATION ANALYSIS (EAA)

(or genotype x environment, GEA)

Seeks a correlative indication of evolutionary responses

to spatial heterogeneity

For e.g. in the white-mice study, an EAA study might test explicit hypotheses with data on:

- % urban cover for each sample
- Composition of food types in each habitat
- Other abiotic or biotic variables

ENVIRONMENTAL ASSOCIATION ANALYSIS (EAA)

(or genotype x environment, GEA)

Seeks a correlative indication of evolutionary responses

to spatial heterogeneity

Tools for EAA differ by:

- the type of model (e.g. logistic regression, matrix correlation,, mixed-effect models)
- the statistical procedure used (e.g. FDR, p-values, GIF)
- Univariate vs. Multivariate
- the way population structure is dealt with

ENVIRONMENTAL ASSOCIATION ANALYSIS (EAA)
(or genotype x environment, GEA)

Seeks a correlative indication of evolutionary responses

to spatial heterogeneity

Tools for EAA differ by:

- the type of model (e.g. logistic regression, matrix correlation,, mixed-effect models)
- the statistical procedure used (e.g. FDR, p-values, GIF)
- Univariate vs. Multivariate
- the way population structure is dealt with

Since 1977!

Single locus correlation with slope aspect

Mitton et al. (1977)

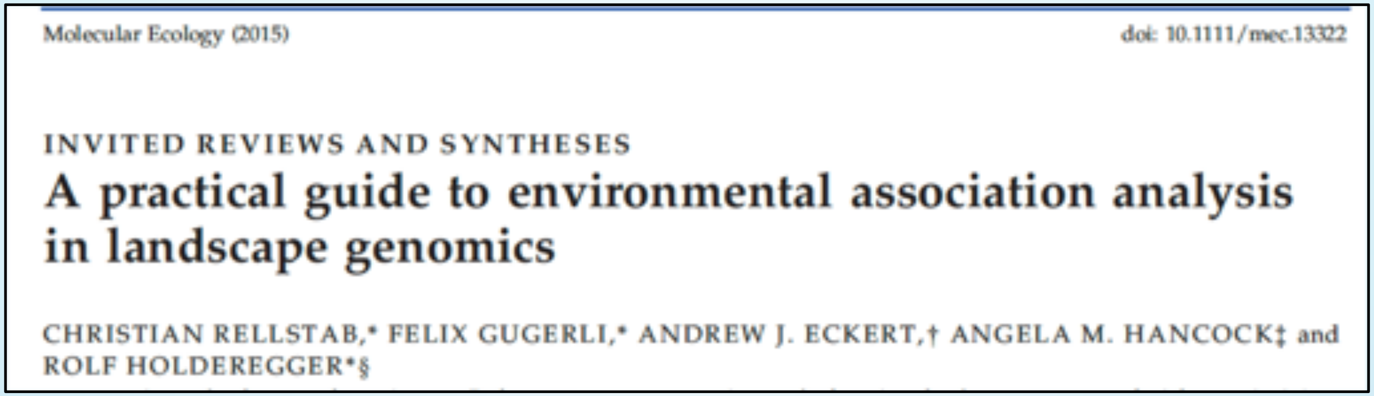
Theoretical and Applied Genetics

January 1977, Volume 51, Issue 1, pp 5–13 | Cite as

Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range

Authors | Authors and affiliations

J. B. Mitton, Y. B. Linhart, J. L. Hamrick, J. S. Beckman

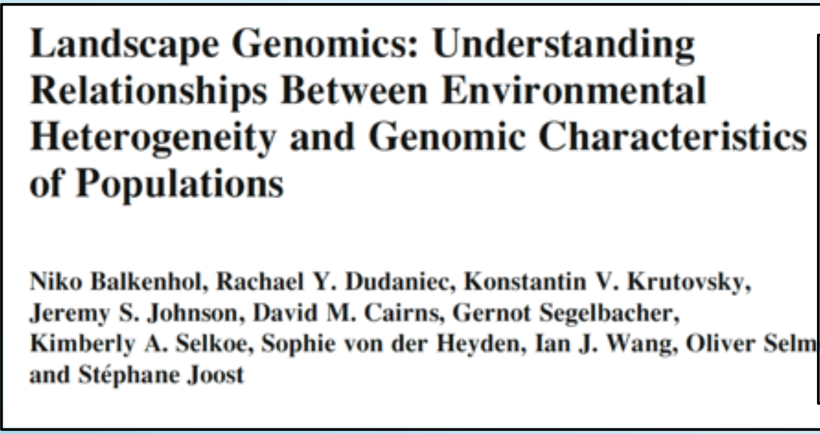


Molecular Ecology (2015) doi: 10.1111/mec.13322

INVITED REVIEWS AND SYNTHESES

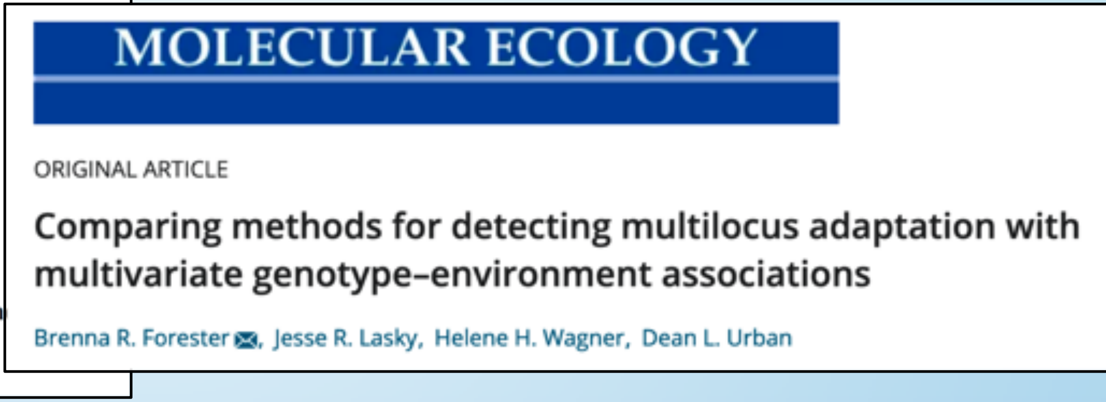
A practical guide to environmental association analysis in landscape genomics

CHRISTIAN RELLSTAB,* FELIX GUGERLI,* ANDREW J. ECKERT,† ANGELA M. HANCOCK‡ and ROLF HOLDEREGGER*§



Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations

Niko Balkenhol, Rachael Y. Dudaniec, Konstantin V. Krutovsky, Jeremy S. Johnson, David M. Cairns, Gernot Segelbacher, Kimberly A. Selkoe, Sophie von der Heyden, Ian J. Wang, Oliver Selm
and Stéphane Joost



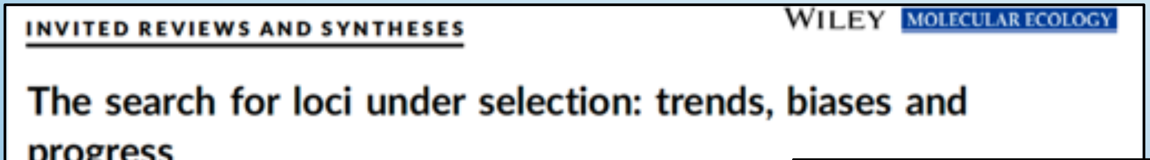
MOLECULAR ECOLOGY

ORIGINAL ARTICLE

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

Brenna R. Forester ✉, Jesse R. Lasky, Helene H. Wagner, Dean L. Urban

---



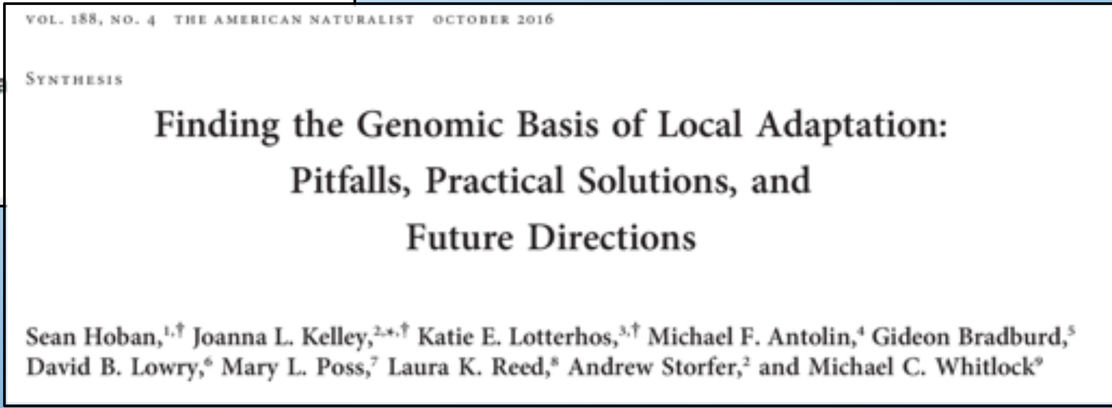
INVITED REVIEWS AND SYNTHESES

WILEY MOLECULAR ECOLOGY

The search for loci under selection: trends, biases and progress

Collin W. Ahrens[1] | Paul D. Rymer[1] | Adam Stow[2] | Jason Bra
Kate D. L. Umbers[1,5] | Rachael Y. Dudaniec[2]

[1]Hawkesbury Institute for the Environment



VOL. 188, NO. 4 THE AMERICAN NATURALIST OCTOBER 2016

SYNTHESIS

Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions

Sean Hoban,[1,†] Joanna L. Kelley,[2,*,†] Katie E. Lotterhos,[3,†] Michael F. Antolin,[4] Gideon Bradburd,[5] David B. Lowry,[6] Mary L. Poss,[7] Laura K. Reed,[8] Andrew Storfer,[2] and Michael C. Whitlock[9]

SAMPLING PRINCIPLES
for EAA

Environmental variables

- Choose variables hypothesized to drive selection
- Capture the range of environmental variables and values occupied by your study species.

SAMPLING PRINCIPLES for EAA

Environmental variables

- Choose variables hypothesized to drive selection
- Capture the range of environmental variables and values occupied by your study species.

Match spatial and temporal scales with genetic data (EAA)

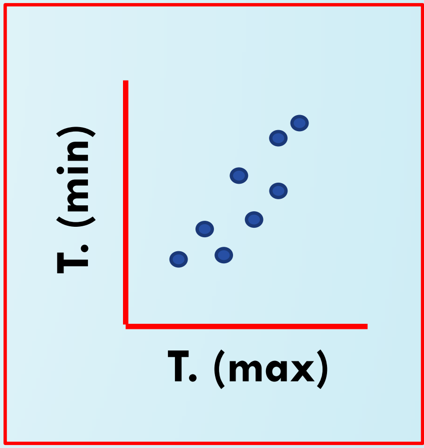
- Cell size resolution – be relevant to study species' gene flow

  sampling at a 2.5 km scale would be inappropriate for a slug!
- Timing of data (daily, monthly, annual, seasonal)
- Treatment of data (total? average? variability?)

Data format - EAA

Environmental data

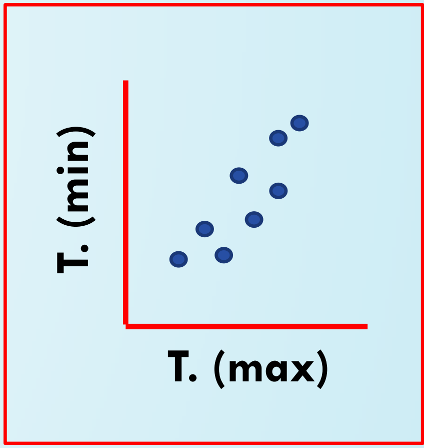
Sample	Lat	Lon	Precip	T. (max)	T. (min)
S1	-15.1	132.3	328.4	29.6	23.6
S2	-15.2	132.5	329.2	29.3	23.4
S3	-15.4	132.1	326.3	30.0	25.5
S4	-15.4	132.7	324.4	28.4	22.4
S5	-15.5	133.6	318.6	28.6	24.1
S6	-15.7	134.2	314.8	27.8	20.6
S7	-15.7	135.4	312.8	27.2	22.1



Data format - EAA

Environmental data

Sample	Lat	Lon	Precip	T. (max)	T. (min)
S1	-15.1	132.3	328.4	29.6	23.6
S2	-15.2	132.5	329.2	29.3	23.4
S3	-15.4	132.1	326.3	30.0	25.5
S4	-15.4	132.7	324.4	28.4	22.4
S5	-15.5	133.6	318.6	28.6	24.1
S6	-15.7	134.2	314.8	27.8	20.6
S7	-15.7	135.4	312.8	27.2	22.1



PCA

Sample	Lat	Lon	PC1	PC2
S1	-15.1	132.3	0.84	0.76
S2	-15.2	132.5	0.92	0.83
S3	-15.4	132.1	0.63	0.85
S4	-15.4	132.7	0.44	0.64
S5	-15.5	133.6	0.36	0.70
S6	-15.7	134.2	0.32	0.61
S7	-15.7	135.4	0.30	0.64

Or exclude highly correlated variables <0.70 general rule of thumb

Data format- EAA

population sampling

Sample	Pop	Lat	Lon	Precip	Temp
S1	P1	-15.1	132.3	328.4	23.6
S2	P1	-15.1	132.3	328.4	23.6
S3	P1	-15.1	132.3	328.4	23.6
S4	P1	-15.1	132.3	328.4	23.6
S5	P2	-15.7	135.4	312.8	21.2
S6	P2	-15.7	135.4	312.8	21.2
S7	P2	-15.7	135.4	312.8	21.2

+

individual sampling

Sample	Lat	Lon	Precip	Temp
S1	-15.1	132.3	328.4	23.6
S2	-15.2	132.5	329.2	23.3
S3	-15.4	132.1	326.3	24.0
S4	-15.4	132.7	324.4	22.4
S5	-15.5	133.6	318.6	22.6
S6	-15.7	134.2	314.8	21.8
S7	-15.7	135.4	312.8	21.2

Genetic data –SNP

	L1	L2	L3	L5
S1	0	0	0	2
S2	2	1	0	0
S3	1	0	0	1
S4	0	1	0	0
S5	1	0	0	0
S6	0	1	1	1
S7	2	0	0	0

SAMPLING PRINCIPLES for EAA

Individual-based analyses work best when:

- you have many coordinates
- environmental data has high variation across sampling area
- Local *Ne* is low (e.g. mammals)

Population-based analyses work best when:

- You have clusters of sampled individuals per site
- Environmental variables change across broader spatial scales
- Local *Ne* is higher (e.g. insects)

SAMPLING PRINCIPLES for EAA

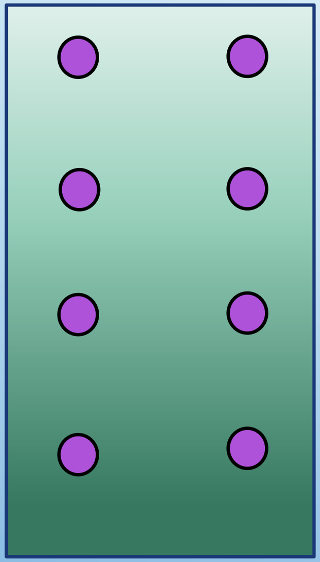
Paired-gradient sampling is best to detect weaker selection.

'Clustered' sampling ok for stronger selection

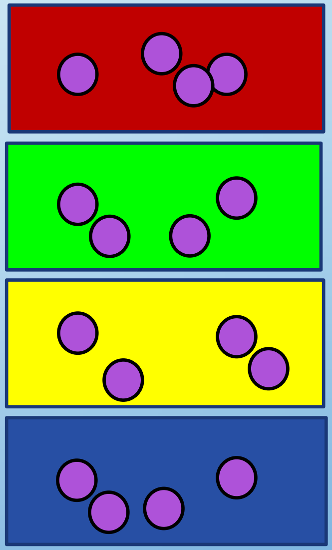
= *maximise environmental differences*

= *minimize differences in evolutionary history*

gradient

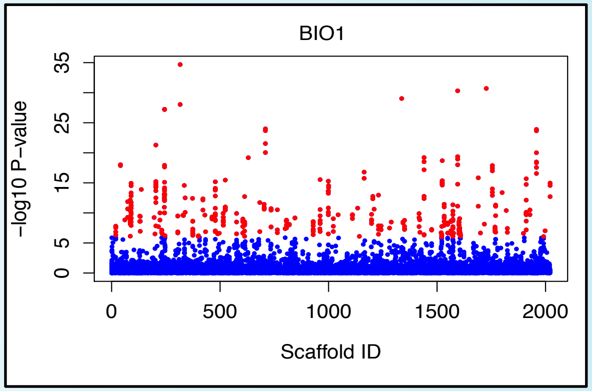


Island model



ENVIRONMENTAL ASSOCIATION ANALYSIS

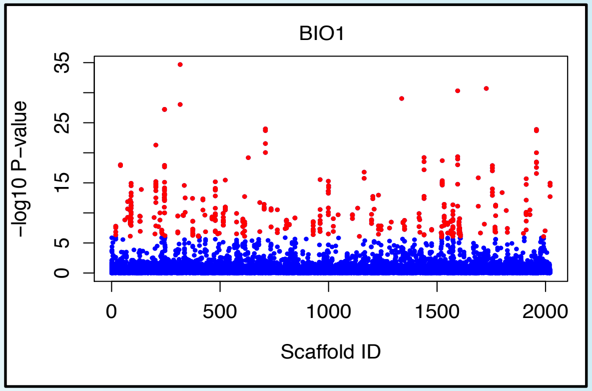
- **LFMM2 -Univariate:** LMM that uses K as latent factors (representing random effects);the environment is used as a fixed effect - very fast (Caye et al. 2019).

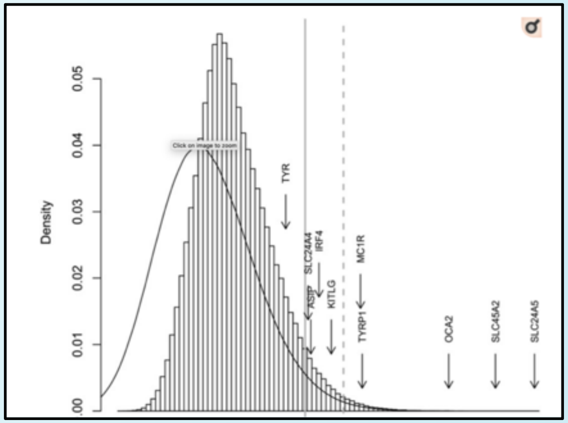


ENVIRONMENTAL ASSOCIATION ANALYSIS

- **LFMM2 -Univariate:** LMM that uses K as latent factors (representing random effects);the environment is used as a fixed effect - very fast (Caye et al. 2019).

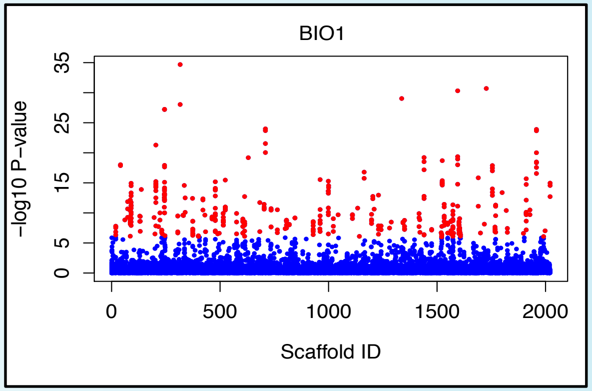
- **BAYENV2- Univariate:** – Bayesian method, generates a kinship matrix from allelic data, to estimate a null model of the demographic structure to test each SNP (Gunther & Coop 2013)- often similar to LFMM



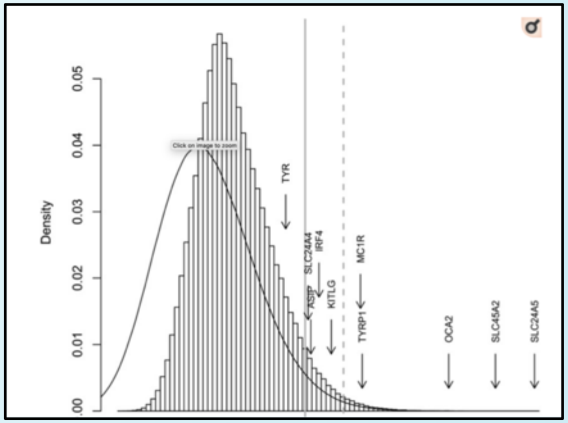


ENVIRONMENTAL ASSOCIATION ANALYSIS

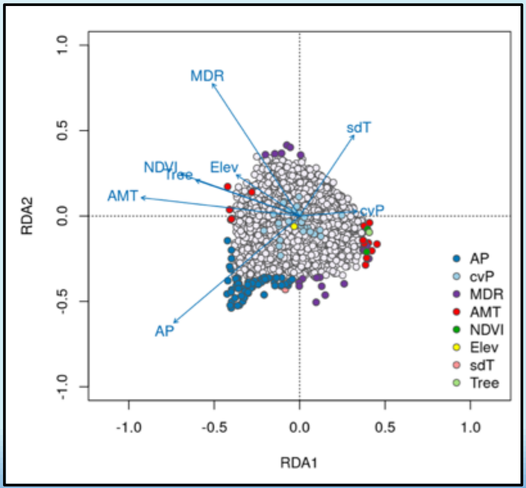
- **LFMM2 - Univariate:** LMM that uses K as latent factors (representing random effects);the environment is used as a fixed effect - very fast (Caye et al. 2019).



- **BAYENV2- Univariate:** – Bayesian method, generates a kinship matrix from allelic data, to estimate a null model of the demographic structure to test each SNP (Gunther & Coop 2013)- often similar to LFMM

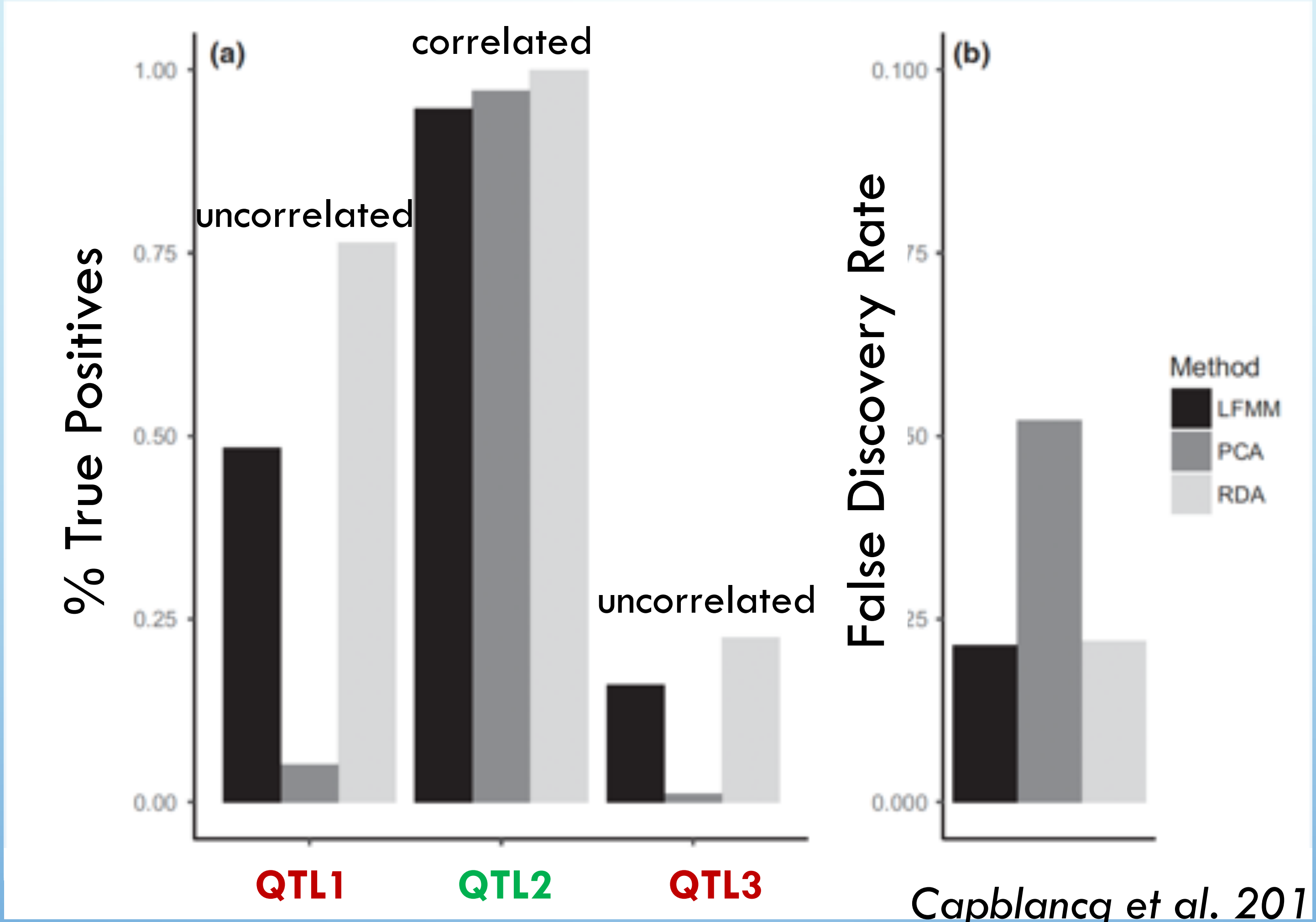


- **RDA -Multivariate:** Redundancy Analysis (+ pRDA) uses ordination to to identify environmental gradients most correlated with adaptive variation (Capblancq et al. 2018), high true positive rate (Forester et al. 2018)



RDA + LFMM + pcadapt comparison

LFMM +RDA have similar detection power – *pcadapt* fails when the environmental gradient is **not** correlated with population structure (a) and has a higher FDR (b)



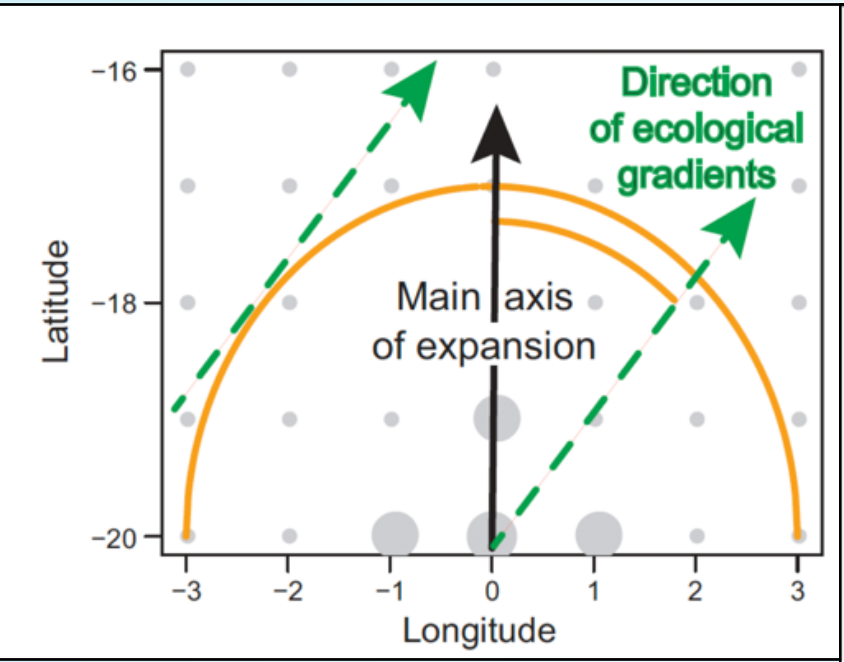
Capblancq et al. 2018

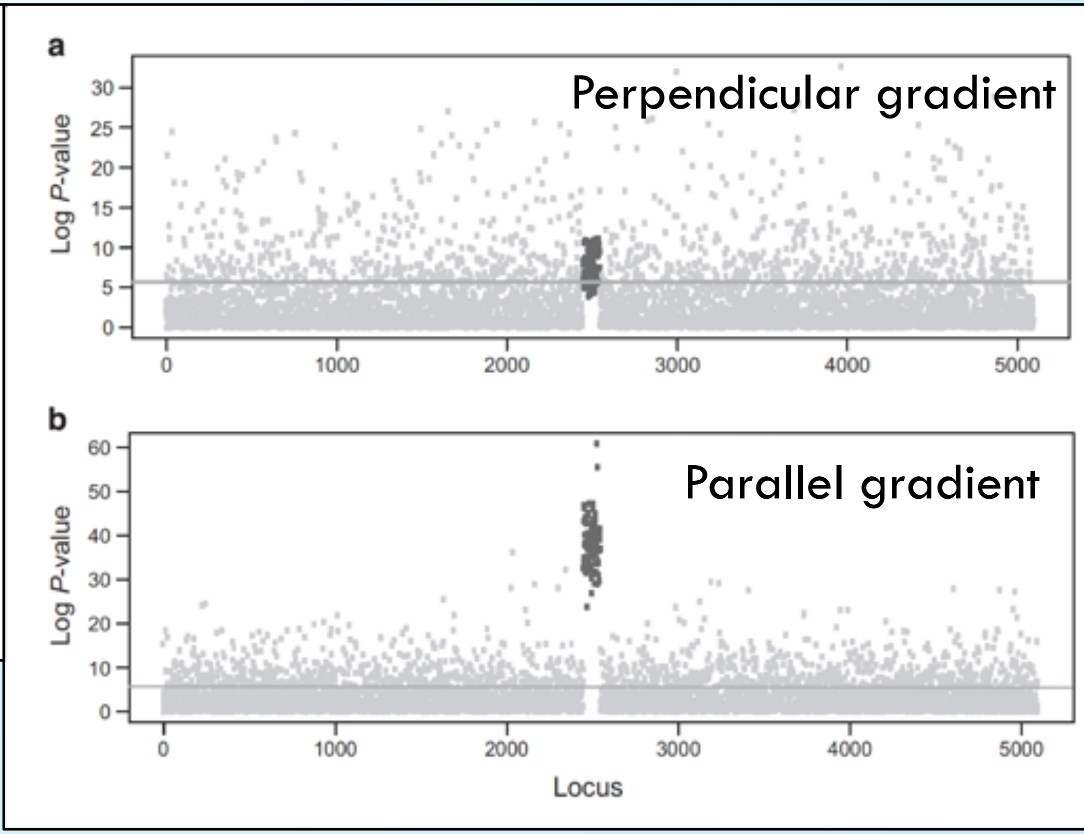
Frichot et al. 2015, Heredity

ORIGINAL ARTICLE

Detecting adaptive evolution based on association with ecological gradients: Orientation matters!

