Applications to modelling and landscape genomics

Juliette Archambeau

juli.archambeau@orange.fr

Lisbon, June 22, 2022

Current issues in the context of climate change

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  $\rightarrow$ Identifying which populations (natural or planted) may be at higher risk under climate change.

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  $\rightarrow$ Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  → Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

  → Identifying optimal genotypes to move to increase the fitness of a threatened population in **assisted gene flow strategies**.

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  → Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

  → Identifying optimal genotypes to move to increase the fitness of a threatened population in **assisted gene flow strategies**.

  → Identifying **optimal** source populations to **restore** degraded ecosystems, to plant in a managed area.

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  → Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

  → Identifying optimal genotypes to move to increase the fitness of a threatened population in **assisted gene flow strategies**.

  → Identifying **optimal** source populations to **restore** degraded ecosystems, to plant in a managed area.

- How to determine **which populations** to plant in a **given location**, or in **which area** to plant a **given population**?

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  → Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

  → Identifying optimal genotypes to move to increase the fitness of a threatened population in **assisted gene flow strategies**.

  → Identifying **optimal** source populations to **restore** degraded ecosystems, to plant in a managed area.

- How to determine **which populations** to plant in a **given location**, or in **which area** to plant a **given population**?

  → Selecting the appropriate experiments and statistical tools.

Current issues in the context of climate change

- Will forest tree populations be able to persist in their **current locations** under **future climates**?

  → Identifying which populations (natural or planted) may be at higher risk under climate change.

- Or in the **locations** where they will be **planted**?

  → Identifying optimal genotypes to move to increase the fitness of a threatened population in **assisted gene flow strategies**.

  → Identifying **optimal** source populations to **restore** degraded ecosystems, to plant in a managed area.

- How to determine **which populations** to plant in a **given location**, or in **which area** to plant a **given population**?

  → Selecting the appropriate experiments and statistical tools.

  → Role of the common gardens

How will populations respond to new environments?

How will populations respond to new environments?

Different ways to address the question...

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

  → Active area of research and method development, major advances within the landscape genomics framework, still in need of validation.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

  → Active area of research and method development, major advances within the landscape genomics framework, still in need of validation.

- Predicting the **long-term responses** of the populations, i.e. ecological and evolutionary trajectories.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

  → Active area of research and method development, major advances within the landscape genomics framework, still in need of validation.

- Predicting the **long-term responses** of the populations, i.e. ecological and evolutionary trajectories.

  → Relies on multiple and interacting processes.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic responses** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → **Trait-based methods** already available, built upon the framework of **quantitative genetics**, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary response to changing conditions.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

  → Active area of research and method development, major advances within the landscape genomics framework, still in need of validation.

- Predicting the **long-term responses** of the populations, i.e. ecological and evolutionary trajectories.

  → Relies on multiple and interacting processes.

  → Highly complex modeling approaches, hardly testable, and whose predictions therefore remain highly uncertain.

How will populations respond to new environments?

Different ways to address the question...

- Predicting the **short-term phenotypic response** of the populations, i.e. without evolution.

  → Relies on **phenotypic plasticity** and **standing genetic variation**.

  → Trait-based methods already available, built upon the framework of quantitative and population genetics, integration of genomic data under development but promising.

- Predicting the **adaptive potential** of the populations, i.e. the potential for an evolutionary change in the next few generations.

  → Relies on the levels of **heritable variation**, the **phenotypic covariance** between **fitness** and the **traits under selection**, the extent of gene flow and the anecdotal onset of new advantageous mutations.

  → Active area of research and method development, major advances within the landscape genomics framework, still in need of validation.

- Predicting the **long-term responses** of the populations, i.e. ecological and evolutionary trajectories.

  → Relies on multiple and interacting processes.

  → Highly complex modeling approaches, hardly testable, and whose predictions therefore remain highly uncertain.

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

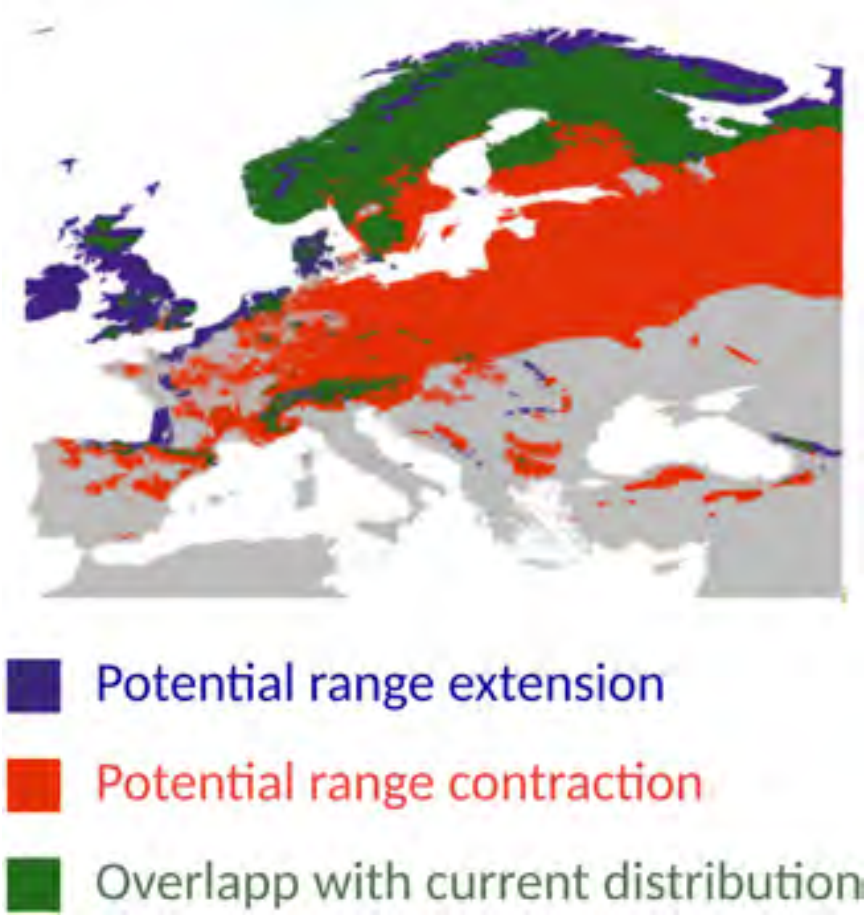
Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Species distribution models (SDMs)

Based on occurrence data.

