

An introduction to genotype-environment associations and genomic offset

Christian Rellstab
WSL, Birmensdorf, Switzerland

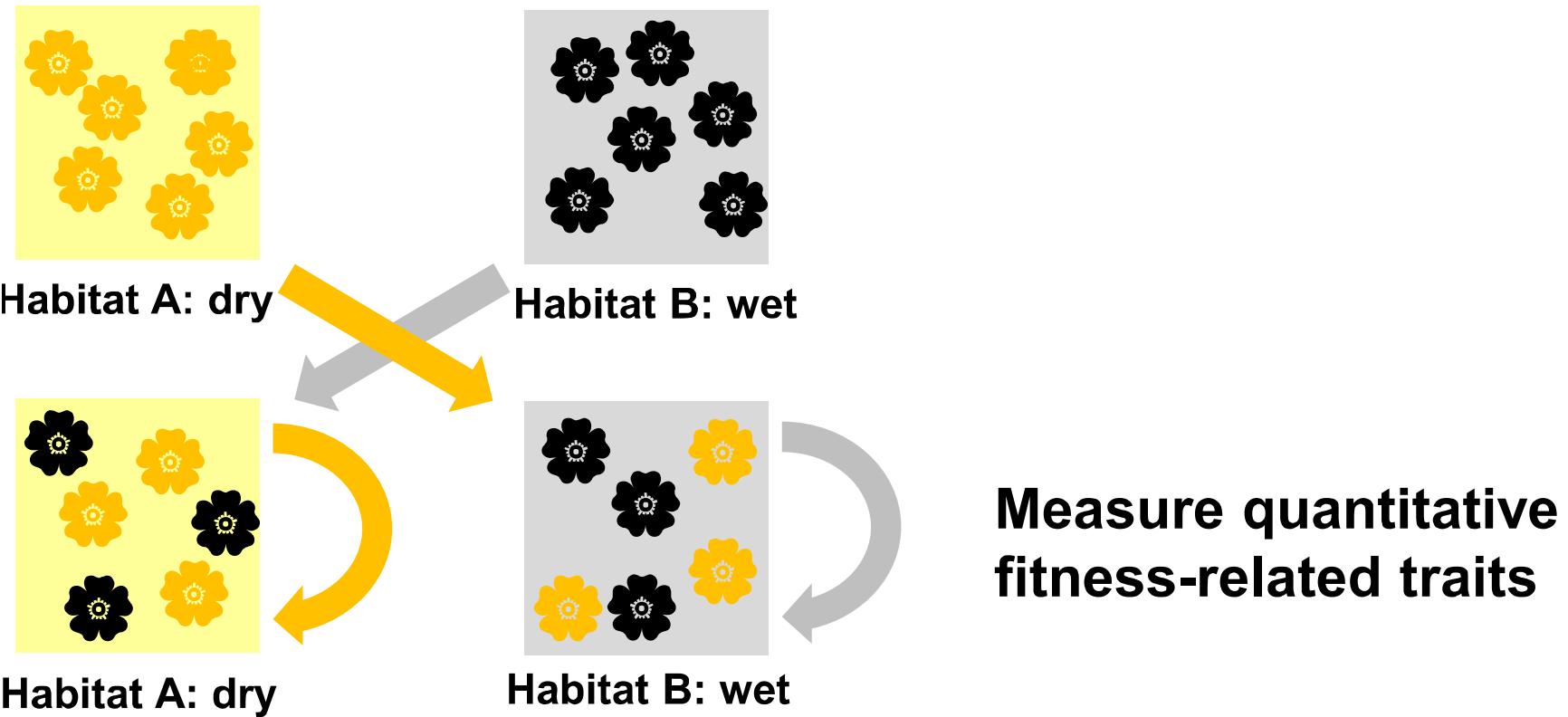


SSMPG 2025, Aussois, France
9.9.2025

Local adaptation and reciprocal transplant experiments

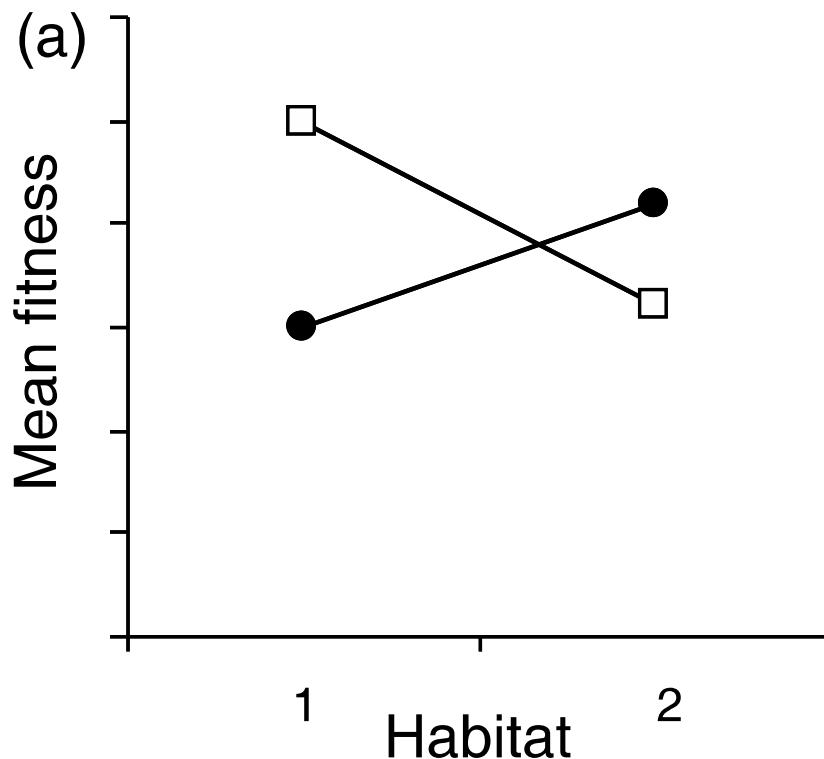
A test for local (!) adaptation

- Phenotype as a measurement of Genotype and Environment
- Grow genotypes in local/foreign habitats



Reciprocal transplant experiment

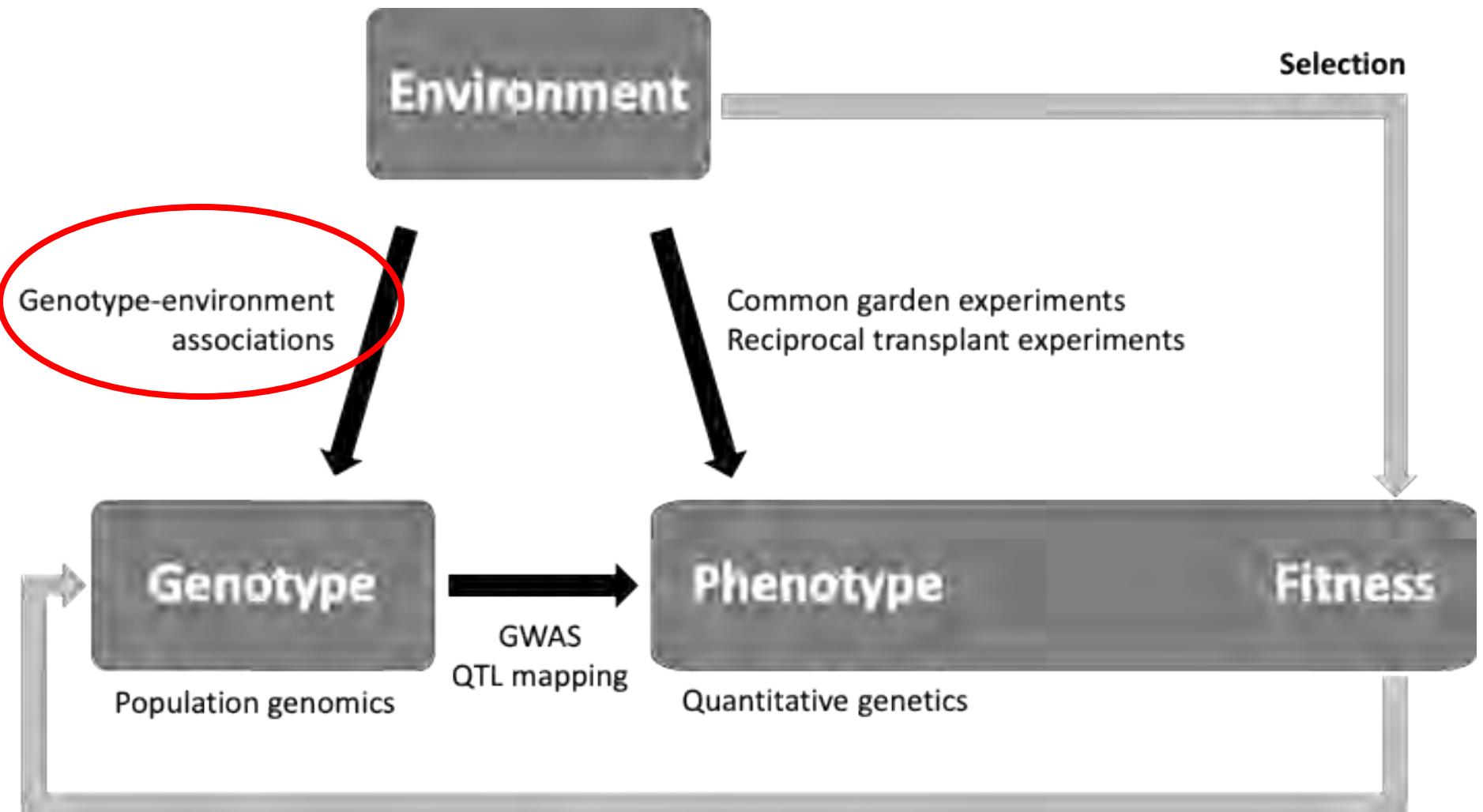
- Local/foreign and home/away principle



Genotype-environment associations

Genotype-environment associations (GEA)
Environmental association analysis (EAA)

The triangle

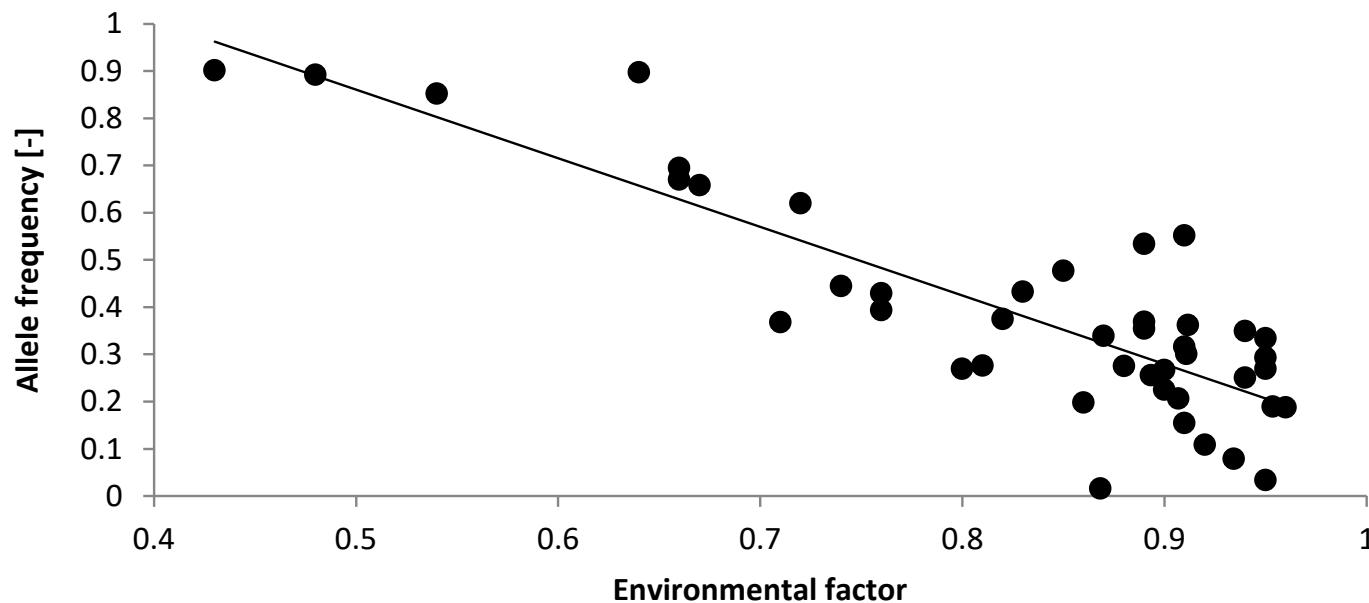


Genotype-environment associations

- GEA, also called environmental association analyses (EAA)
- Associate genetic variation with environmental conditions:

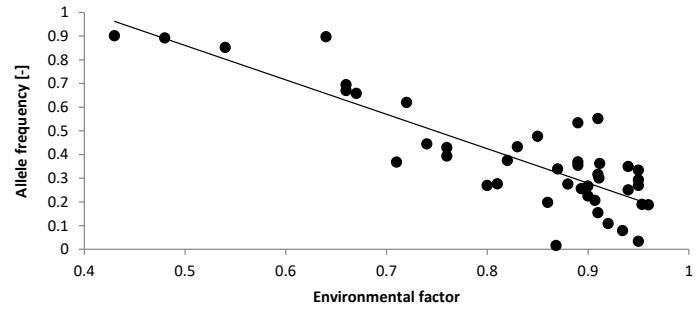
$$G \text{ (genotype)} \sim E \text{ (environment)}$$