E Frichot[1], SD Schoville[2], P de Villemereuil[3], OE Gaggiotti[3,4] and O François[1]





Perpendicular gradient

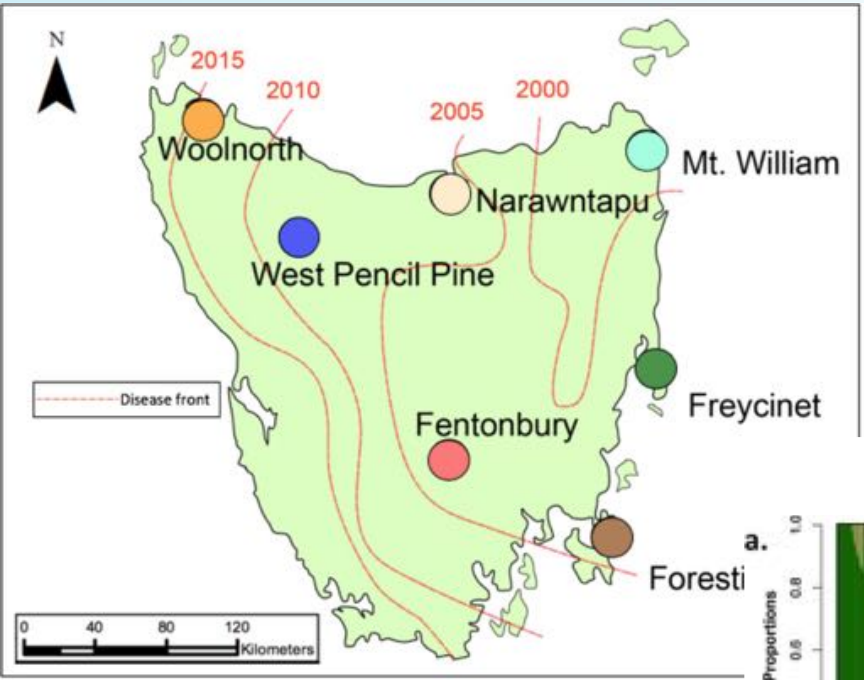
Parallel gradient

LFMM

Fewer false positives when expansion axis is parallel!

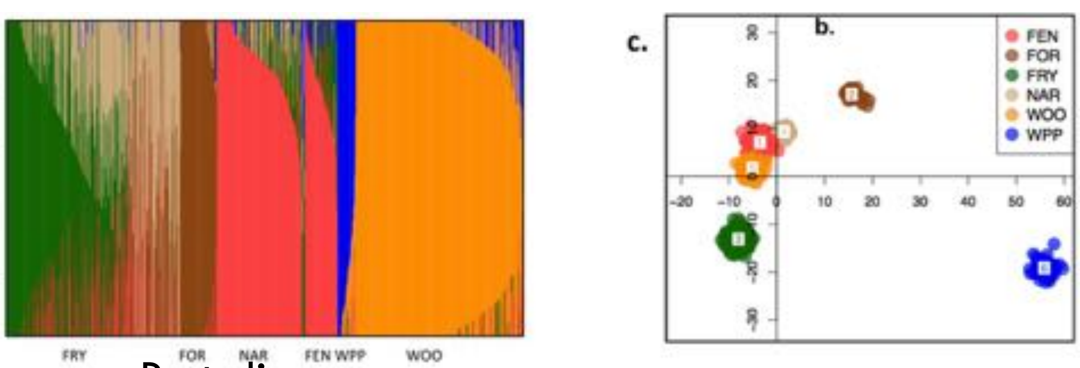
DISEASE VS ABIOTIC SELECTION IN TASMANIAN DEVILS WITH DFTD

Fraik et al. 2019 BioRxiv

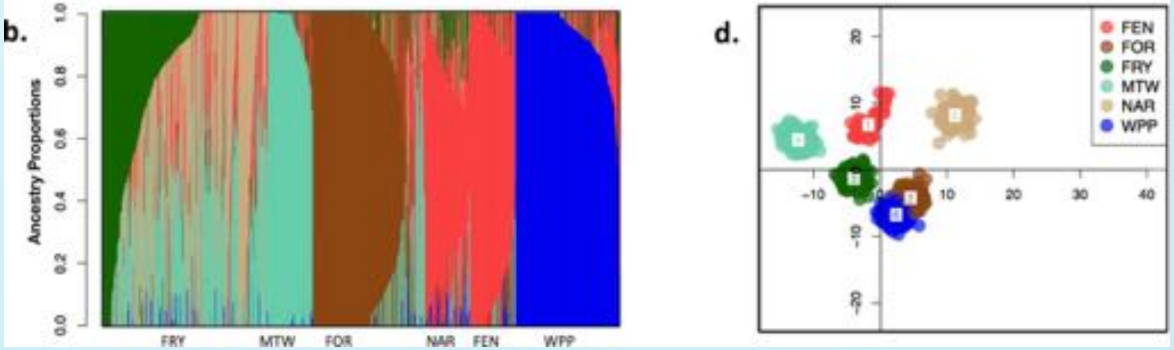


- Sampled pre-and post disease
- Shifts in genetic structure
- LFMM and BAYENV2: RADCapture of immune response + random SNPs
- 3,568 individuals

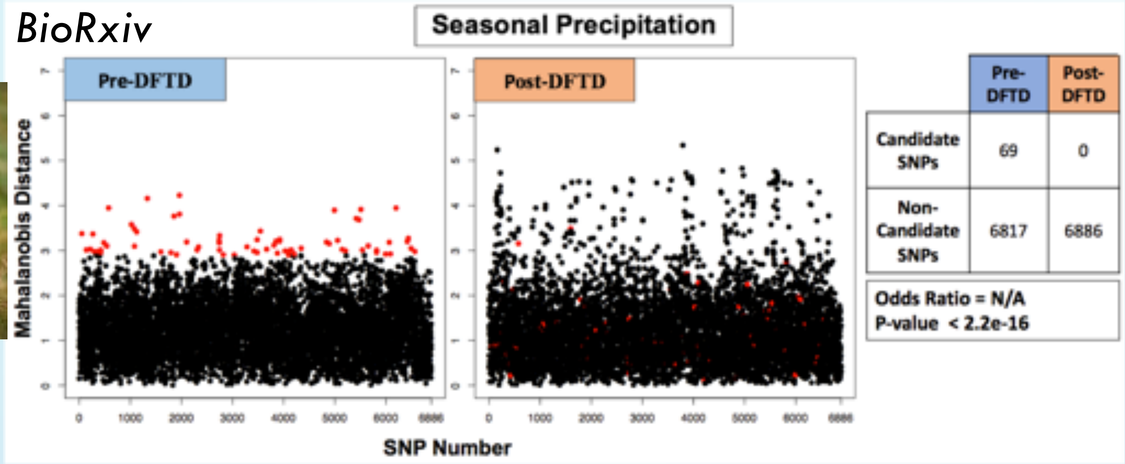
Pre-disease



Post-disease

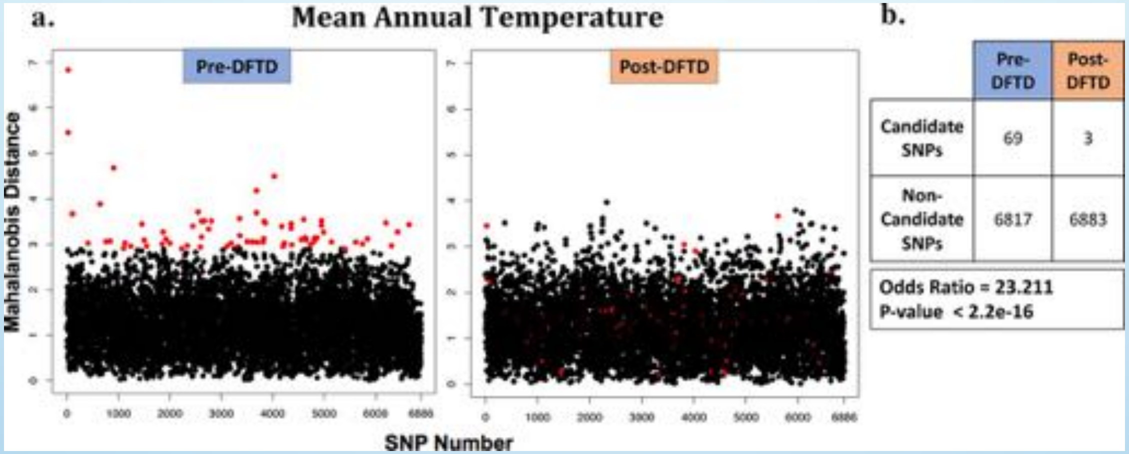


Fraik et al. 2019 BioRxiv



	Pre-DFTD	Post-DFTD
Candidate SNPs	69	0
Non-Candidate SNPs	6817	6886

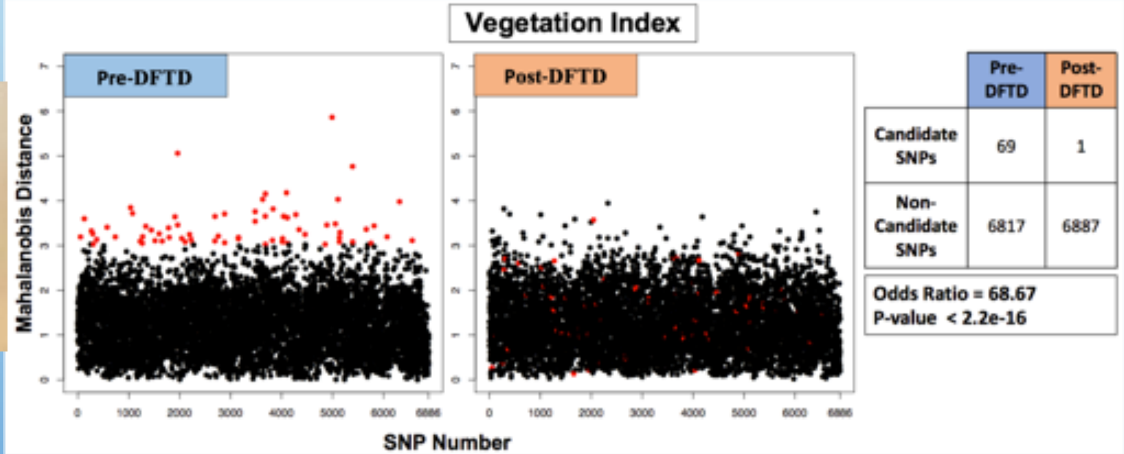
Odds Ratio = N/A
P-value < 2.2e-16



	Pre-DFTD	Post-DFTD
Candidate SNPs	69	3
Non-Candidate SNPs	6817	6883

Odds Ratio = 23.211
P-value < 2.2e-16

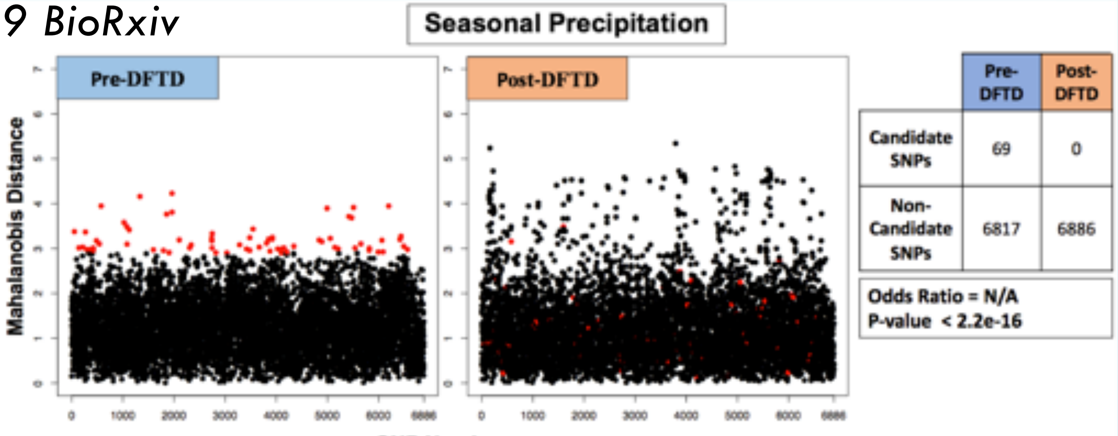
3 of 8 variables shown



	Pre-DFTD	Post-DFTD
Candidate SNPs	69	1
Non-Candidate SNPs	6817	6887

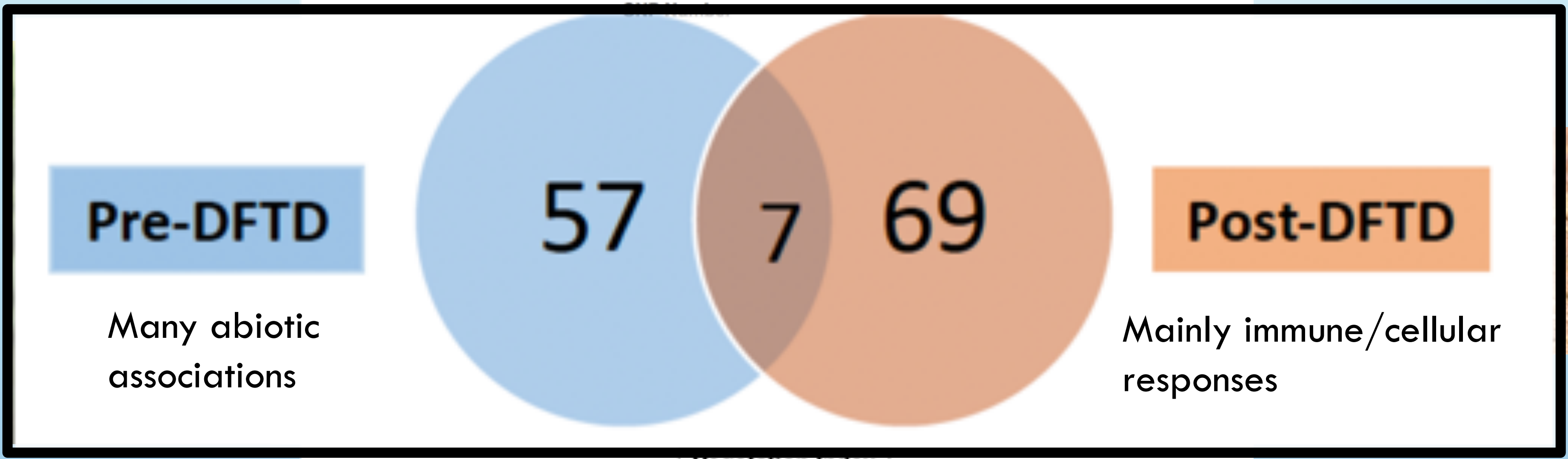
Odds Ratio = 68.67
P-value < 2.2e-16

Fraik et al. 2019 BioRxiv



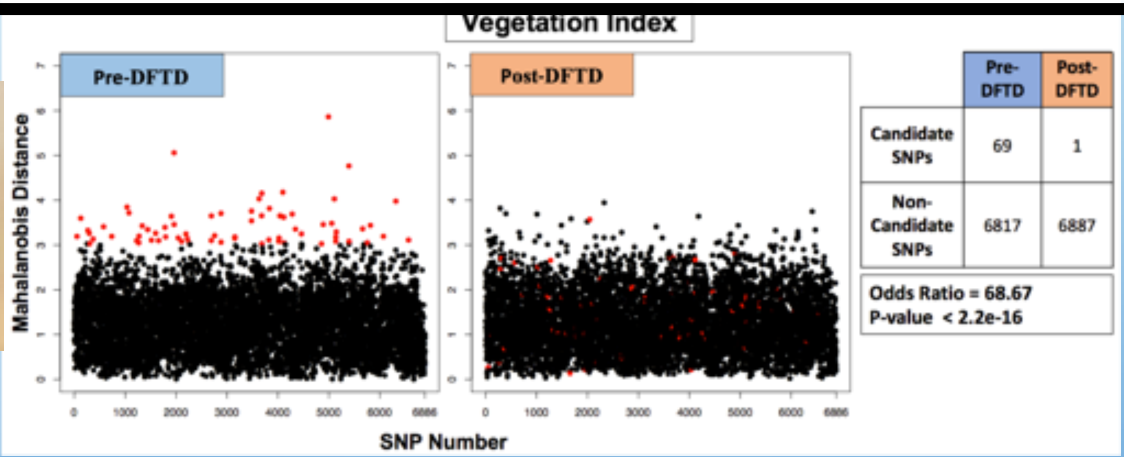
	Pre-DFTD	Post-DFTD
Candidate SNPs	69	0
Non-Candidate SNPs	6817	6886

Odds Ratio = N/A
P-value < 2.2e-16



Many abiotic associations

Mainly immune/cellular responses



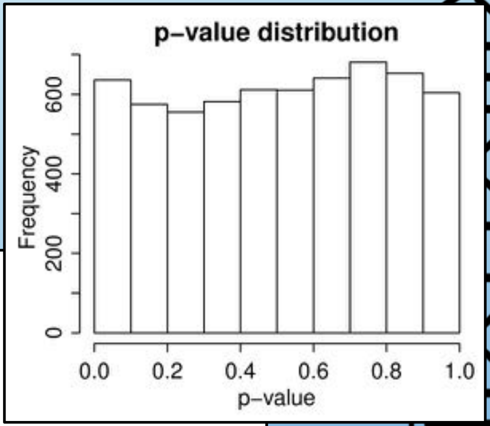
	Pre-DFTD	Post-DFTD
Candidate SNPs	69	1
Non-Candidate SNPs	6817	6887

Odds Ratio = 68.67
P-value < 2.2e-16

FALSE DISCOVERY RATES

- The expected proportion of false positives among the list of positive tests.
- So, an FDR = 0.05 means that the list of candidates is expected to contain up to 5% false positives
- The underlying principle of FDR control relies on significance values corresponding to the null hypothesis.

(i.e. that P-values are uniformly distributed)



INVITED REVIEWS AND SYNTHESES

Controlling false discoveries in genome scans for selection

OLIVIER FRANÇOIS,* HELENA MARTINS,* KEVIN CAYE* and SEAN D. SCHOVILLE†

GENOMIC INFLATION FACTOR (GIF)

Used to recalibrate z-scores and incorrect P-values to control FDR in GWAS, LFMM, RDA.

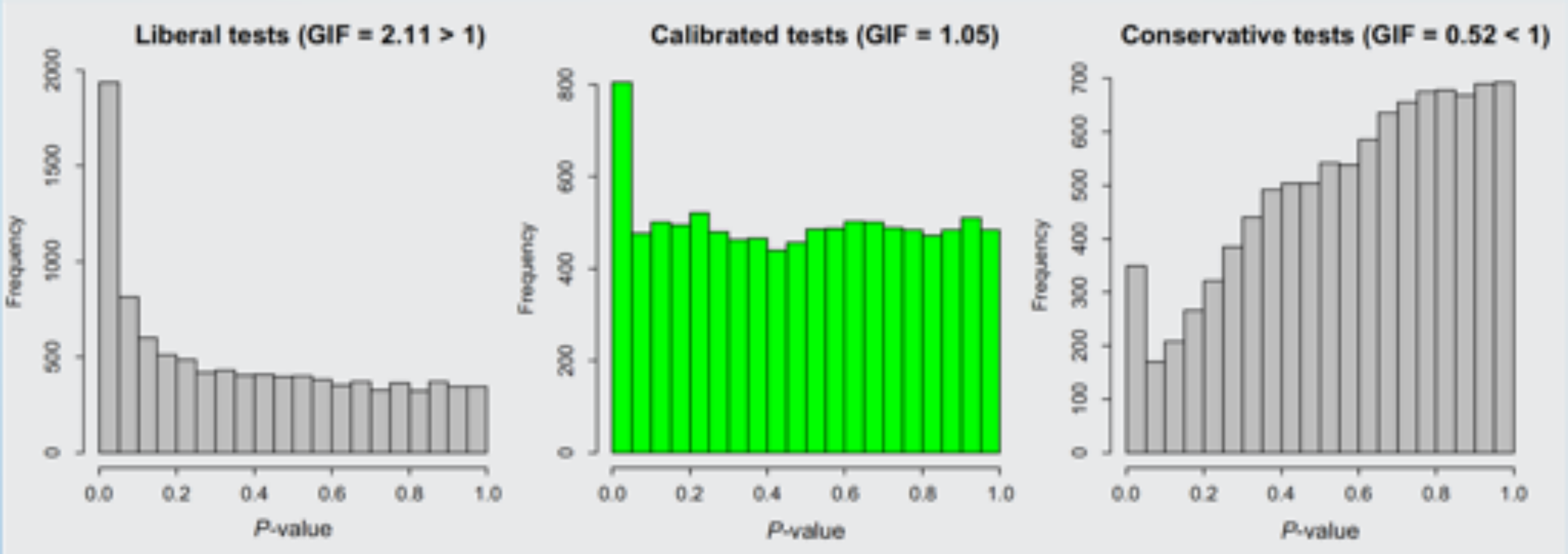
$$GIF = \frac{\textit{median of squared z-scores}}{\textit{median of the chi-squared distribution}}$$

Expresses the **deviation** of the distribution of the observed test statistic from the distribution of the expected test statistic, **i.e. inflation of scores**

GIF magnitude depends on sample size, relatedness, LD, population substructure, and the number of causal variants.

MODIFYING THE GIF

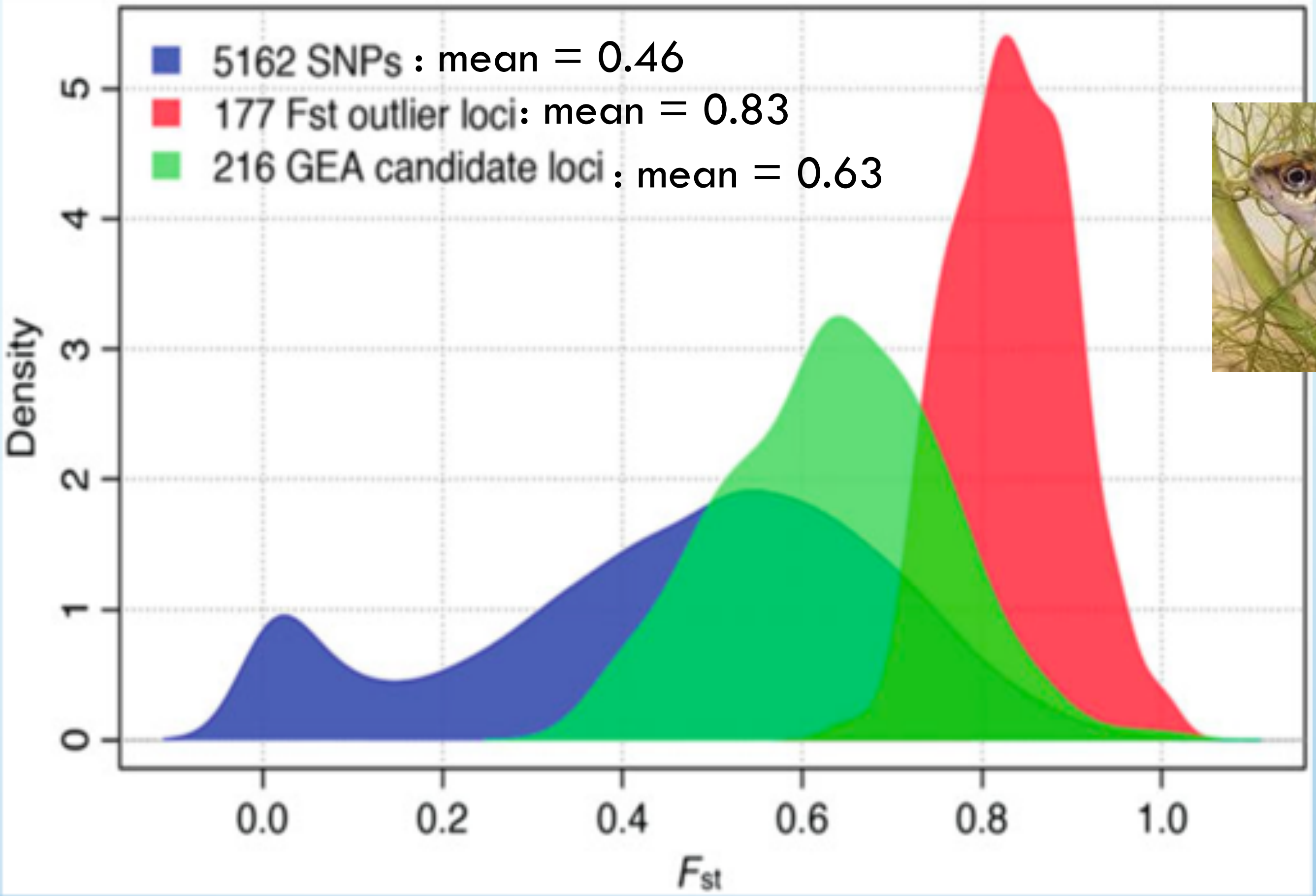
- Calibrating P-values using GIF attempts to flatten the histogram of P-values when the null hypothesis is true.
- FDR cut-offs only makes sense if your data fit this underlying null model (a uniform P-value distribution)
- Some data sets do not, no matter how much you modify GIF!



Box Fig. 1. Histograms of test significance values (*P*-values) prior to the application of FDR control algorithms (artificial data). GIF is the genomic inflation factor for each data set.