Assume **uniformity of climate responses** below the species level.



Predicted future distribution of Scots pine under the (pessimistic) scenario RCP8.5.

Dyderski *et al.* 2017

Climate-based population response functions

Climate-based population response functions

Population-level models that:
- integrate **intraspecific trait variation**,

Climate-based population response functions

Population-level models that:

- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Climate-based population response functions

Population-level models that:

- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Climate-based population response functions

Population-level models that:

- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007

Climate-based population response functions

Population-level models that:

- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007

Climate-based population response functions

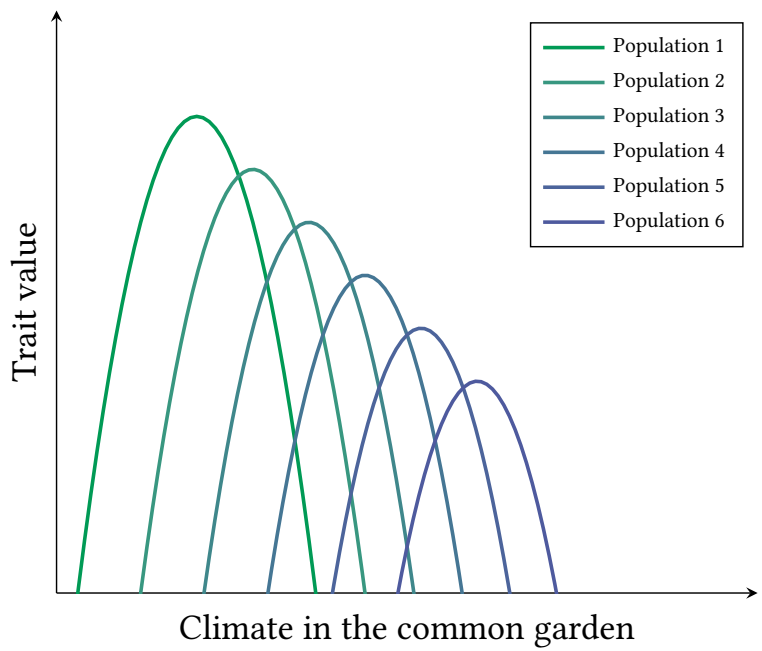
Population-level models that:
- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007



Separate the **genetic** and **plastic** components of trait variation.

Climate-based population response functions

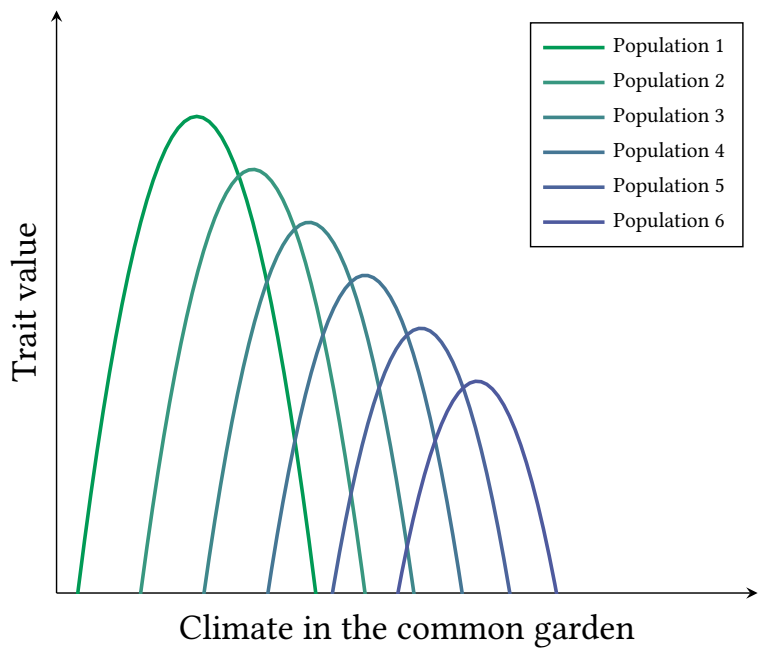
Population-level models that:
- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007



Separate the **genetic** and **plastic** components of trait variation.

Identify **potential drivers** of the genetic and plastic components.

Climate-based population response functions

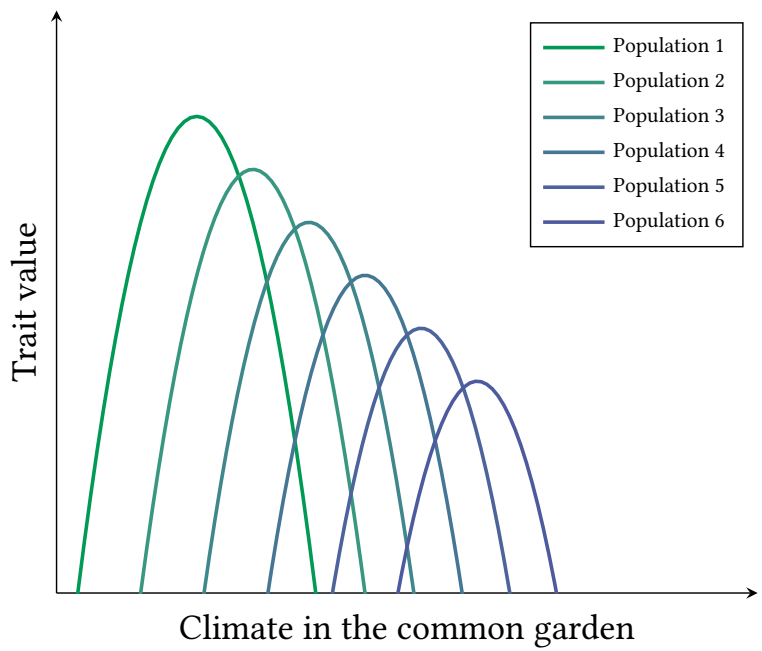
Population-level models that:
- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007



Separate the **genetic** and **plastic** components of trait variation.