- Main tools of [landscape genomics](#)
- Identify alleles/SNPs/genes/gene regions associated with particular environmental factors
- Identify environmental factors that drive local adaptation



Assumptions of GEA

- Environment is driving selection
- Populations are locally adapted
- Fitness is genetically controlled
- Allele frequency clines along environmental gradients
 - Antagonistic pleiotropy
 - Additive effects
- Environmental data accurately describes selection pressures
- Sampled individuals are representative of the study populations



Signatures of adaptation in *Pinus ponderosa*

- One of the earliest examples of GEA
- One allozyme locus (peroxidase locus)
- Association with aspect, drought, altitude

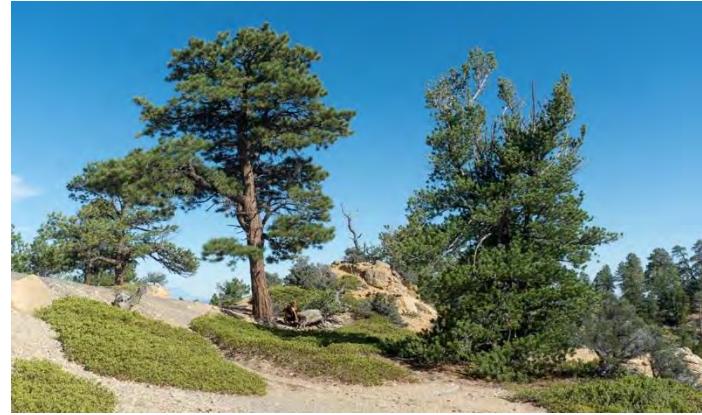
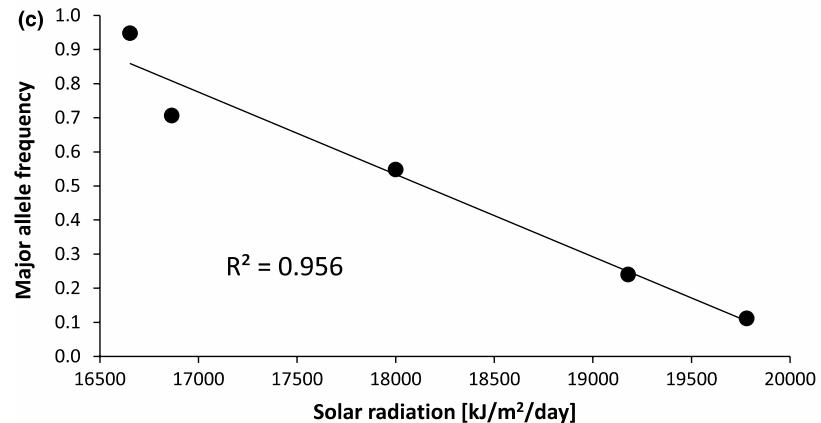
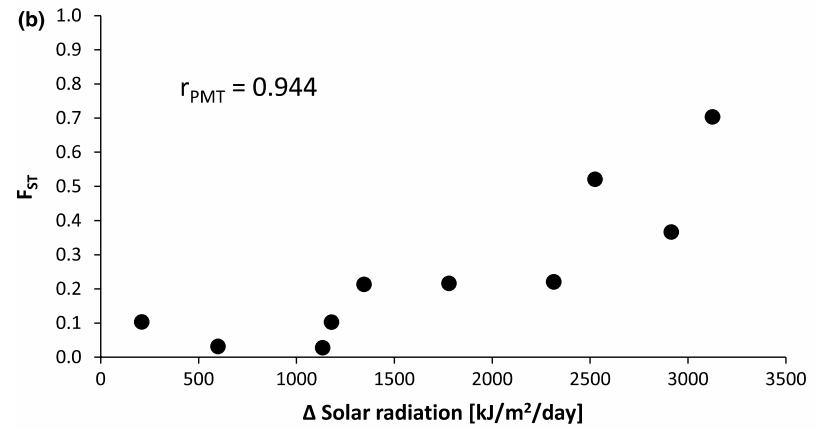
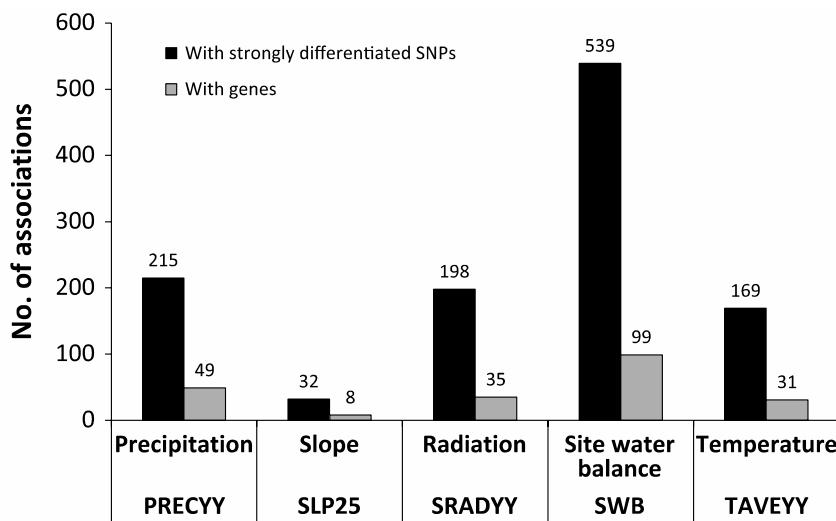
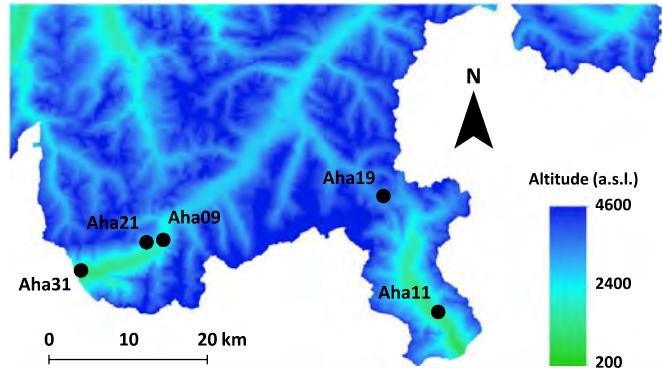


Table 2. Allele and genotypic frequencies of a peroxidase locus at several localities sampled

Sample locality	Aspect	Elevation (m)	Genotypes			N	Per χ^2 freq. \pm S.E.	χ^2_A	χ^2_B	Per χ^2 freq. \pm S.E.	χ^2_C
			22	23	33						
Glacier Lake		2590	135 (129.7)	44 (54.5)	11 (5.7)	190	.826 \pm .019	7.1**		.826 \pm .019	
Left Hand Canyon	North Facing	\approx 2440	60 (60.4)	11 (10.1)	0 (0.4)	71	.923 \pm .022	0.5			
	South Facing	\approx 2440	44 (45.1)	26 (23.8)	2 (3.1)	72	.792 \pm .034	0.6		10.0***	.857 \pm .021
Coal Creek Canyon	North Facing	\approx 2440	62 (61.8)	26 (26.4)	3 (2.8)	91	.824 \pm .028	0.1			
	South Facing	\approx 2440	55 (56.1)	24 (21.8)	1 (2.1)	80	.837 \pm .029	0.8		0.1	.830 \pm .020
Lower Sugarloaf Mountain	North Facing	2130	41 (42.2)	29 (26.6)	3 (4.2)	73	.760 \pm .035	0.6			
	South Facing	2130	23 (31.6)	52 (34.8)	1 (9.6)	76	.645 \pm .039	18.5***		4.7*	.701 \pm .027
Boulder Canyon	North Facing	1738	54 (56.5)	37 (32.0)	2 (4.5)	93	.780 \pm .030	2.3			
	South Facing	1738	40 (45.6)	51 (39.7)	3 (8.6)	94	.697 \pm .034	7.6**		3.3x	.738 \pm .023
Eldorado Springs		1760	26 (35.0)	64 (45.9)	6 (15.1)	96	.604 \pm .035	14.9***			.604 \pm .035
Cheyenne Lookout		1770	2 (7.5)	31 (20.1)	8 (13.4)	41	.427 \pm .055	12.2***			.427 \pm .055