Francois et al. 2016

Fst Outlier vs EAA tests – strong vs weak selection



Fst Outliers:
Bayescan + Arlequin

EAA:
glNland + RDA

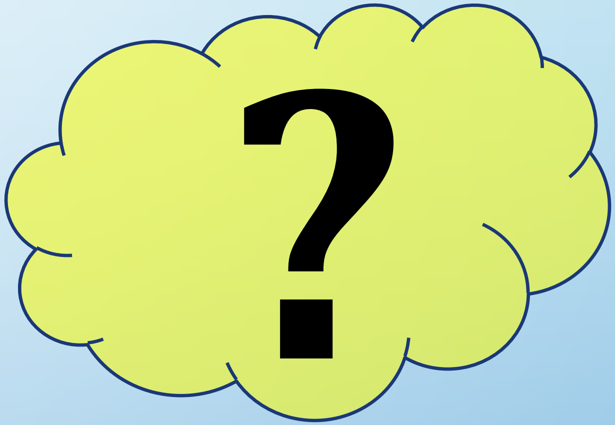
EAA candidates can cover a lower, broader Fst range – loci of smaller effect

Brauer et al. 2016, Mol Ecol

COMPARING Fst outlier TESTS WITH EAA

How is the number of outliers detected affected by:

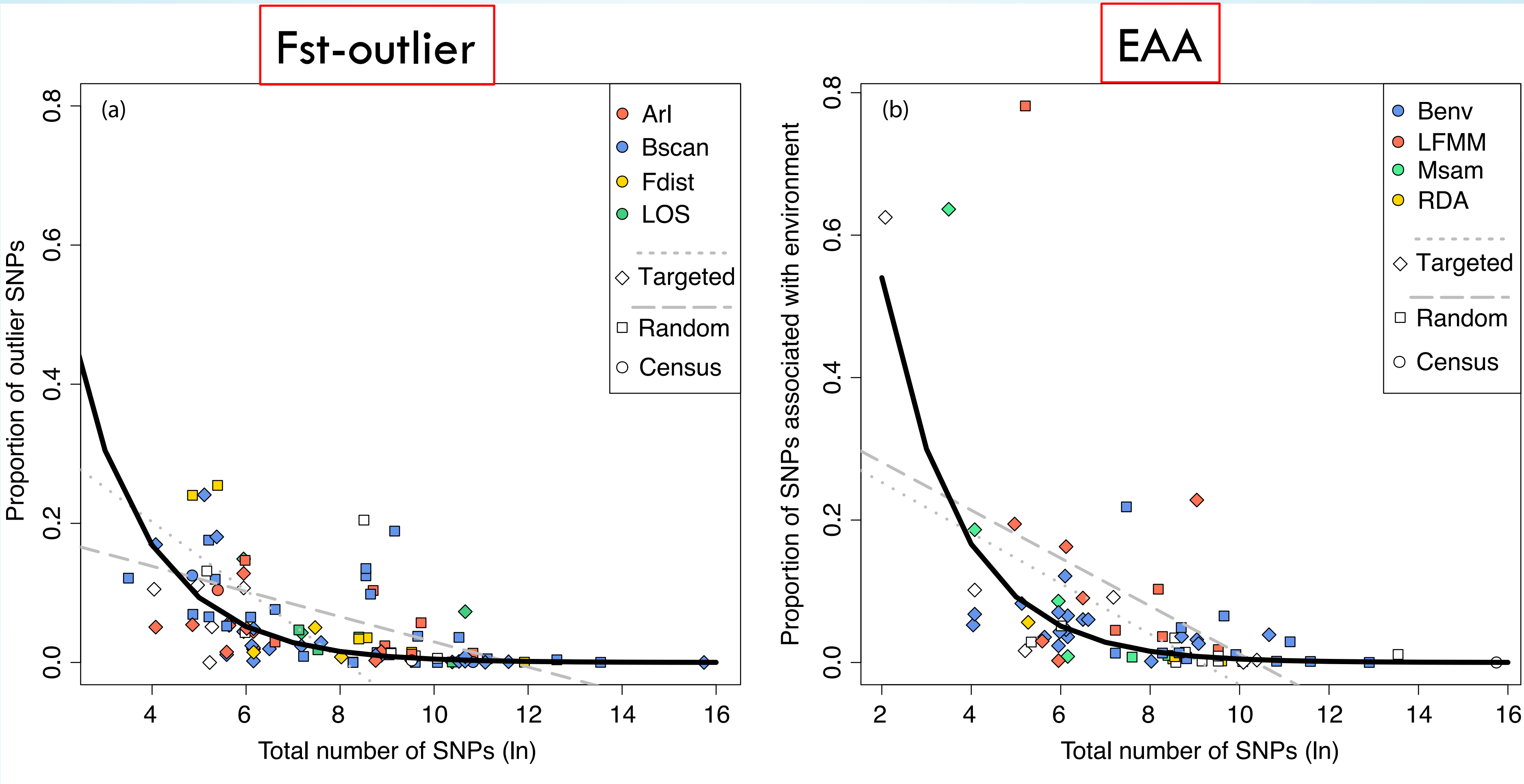
- N SNPs
- N individuals
- N genetic clusters
- global Fst
- environmental gradient steepness



EAA versus FST OUTLIER TESTS

More SNPs does not equal more outliers

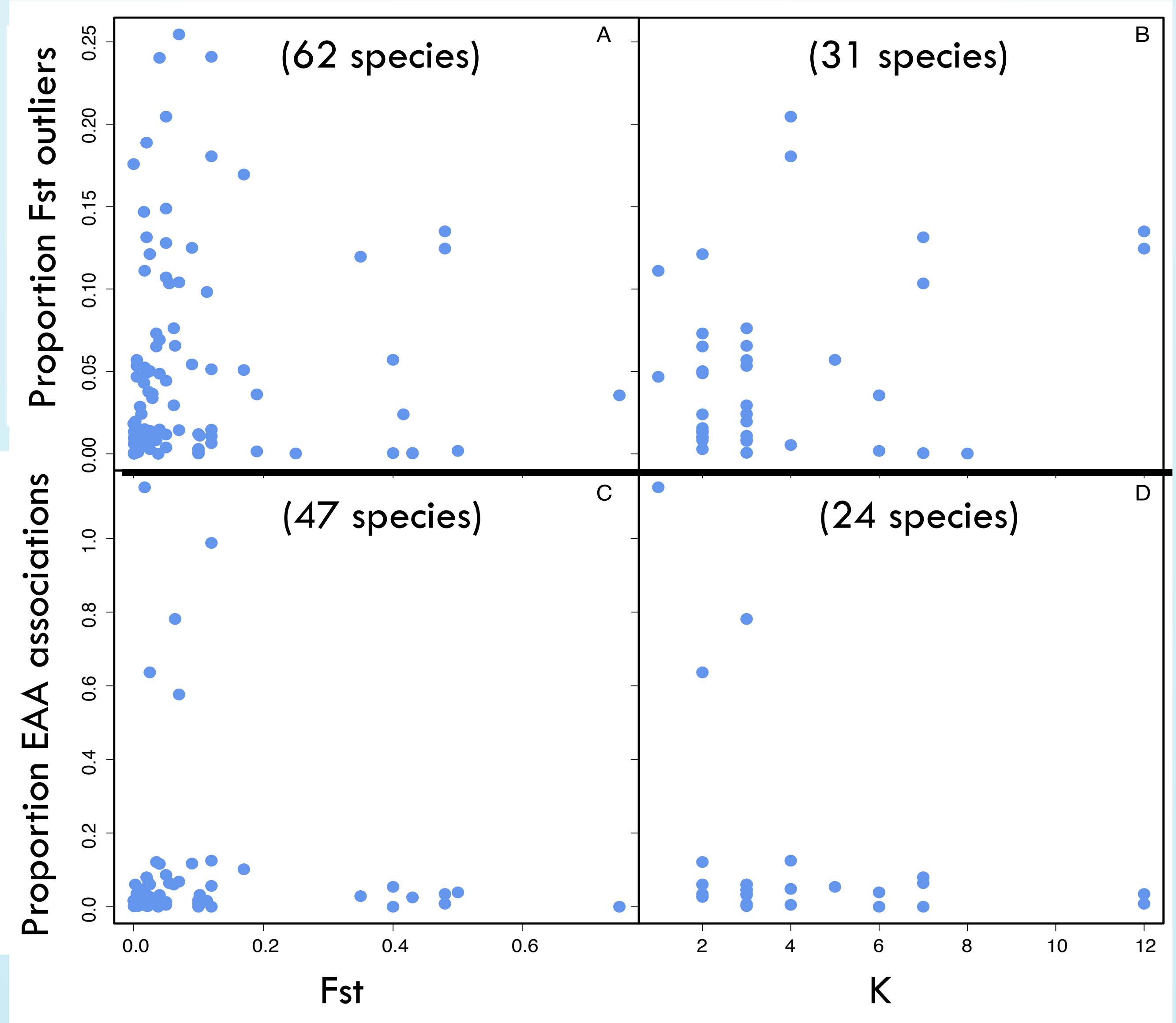
Detection of outliers level out at ~3K SNPS



N = 66 studies – Ahrens et al. 2018

EAA versus FST OUTLIER TESTS

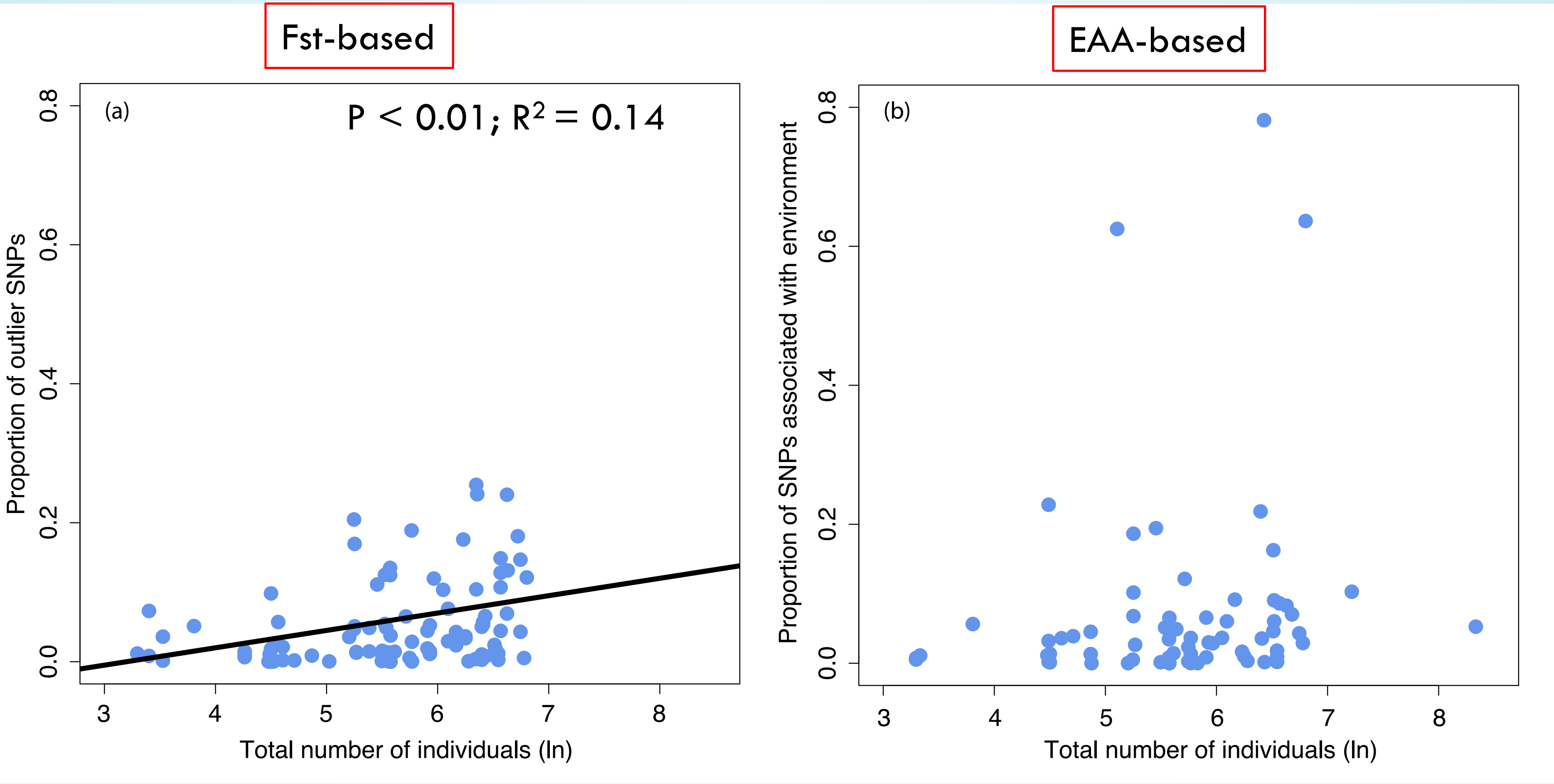
No change in # outliers with global Fst or N genetic clusters (K)



Ahrens et al. 2018

EAA versus FST OUTLIER TESTS

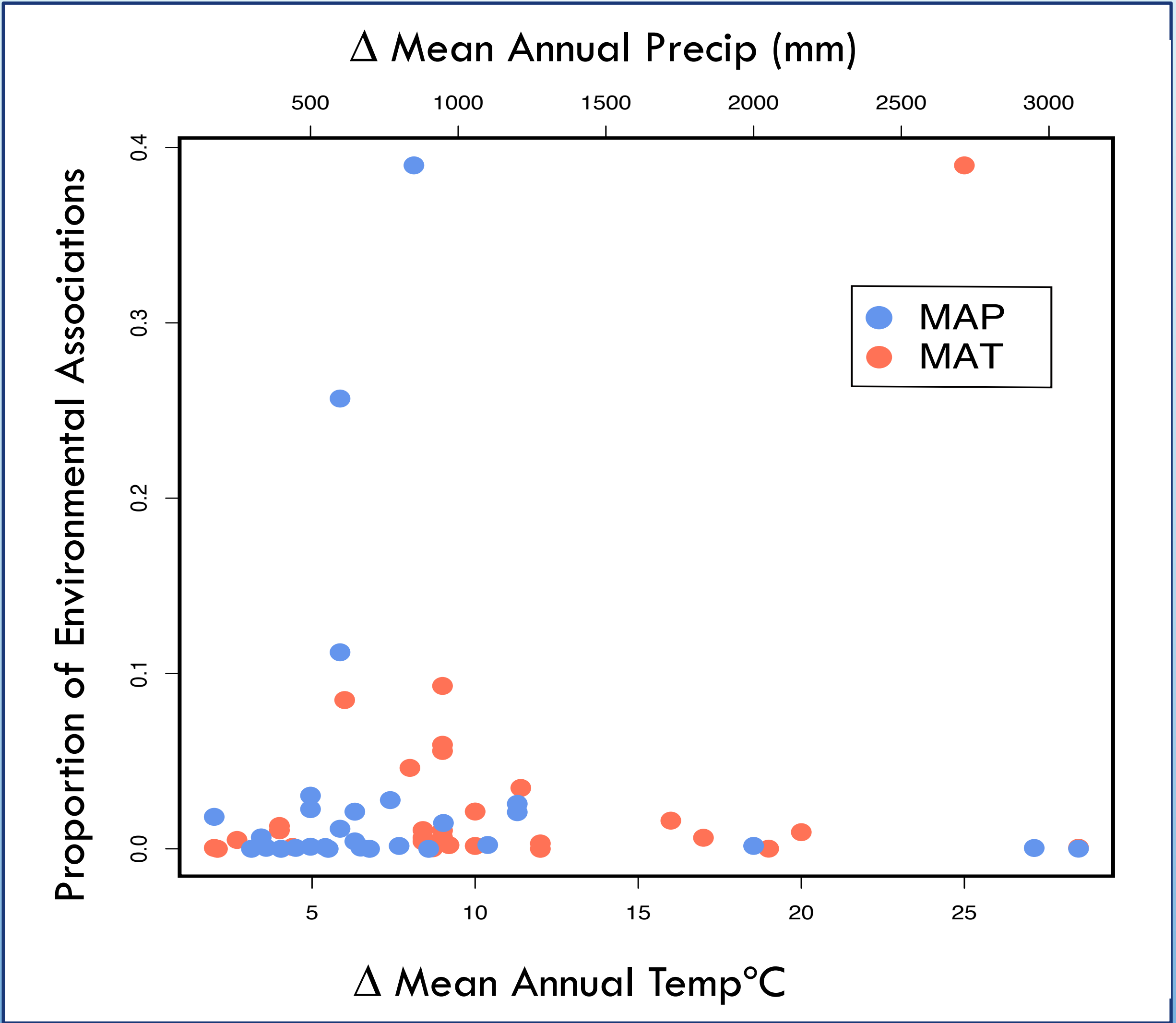
Number of individuals genotyped



N = 66 studies – Ahrens et al. 2018

ENVIRONMENTAL STEEPNESS DOES NOT AFFECT OUTLIER DETECTION

Environmental steepness – maximum change in temperature and precipitation between sampling sites (for 52 EAA papers)

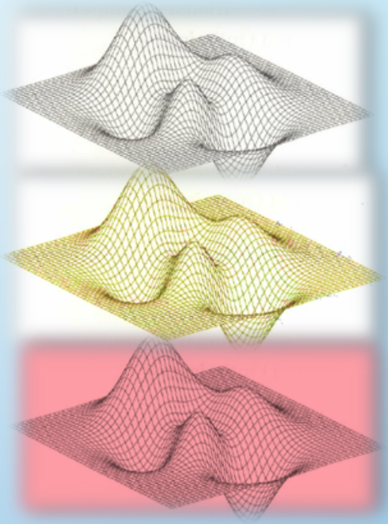


Ahrens et al. 2018, Mol Ecol

Questions?

PART 2: EXAMPLES + DOWNSTREAM ANALYSIS

- **Case Study 1** – Grasshoppers
  - morphological associations, annotation
- **Case Study 2** – Damselflies
  - Identifying strong vs weak outliers, GDM



- Stages of analysis overview
- Improving inference
- Combining approaches and the future

CASE STUDY

DETECTING SELECTION
TWO INSECTS. TWO ENVIRONMENTAL GRADIENTS.

1. Grasshoppers

(*Phaulacridium vittatum*)

Agricultural pest in Australia

MACQUARIE University SYDNEY·AUSTRALIA

UNIVERSITY OF TASMANIA

Sonu Yadav
(PhD graduate)

CASE STUDY

DETECTING SELECTION
TWO INSECTS. TWO ENVIRONMENTAL GRADIENTS.

1. Grasshoppers

(*Phaulacridium vittatum*)

Agricultural pest in Australia

Sonu Yadav
(PhD graduate)

2. Damselflies

(*Ischnura elegans*)

Range expanding in Sweden

CJ Yong
(Masters)

Alex Carey
(Masters)

CASE STUDY

DETECTING SELECTION
TWO INSECTS. TWO ENVIRONMENTAL GRADIENTS.

Two environmental gradients: 500-900km (5.0-8.0° latitude)

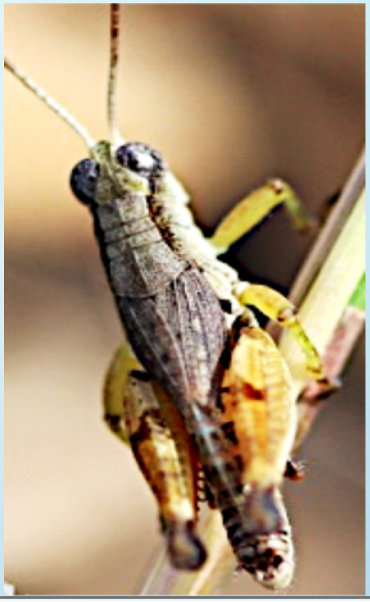
1. Are patterns of selection associated with environmental and morphological gradients?
2. Are environmental drivers of neutral genetic connectivity and adaptive genetic variation similar?

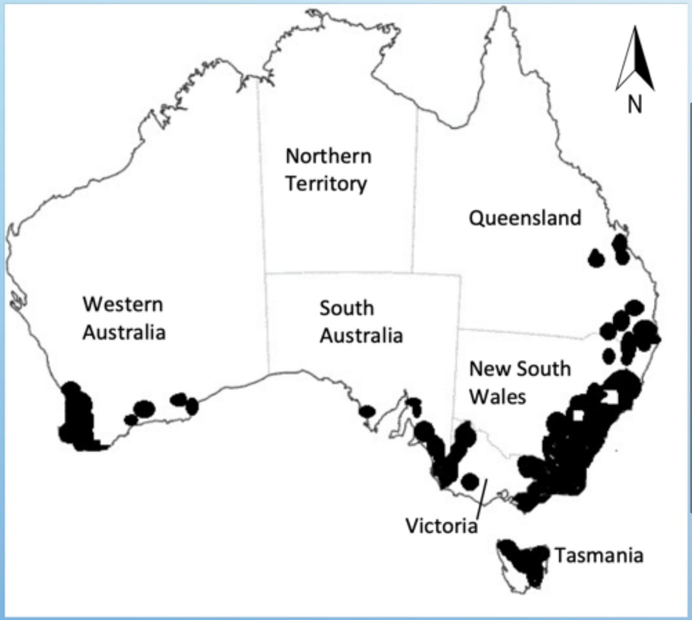
Yadav et al. 2019, Mol Ecol

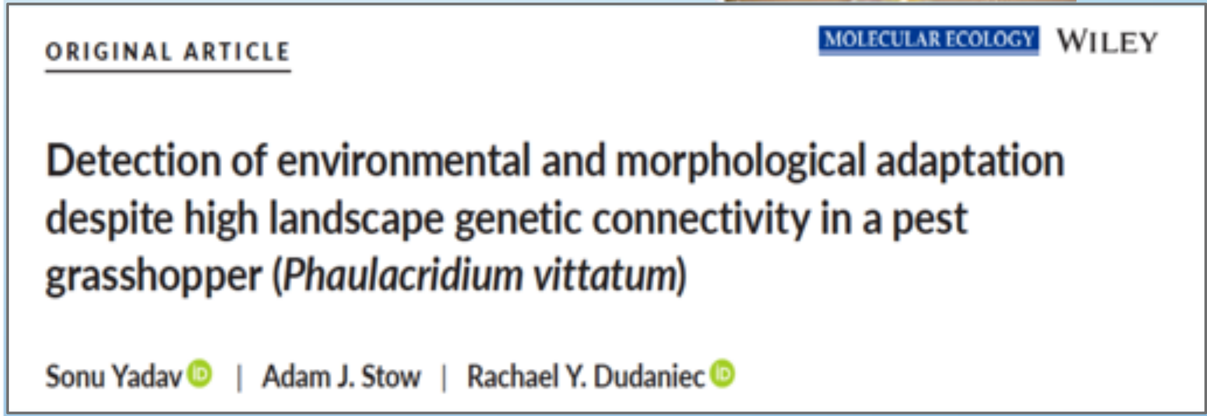
Dudaniec et al. 2018, Mol Ecol

(1) The Wingless Grasshopper (*Phaulacridium vittatum*)

- Endemic agricultural pest of pasture/herbaceous crops
- Outbreaks every 4-5 years + local annual outbreaks
- Economic costs from crop/pasture loss







ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

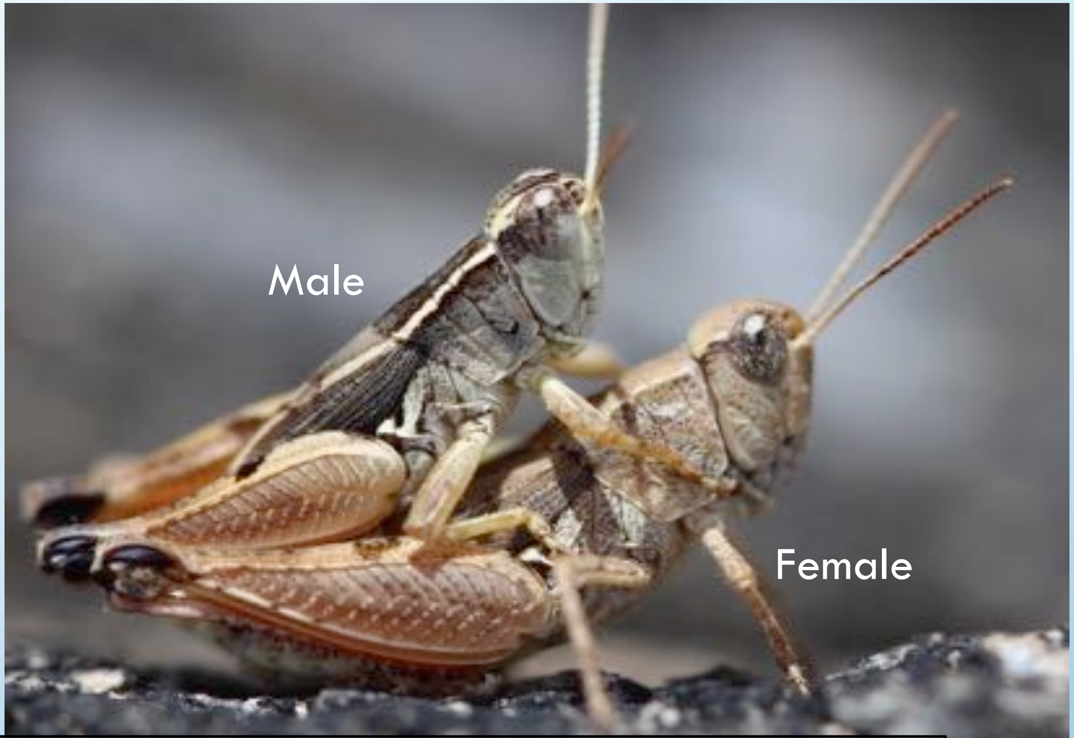
Detection of environmental and morphological adaptation despite high landscape genetic connectivity in a pest grasshopper (*Phaulacridium vittatum*)

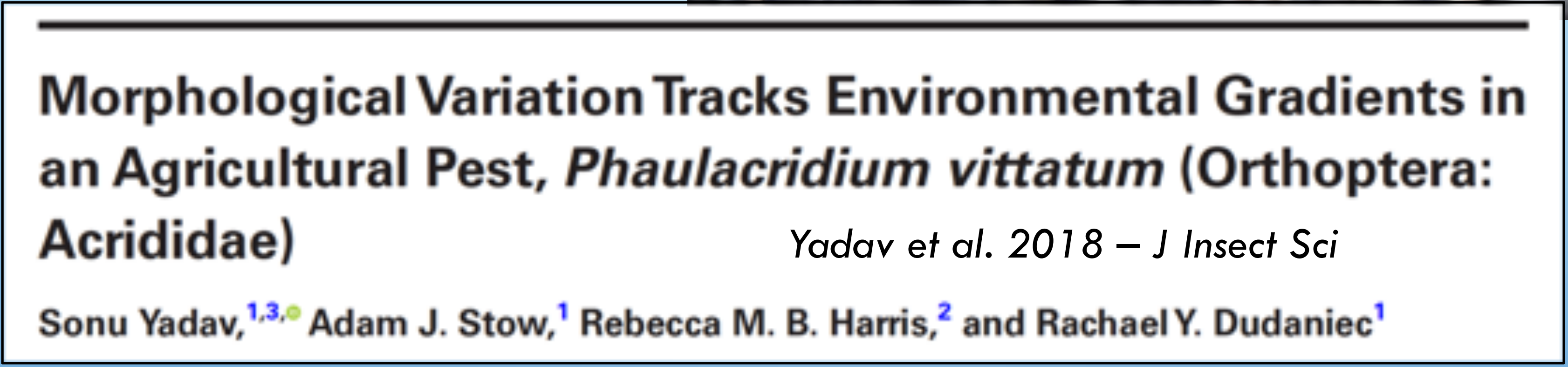
Sonu Yadav | Adam J. Stow | Rachael Y. Dudaniec

Morphological variation in *P. vittatum*

Body size (femur length) varies 6-13mm between sexes and individuals

Positively associated with solar radiation (Yadav et al. 2018)