Identify **potential drivers** of the genetic and plastic components.

Predict trait variation under **climate change** scenarios.

Climate-based population response functions

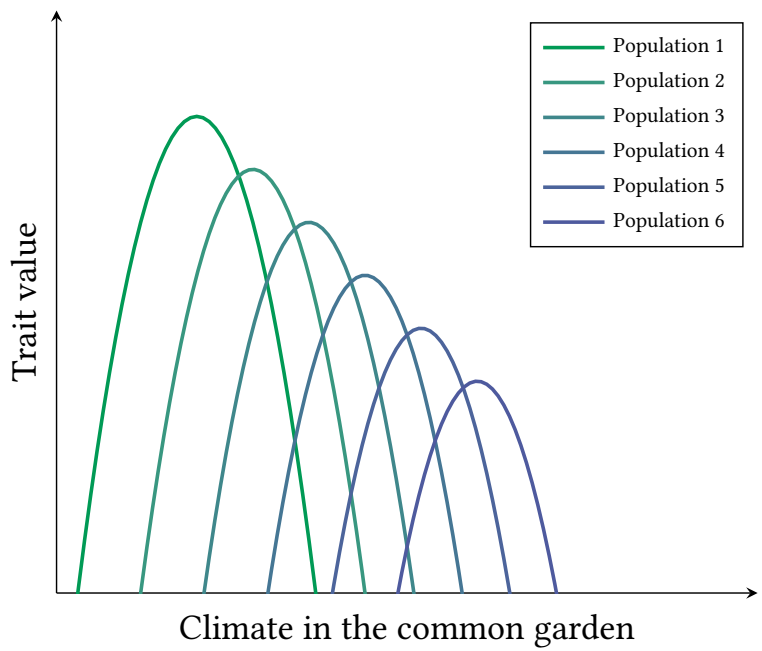
Population-level models that:
- integrate **intraspecific trait variation**,
- use the **climate-of-origin** of the populations as a proxy of **genetic adaptation** to climate.

Commonly-used in forest trees.

Rehfeldt *et al.* 1999, Wang *et al.* 2006, 2010, Leites *et al.* 2012, Benito Garzón *et al.* 2019

Based on a long history of provenance trials.

Langlet 1971, Morgenstern 1996, Savolainen *et al.* 2007



Separate the **genetic** and **plastic** components of trait variation.

Identify **potential drivers** of the genetic and plastic components.

Predict trait variation under **climate change** scenarios.

Rely on **multi-site common gardens**.

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al*. 2016, Josephs *et al*. 2019

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al*. 2016, Josephs *et al*. 2019

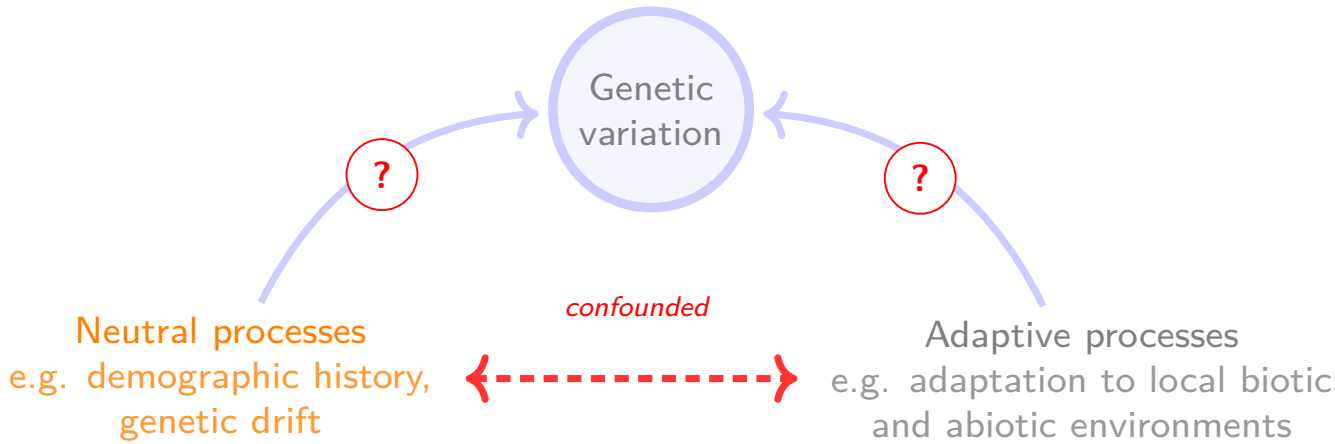
- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al.* 2016, Josephs *et al.* 2019

- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.

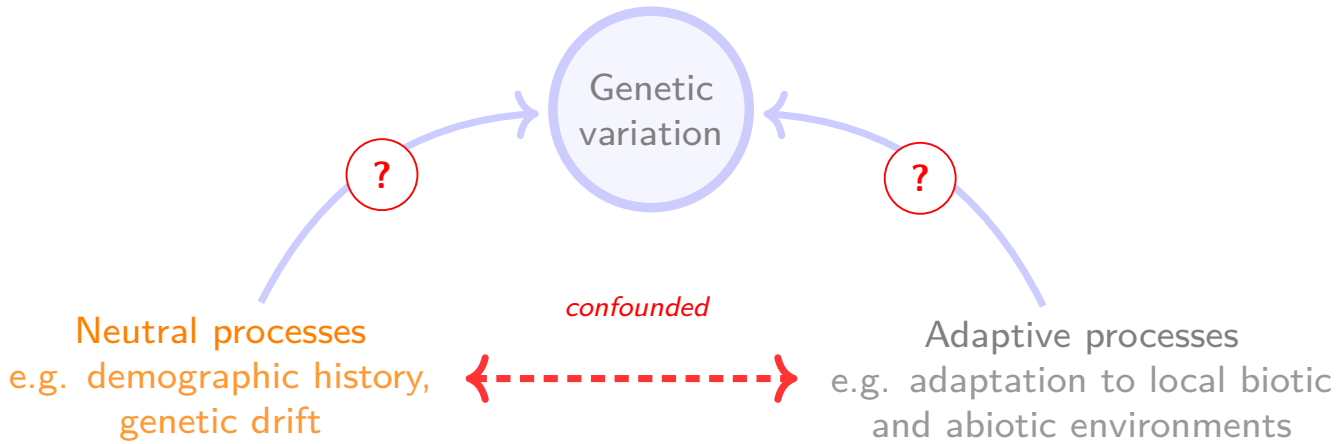


Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al.* 2016, Josephs *et al.* 2019

- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.