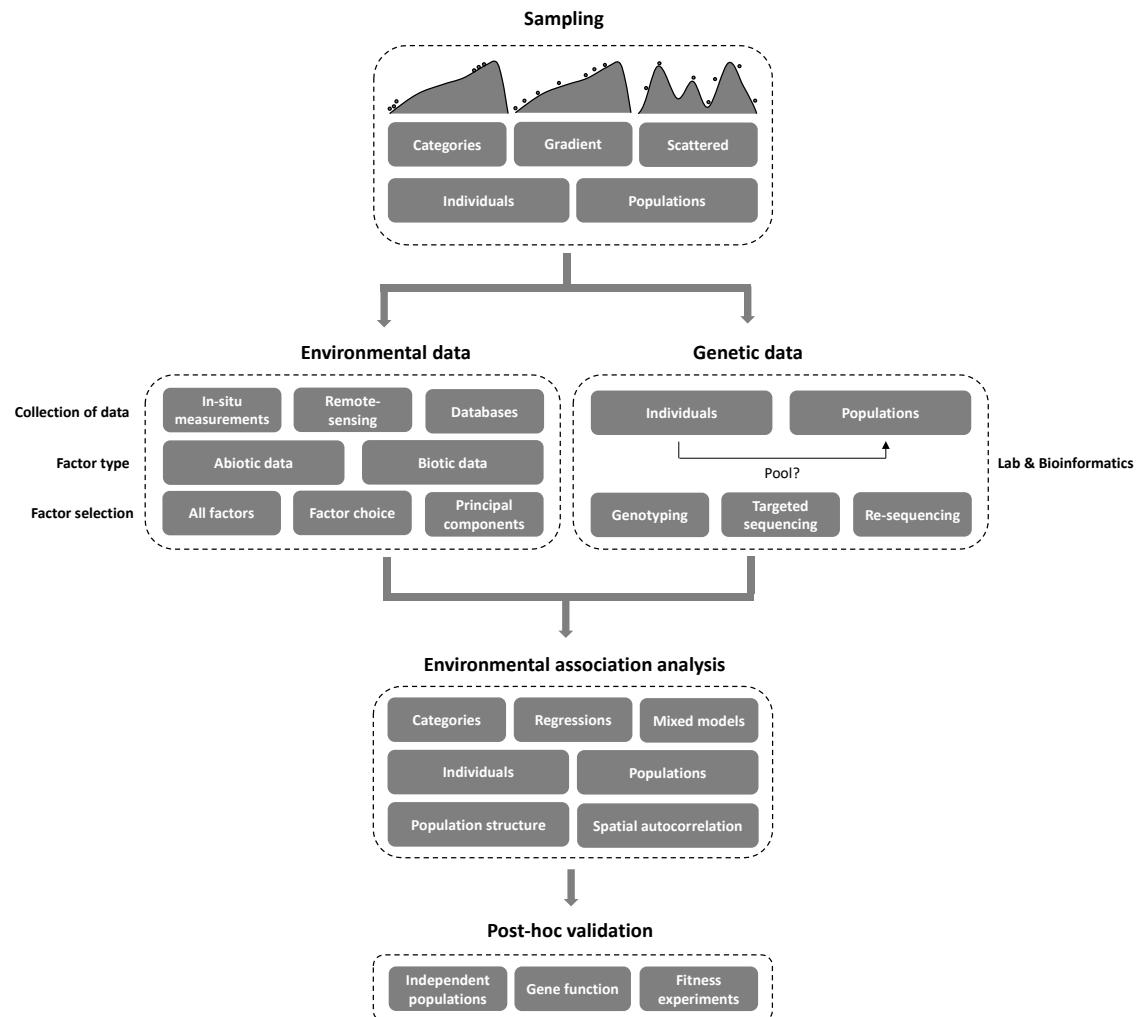
Mitton et al. 1977

Whole-genome analyses in *Arabidopsis halleri*

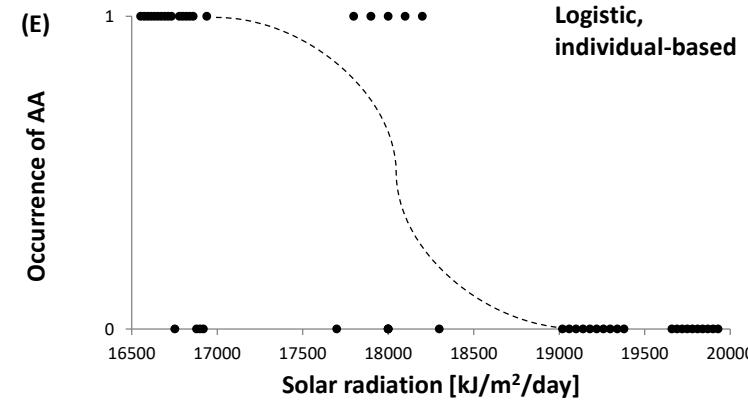
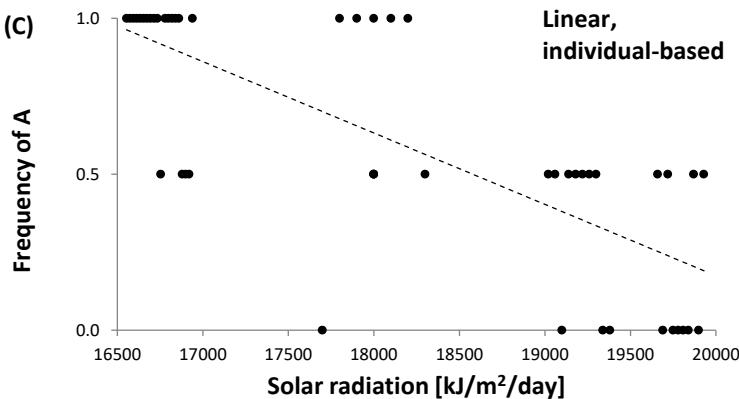
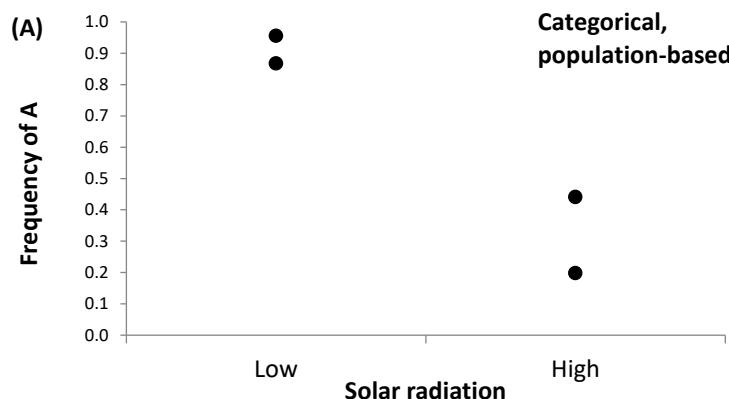
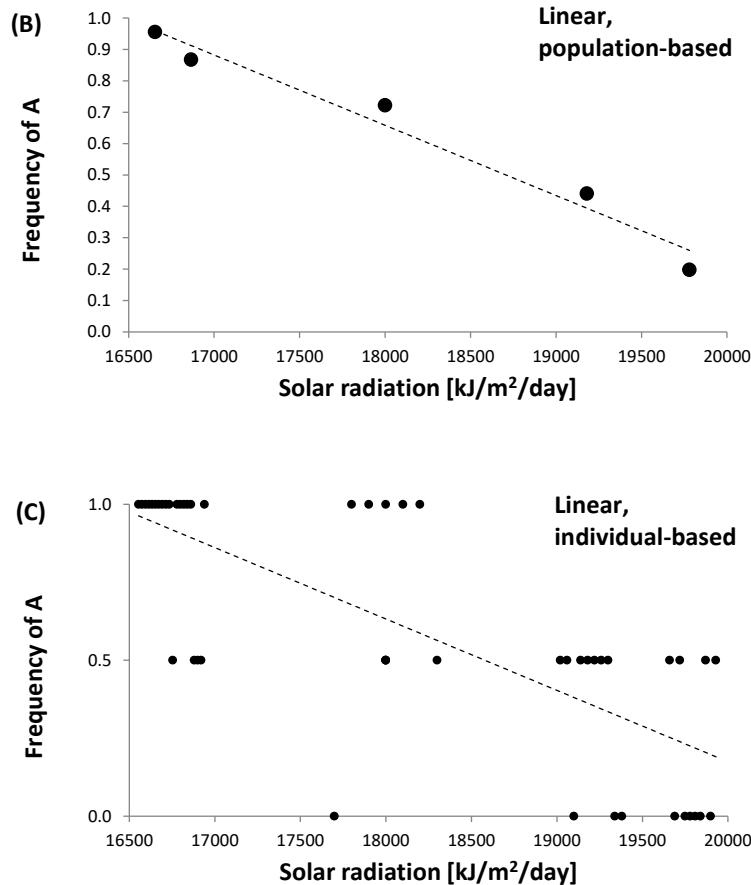


What do we need?

- Smart sampling design
- Environmental data
- Genomic data
- Analytical pipelines
- Validation



Different approaches & response curves



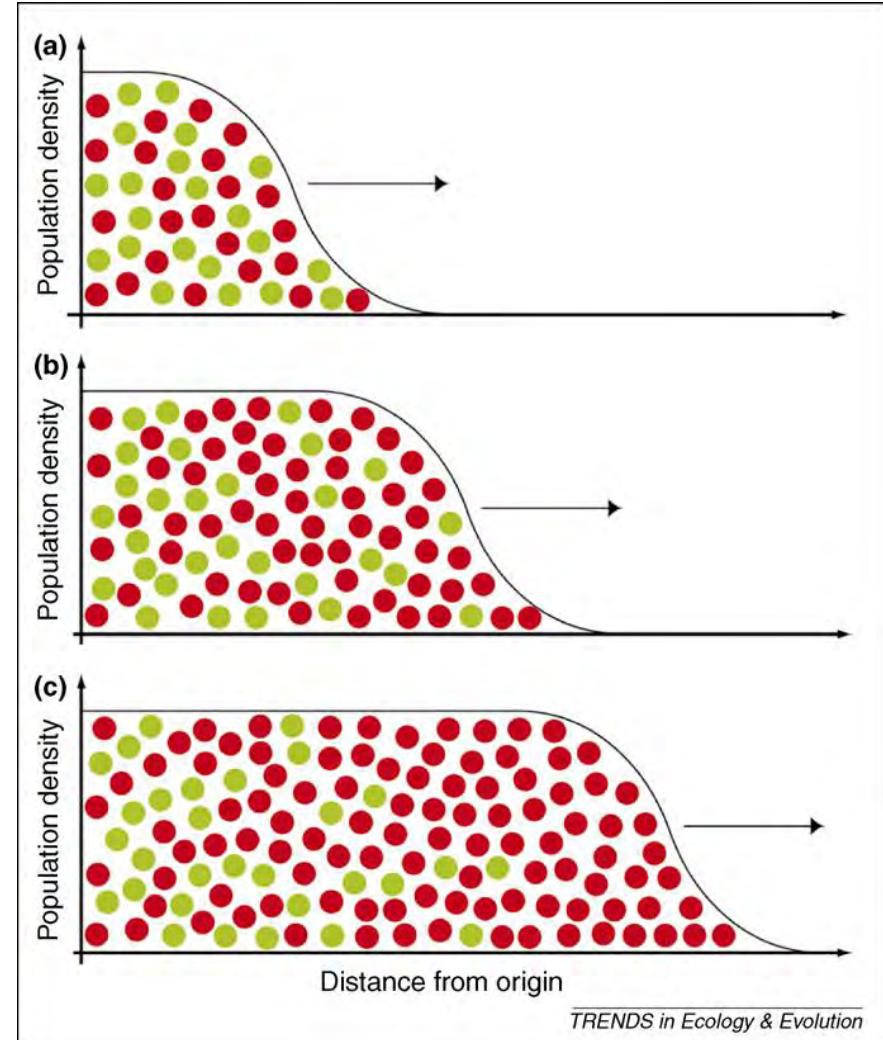
GEA: what is missing?

Neutral population structure

- Neutral population structure can confound/mimic genomic signatures of adaptation
 - Drift
 - Species expanding their distribution range (Excoffier 2008)
 - When past demographic processes are highly correlated with environmental structure
 - Re-colonization (south --> north) after the last glacial maximum (LGM)
 - Isolation by distance
 - Hierarchical structure (Excoffier 2009)
- Especially the case in
 - large, scattered sampling designs
 - gradient sampling designs without replication
- We have to control for population structure!
- We have to choose smart sampling designs!

Surfing alleles at the expansion front

- Large populations at the origin
- Small populations at the expansion front
- Alleles at the expansion front increase in frequency due to genetic **drift**
- High genetic diversity in the original populations
- Low genetic diversity at the expansion front
- Can mimic allele frequencies shaped by selection

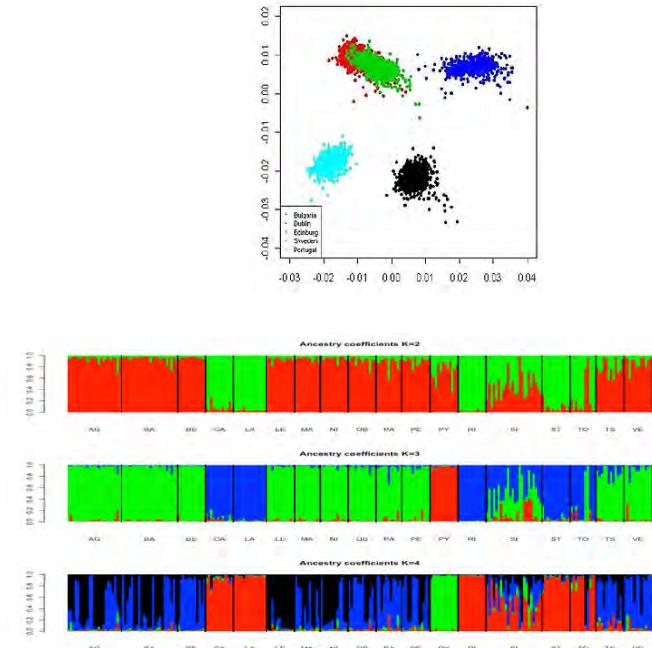


Solution 1: Include neutral genetic structure in GEA

- Neutral processes affect **ALL loci** across the genome
- Selection (non-neutral) affects **only a subset of loci**
- Adaptive loci: Those associated to environment after accounting for the neutral background structure

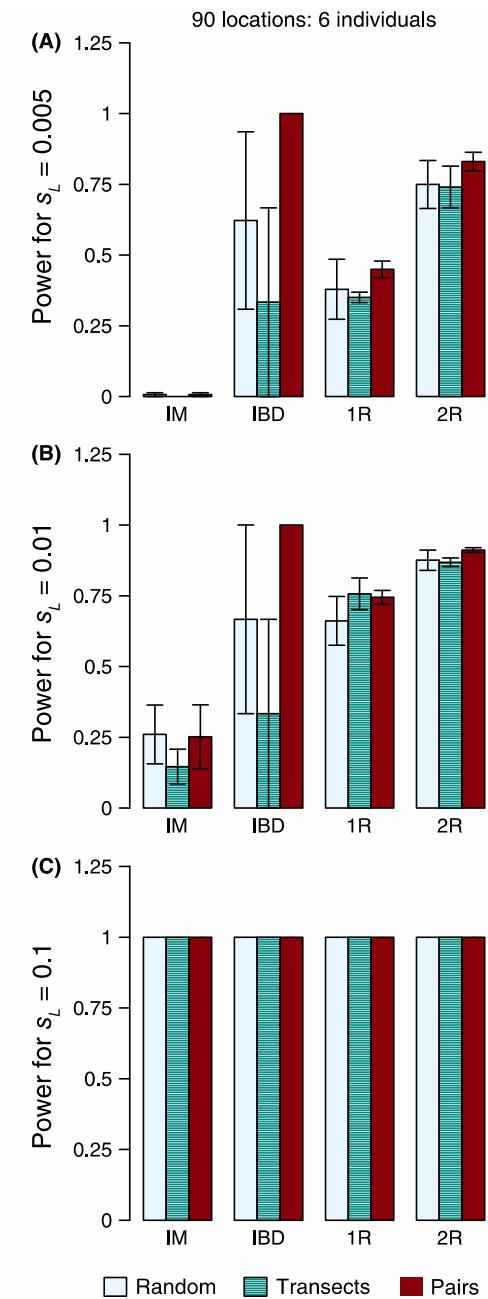
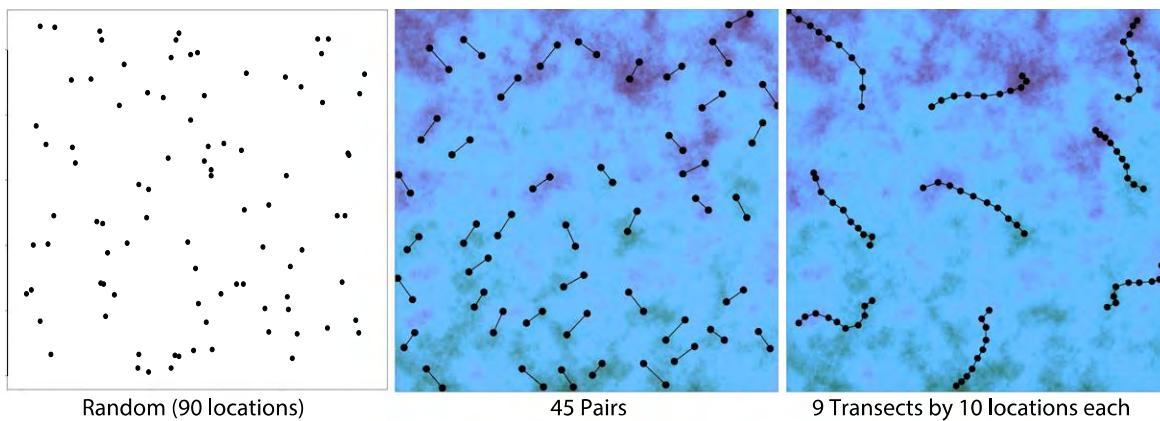
Genotype \sim Environment + Structure

- Structure can be implemented as
 - Covariate (test performed on residuals)
 - Random factor
- Structure can be described using
 - Principal components
 - Q-values of STRUCTURE/ADMIXTURE/DAPC/sNMF
 - Covariance matrix of (neutral) SNPs
 - Latent factors



Solution 2: Choose a smart sampling design

- Choose a sampling design where genetic structure is decoupled from environmental structure
- Large environmental differences at small spatial scale (= high gene flow) -> **population pairs with contrasting habitats**
- Replicated environmental gradients