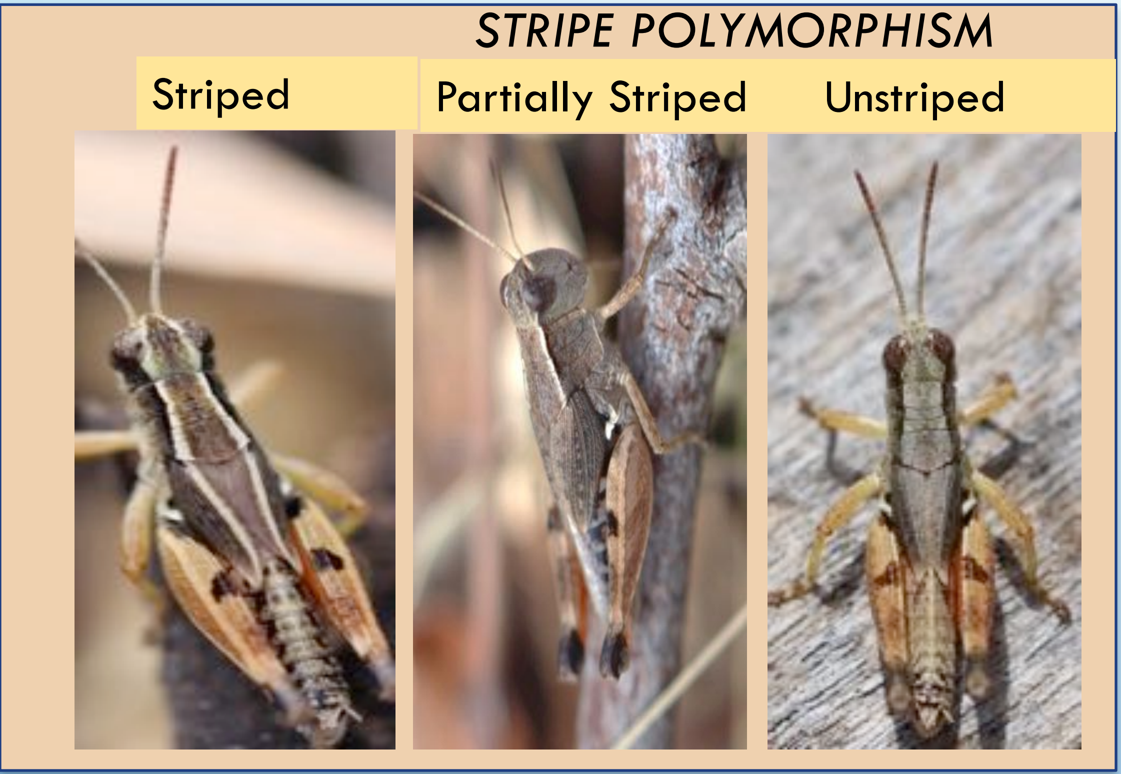
Morphological Variation Tracks Environmental Gradients in an Agricultural Pest, *Phaulacridium vittatum* (Orthoptera: Acrididae)

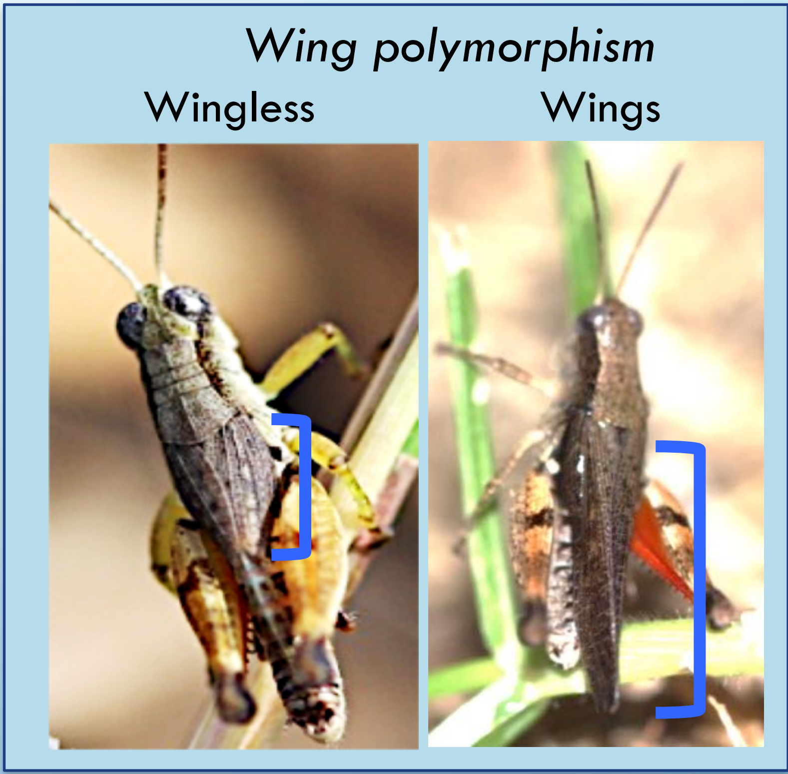
Sonu Yadav,[1,3] Adam J. Stow,[1] Rebecca M. B. Harris,[2] and Rachael Y. Dudaniec[1]

Yadav et al. 2018 – J Insect Sci

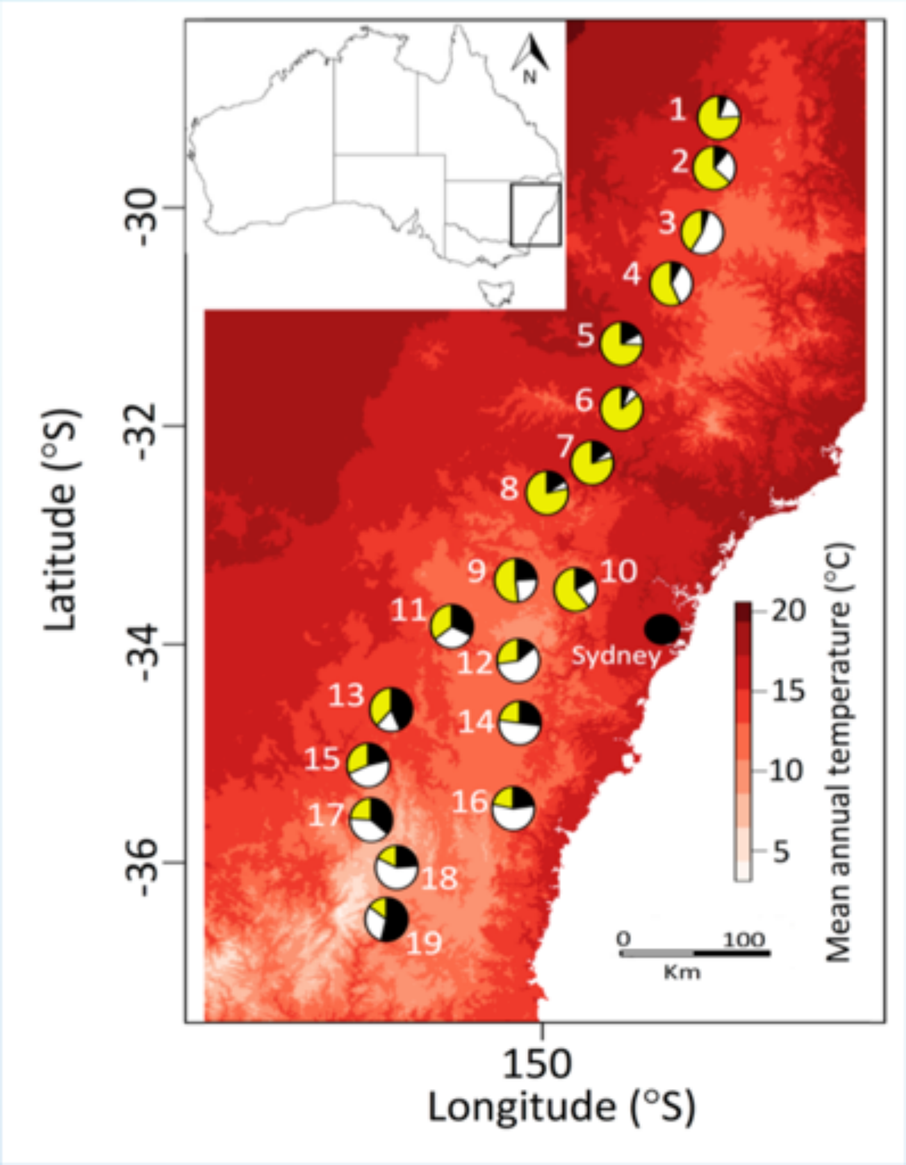
Morphological variation in *P. vittatum*

- Striped morph increases with solar radiation,
- Winged morph increases with high forest cover (*Yadav et al. 2018, J Insect Sci)*



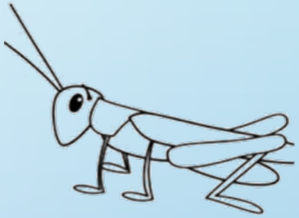


Gradient sampling and neutral genetic structure



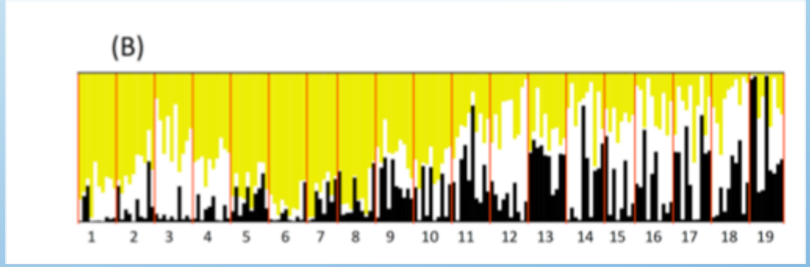
New South Wales, Australia

- 185 *P. vittatum* (2017)
- ddRADseq = 11,464 SNPs
- 900km, spanning 6.5° latitude
- 6-17°C/ 130-1600m



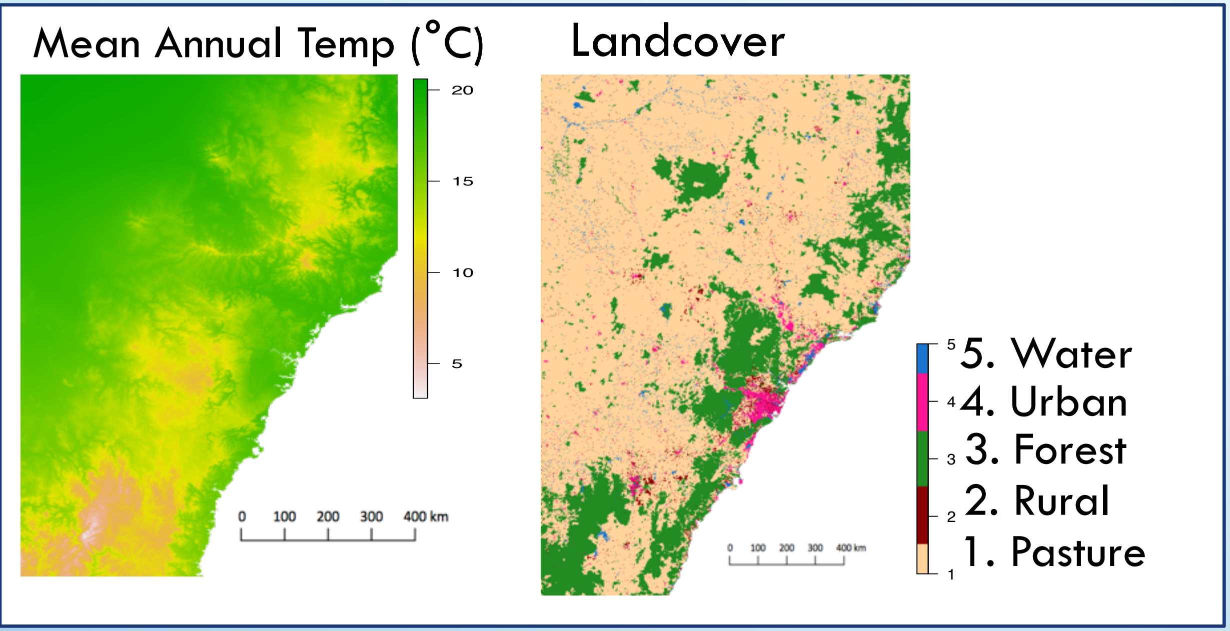
High admixture

Pairwise site Fst = 0.0003-0.08



Yadav et al. 2019, Mol Ecol.

Landscape genetic connectivity – resistance surfaces

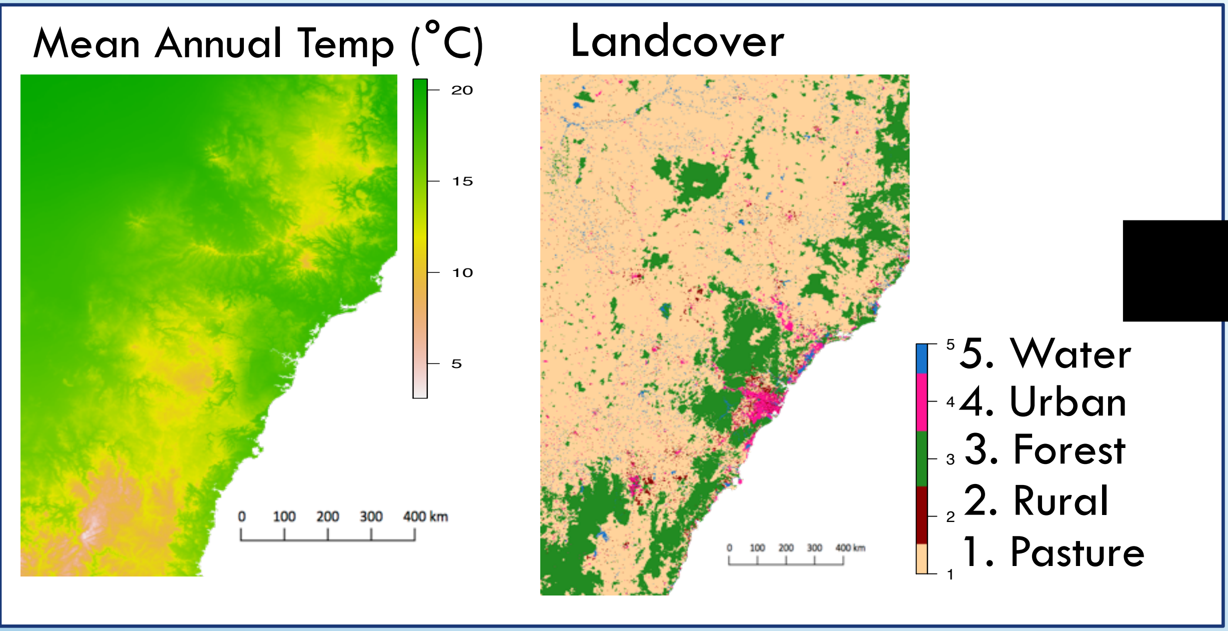


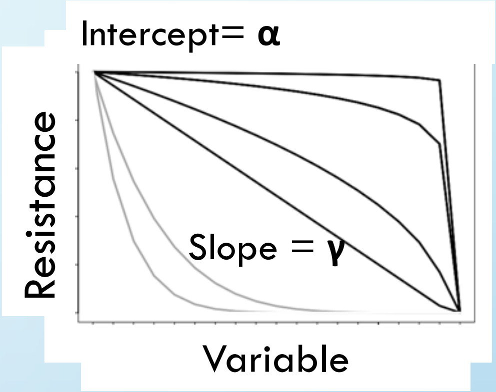
Mean Annual Temp (°C)

Landcover

Yadav et al. 2019, Mol Ecol.

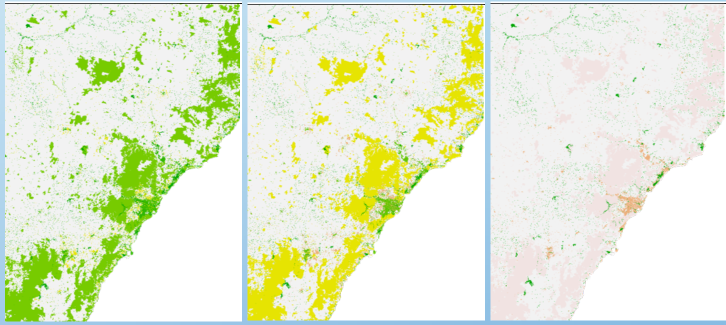
Landscape genetic connectivity – resistance surfaces





Resistance = 1+α (cell value-1/n-1) ^γ

$$\text{Resistance} = 1+\alpha\,(\text{cell value}-1/n-1)^{\gamma}$$



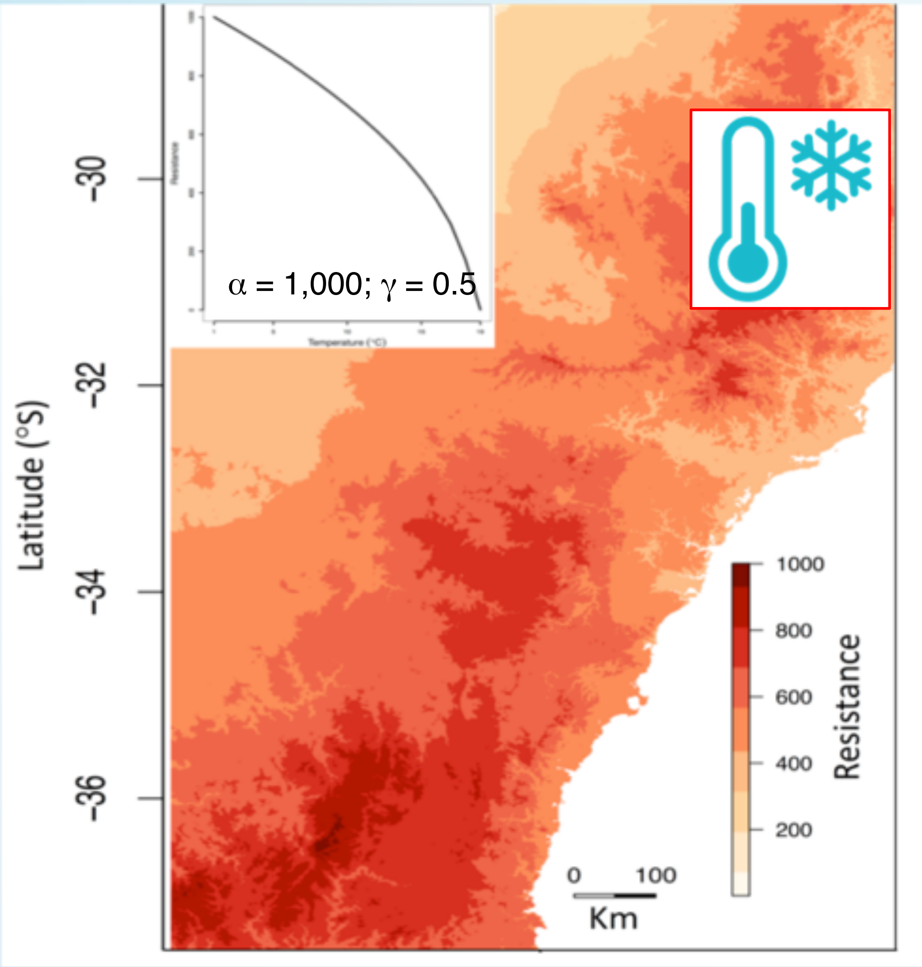
2x 35 resistance distance (Circuitscape)
~genetic distance (AIC model selection)

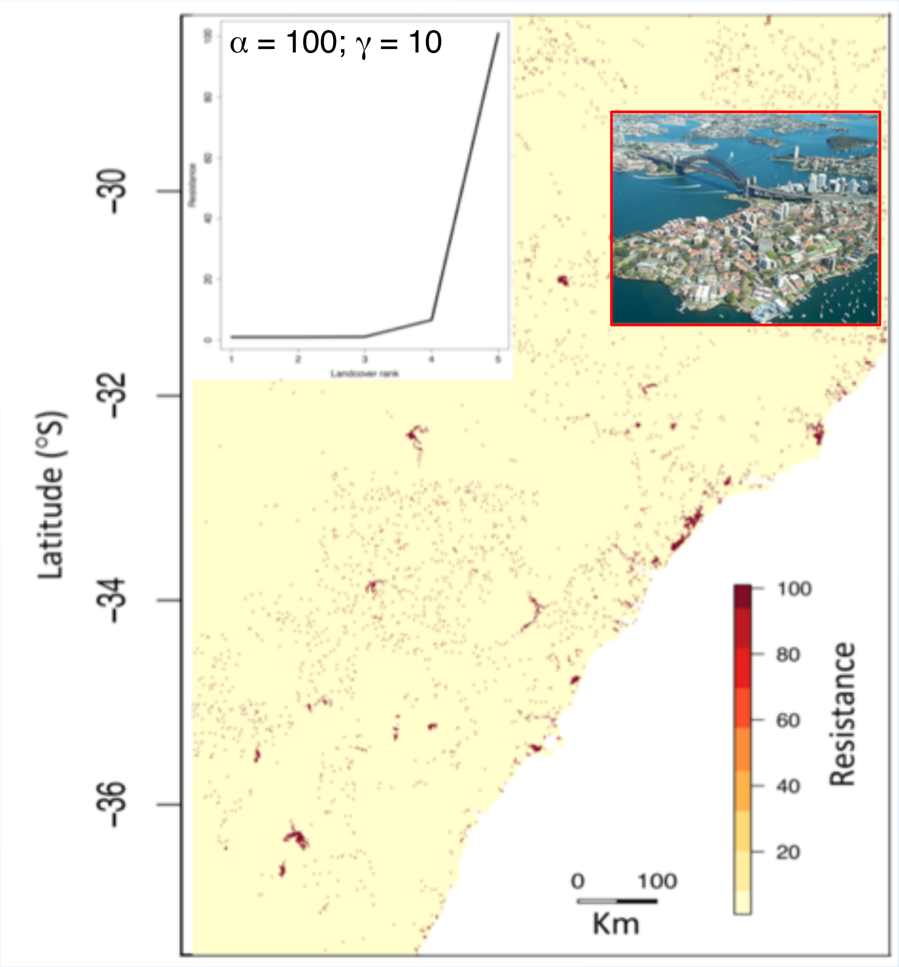
Yadav et al. 2019, Mol Ecol.

Result: Temperature best predicts genetic distance

Higher resistance
= low temperature

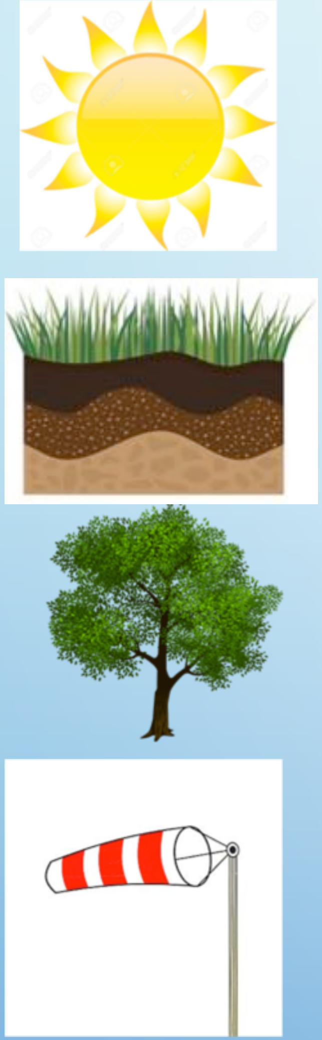
Higher resistance = urban + water
10x less resistant than temperature



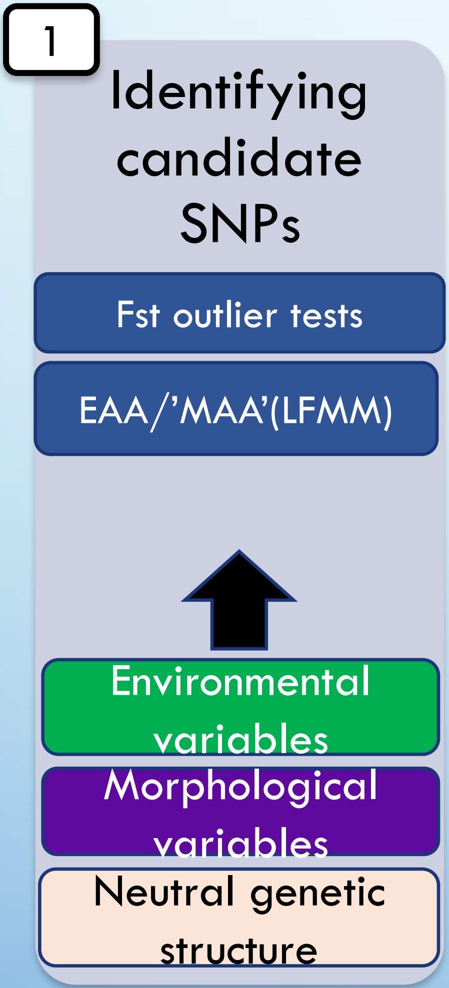


Yadav et al. 2019, Mol Ecol.