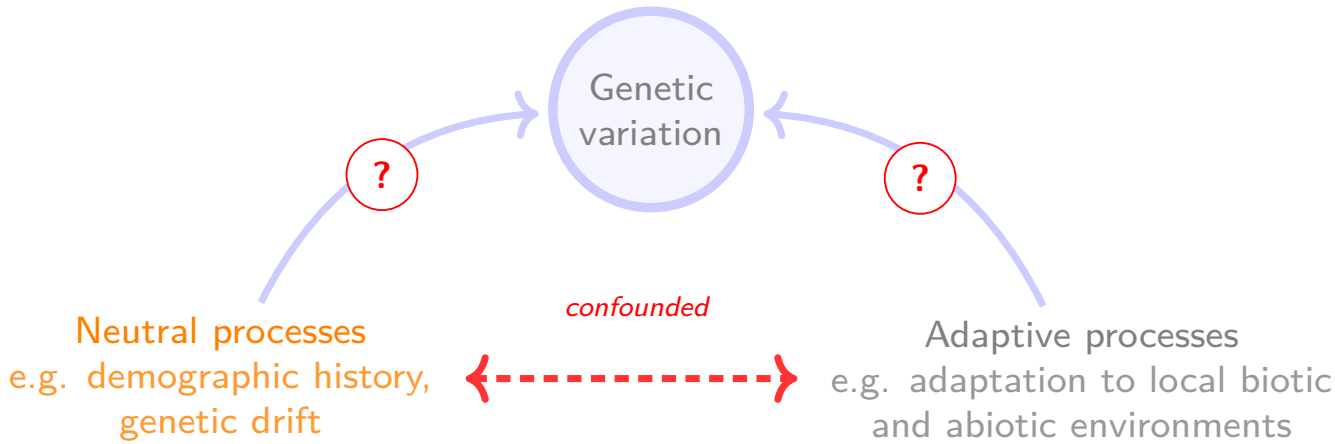
- detecting **adaptive variants/traits**.

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al.* 2016, Josephs *et al.* 2019

- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.



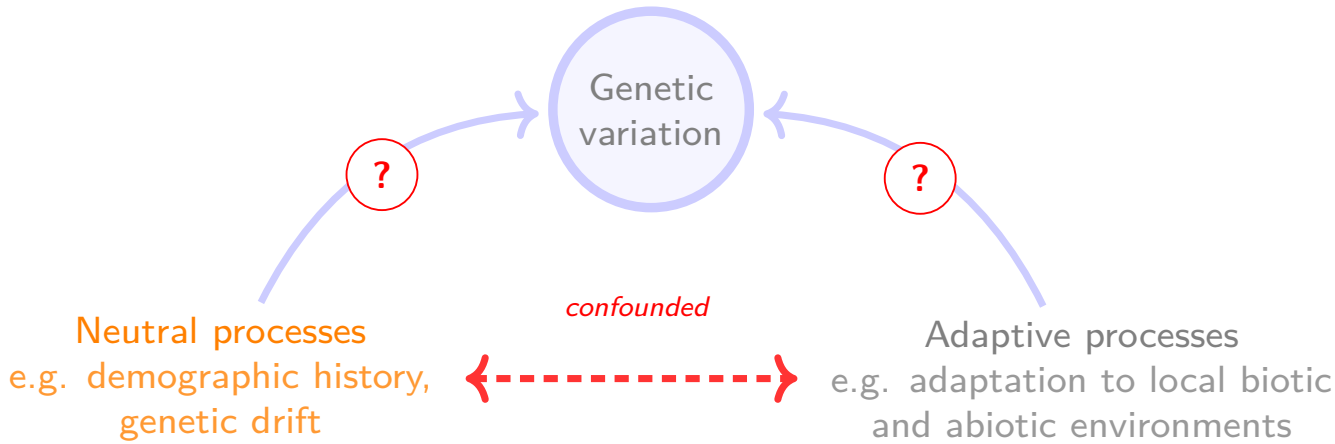
- detecting **adaptive variants/traits**.
- improving **individual-level predictions** of quantitative traits

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al.* 2016, Josephs *et al.* 2019

- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.



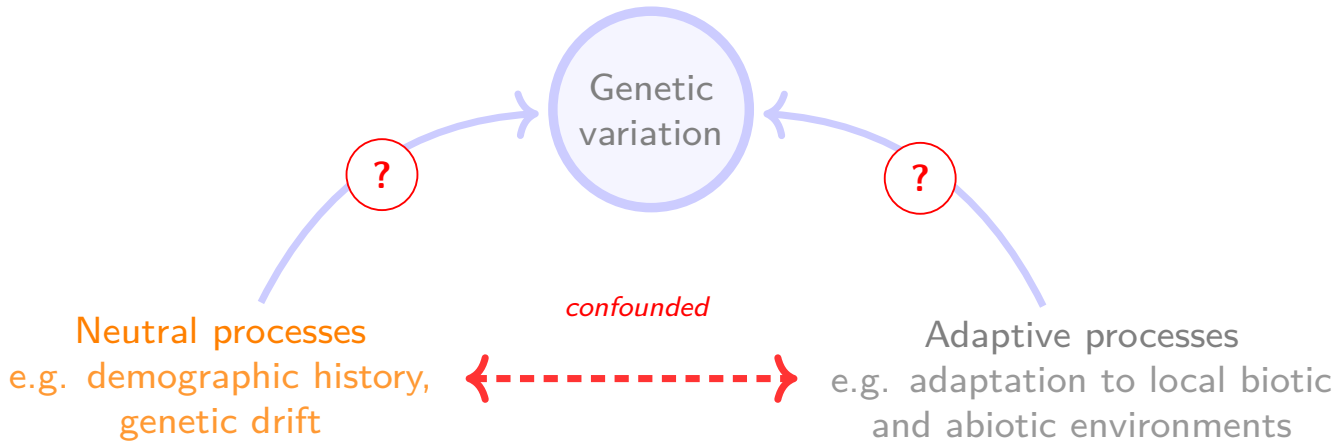
- detecting **adaptive variants/traits**.
- improving **individual-level predictions** of quantitative traits
- broadening the predictions to a **larger** number of **species** and **populations**.