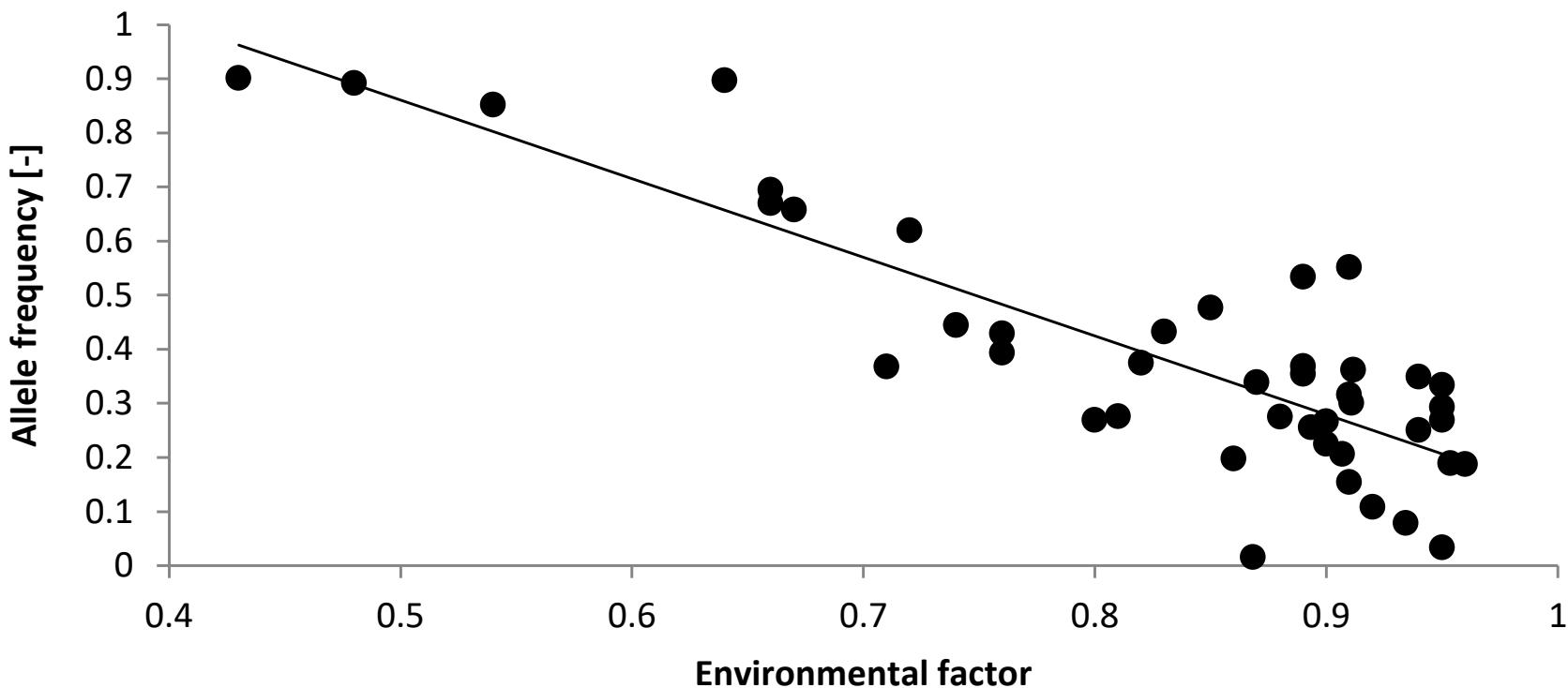


GEA – Overview of existing approaches

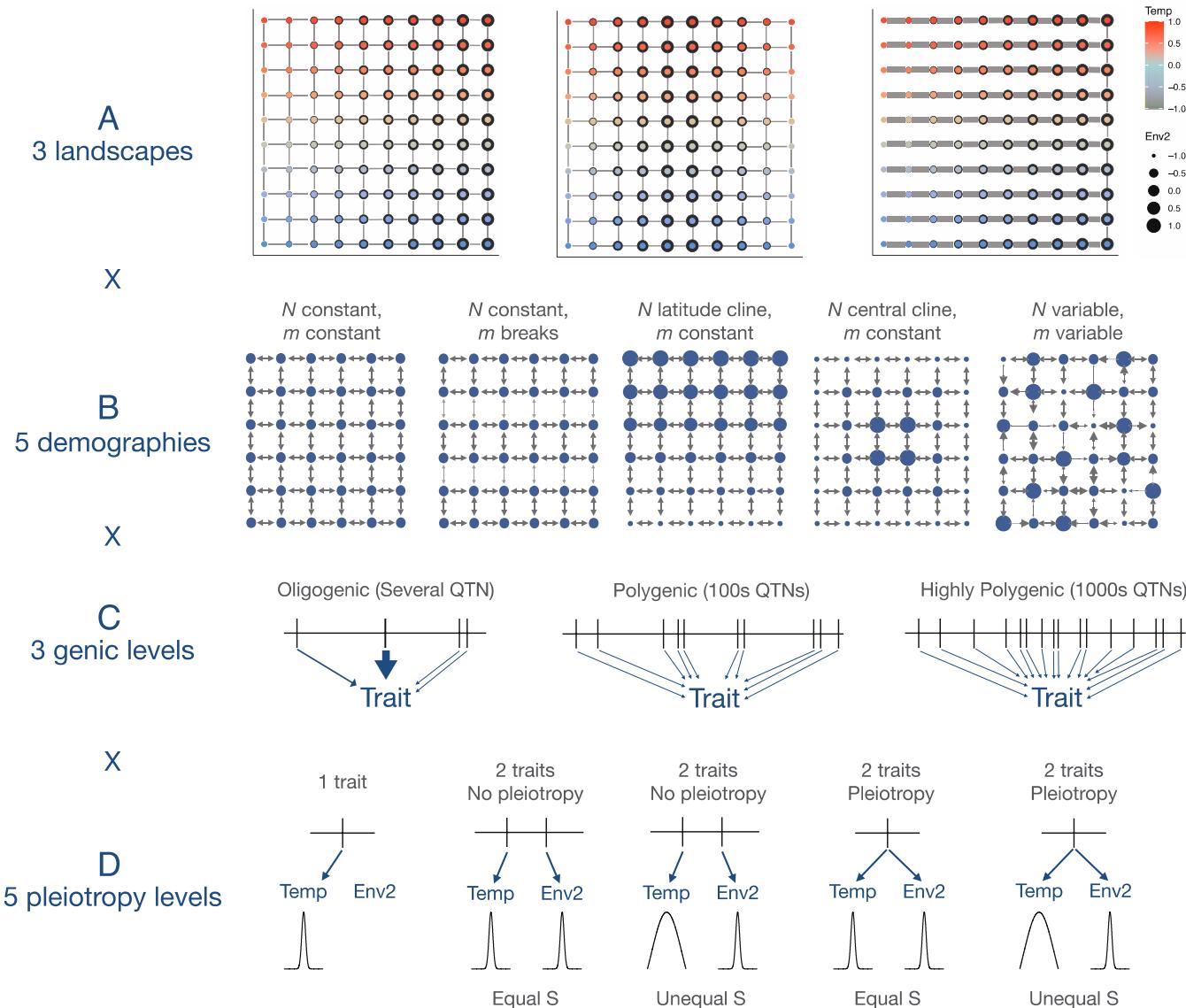
Name	Model type	Neutral structure	Multivariate X (Environment)	Multivariate Y (Genotypes)	Ind./pop. data	PoolSeq depth	Sample size	References
LFMM	Linear mixed	Yes	Both	No	Both	No	No	Fritchot et al. 2013 Caye et al. 2019 Gain & Francois 2021
BayEnv	Linear mixed, rank-linear	Yes	No	No	Population	Yes	Yes	Coop et al. 2010 Günther & Coop 2013
BayPass	Linear mixed, rank-linear	Yes	Both	No	Population***	Yes	Yes	Gautier 2015
(p)RDA	Linear	Both	Yes	Yes	Both	No	No	Legendre & Legendre 2012 Capblancq & Forester 2021
SamBada	Logistic	Both	Both	No	Individual	No	No	Joost et al. 2007 Stucki et al. 2017
Gradient Forest	Machine learning (monotonic)	No*	Yes	No	Both	No	No	Fitzpatrick & Keller 2015
BayScEnv	F	Yes	No	No	Both	No	Yes	de Villemereuil & Gaggiotti 2015
WZA	Summary statistic	-	-	Windows	-	-	-	Booker et al. 2024
Local Score	Summary statistic	-	-	Windows	-	-	-	Fariello et al. 2017

Can we expect allele frequency clines?

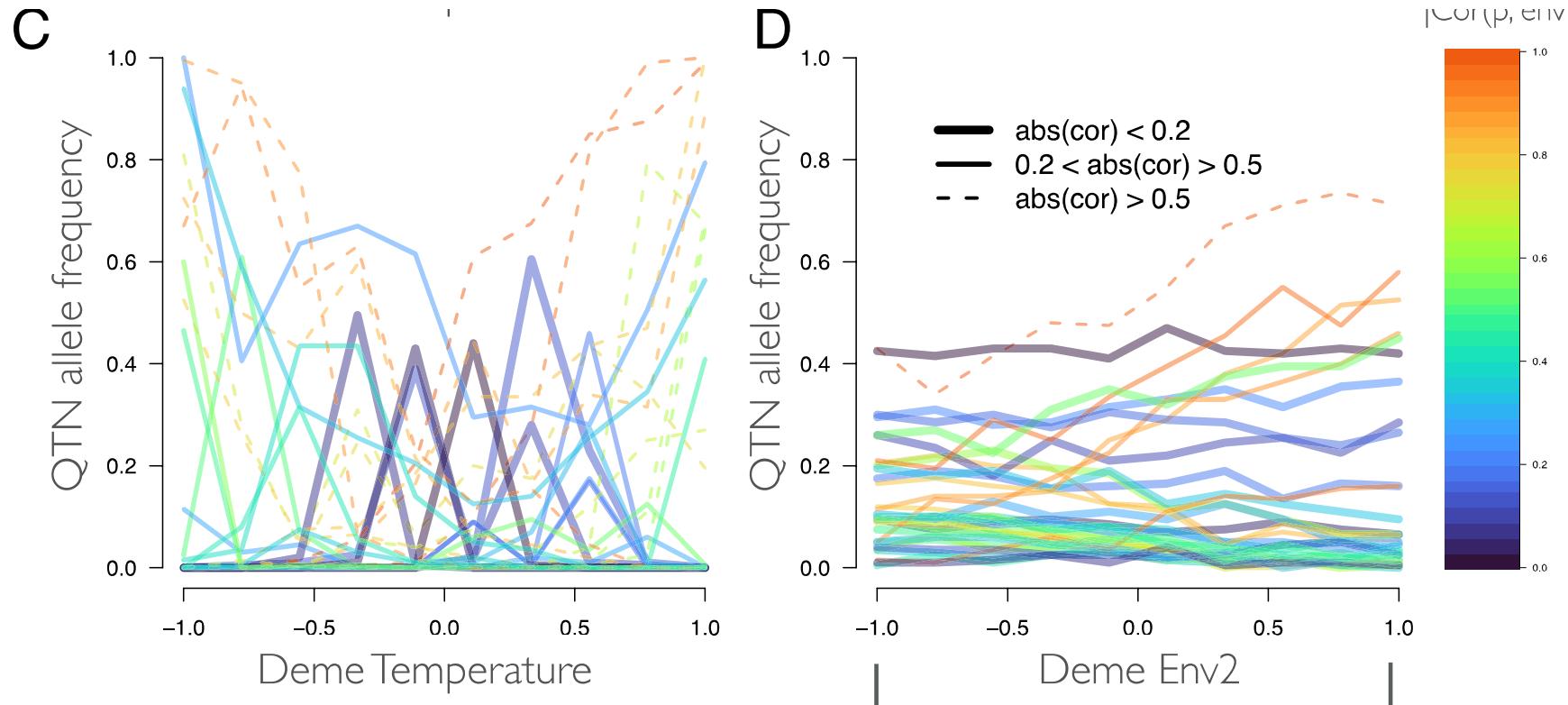
- GEA assumes that allele frequencies in- or decrease along an environmental gradient (antagonistic pleiotropy with additive effects)
- But is this realistic?