Methods: Detecting env/morph selection



+ LATITUDE



1

Identifying candidate SNPs

Fst outlier tests

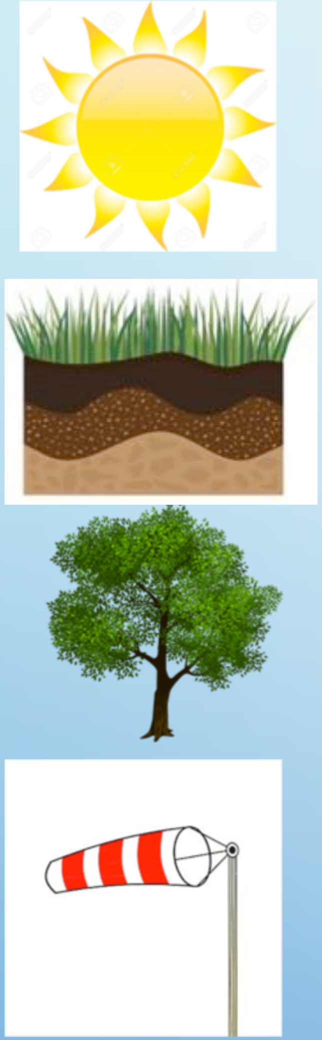
EAA/'MAA'(LFMM)

Environmental variables

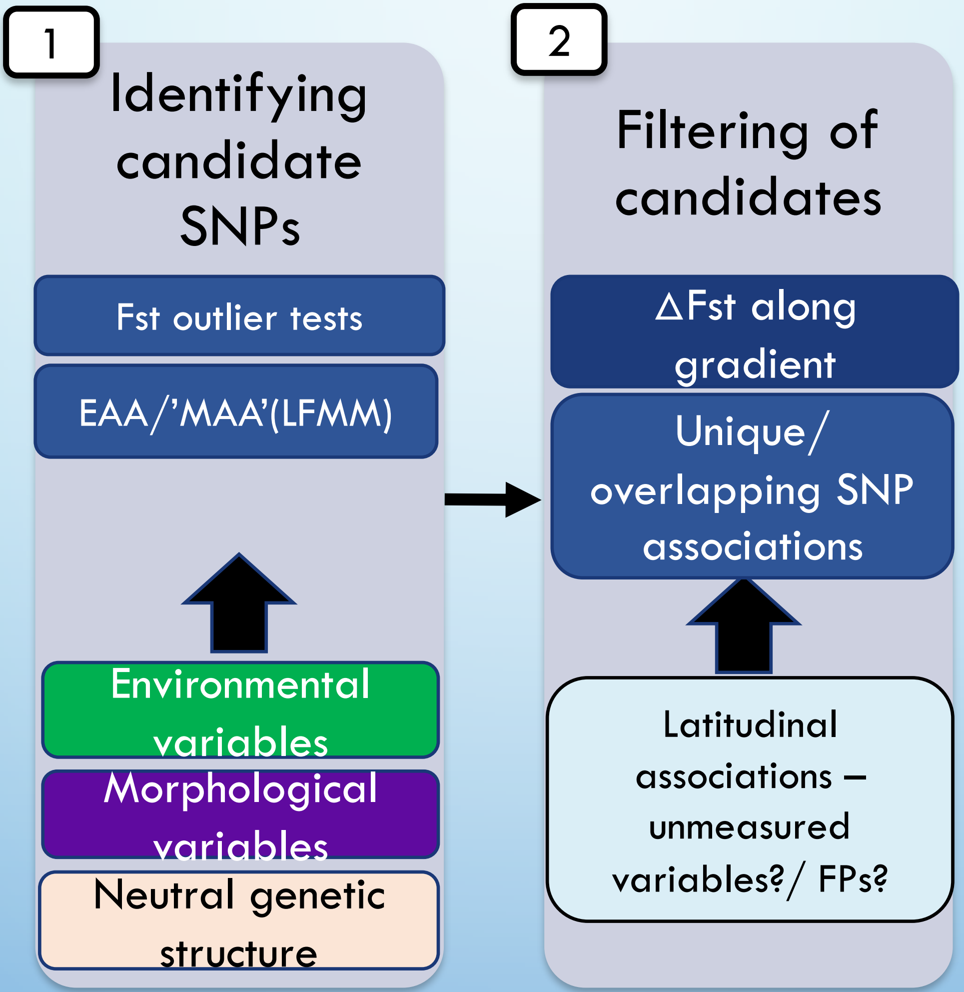
Morphological variables

Neutral genetic structure

Methods: Detecting env/morph selection



+ LATITUDE



1

Identifying candidate SNPs

Fst outlier tests

EAA/'MAA'(LFMM)

Environmental variables

Morphological variables

Neutral genetic structure

2

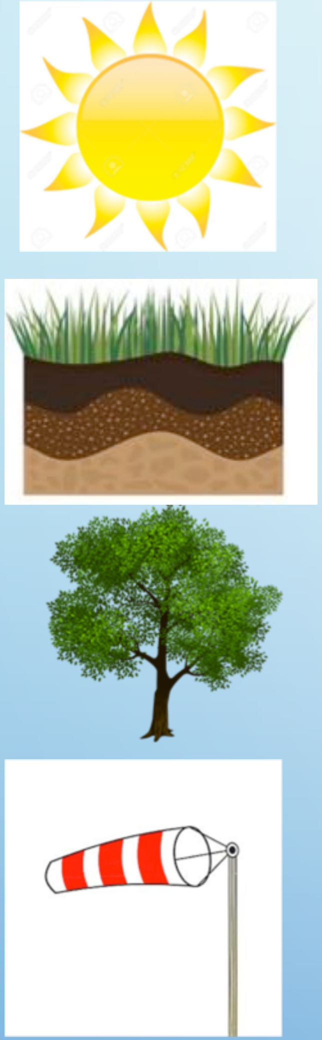
Filtering of candidates

ΔFst along gradient

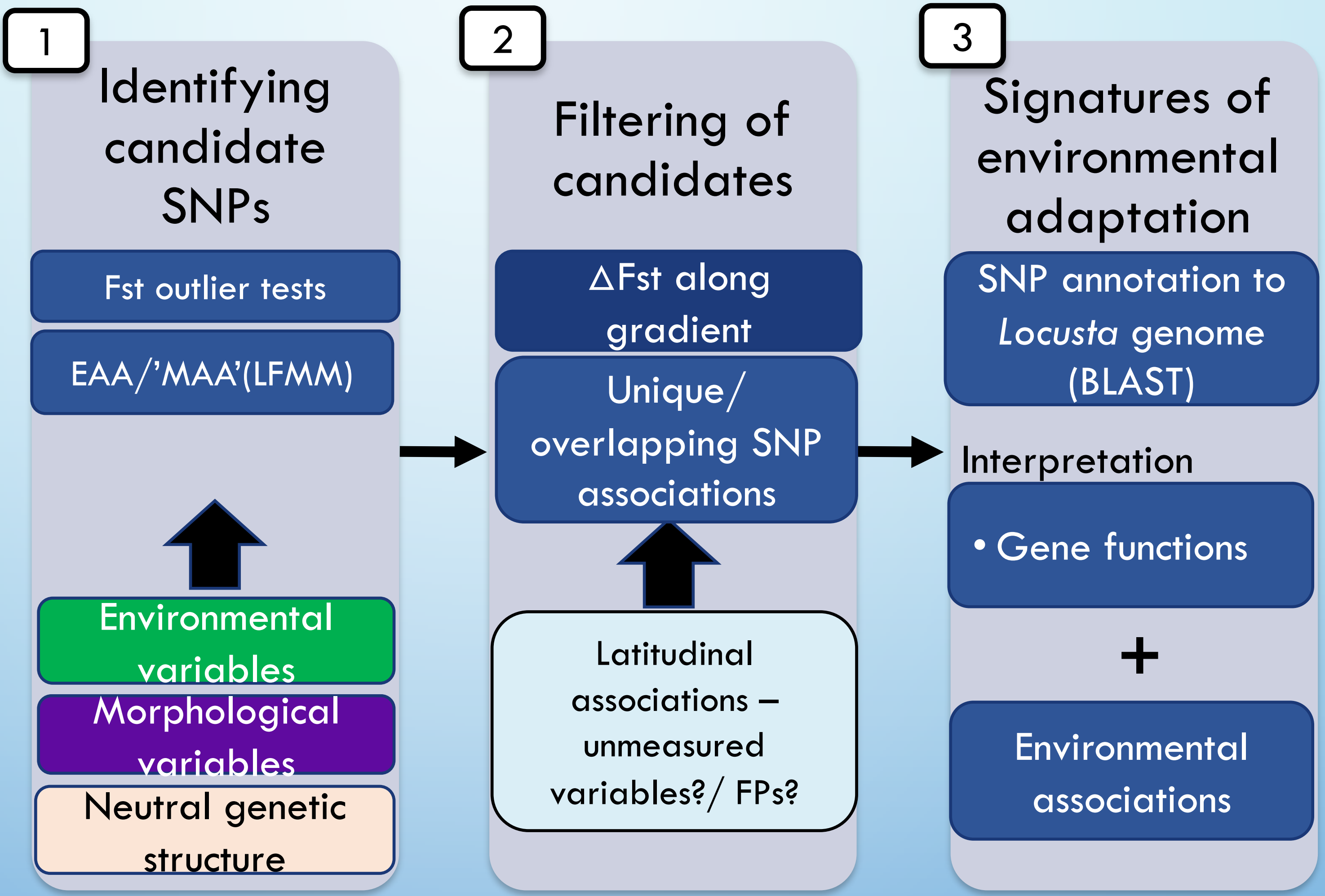
Unique/ overlapping SNP associations

Latitudinal associations – unmeasured variables?/ FPs?

Methods: Detecting env/morph selection



+ LATITUDE



1

Identifying candidate SNPs

- Fst outlier tests
- EAA/'MAA'(LFMM)

Environmental variables

Morphological variables

Neutral genetic structure

2

Filtering of candidates

- ΔFst along gradient
- Unique/ overlapping SNP associations

Latitudinal associations – unmeasured variables?/ FPs?

3

Signatures of environmental adaptation

SNP annotation to *Locusta* genome (BLAST)

Interpretation

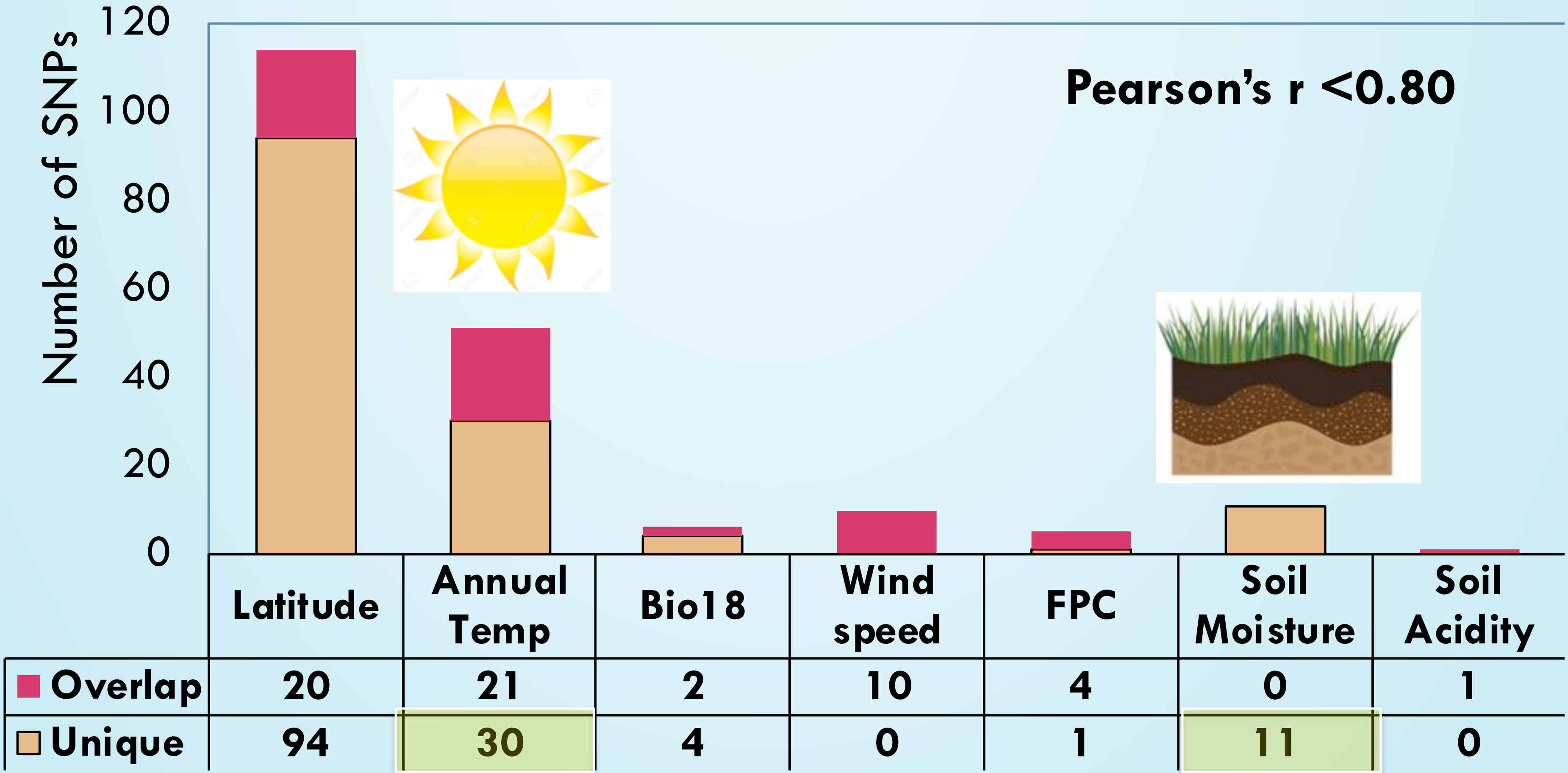
- Gene functions

\+

Environmental associations

Result: Environmental adaptation signatures

- Plus 17 Fst outliers Bayescan + OutFLANK
- Latitude: other variables? Confounded by structure?

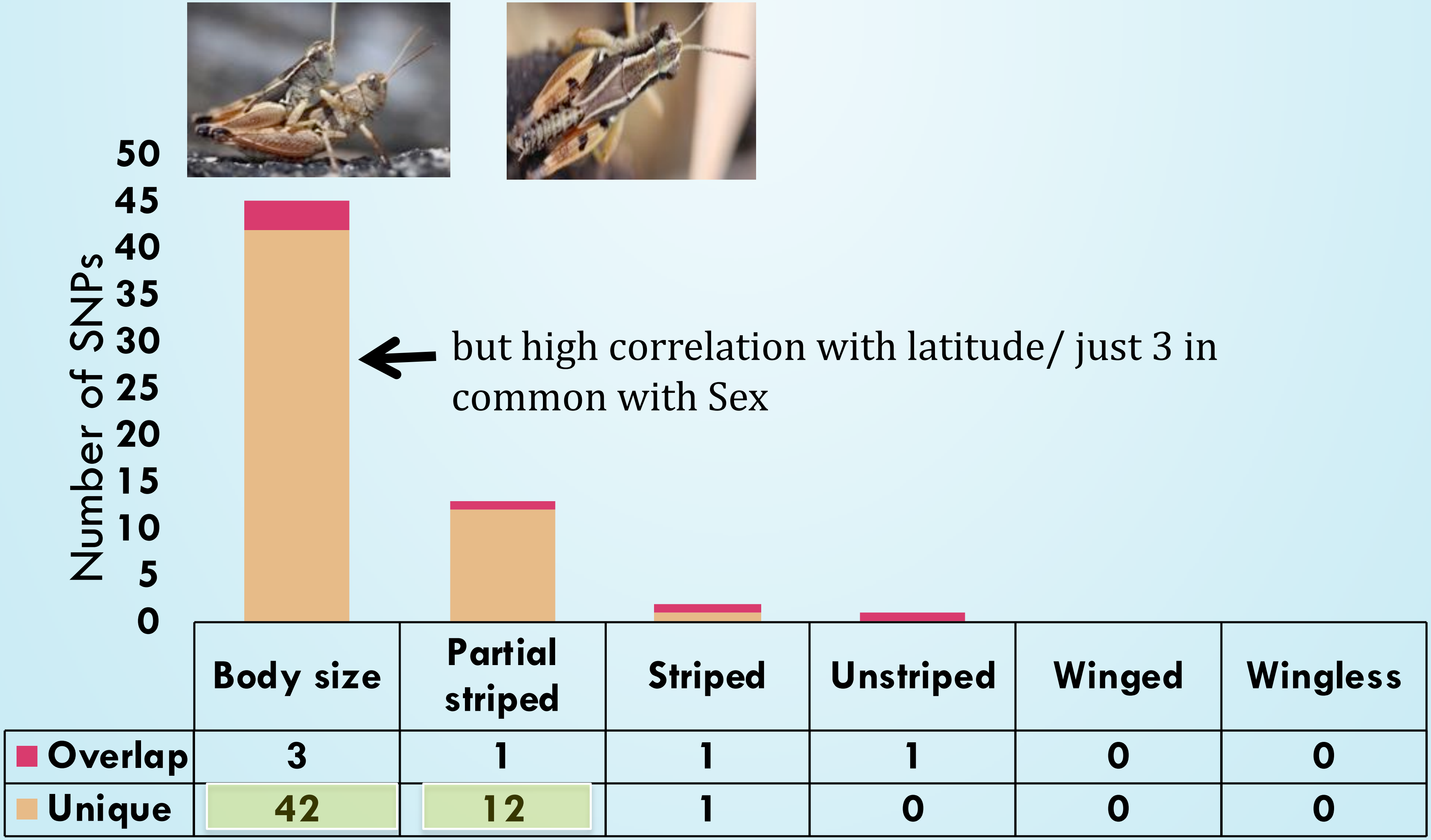


Pearson's r <0.80

	Latitude	Annual Temp	Bio18	Wind speed	FPC	Soil Moisture	Soil Acidity
🟥 Overlap	20	21	2	10	4	0	1
🟧 Unique	94	30	4	0	1	11	0

Yadav et al. 2019, Mol Ecol

Result: Morphological adaptation signatures

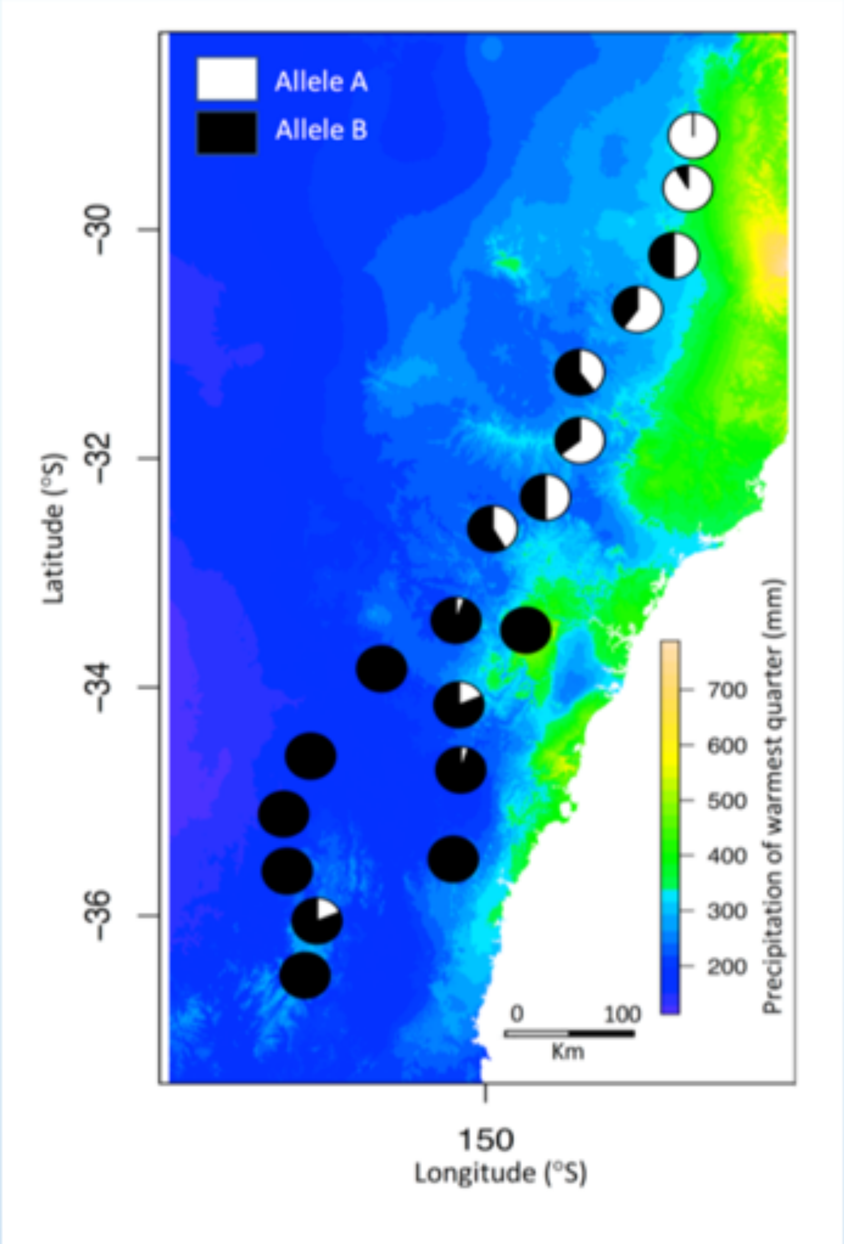


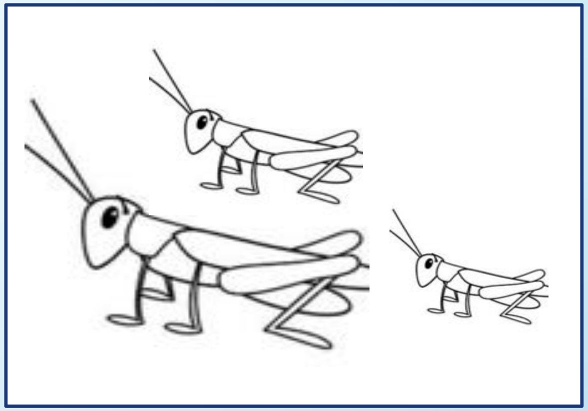
but high correlation with latitude/ just 3 in common with Sex

	Body size	Partial striped	Striped	Unstriped	Winged	Wingless
🟥 Overlap	3	1	1	1	0	0
🟧 Unique	42	12	1	0	0	0

Yadav et al. 2019, Mol Ecol.

Result: Allele frequency change along gradient



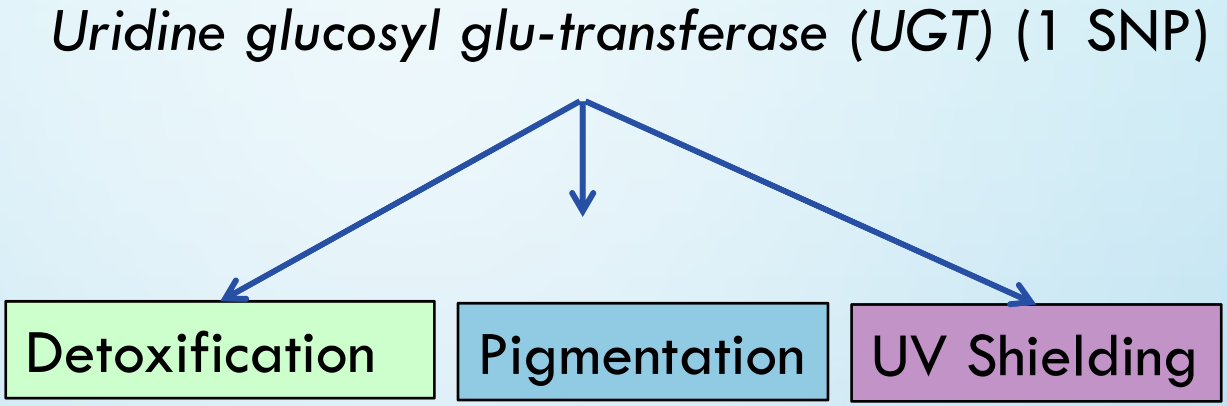


1 SNP with high allelic turnover for **body size**. Detected in both Fst outlier tests + morph EAA

Yadav et al. 2019, Mol Ecol.

Result: Gene annotation *(to locust genome)*

Temperature



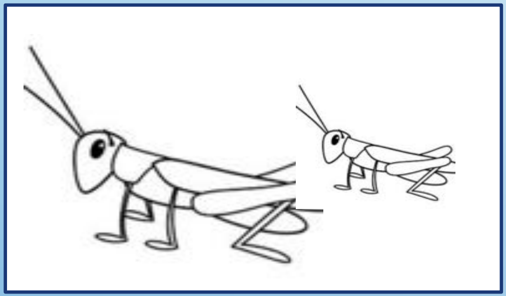
Uridine glucosyl glu-transferase (UGT) (1 SNP)

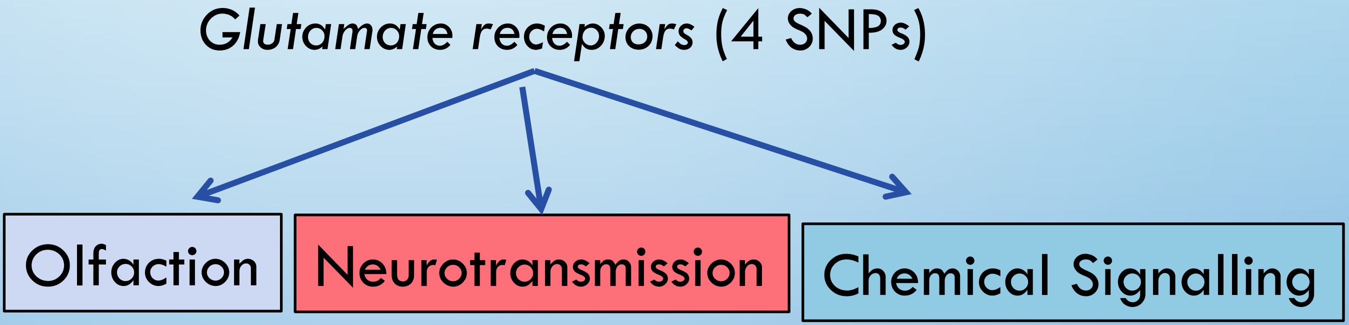
Detoxification

Pigmentation

UV Shielding

Body size





Glutamate receptors (4 SNPs)

Olfaction

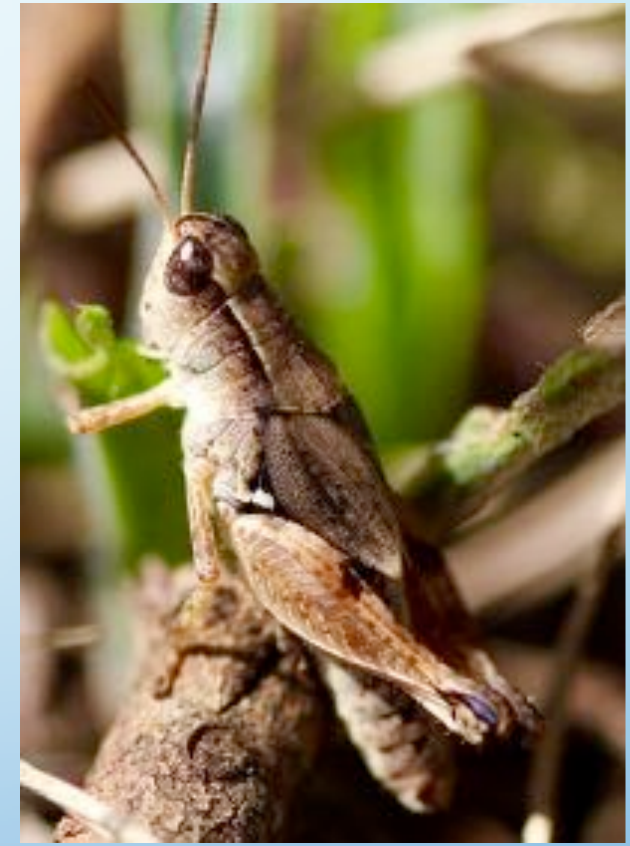
Neurotransmission

Chemical Signalling

Yadav et al. 2019, Mol Ecol

Case study 1: Conclusions

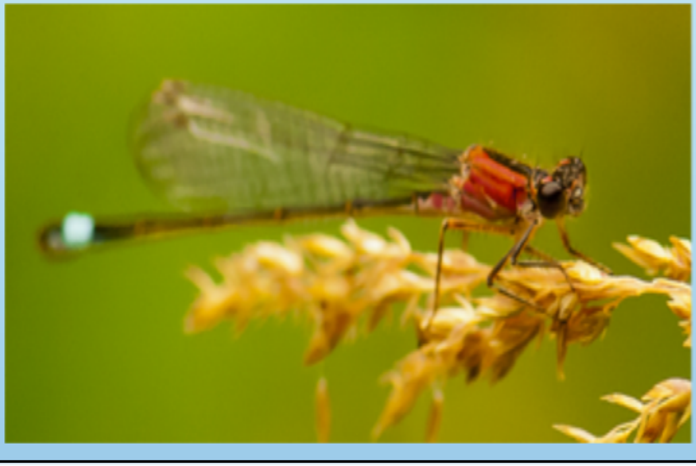
- Temperature has greatest effect on both neutral connectivity and local adaptation
- Selection on morphology may facilitate local adaptation
- Warming temperatures and agricultural expansion will facilitate future spread and outbreaks



CASE STUDY

(2) Range expanding damselflies

Ischnura elegans
Blue-tailed damselfly
(Odonata)



ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*)

Rachael Y. Dudaniec[1] | Chuan Ji Yong[1] | Lesley T. Lancaster[2] | Erik I. Svensson[3] | Bengt Hansson[3]