Towards genome-informed predictions

Integrating **genomic data** within the modeling of quantitative **trait** variation provides new opportunities.

de Villemereuil *et al.* 2016, Josephs *et al.* 2019

- understanding the **evolutionary processes** behind the genetic component of quantitative trait variation.

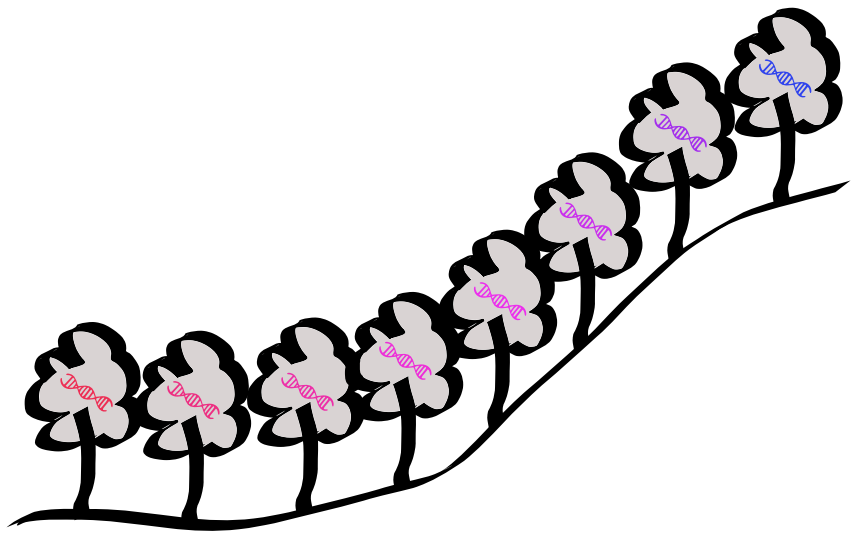


- detecting **adaptive variants/traits**.
- improving **individual-level predictions** of quantitative traits
- broadening the predictions to a **larger** number of **species** and **populations**.
- increasing the chances of **fast-adaptive decision** making.

Towards genome-informed predictions

Landscape genomics: the study of the processes shaping the geographical patterns of adaptive genetic variation across the landscape.

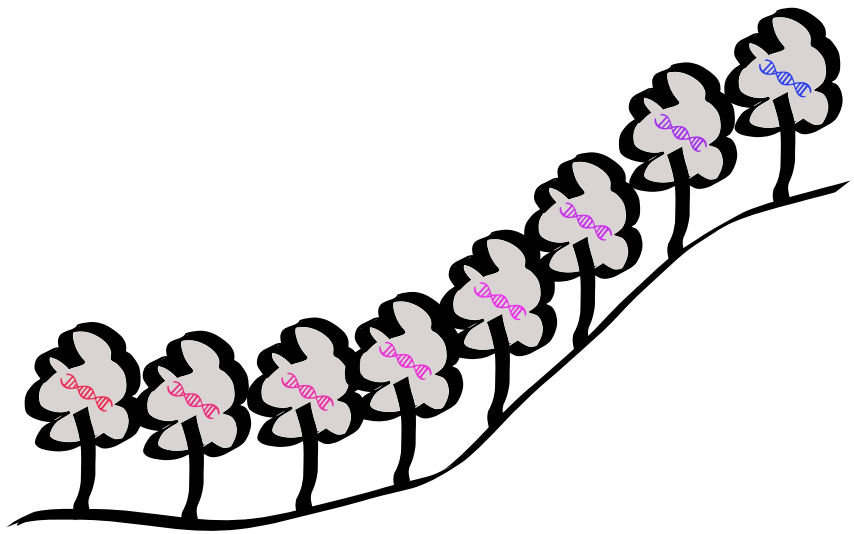
Storfer *et al.* 2018



Towards genome-informed predictions

Landscape genomics: the study of the processes shaping the geographical patterns of adaptive genetic variation across the landscape.

Storfer *et al.* 2018

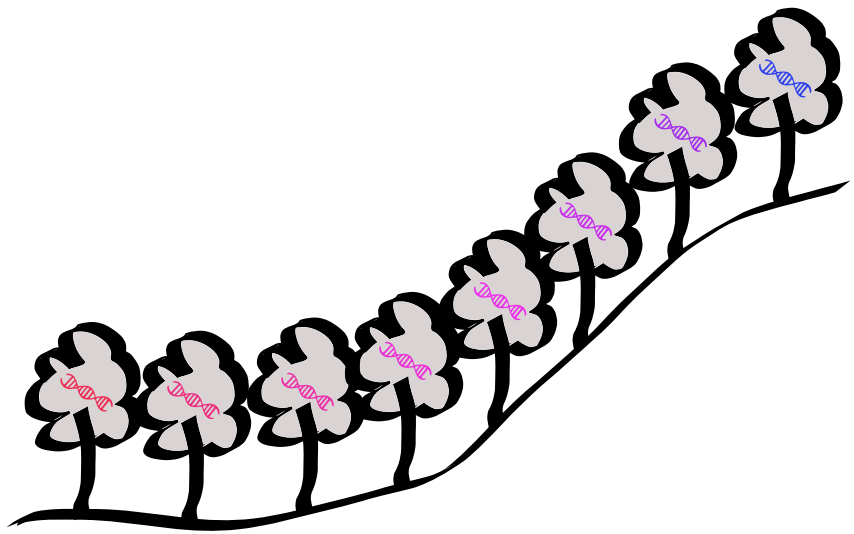


Phenotype-free approaches avoid the potential biases arising from:

Towards genome-informed predictions

Landscape genomics: the study of the processes shaping the geographical patterns of adaptive genetic variation across the landscape.