Allele frequency clines – Simulations

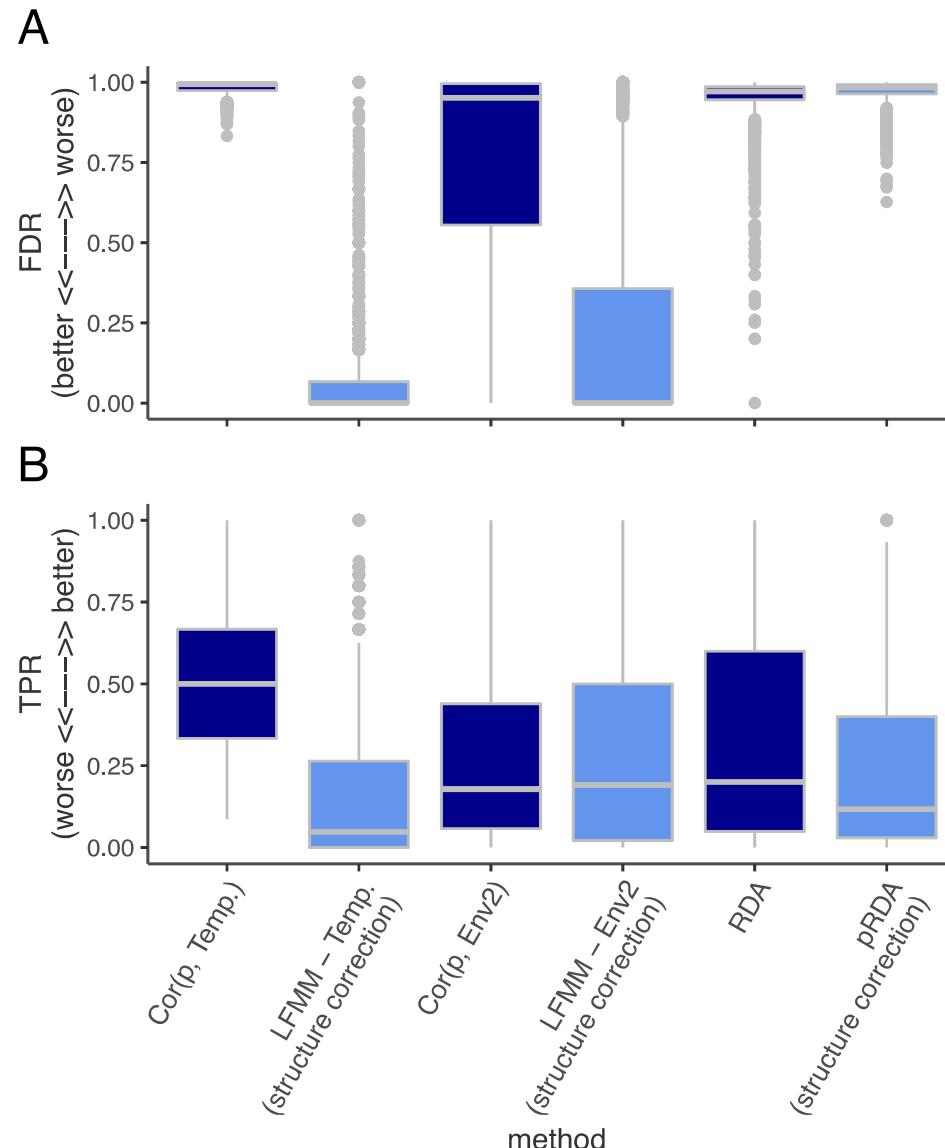


Allele frequency clines - Simulations



- The likelihood of an allele frequency cline decreases with
 - degree of polygenicity
 - degree of pleiotropy

Allele frequency clines - Simulations



Strengths & limitations of GEA

(Answers from SSMPG 2025 participants)

- **Strengths**

- No phenotyping needed
- Explorative approach, no a-priori hypotheses needed
- Prediction over space and time
- Fast computation
- Computationally straight forward
- Exploitation of databases
- Studies the environmental selective pressure

- **Limitations**

- Power of detection: polygenicity, population structure, multiple testing
- Local adaptation assumed
- No phenotypes included
- Epigenetic effects ignored
- Relies on good databases
- Microclimate ignored
- Biotic data limited

GEA: Summary

- GEA includes the actual driver (=environment) into its analysis
- GEA is a promising approach to study genomic signatures of environmental adaptation
- GEA does not need traits or fitness measurements
- GEA needs, if possible, to account for neutral population structure

$$G \sim E + S$$

- Mind the many limitations!
- Combine GEA with other approaches to confirm results

Literature – Reviews

- **Dauphin**, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F. & Manel, S. (2023). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, 38, 261-274.
- **François**, O., Martins, H., Caye, K. & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, 25, 454-469.
- **Hoban**, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *American Naturalist*, 188, 379-397.
- **Lasky**, J. R., Josephs, E. B. & Morris, G. P. (2023). Genotype-environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell*, 35, 125-138.
- **Rellstab**, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24, 4348-4370.

Genomic offset

Genetic offset

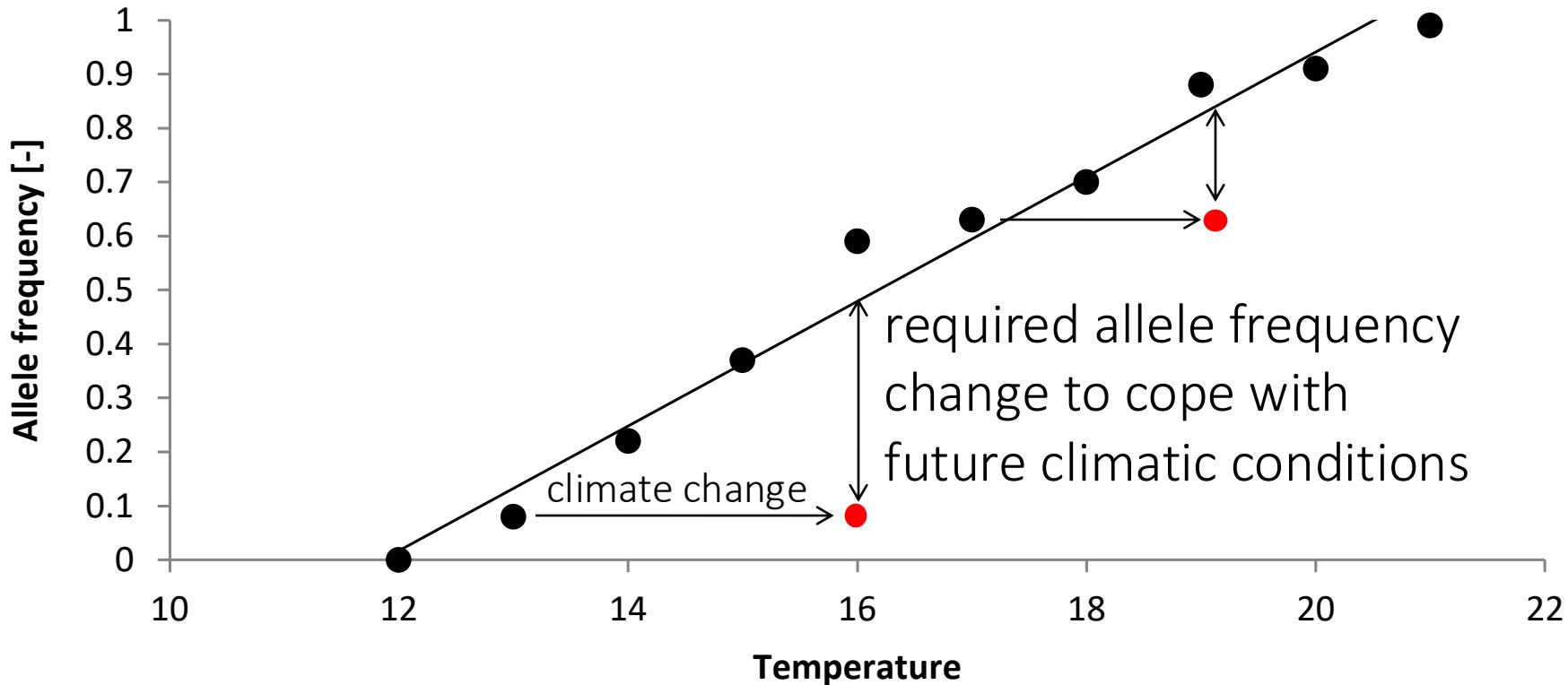
Genomic vulnerability

Risk of non-adaptedness (RONA)

Genetic gap

Risk of non-adaptedness (RONA)

to future climatic conditions
(calculated using linear regressions)

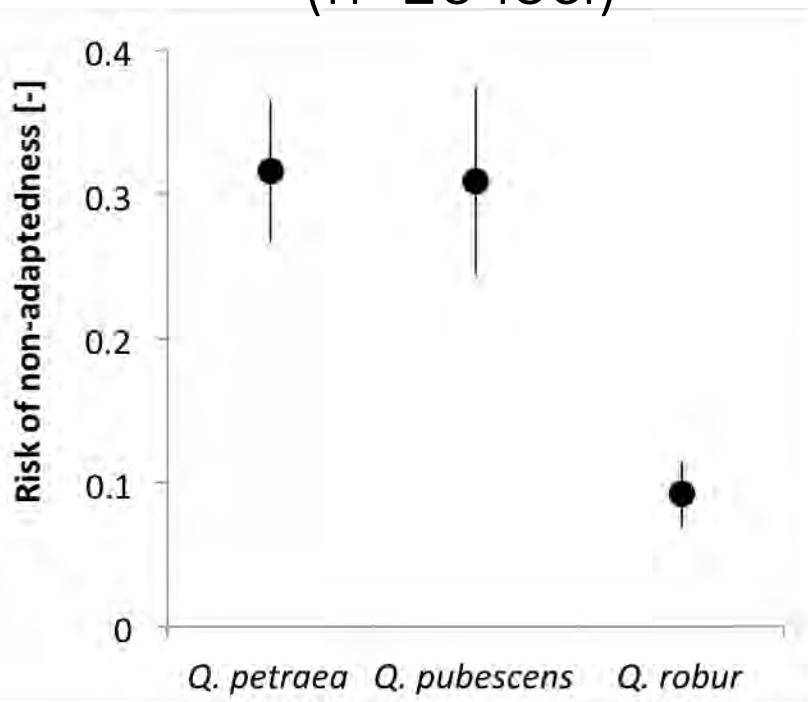


RONA in white oaks (*Quercus* spp.)

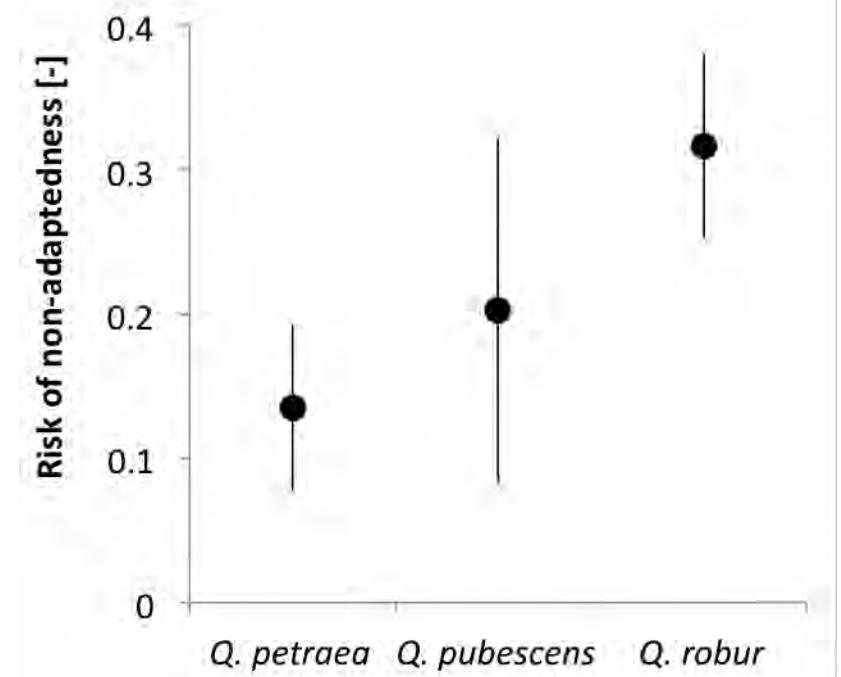
Average change in allele frequency required
to match future climatic conditions



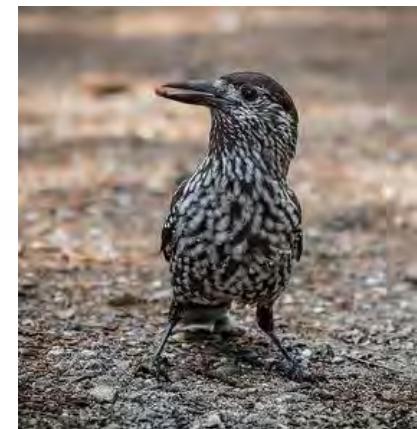
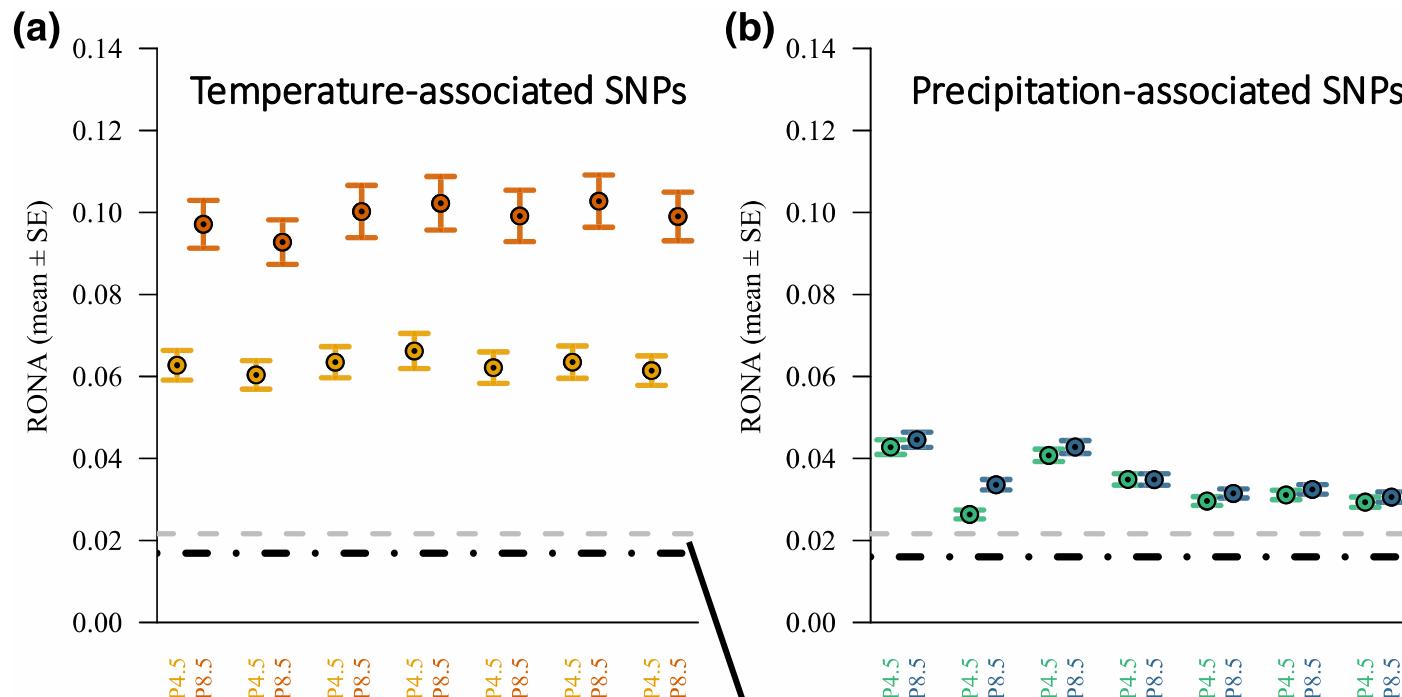
Mean temperature (\pm SD)
(n=20 loci)