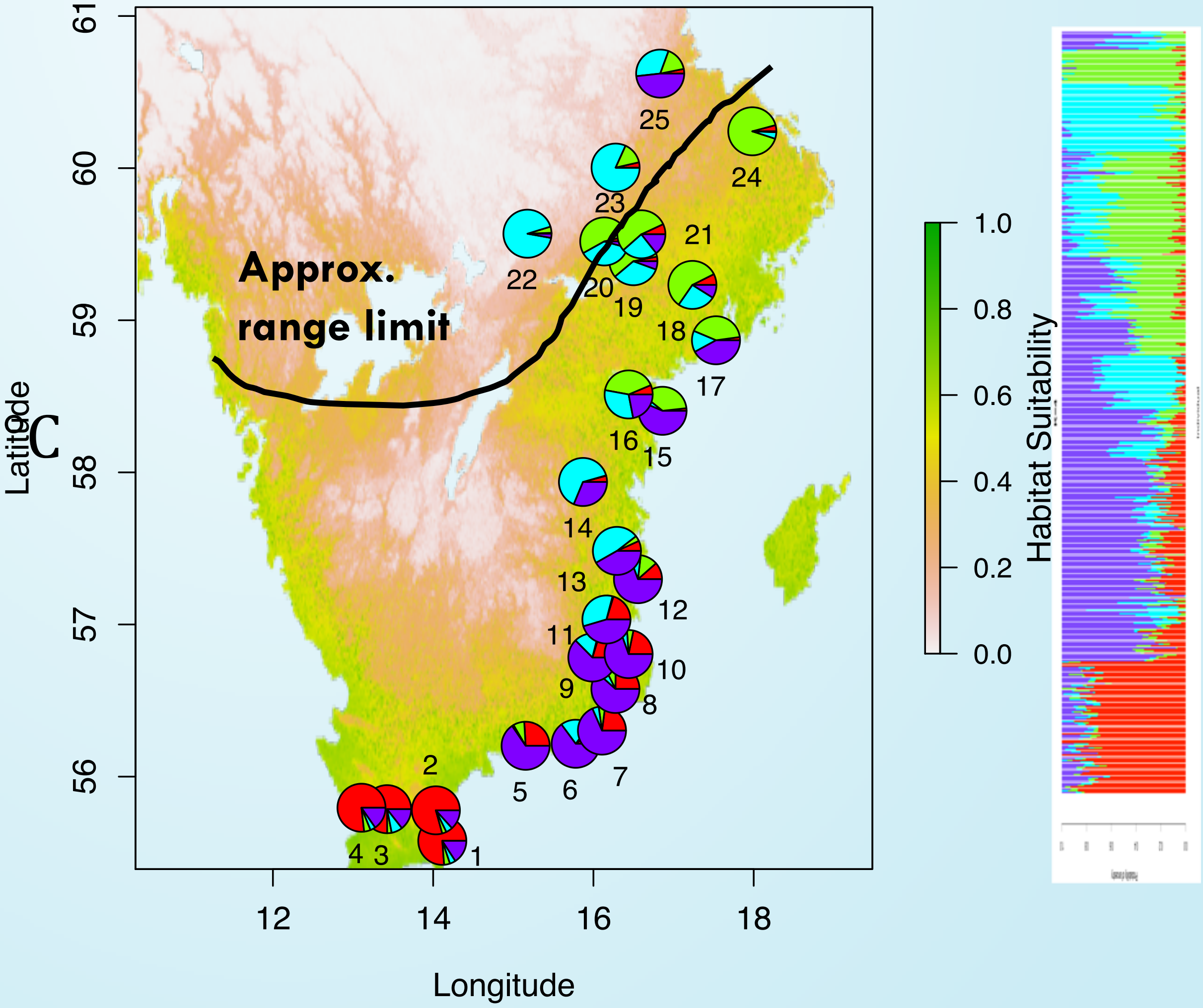
Dudaniec et al. 2018, Mol Ecol

Gradient sampling and neutral genetic structure

Southern Sweden

- 426 *I. elegans* (2013)
- RADseq/13,612 SNPs
- 500km gradient/3°C change

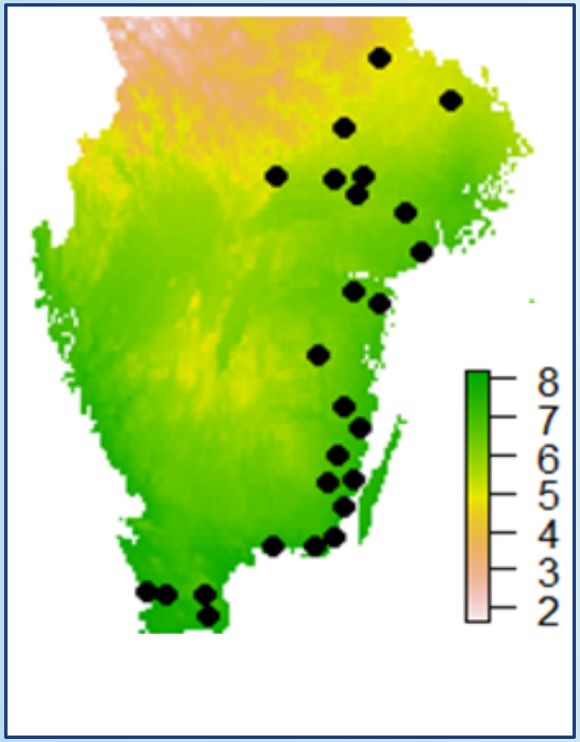
4 admixed clusters



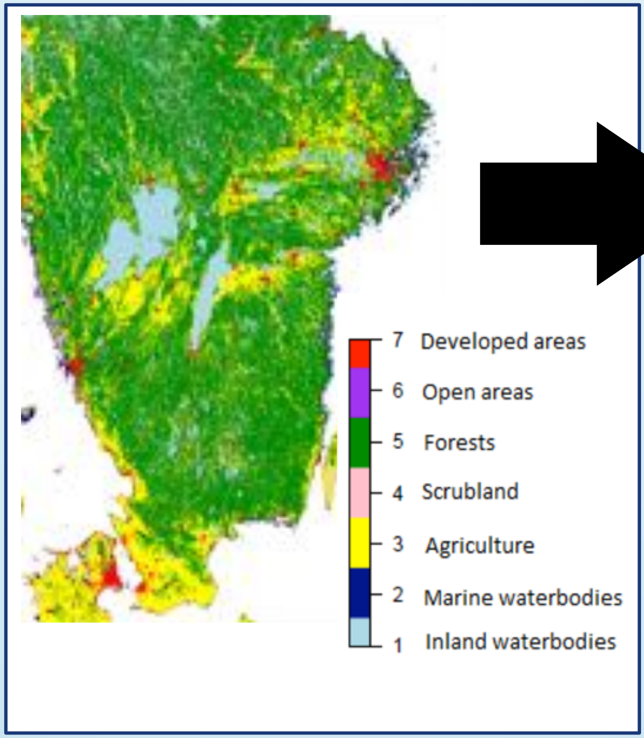
Dudaniec et al. 2018, Mol Ecol

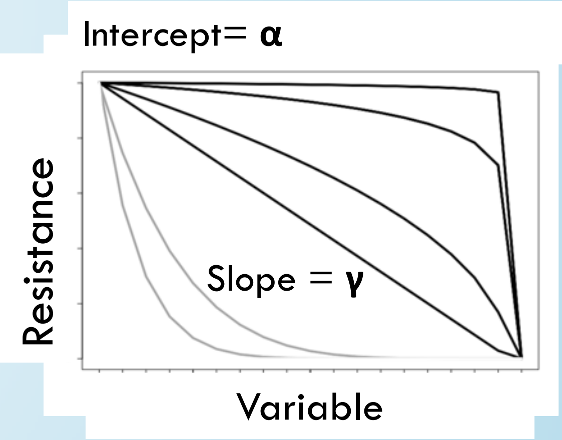
Landscape genetic connectivity – resistance surfaces

Mean annual temperature °C

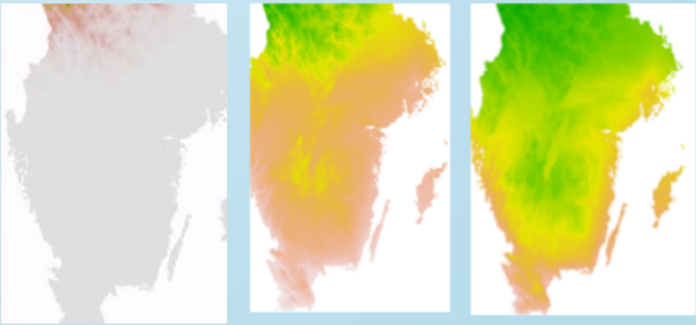


Landcover





29x2 resistance surfaces each for
annual temp and land cover data

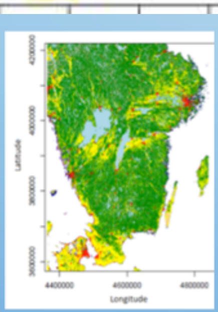


Genetic distance~Resistance distance
(mixed effects model [MLPE.] +AIC).

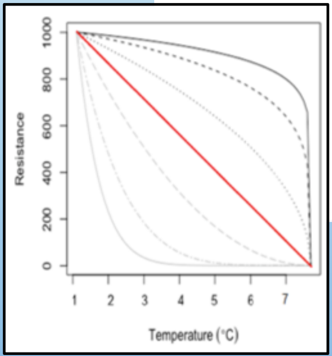
Carey et al. (unpublished)

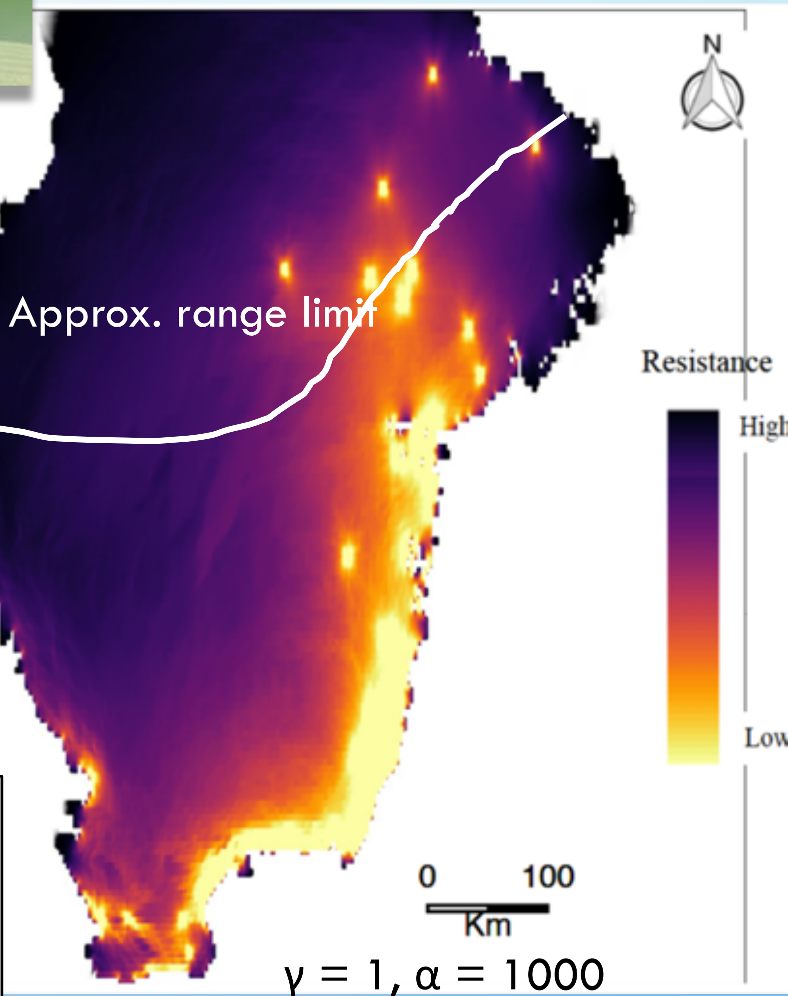
Result: Temperature drives neutral gene flow

IBD outperformed all land cover models



Mean annual temperature affects genetic distance linearly

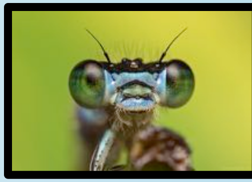


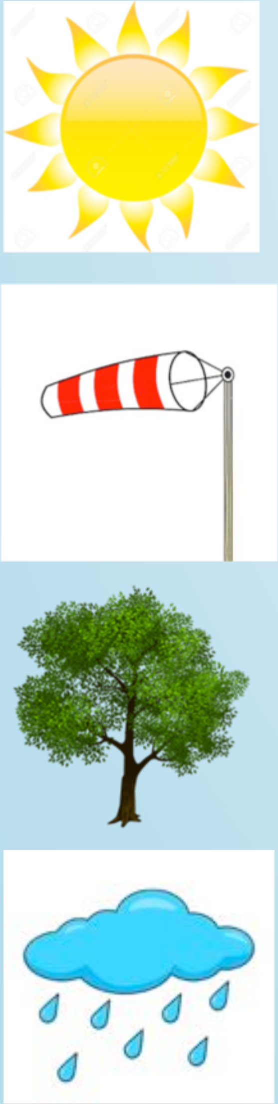


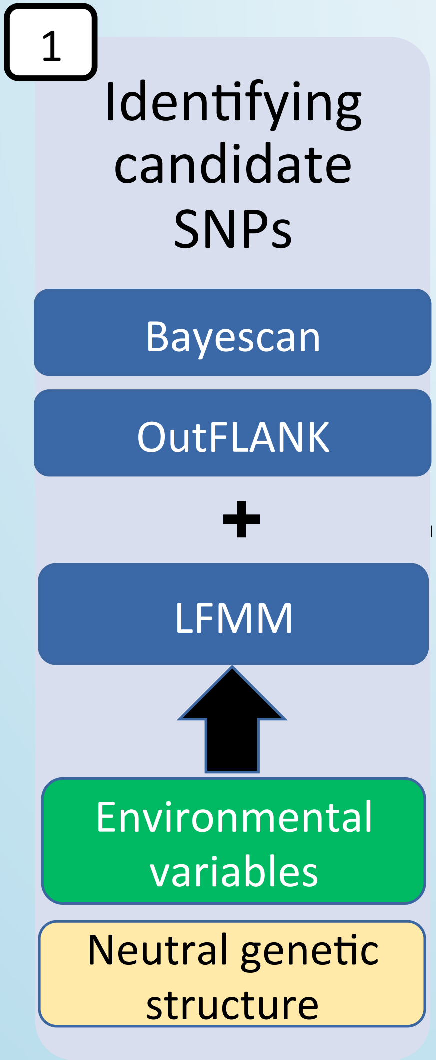
γ = 1, α = 1000

Carey et al. (unpublished)

Methods: Detecting adaptive genetic variation







1

Identifying candidate SNPs

Bayescan

OutFLANK

+

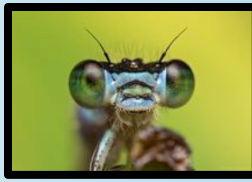
LFMM

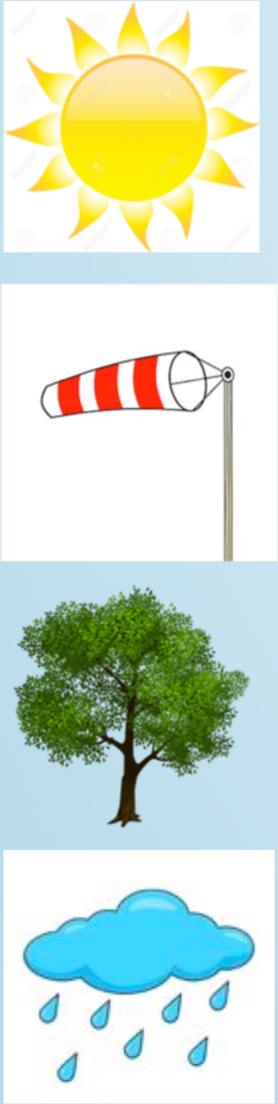
Environmental variables

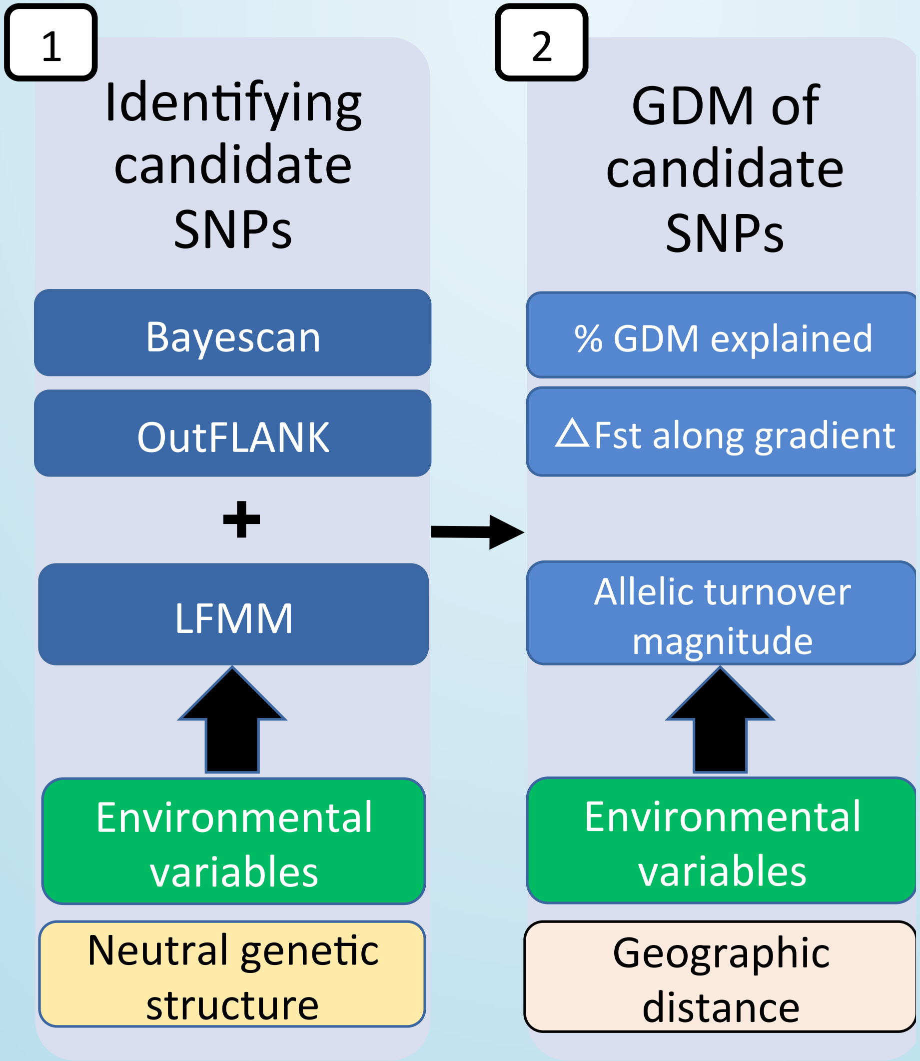
Neutral genetic structure

Dudaniec et al. 2018, Mol Ecol

Methods: Detecting adaptive genetic variation







1

Identifying candidate SNPs

Bayescan

OutFLANK

+

LFMM

Environmental variables

Neutral genetic structure

2

GDM of candidate SNPs

% GDM explained

△Fst along gradient

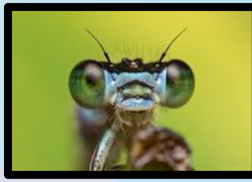
Allelic turnover magnitude

Environmental variables

Geographic distance

Dudaniec et al. 2018, Mol Ecol

Methods: Detecting adaptive genetic variation



1 Identifying candidate SNPs

- Bayescan
- OutFLANK

+

- LFMM

↑

- Environmental variables
- Neutral genetic structure

→

2 GDM of candidate SNPs

- % GDM explained
- △Fst along gradient

- Allelic turnover magnitude

↑

- Environmental variables
- Geographic distance

→

3 Signatures of environmental adaptation

SNP annotation to transcriptome (BLAST2GO)

Interpretation
- Gene function
- Experimental data
- Prior knowledge

Environmental associations

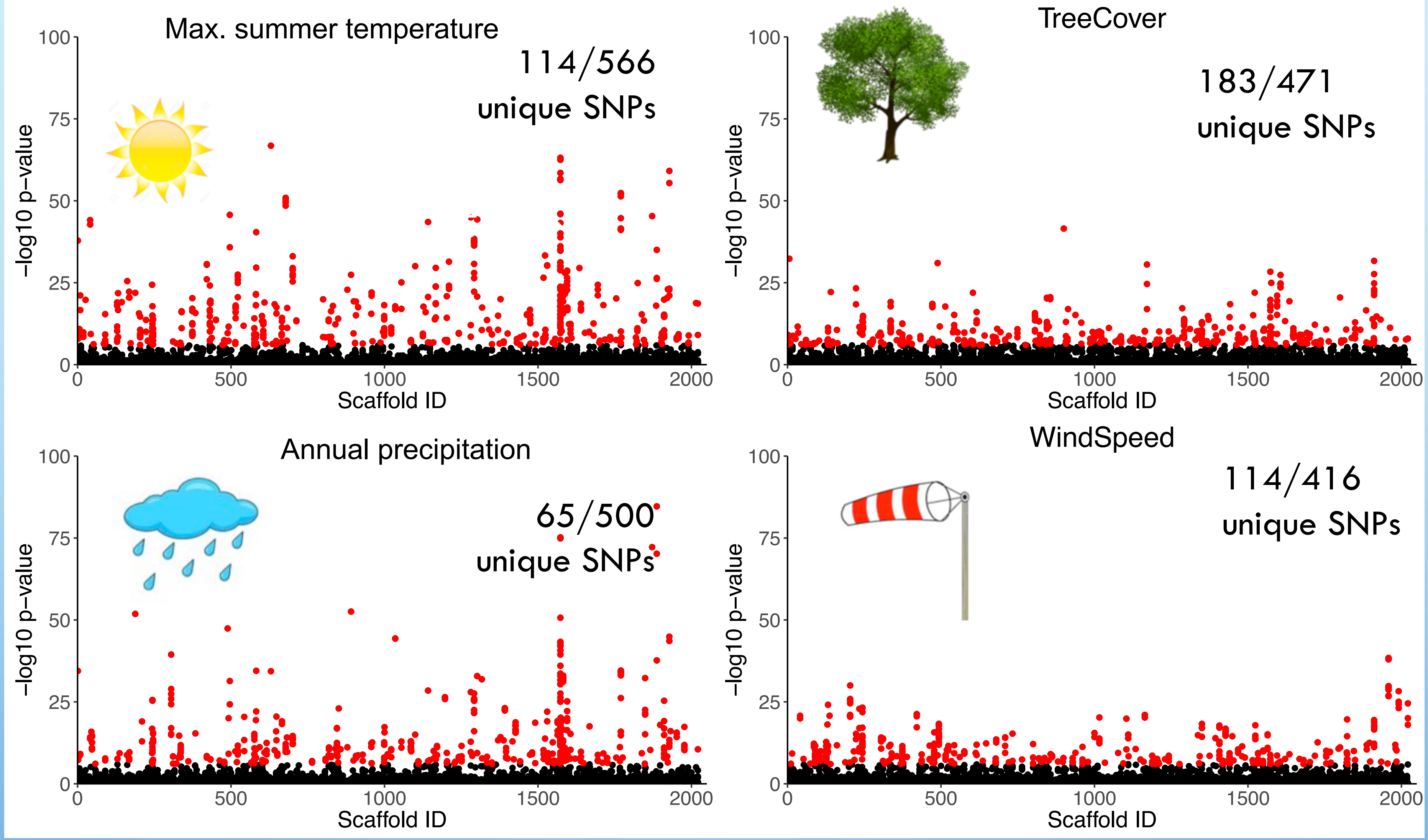
+

Allelic turnover relationships

Dudaniec et al. 2018, Mol Ecol

Result: Environmental selection signatures

1251 total SNP associations (LFMM)



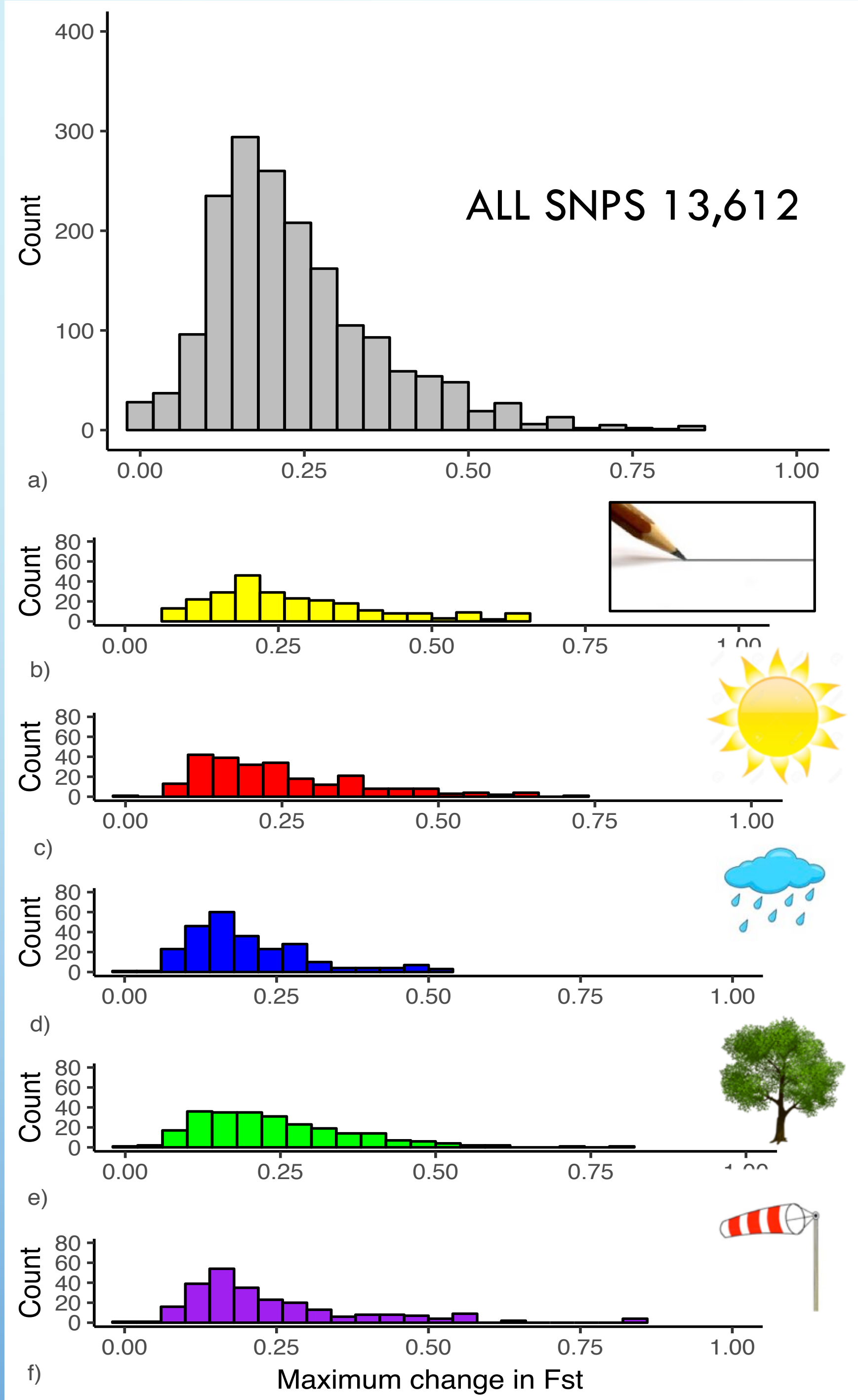
114/566 unique SNPs

183/471 unique SNPs

65/500 unique SNPs

114/416 unique SNPs

Dudaniec et al. 2018, Mol Ecol



F_{ST} DISTRIBUTIONS OF ALL CANDIDATES PER VARIABLE

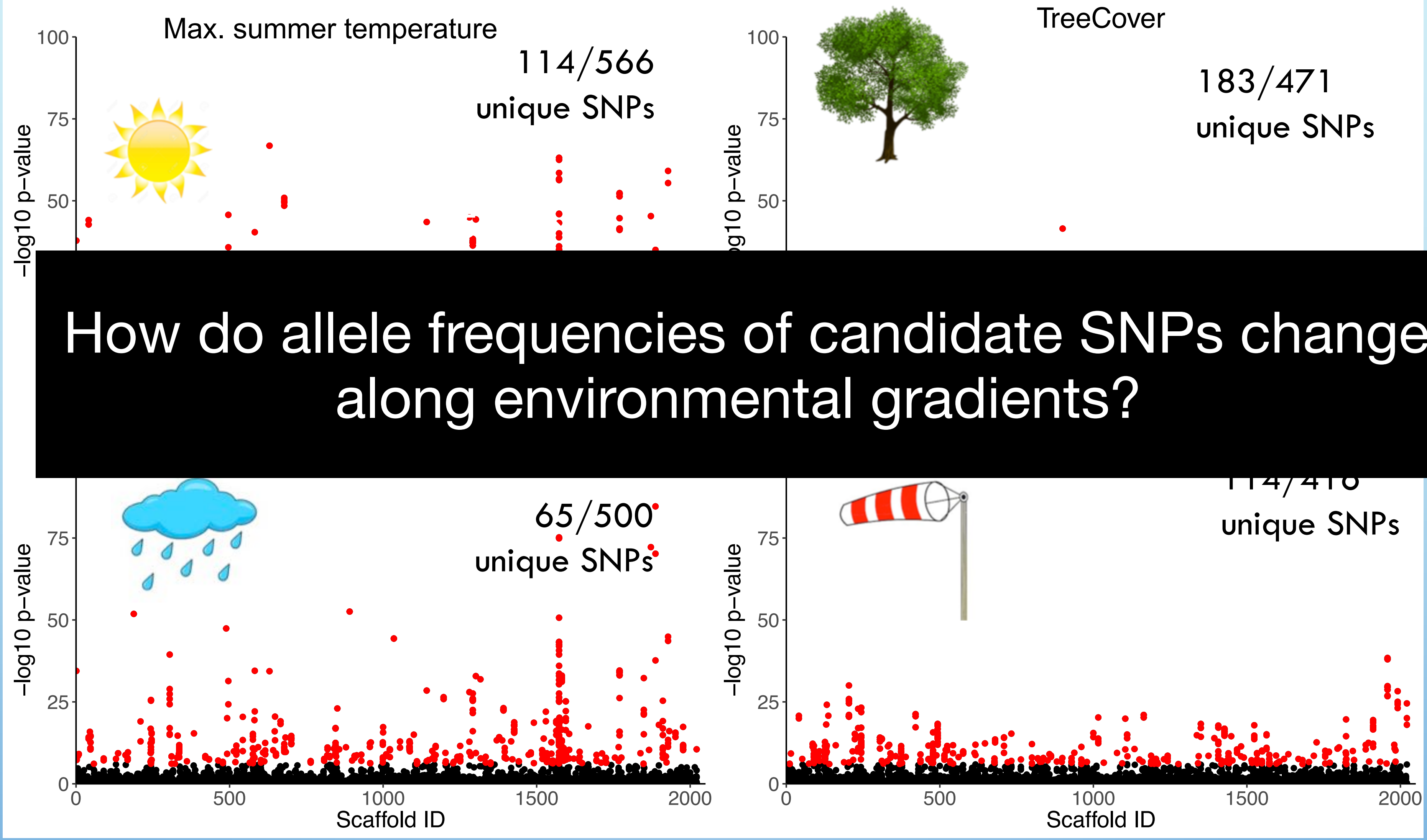
Similar to all SNPs combined, variable, but mostly on lower end of Fst

Weak evidence for higher Fst among candidates

Many loci of smaleer effect?