Storfer *et al.* 2018



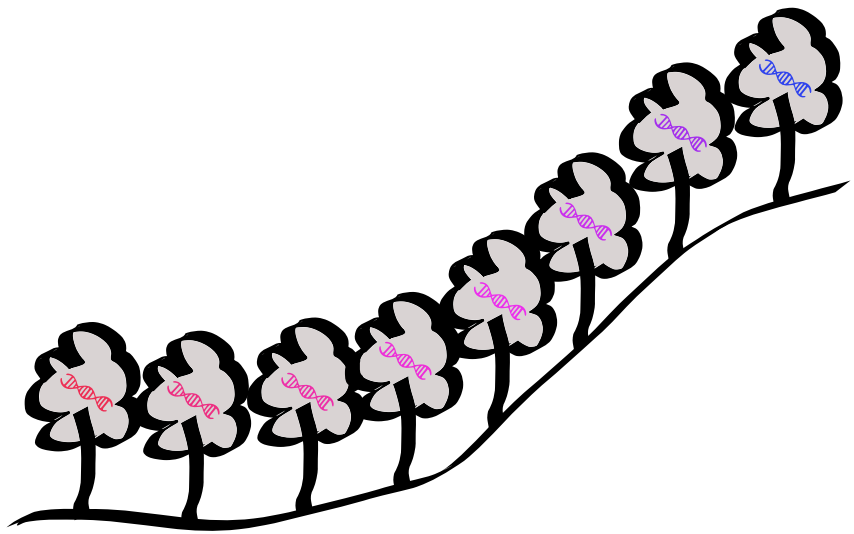
Phenotype-free approaches avoid the potential biases arising from:

- the choice of which phenotypes to measure, in which ontogenic stage or environment, etc.

Towards genome-informed predictions

Landscape genomics: the study of the processes shaping the geographical patterns of adaptive genetic variation across the landscape.

Storfer *et al.* 2018



Phenotype-free approaches avoid the potential biases arising from:

- the choice of which phenotypes to measure, in which ontogenic stage or environment, etc.
- the uncertain link between fitness and the phenotypes.

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: two main types of statistical models

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

- **Animal model**

Henderson 1975

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

• Animal model

Henderson 1975

Breeding values are estimated from the **covariance among relatives** in **additive genetic effects**, such as:

$$a_i \sim \mathcal{N}(0, \mathbf{A}V_A)$$

V_A is the additive genetic variance.
A is the relatedness matrix.

A can either be estimated from a pedigree or from genomic markers (genomic relationship matrix).

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

• Animal model

Henderson 1975

Breeding values are estimated from the **covariance among relatives** in **additive genetic effects**, such as:

$$a_i \sim \mathcal{N}(0, \mathbf{A}V_A)$$

V_A is the additive genetic variance.
A is the relatedness matrix.

A can either be estimated from a pedigree or from genomic markers (genomic relationship matrix).

• Genome regression

Meuwissen *et al.* 2001

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

• Animal model

Henderson 1975

Breeding values are estimated from the **covariance among relatives** in **additive genetic effects**, such as:

$$a_i \sim \mathcal{N}(0, \mathbf{A}V_A)$$

V_A is the additive genetic variance.
A is the relatedness matrix.

A can either be estimated from a pedigree or from genomic markers (genomic relationship matrix).

• Genome regression

Meuwissen *et al.* 2001

Breeding values are calculated as the **sum of the additive effects** of the markers (previously estimated), such as:

$$a_i = \sum (SNP_{ij}\gamma_j)$$

SNP_{ij} is the genotype of the individual i at the locus j.
γ_j is the effect of the SNP on the phenotype.

Estimating **breeding values**: two main types of statistical models

$$y_i = \mu + a_i + e_i$$

y_i is the phenotype of the individual i.
μ is the average population phenotype.
a_i is the **breeding value** of the individual i, i.e. its **aggregate additive genetic value**, of which the individual passes on average half to its offspring.
e_i is the residual variation.

• Animal model

Henderson 1975

Breeding values are estimated from the **covariance among relatives** in **additive genetic effects**, such as:

$$a_i \sim \mathcal{N}(0, \mathbf{A}V_A)$$

V_A is the additive genetic variance.
A is the relatedness matrix.

A can either be estimated from a pedigree or from genomic markers (genomic relationship matrix).

• Genome regression

Meuwissen *et al.* 2001

Breeding values are calculated as the **sum of the additive effects** of the markers (previously estimated), such as:

$$a_i = \sum (SNP_{ij}\gamma_j)$$

SNP_{ij} is the genotype of the individual i at the locus j.
γ_j is the effect of the SNP on the phenotype.

Known as **polygenic risk scores** (PRS) in humans.

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: from marker-assisted to genomic selection

Jannink *et al.* 2010, Meuwissen *et al.* 2016

- **Marker-assisted selection (MAS)**

Using a **subset of SNPs** the most associated with the phenotypes.

Likely more accurate for **oligonenic** traits.

Often overpredicts phenotypes because QTL effects are estimated independently of each other.

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: from marker-assisted to genomic selection

Jannink *et al.* 2010, Meuwissen *et al.* 2016

• Marker-assisted selection (MAS)

Using a **subset of SNPs** the most associated with the phenotypes.

Likely more accurate for **oligonenic** traits.

Often overpredicts phenotypes because QTL effects are estimated independently of each other.

• Genomic selection (GS)

Using **all SNPs**, regardless of the magnitude of their association with the phenotypes.

Likely more accurate for **polygenic** traits.