Site water balance (\pm SD)
(n=17 loci)



RONA in Swiss stone pine (*Pinus cembra*)



Past allele frequency change

Genomic offset

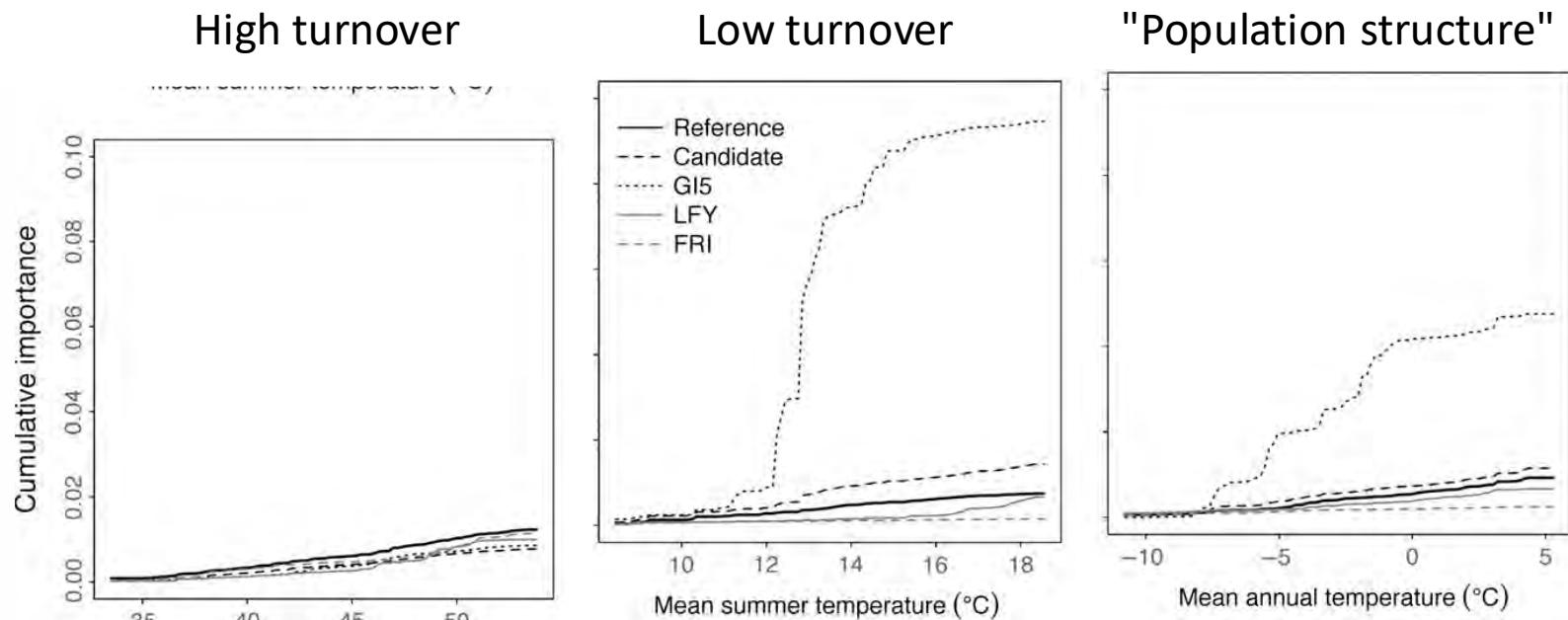
- A genetic distance between the present and required future state of a population
- In RONA = average required allele frequency change to be adapted to future conditions
- Magnitude that current GEAs would be disrupted by environmental change
- **Relative risk for future maladaptation from the point of today**
- Populations can counteract to this risk by, e.g.,
 - Adaptation based on standing genetic variation
 - Gene flow and hybridization/backcrossing
 - Migration
 - Phenotypic plasticity
 - Epigenetic changes

Original method: Genetic offset

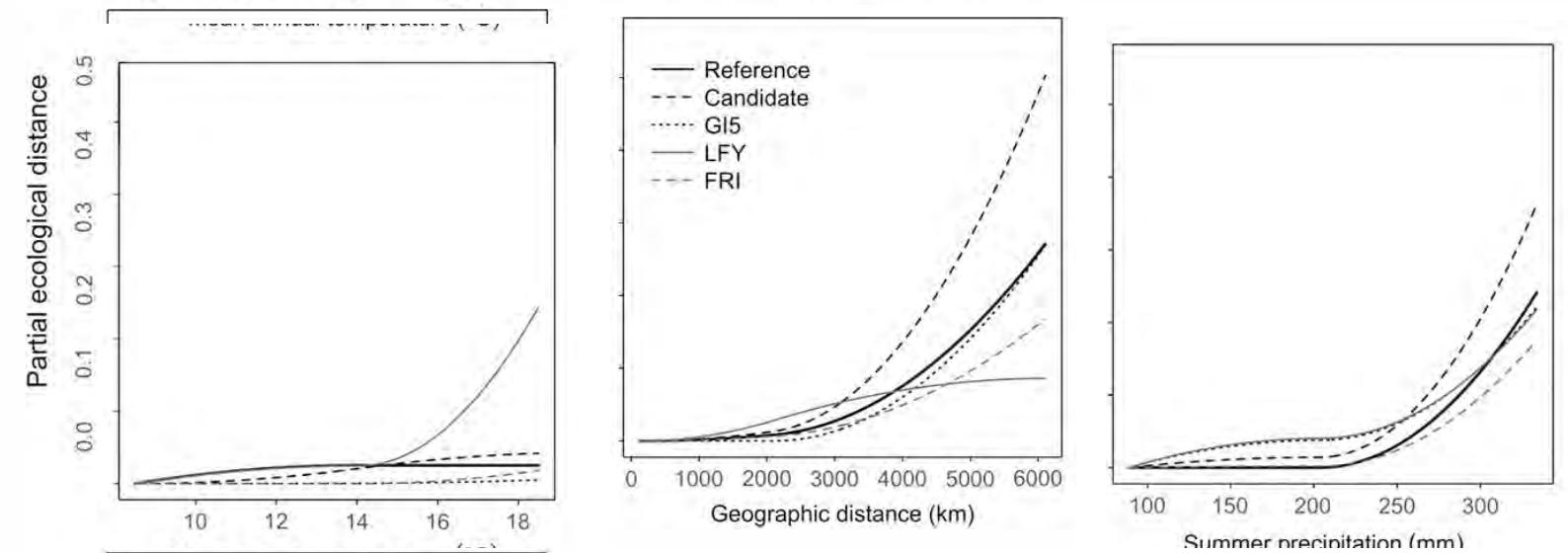
- The **original method** to assess possible genetic maladaptation to new/future environmental conditions
- Allele frequency **turnovers** along environmental gradients
- Originally from community modelling looking at species turnover = changes in species composition
- Two different methods
 - **Gradient forest (GF)**: Based on the machine-learning random forest method
 - **Generalized dissimilarity modelling (GDM)**: Based on pairwise dissimilarity matrices of genetic, environmental, and spatial components

Genetic offset

GF



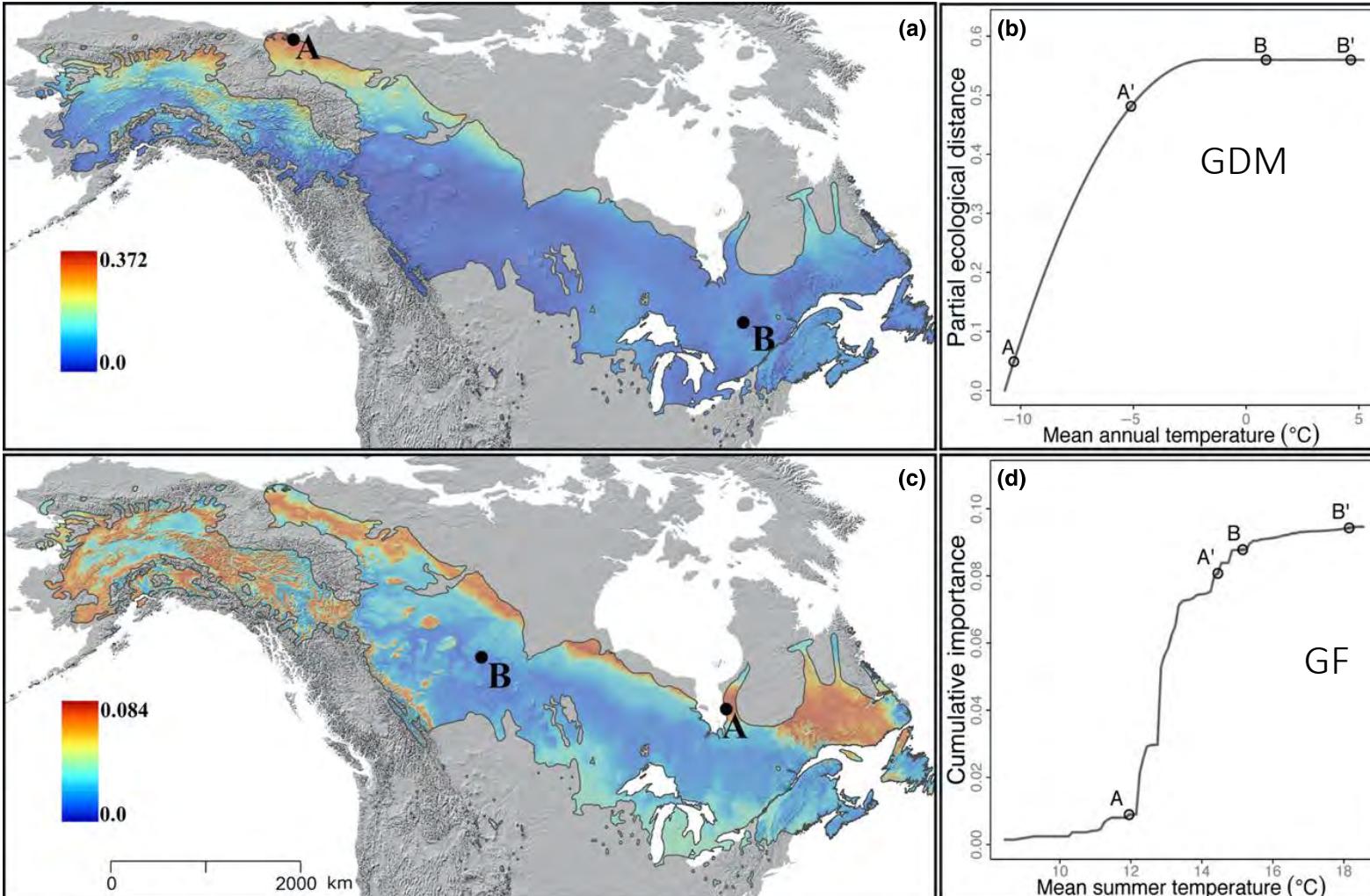
GDM



Genetic offset



In candidate genes of *Populus balsamifera* (Balsam poplar)



Genomic Offset – Overview of existing approaches

Name	Model type	Neutral structure	Multivariate X (Environment)	Multivariate Y (Genotypes)	Ind./pop. data	References
Genetic Gap	Linear mixed	Yes	Both	Yes	Both	Gain & Francois 2021
BayPass	Linear mixed	Yes	Both	Yes	Population ***	Gautier 2015 Camus et al. 2024
(p)RDA	Linear	Both	Yes	Yes	Both	Legendre & Legendre 2012 Capblancq & Forester 2021
Gradient Forest	Machine learning (monotonic)	Both*	Yes	No	Both	Fitzpatrick & Keller 2015
GDM	Non-linear dissimilarity	Both*	Yes	No	Both	Fitzpatrick & Keller 2015
RONA	Linear	Both*/**	Both**	Yes**	Both	Rellstab et al. 2016