Result: Environmental selection signatures

1251 total SNP associations (LFMM)

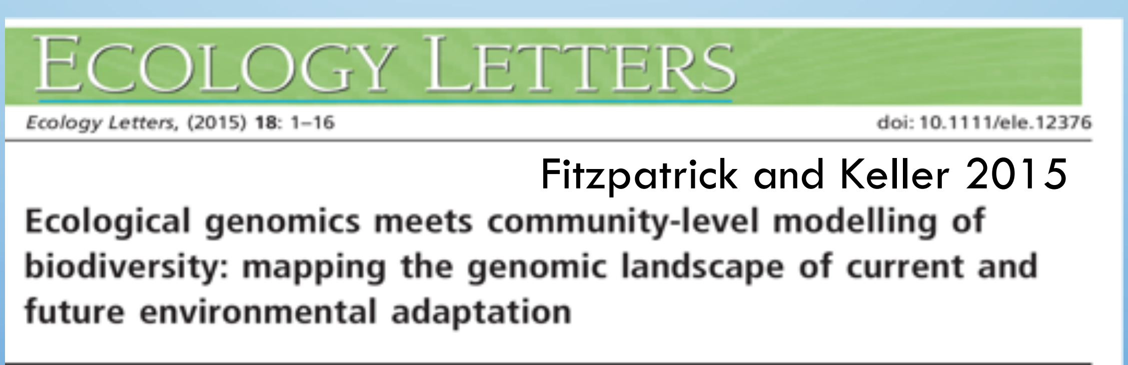


How do allele frequencies of candidate SNPs change along environmental gradients?

Dudaniec et al. 2018, Mol Ecol

GENERALISED DISSIMILARITY MODELING

- Based on the concept of species community-level compositional turnover functions (Ferrier et al. 2007).
- Non-linear, threshold responses of allele frequencies to environmental gradients – ‘allelic turnover’
- Uses distance matrices of env + genetic data, with splines and GLM to account for non-linearity
- *Gradient Forest* – similar outcomes/ turnover functions built differently



ECOLOGY LETTERS

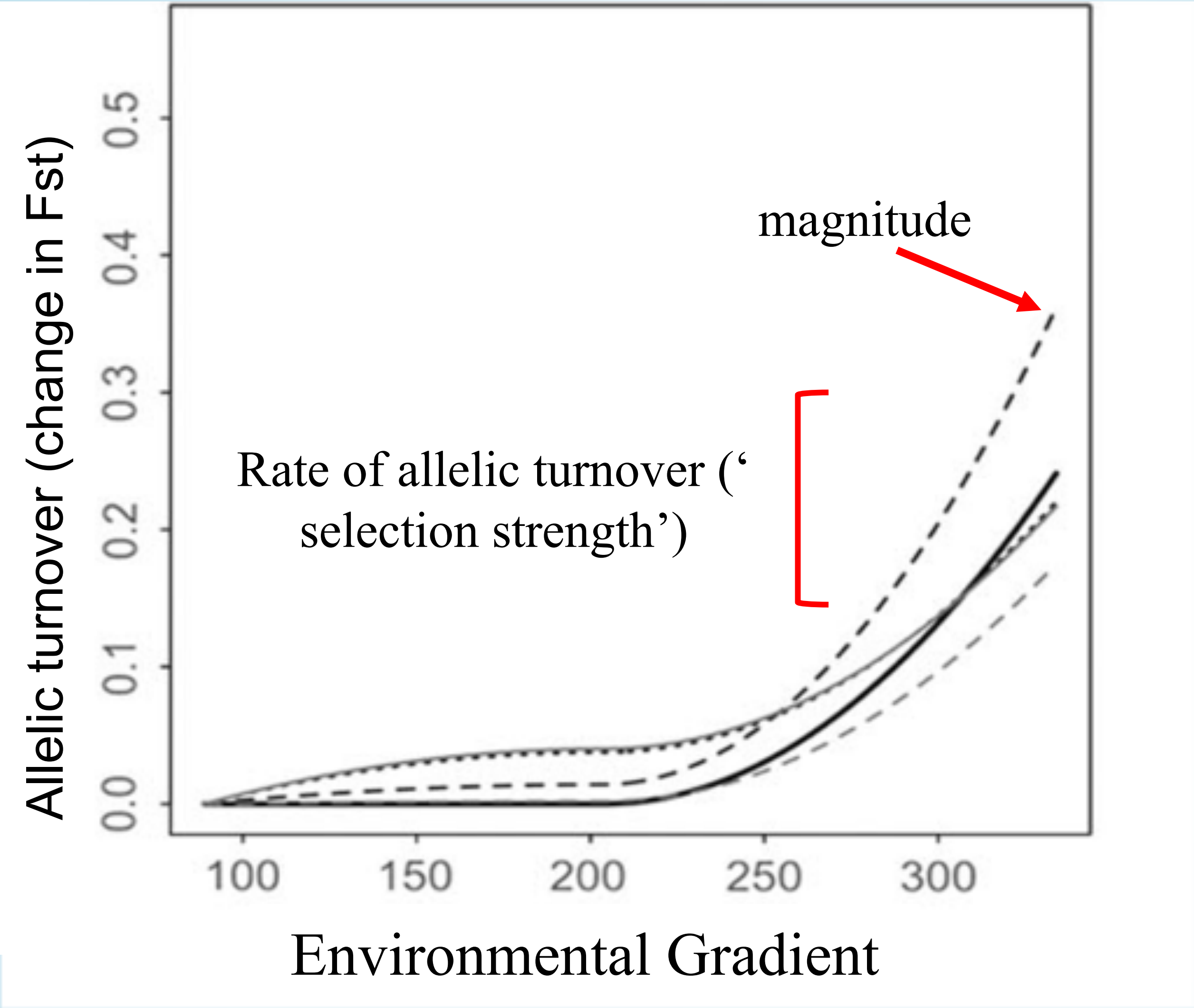
Ecology Letters, (2015) **18**: 1–16

doi: 10.1111/ele.12376

Fitzpatrick and Keller 2015

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

GENERALISED DISSIMILARITY MODELING



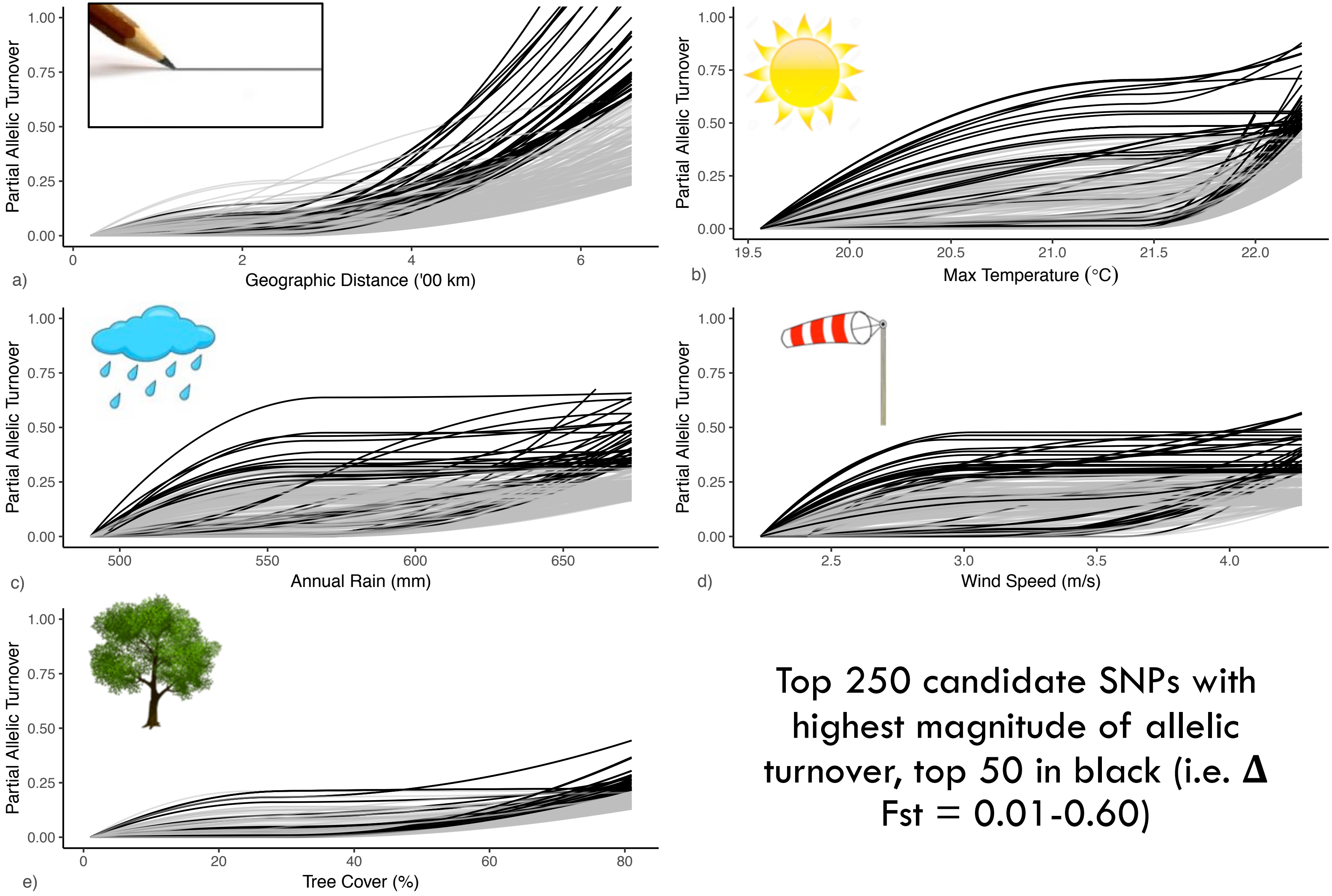
magnitude

Rate of allelic turnover (‘ selection strength’)

Overall magnitude = relative importance of predictor

Fitzpatrick and Keller(2015) –Ecology Letters

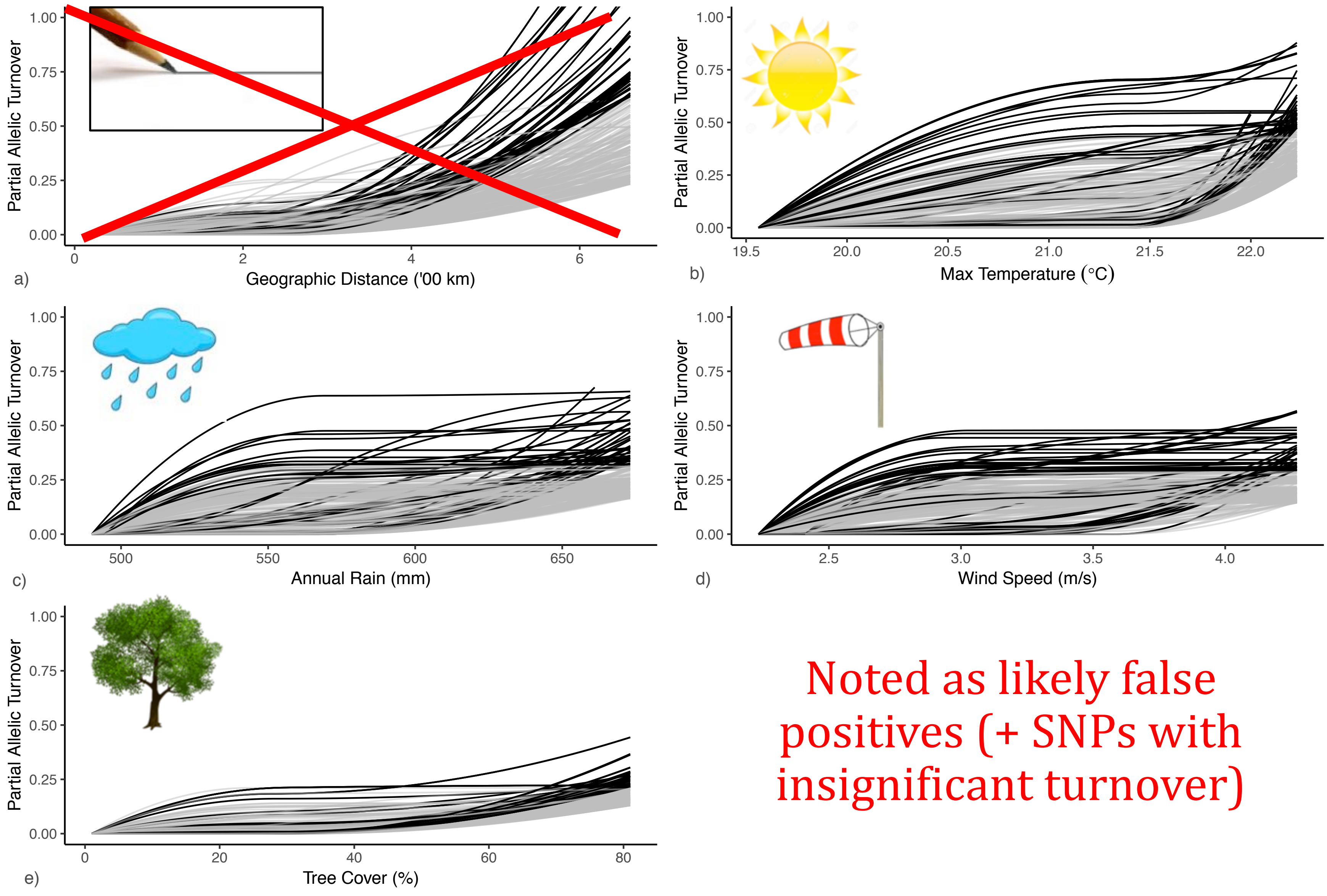
Result: Selection thresholds via allelic turnover (GDM)



Top 250 candidate SNPs with highest magnitude of allelic turnover, top 50 in black (i.e. Δ Fst = 0.01-0.60)

Dudaniec et al. 2018, Mol Ecol

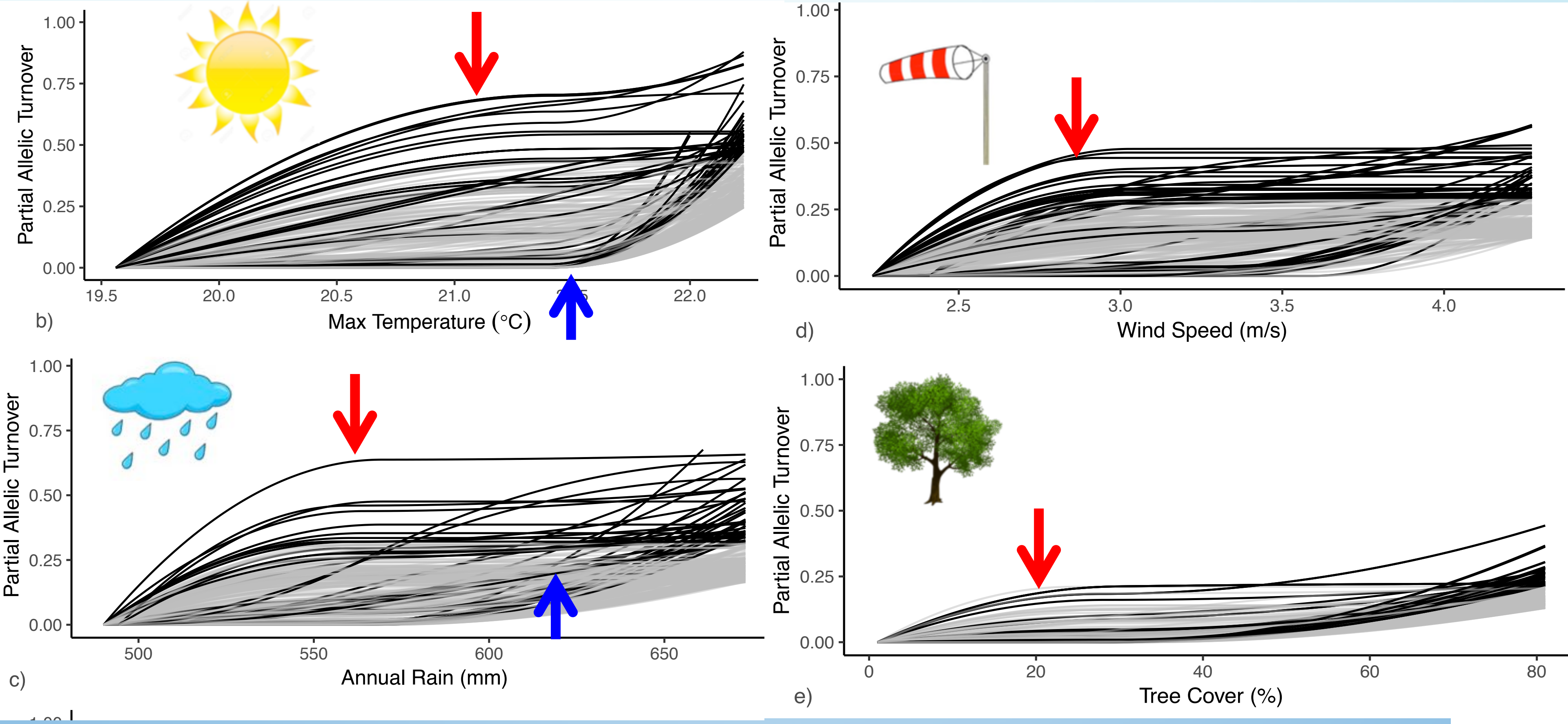
Result: Selection thresholds via allelic turnover (GDM)



Noted as likely false positives (+ SNPs with insignificant turnover)

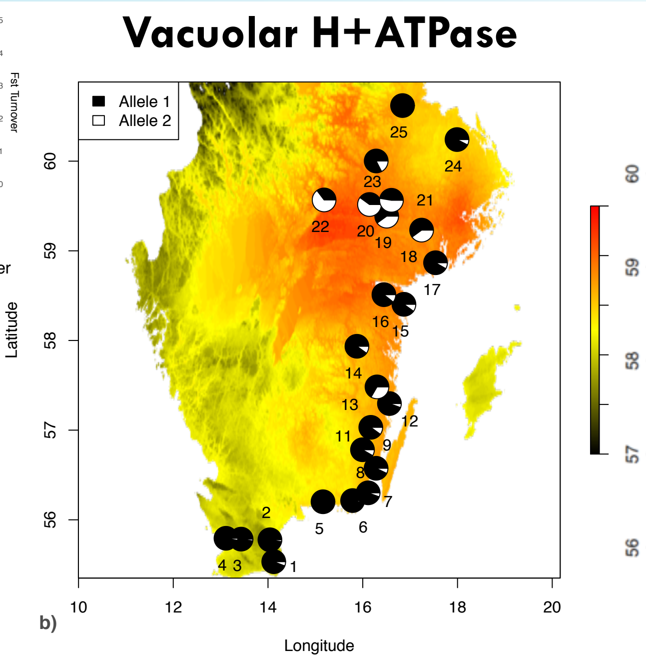
Dudaniec et al. 2018, Mol Ecol

Result: Selection thresholds via allelic turnover (GDM)



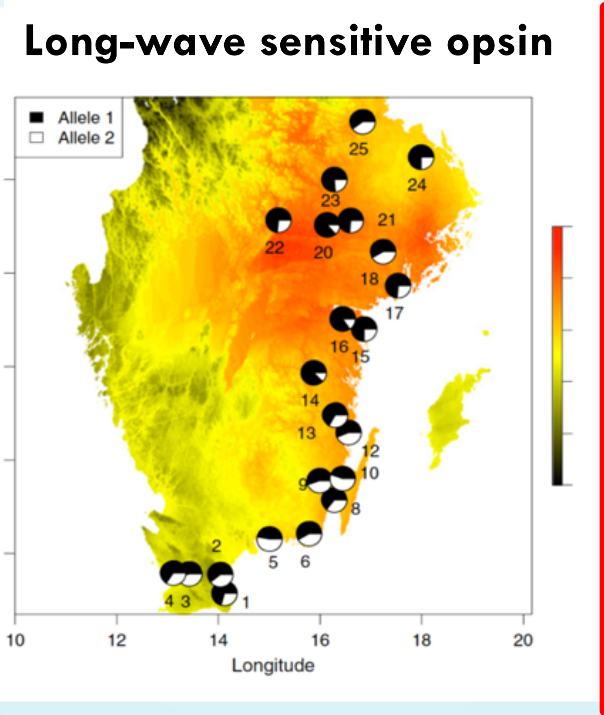
Dudaniec et al. 2018, Mol Ecol

Result: Allelic turnovers of annotated genes



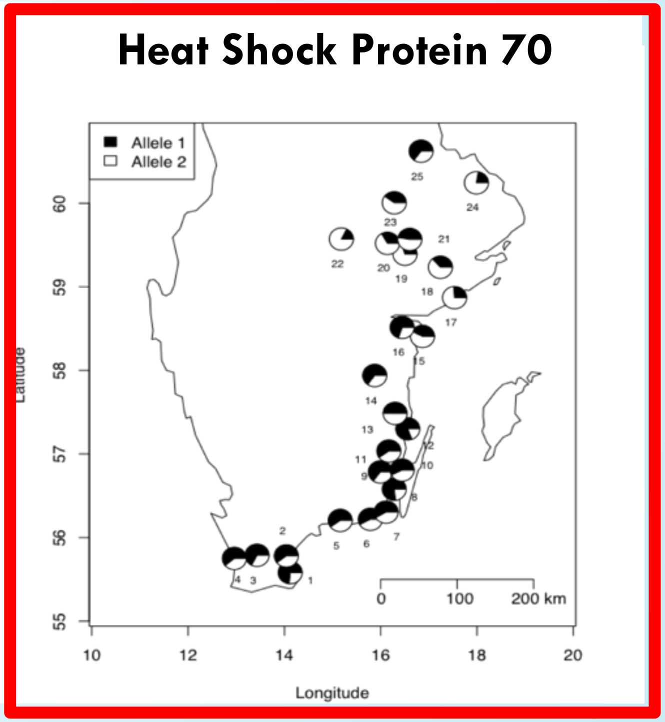
Vacuolar H+ATPase

Proton pump - aids in osmosis + cold tolerance



Long-wave sensitive opsin

Visual processing



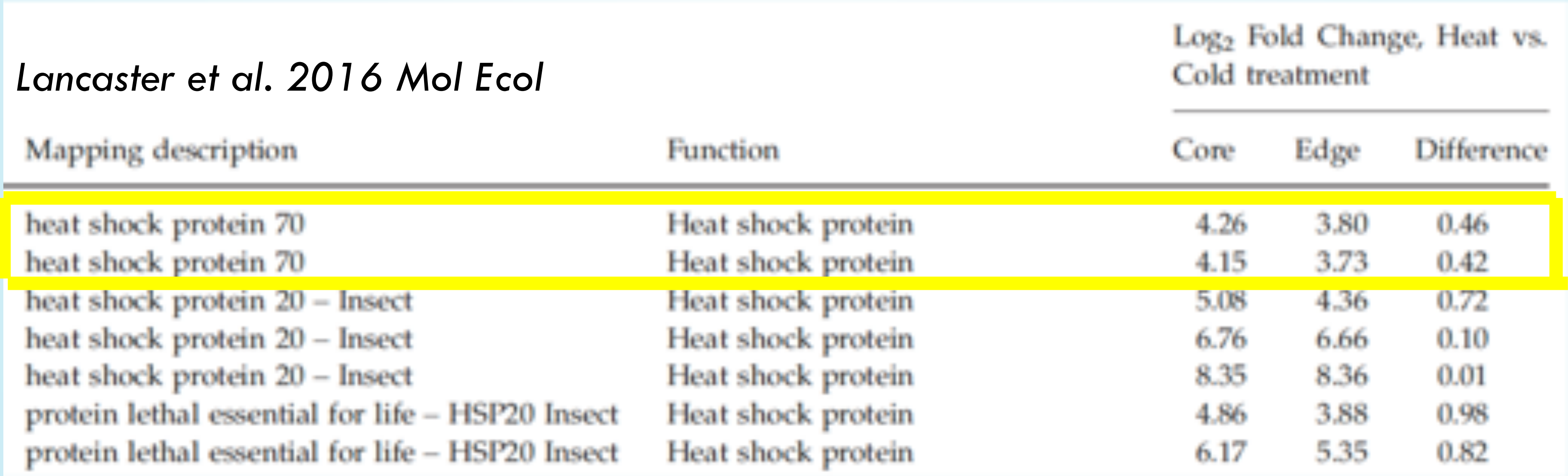
Heat Shock Protein 70

Thermal tolerance

Dudaniec et al. 2018, Mol Ecol

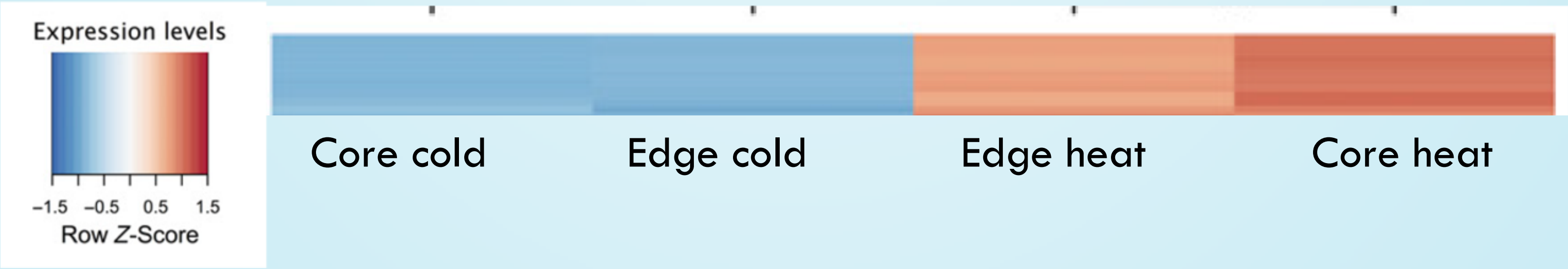
SUPPORTED BY GENE EXPRESSION DATA

- Heat Shock Protein 70 was differentially expressed in response to heat stress in the core compared with the range edge



Lancaster et al. 2016 Mol Ecol

Mapping description	Function	Log$_2$ Fold Change, Heat vs. Cold treatment: Core	Edge	Difference
heat shock protein 70	Heat shock protein	4.26	3.80	0.46
heat shock protein 70	Heat shock protein	4.15	3.73	0.42
heat shock protein 20 – Insect	Heat shock protein	5.08	4.36	0.72
heat shock protein 20 – Insect	Heat shock protein	6.76	6.66	0.10
heat shock protein 20 – Insect	Heat shock protein	8.35	8.36	0.01
protein lethal essential for life – HSP20 Insect	Heat shock protein	4.86	3.88	0.98
protein lethal essential for life – HSP20 Insect	Heat shock protein	6.17	5.35	0.82