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: from marker-assisted to genomic selection

Jannink *et al.* 2010, Meuwissen *et al.* 2016

- **Marker-assisted selection (MAS)**

Using a **subset of SNPs** the most associated with the phenotypes.

Likely more accurate for **oligonenic** traits.

Often overpredicts phenotypes because QTL effects are estimated independently of each other.

- **Genomic selection (GS)**

Using **all SNPs**, regardless of the magnitude of their association with the phenotypes.

Likely more accurate for **polygenic** traits.

In **small breeding populations**, GS methods now predict breeding values **better than** pedigree-based approaches.

Jannink *et al.* 2010

Methods and limits in predicting phenotypes based on genotypes

Estimating **breeding values**: from marker-assisted to genomic selection

Jannink *et al.* 2010, Meuwissen *et al.* 2016

- **Marker-assisted selection (MAS)**

Using a **subset of SNPs** the most associated with the phenotypes.

Likely more accurate for **oligonenic** traits.

Often overpredicts phenotypes because QTL effects are estimated independently of each other.

- **Genomic selection (GS)**

Using **all SNPs**, regardless of the magnitude of their association with the phenotypes.

Likely more accurate for **polygenic** traits.

In **small breeding populations**, GS methods now predict breeding values **better than** pedigree-based approaches.

Jannink *et al.* 2010

In natural populations of **large effective population size**, GS methods still have slightly lower predictive accuracy than pedigree-based approaches.

e.g. in Beaulieu *et al.* 2014. When predicting wood and growth traits in a large undomesticated population of related white spruces (within-family predictions, i.e. with known relatedness between training and testing data sets), GS methods achieved between 68 and 89% of the accuracy obtained with pedigree information.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al.* 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

$\rightarrow$ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al.* 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

→ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

e.g. in Beaulieu *et al.* 2014. In white spruce, GS models have only 30% reductions in predictive accuracy between sites compared to predictions within sites for wood density and microfibril angle.

→ Can be explained by the generally low **genotype-by-environment interactions** in white spruce for growth and wood traits.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al*. 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al*. 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

→ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

e.g. in Beaulieu *et al*. 2014. In white spruce, GS models have only 30% reductions in predictive accuracy between sites compared to predictions within sites for wood density and microfibril angle.

→ Can be explained by the generally low **genotype-by-environment interactions** in white spruce for growth and wood traits.

- populations

e.g. in Resende *et al*. 2012. In two large genetically unrelated *Eucalyptus* breeding populations, GS models predicted poorly **across populations** for tree circumference, height growth, wood specific gravity and pulp yield.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al.* 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

→ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

e.g. in Beaulieu *et al.* 2014. In white spruce, GS models have only 30% reductions in predictive accuracy between sites compared to predictions within sites for wood density and microfibril angle.

→ Can be explained by the generally low **genotype-by-environment interactions** in white spruce for growth and wood traits.

- populations

e.g. in Resende *et al.* 2012. In two large genetically unrelated *Eucalyptus* breeding populations, GS models predicted poorly **across populations** for tree circumference, height growth, wood specific gravity and pulp yield.

→ Can be explained by **genotype-by-environment interactions**, **genetic interactions**, **variable patterns of linkage disequilibrium** and **inconsistent allelic effects**.

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al.* 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

→ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

e.g. in Beaulieu *et al.* 2014. In white spruce, GS models have only 30% reductions in predictive accuracy between sites compared to predictions within sites for wood density and microfibril angle.

→ Can be explained by the generally low **genotype-by-environment interactions** in white spruce for growth and wood traits.

- populations

e.g. in Resende *et al.* 2012. In two large genetically unrelated *Eucalyptus* breeding populations, GS models predicted poorly **across populations** for tree circumference, height growth, wood specific gravity and pulp yield.

→ Can be explained by **genotype-by-environment interactions**, **genetic interactions**, **variable patterns of linkage disequilibrium** and **inconsistent allelic effects**.

e.g. in Beaulieu *et al.* 2014. Predictive accuracy **slightly lower** for **between-family** than for **within-family** predictions, and much weaker for **between-population** predictions (though still different from zero for some traits, and still higher than pedigree-based approaches).

Methods and limits in predicting phenotypes based on genotypes

But, **weak predictive accuracy** across:

- ages

e.g. in Resende Jr *et al.* 2012. GS models developed for early growth traits (diameter and height at age 1 or 2 yr) had limited accuracy in predicting phenotypes at age 6 yr in loblolly pine.

- sites

e.g. in Resende Jr *et al.* 2012. In loblolly pine, GS models have a **high** predictive accuracy for diameter and height **within breeding zone** (0.64-0.68) but much **weaker** accuracies (0.18-0.32) **across breeding zones**.

→ **Genotype-by-environment interactions** may strongly affect the transferability of models across breeding zones.

e.g. in Beaulieu *et al.* 2014. In white spruce, GS models have only 30% reductions in predictive accuracy between sites compared to predictions within sites for wood density and microfibril angle.

→ Can be explained by the generally low **genotype-by-environment interactions** in white spruce for growth and wood traits.

- populations

e.g. in Resende *et al.* 2012. In two large genetically unrelated *Eucalyptus* breeding populations, GS models predicted poorly **across populations** for tree circumference, height growth, wood specific gravity and pulp yield.

→ Can be explained by **genotype-by-environment interactions**, **genetic interactions**, **variable patterns of linkage disequilibrium** and **inconsistent allelic effects**.

e.g. in Beaulieu *et al.* 2014. Predictive accuracy **slightly lower** for **between-family** than for **within-family** predictions, and much weaker for **between-population** predictions (though still different from zero for some traits, and still higher than pedigree-based approaches).

→ Can be explained by the presence of **shared ancestry** (with the relatedness picked up by the markers) and **historical LD** (likely if the sampled populations are from the same glacial lineage).

So, how to incorporate **genomic information** in phenotypic predictions encompassing **multiple natural populations**?

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Predicting phenotypes based on genotypes in natural populations

Some examples from forest tree research:

Predicting phenotypes based on genotypes in natural populations

Some examples from forest tree research:

- Using **genomic estimated breeding values** (GEBVs; the sum of the effects of genome-wide markers on the trait of interest), i.e. **polygenic scores**.