* by manually adding covariates describing population structure

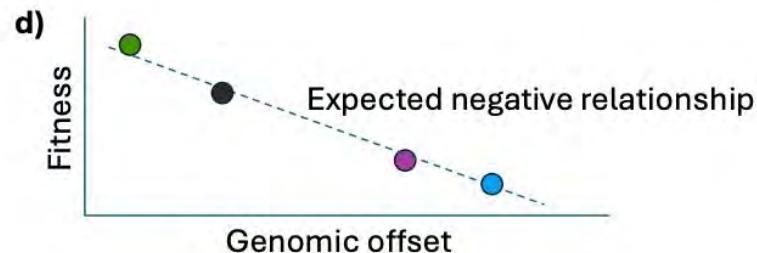
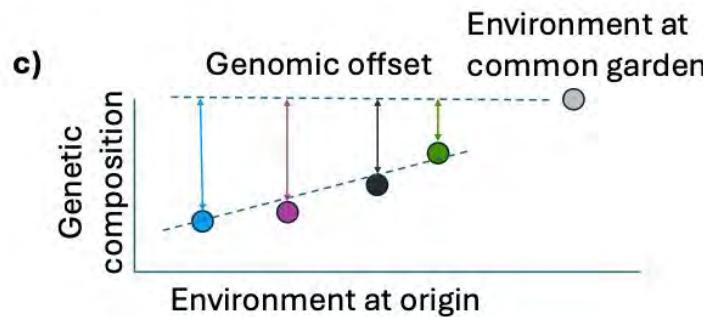
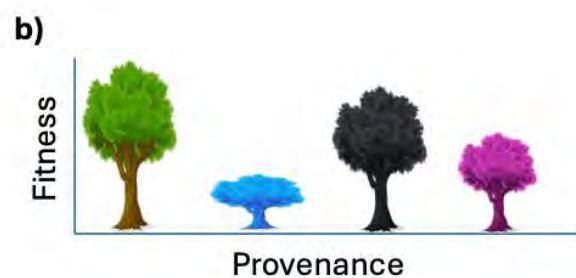
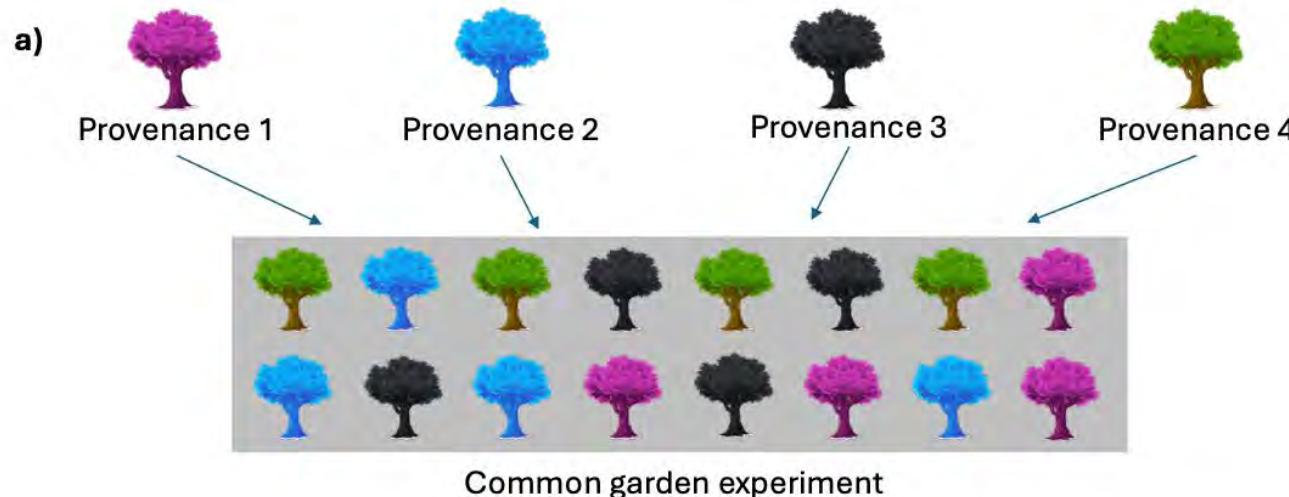
** implemented in GeneticGap (environmental distance) and BayPass (compute.rona=T)

*** can handle individual genotype likelihoods

How to validate genomic offset?

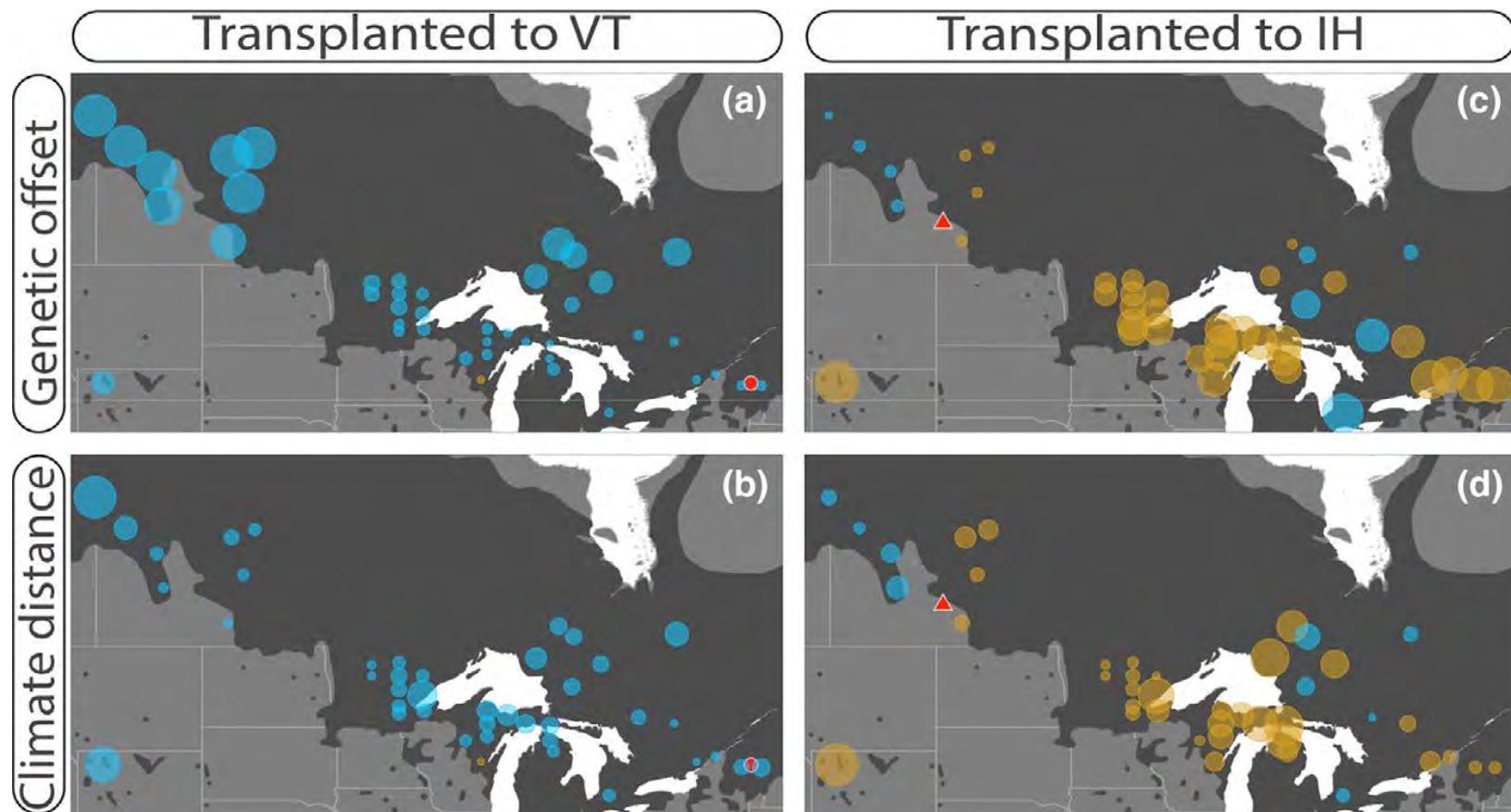
- Validation is difficult due to the predictive nature of genomic offset
- Possibilities:
 - Time-series
 - Additional data/approaches
 - **Common garden experiments**
 - Simulations

Validation in common gardens



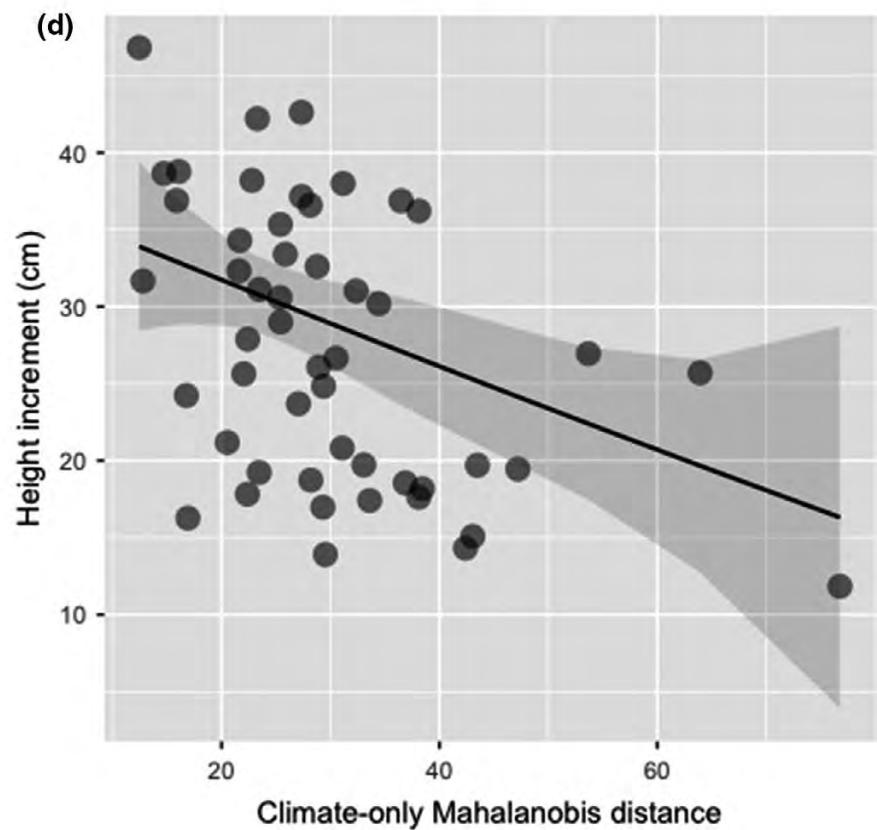
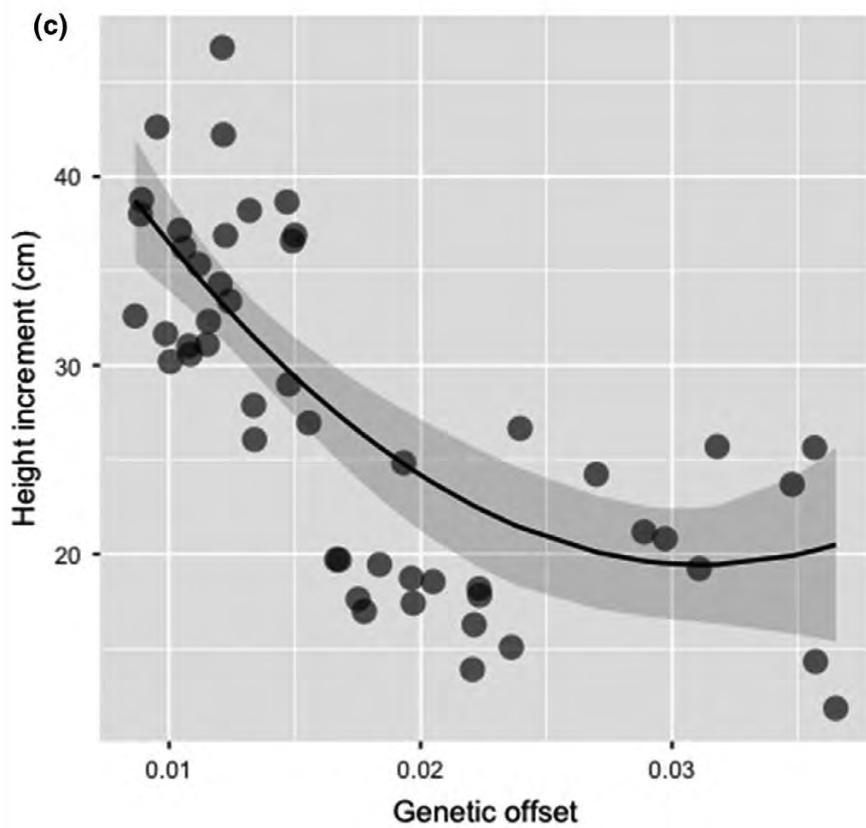
Validation in common gardens

Balsam poplar (*Populus balsamifera* L.).