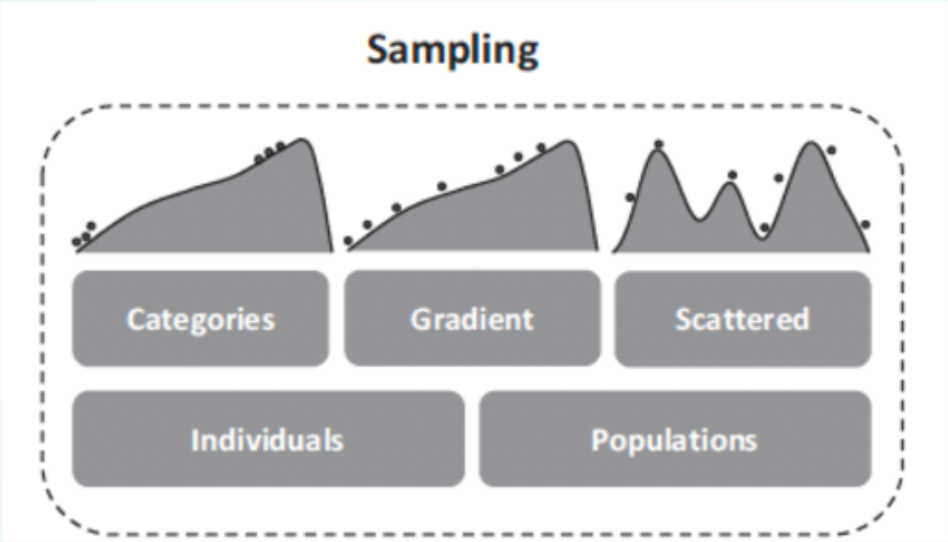


+ 50 other differentially expressed genes found among EAA SNPs

Conclusions – *Ischnura elegans* range expansion

- High adaptive capacity to climate during range expansion,
- Neutral and adaptive variation = temperature key driver
- Functional relevance with respect to environmental variation and stressors
- Evidence for rapid adaptation during ongoing range expansion

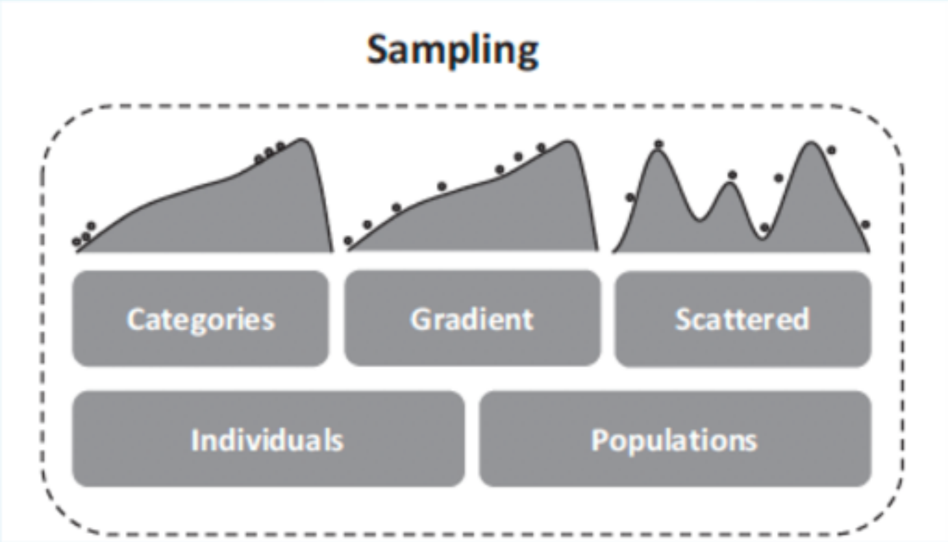
STAGES OF ANALYSIS: (1) SAMPLING



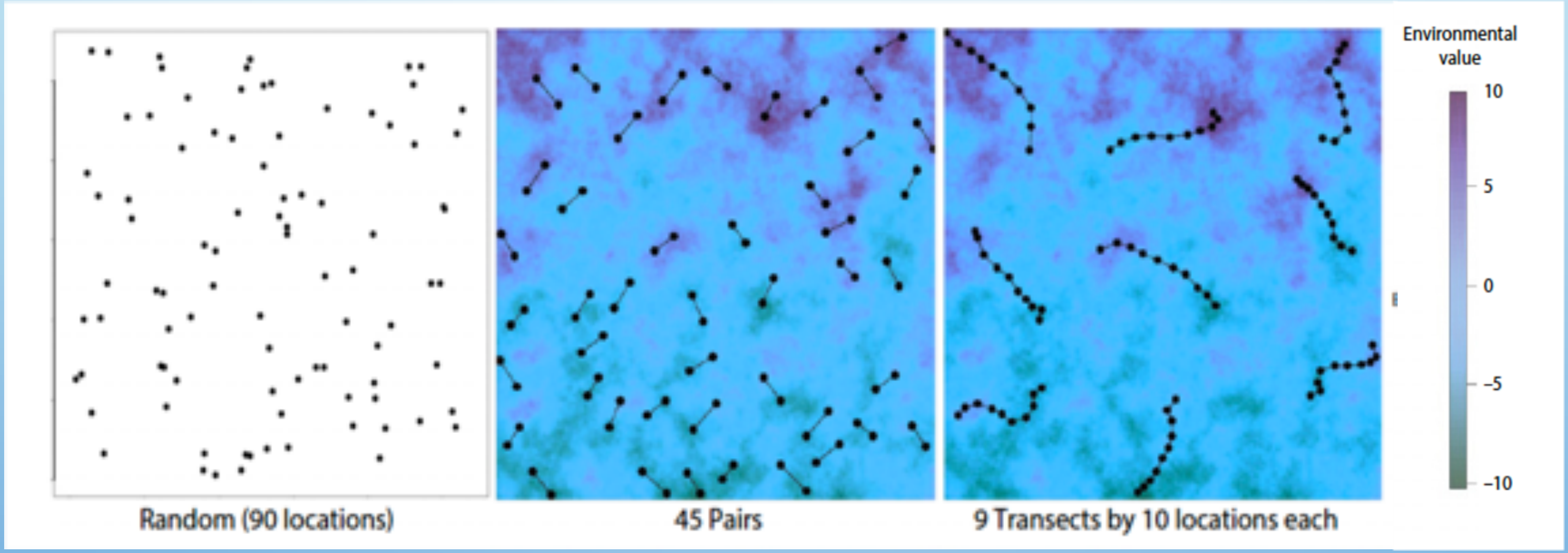
& geographic coordinates!

Rellstab et al. 2015, Lotterhos & Whitlock 2015

STAGES OF ANALYSIS: (1) SAMPLING



& geographic coordinates!



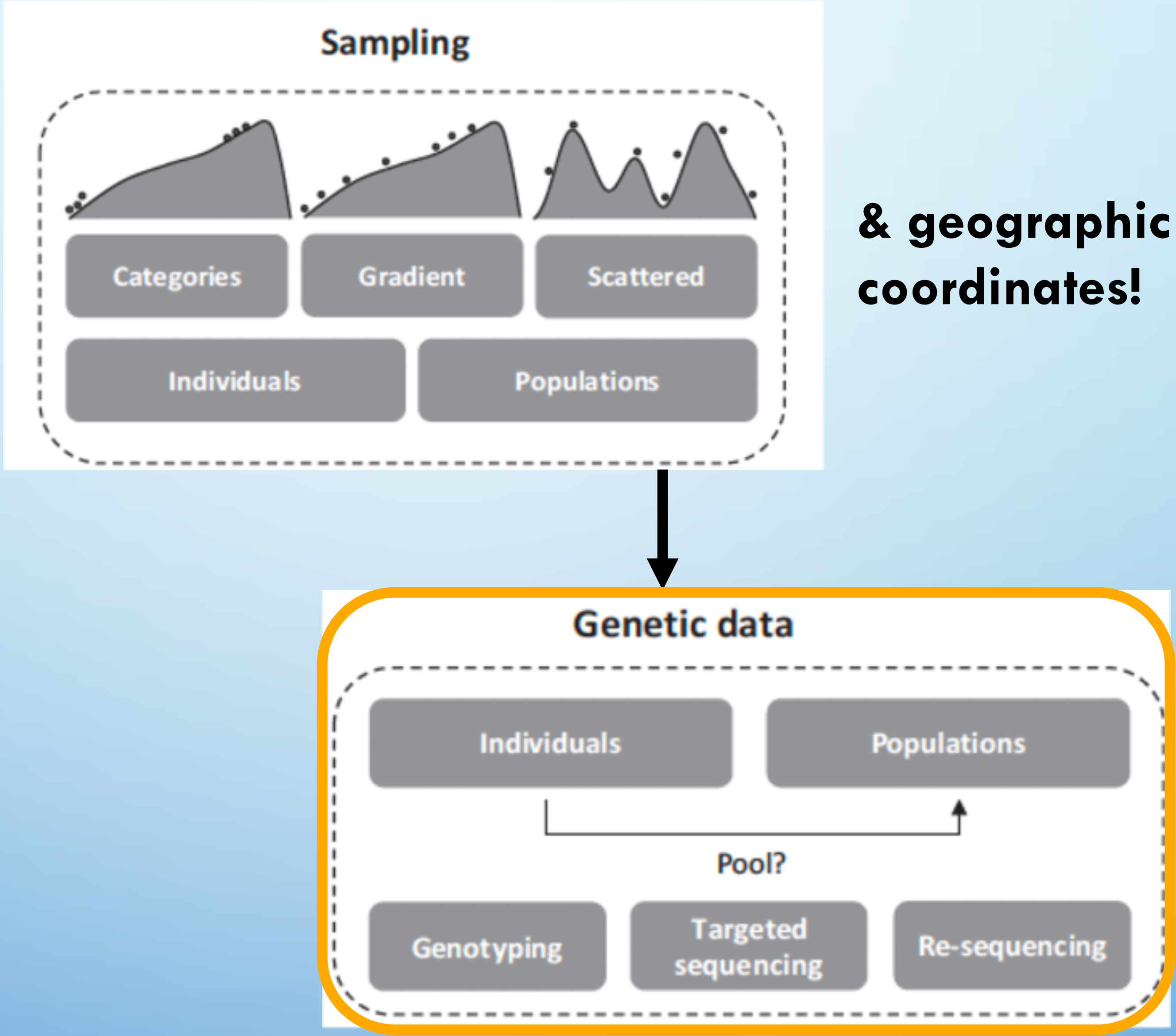
Random (90 locations)

45 Pairs

9 Transects by 10 locations each

Rellstab et al. 2015, Lotterhos & Whitlock 2015

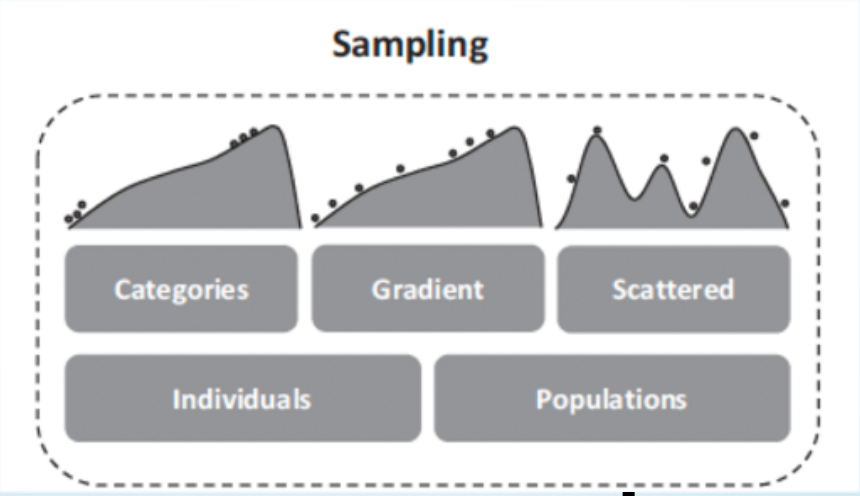
STAGES OF ANALYSIS: (2) GENETIC DATA

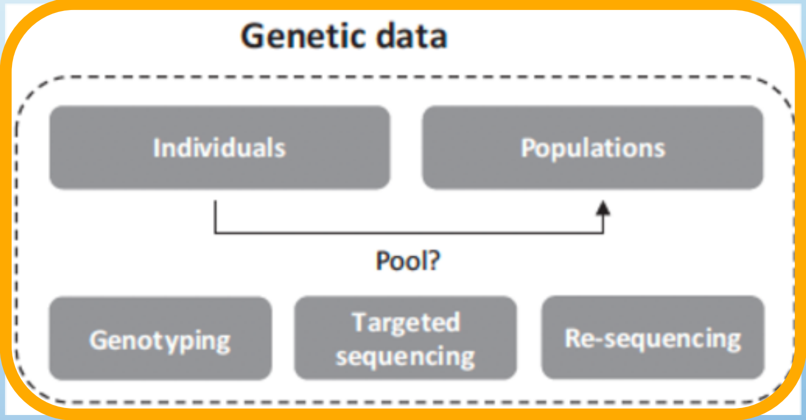


& geographic coordinates!

Rellstab et al. 2015

STAGES OF ANALYSIS: (3) OUTLIER DETECTION





Differentiation-based Analyses

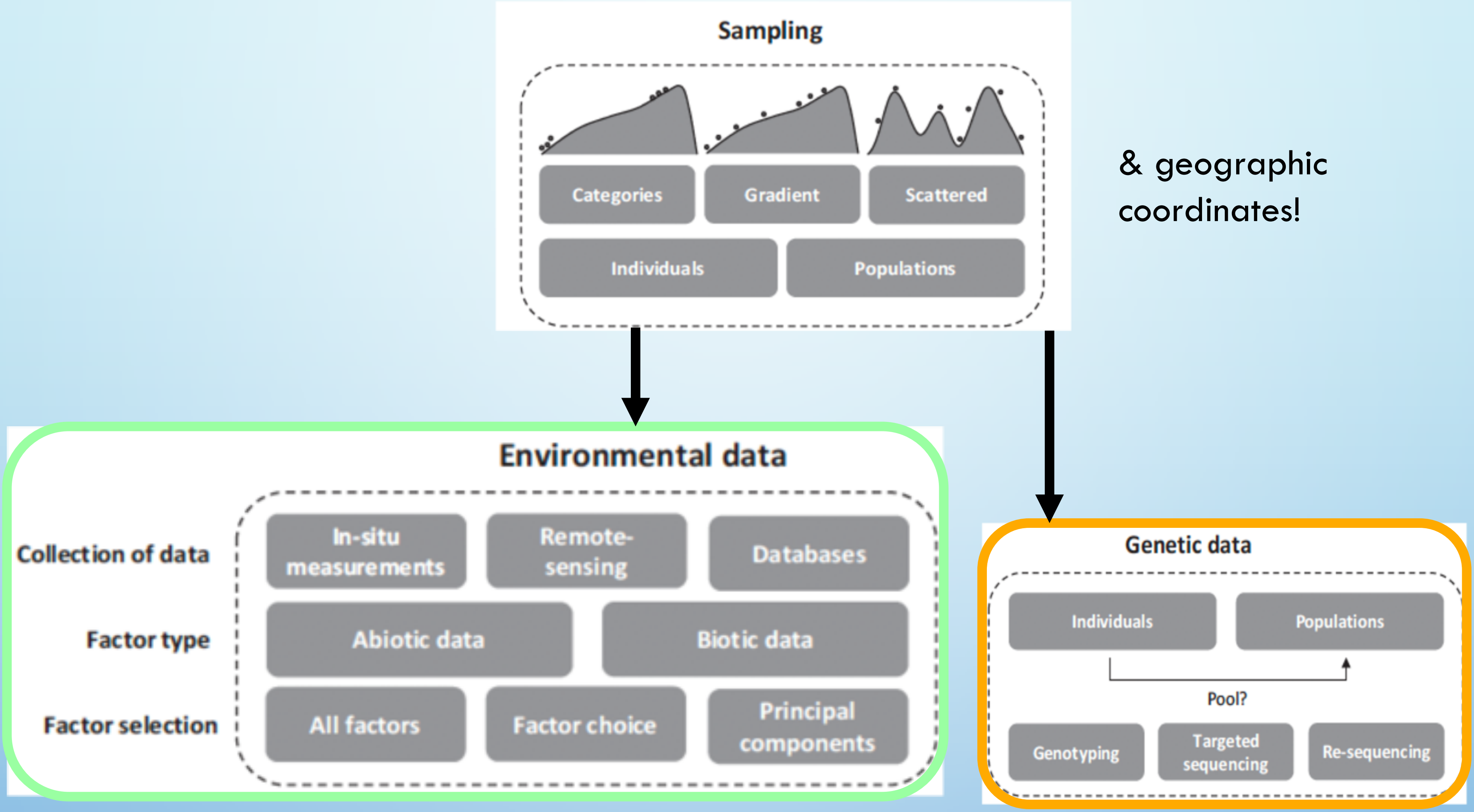
Fst Outlier tests

PCA/ordination-based

No environmental data needed

Rellstab et al. 2015

STAGES OF ANALYSIS: (4) ENVIRONMENTAL DATA



Sampling

Categories
Gradient
Scattered
Individuals
Populations

& geographic coordinates!

Environmental data

Collection of data: In-situ measurements; Remote-sensing; Databases

Factor type: Abiotic data; Biotic data

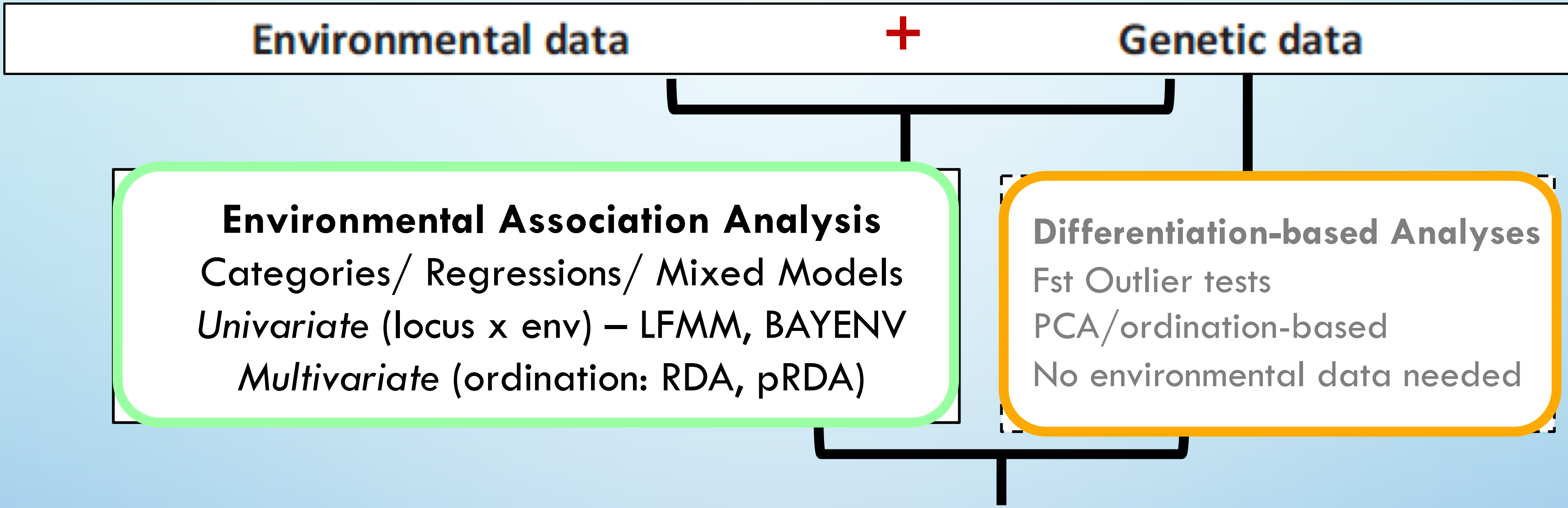
Factor selection: All factors; Factor choice; Principal components

Genetic data

Individuals
Populations
Pool?
Genotyping
Targeted sequencing
Re-sequencing

Rellstab et al. 2015

STAGES OF ANALYSIS: (5) EAA/GEA

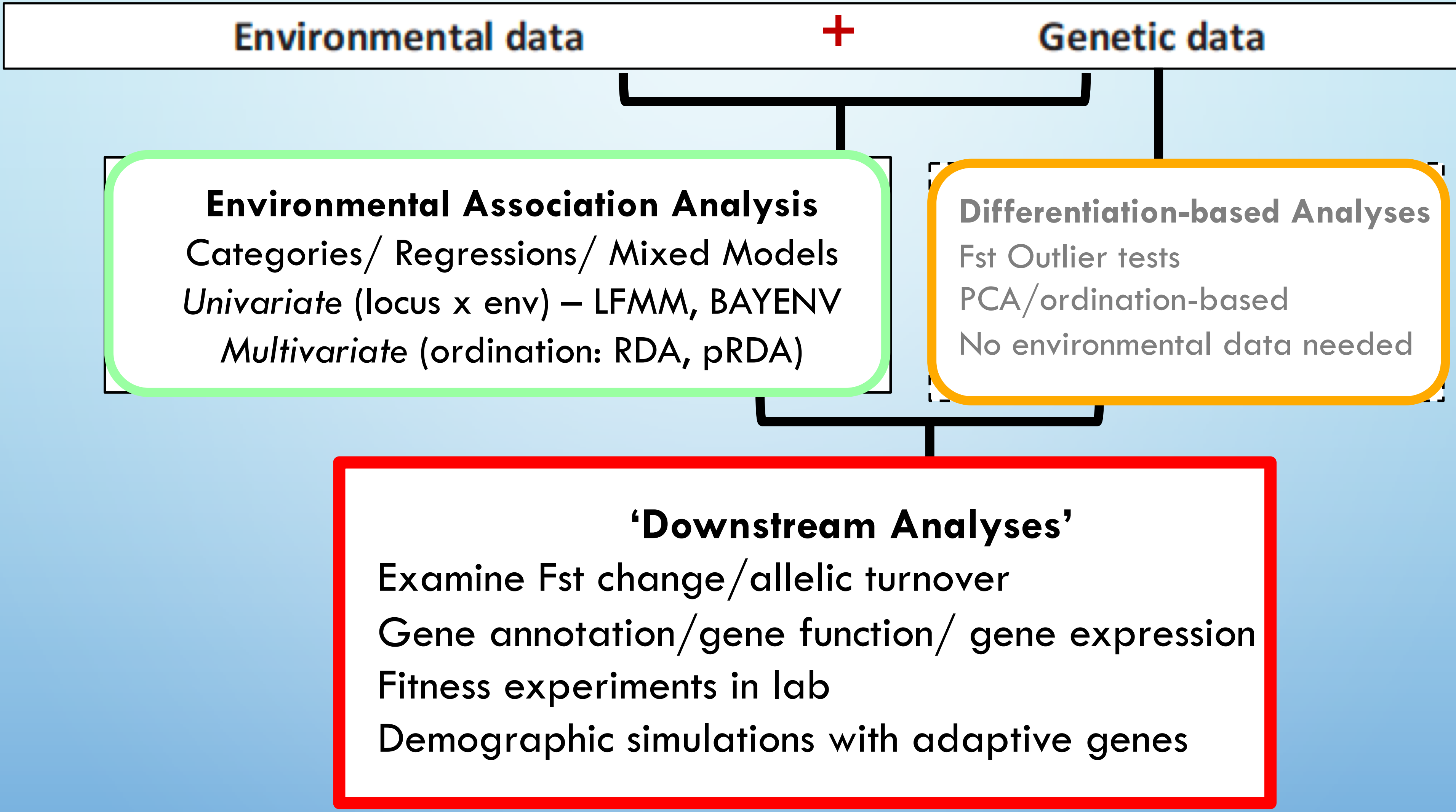


Environmental data + Genetic data

Environmental Association Analysis
Categories/ Regressions/ Mixed Models
Univariate (locus x env) – LFMM, BAYENV
Multivariate (ordination: RDA, pRDA)

Differentiation-based Analyses
Fst Outlier tests
PCA/ordination-based
No environmental data needed

STAGES OF ANALYSIS: (5) EAA/GEA



Environmental data + **Genetic data**

Environmental Association Analysis

Categories/ Regressions/ Mixed Models

Univariate (locus x env) – LFMM, BAYENV

Multivariate (ordination: RDA, pRDA)

Differentiation-based Analyses

Fst Outlier tests

PCA/ordination-based

No environmental data needed

‘Downstream Analyses’

Examine Fst change/allelic turnover

Gene annotation/gene function/ gene expression

Fitness experiments in lab

Demographic simulations with adaptive genes

A note on combining test results

- Multiple analyses might decrease the likelihood of false positives BUT avoid using every method out there!
- Approaches vary in power due to demographic history, type of selection, and study design.
- Composite measures of selection (combining p-values, multivariate distances) can improve detection….
- BUT: power is limited by the power of the univariate statistics summarized (see MINOTAUR R package)

Composite measures of selection can improve the signal-to-noise ratio in genome scans

Katie E. Lotterhos, Daren C. Card, Sara M. Schaal, Liuyang Wang, Caitlin Collins, Bob Verity

First published: 08 June 2017 | https://doi.org/10.1111/2041-210X.12774 | Citations: 9

HOW CAN WE IMPROVE INFERENCE FROM EAA/FST OUTLIER TESTS?

Approach	Improvement
REPLICATE	Identify common candidates across replicate gradients, examine for parallel adaptation
SIMULATE	Optimise sampling design, verify empirical findings, forecast adaptive variation
ANNOTATE	Genes with biological/ecological relevance improve interpretation

Modified from Forester et al. (2018)

HOW CAN WE IMPROVE INFERENCE FROM EAA/FST OUTLIER TESTS?

Approach	Improvement
REPLICATE	Identify common candidates across replicate gradients, examine for parallel adaptation
SIMULATE	Optimise sampling design, verify empirical findings, forecast adaptive variation
ANNOTATE	Genes with biological/ecological relevance improve interpretation
WHOLE GENOME SEQUENCING	All genetic variants identified – large power to detect selection + via structural variants
GWAS	Overlap in candidate loci with EAA can improve strength of inference
COMMON GARDEN + RECIP TRANSPLANTS	Link candidate genes with traits relevant for fitness
TRANSCRIPTOMICS + EPIGENETICS-'EWAS'	Environment-mediated gene expression Plasticity and non-DNA based local adaptation

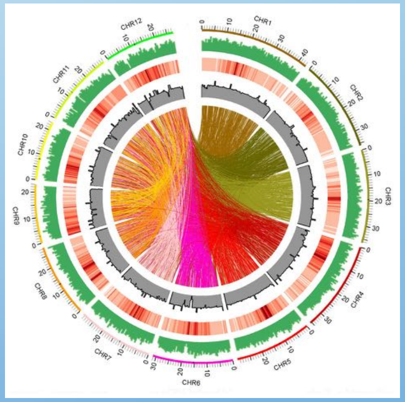
Modified from Forester et al. (2018)

Future directions - whole genome data

- Future analyses will need to integrate genomic architecture in to environmental selection detection

Challenges:

- ‘Coding’ for copy number variants, chromosome inversions, transposable elements (i.e. can they be considered in the same way as alleles?)
- Constructing hierarchical models to integrate sources of error from different data types (e.g. SNPs, SVs).
- Identifying gene modularity and interactions



discussed in Storfer et al. (2018)

ACKNOWLEDGEMENTS

Sonu Yadav (MQ PhD)

Adam Stow (MQ)

Rebecca Harris (CSIRO)

Chuan Ji Yong (MQ Masters)

Alexander Carey (MQ Masters)

Brenna Forester (Colorado U)

MACQUARIE
University

LUND
UNIVERSITY

Bengt Hansson (LU)

Erik Svensson (LU)

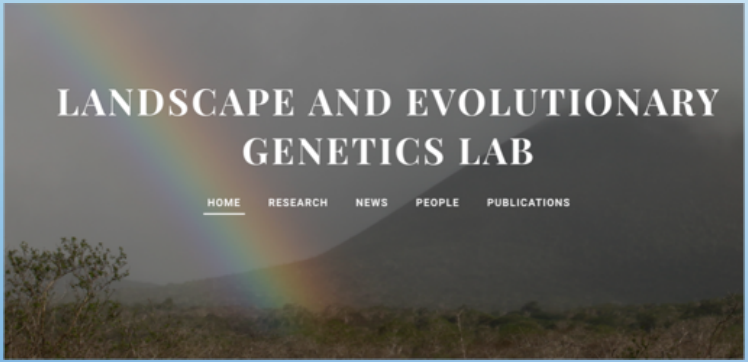
Lesley Lancaster (UAB)

Maren Wellenreuther (U Auckland)

Pallavi Chauhan (LU)

Julian Catchen (U Ilinois)

Collin Ahrens (WSU)



LANDSCAPE AND EVOLUTIONARY GENETICS LAB

1495
UNIVERSITY
OF ABERDEEN

MARIE CURIE
ACTIONS

Wenner-Gren Foundations
Wenner-Gren Stiftelserna