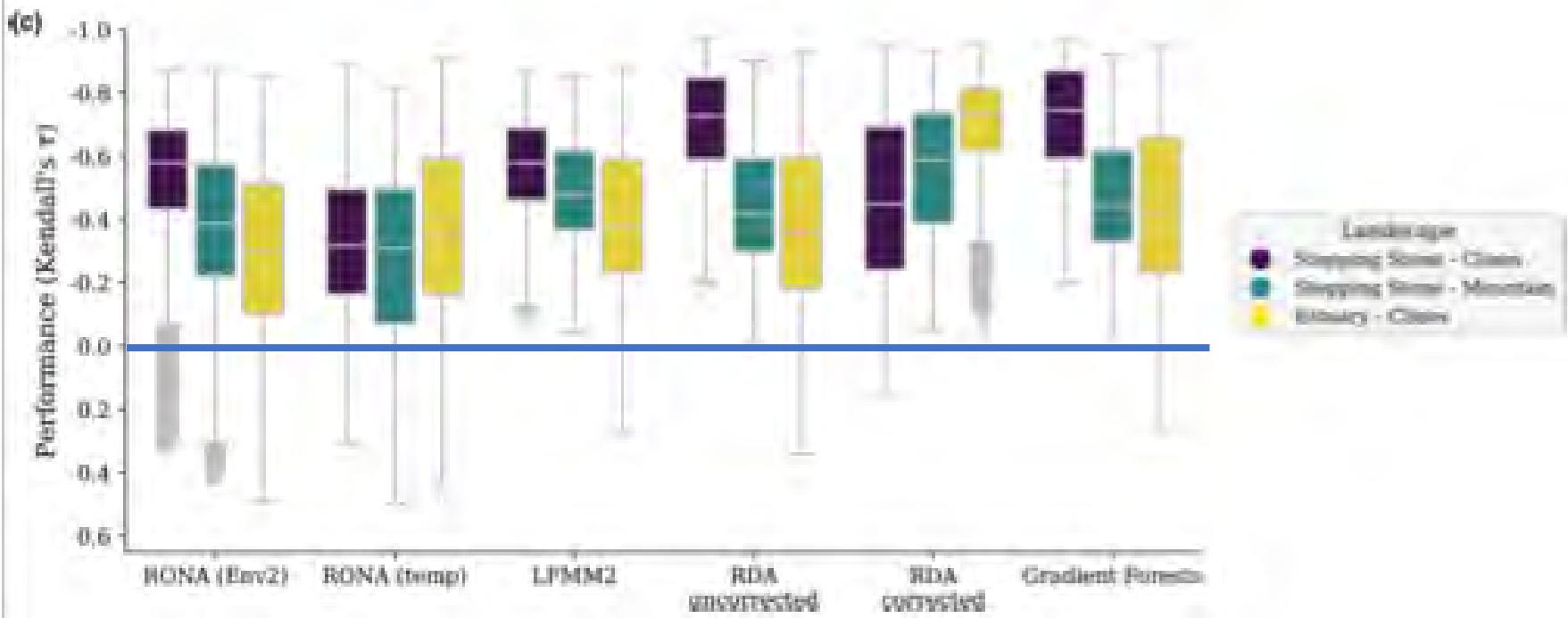
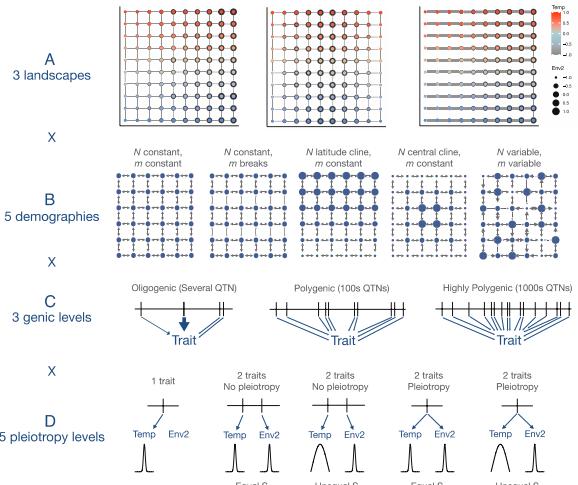


Validation in common gardens

- Genomic offset predicts fitness better than climate distance!
- Random SNP set performed better than candidate SNP set!

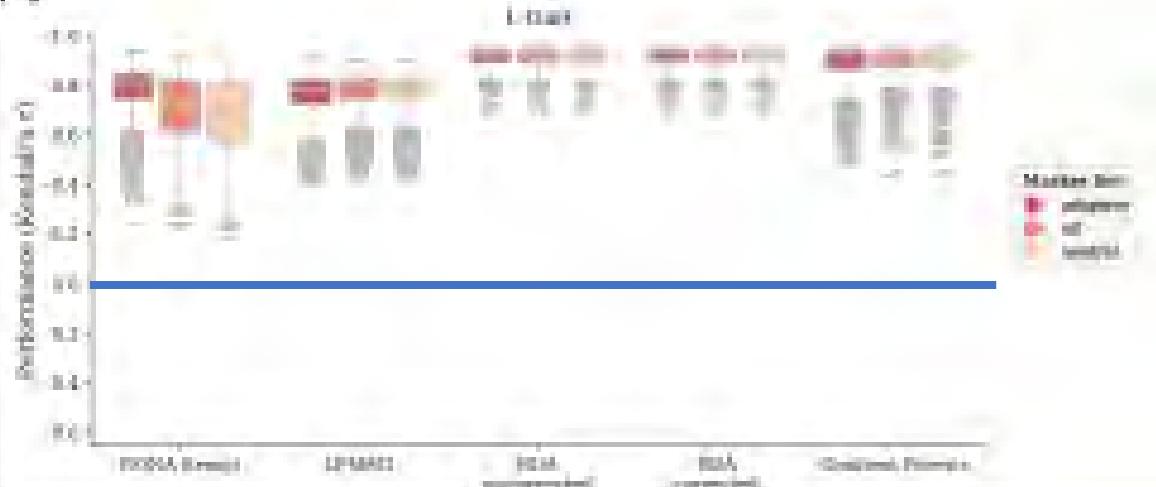


Simulations (and common gardens)



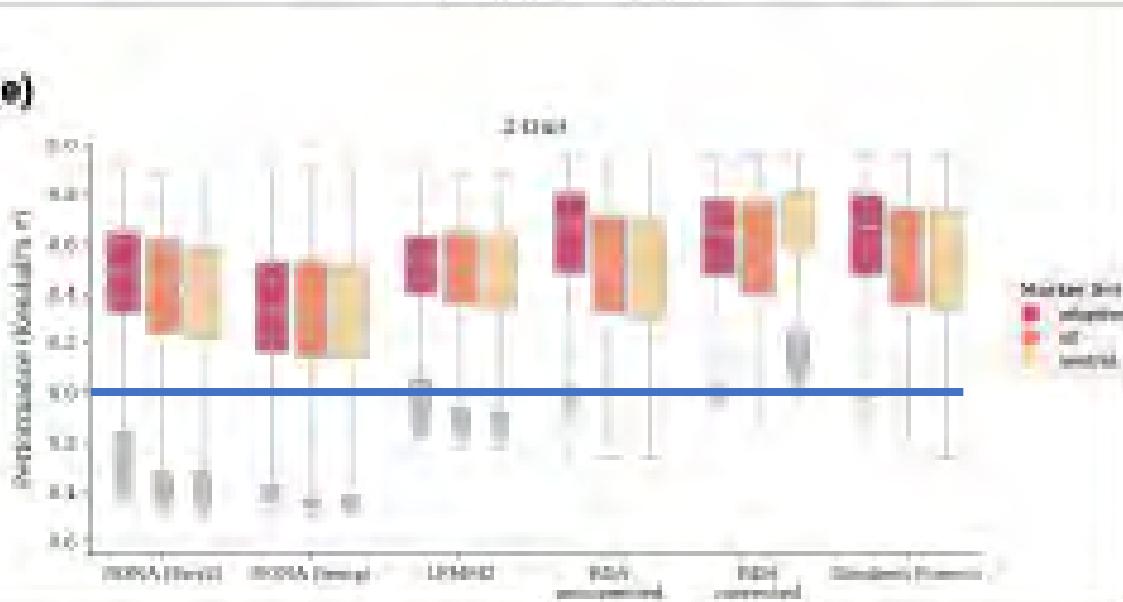
Simulations: SNP set

(d)

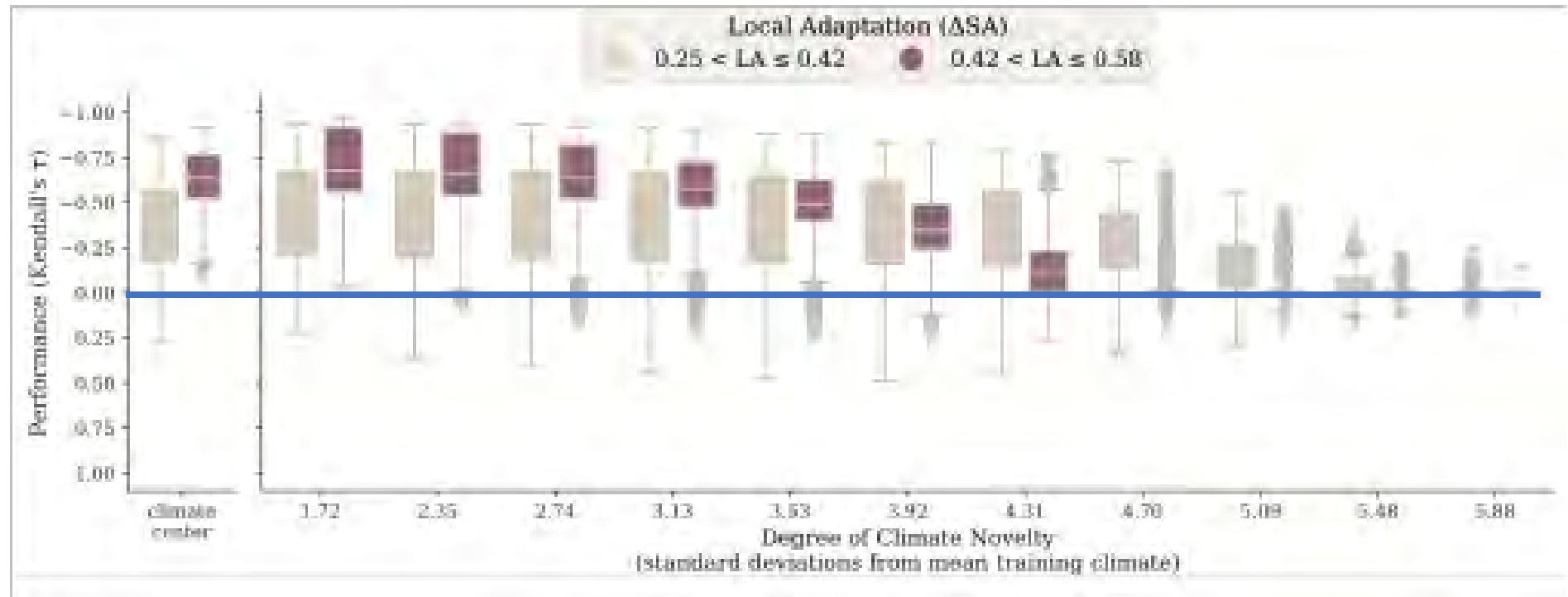


- Using adaptive SNPs only slightly increase performance!
- Performance decreases with pleiotropy and polygenicity

(e)



Simulations: Climate novelty



Performance decreases with climate novelty

Genomic offset – Strengths and limitations

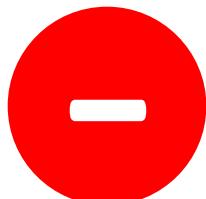
Strengths

- Can assess and compare the risk of genetic maladaptation to new/future conditions
- Without performing experiments and trait measurements
- Can be used in assisted migration/gene flow and conservation genomics



Limitations

- Space-for-time-substitution
- Model fits can be low
- Mostly ignores future neutral population structure
- Mostly based on standing genetic variation and adaptation *in-situ* (no gene flow)
- Assumes that loci are relevant for fitness!



Summary

- Genomic offset is a **3-step process**:
 - Perform a GEA to establish the E-G relationship
 - Predict the expected genomic composition for new/future conditions
 - Calculate the difference between current and required new/future genomic composition
- Genomic offset should be complemented with additional data and, at best, validated

Literature – Reviews/simulations

- **Capblancq**, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M. & Keller, S. R. (2020). Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 51, 245-269.
- **Gain**, C., Rhoné, B., Cubry, P., Salazar, I., Forbes, F., Vigouroux, Y., . . . François, O. (2023). A Quantitative Theory for Genomic Offset Statistics. *Molecular Biology and Evolution*, 40.
- **Hoffmann**, A. A., Weeks, A. R. & Sgrò, C. M. (2021). Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell*, 184, 1420-1425.
- **Lind**, B. M. & Lotterhos, K. E. (2025). The accuracy of predicting maladaptation to new environments with genomic data. *Molecular Ecology Resources*, 25, e14008.
- **Rellstab**, C., Dauphin, B. & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, 14, 1202-1212.

ETH/Evoltree course

"Genomics of environmental adaptation"

- 26.-30.1.2026
- WSL, Birmensdorf, Switzerland
- ETH students and Evoltree members have priority
- Tentative program:
 - Day 1:
 - Genomic markers and NGS techniques
 - Selection, neutral and adaptive genetic diversity, genetic drift, population genetic structure
 - Day 2:
 - Outlier analysis
 - Day 3:
 - Environmental data
 - Genotype-environment associations (GEA)
 - Day 4:
 - Genotype-environment associations (GEA)
 - Genomic offset
 - Day 5:
 - Genome-wide association studies (GWAS)
 - Follow-up analyses



Teachers: Christian Rellstab, Felix Gugerli, Benjamin Dauphin, Christine Grossen, Gözde Cilingir