Browne L., Wright J. W., Fitz-Gibbon S., Gugger P. F., and Sork V. L. (2019). Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences* **116**.50: 25179– 25185.

Predicting phenotypes based on genotypes in natural populations

Some examples from forest tree research:

- Using **genomic estimated breeding values** (GEBVs; the sum of the effects of genome-wide markers on the trait of interest), i.e. **polygenic scores**.

Browne L., Wright J. W., Fitz-Gibbon S., Gugger P. F., and Sork V. L. (2019). Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences* **116**.50: 25179– 25185.

- Using counts of **positive-effect alleles** (PEAs; alleles positively associated with the trait of interest).

Mahony C. R., MacLachlan I. R., Lind B. M., Yoder J. B., Wang T., and Aitken S. N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications* **13**.1: 116– 131.

MacLachlan I. R., McDonald T. K., Lind B. M., Rieseberg L. H., Yeaman S., and Aitken S. N. (2021). Genome-wide shifts in climate-related variation underpin responses to selective breeding in a widespread conifer. *Proceedings of the National Academy of Sciences* **118**.10.

Archambeau, J., Garzón, M. B., Barraquand, F., Miguel, M. D., Plomion, C., & González-Martínez, S. C. (2022). Combining climatic and genomic data improves range-wide tree height growth prediction in a forest tree. *The American Naturalist*.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

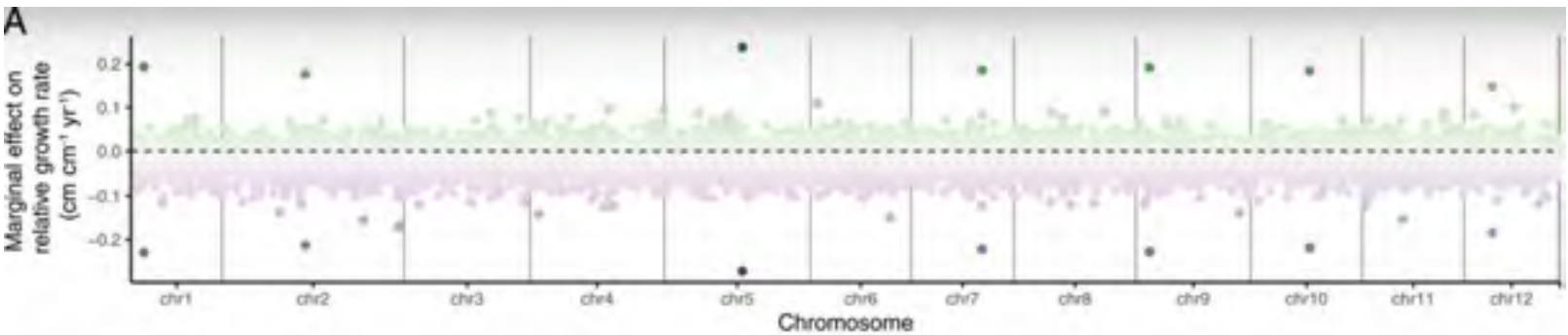
Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

1. A **genome-wide association study** (GWAS) framework was used to estimate allelic effects on **relative growth rates** after controlling for potential confounding factors.



Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

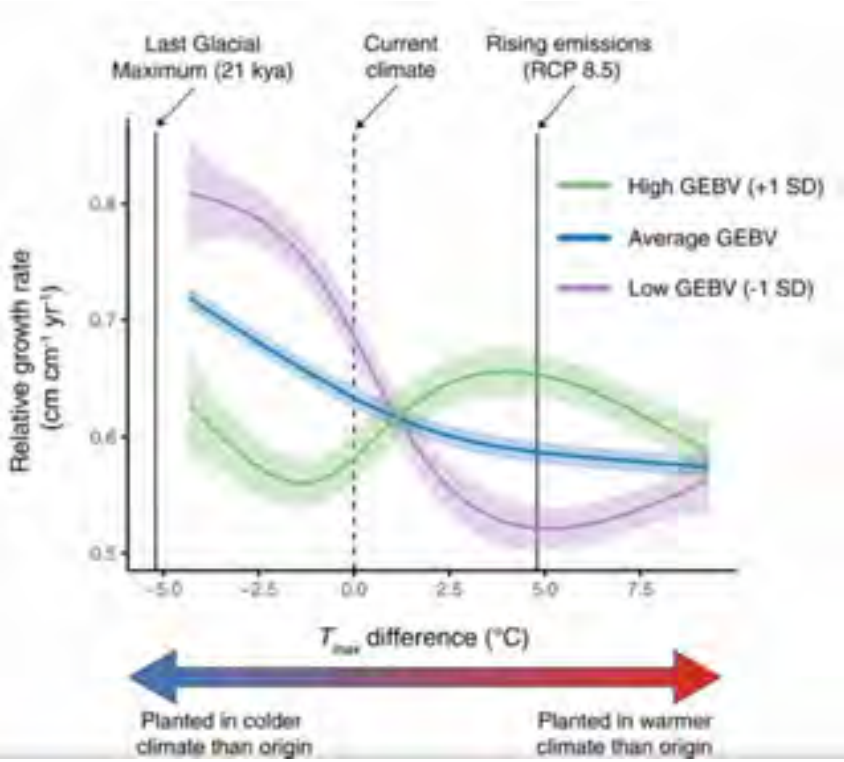
2. Genomic estimated breeding values (GEBVs) indicate the **genetic value** of each maternal tree (i.e. maternal genotype of each seedling) and were estimated as the **average progeny performance** (i.e. the predicted adjusted progeny relative growth rates) when grown in **warmer temperatures**.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

2. Genomic estimated breeding values (GEBVs) indicate the **genetic value** of each maternal tree (i.e. maternal genotype of each seedling) and were estimated as the **average progeny performance** (i.e. the predicted adjusted progeny relative growth rates) when grown in **warmer temperatures**.



Predictions:

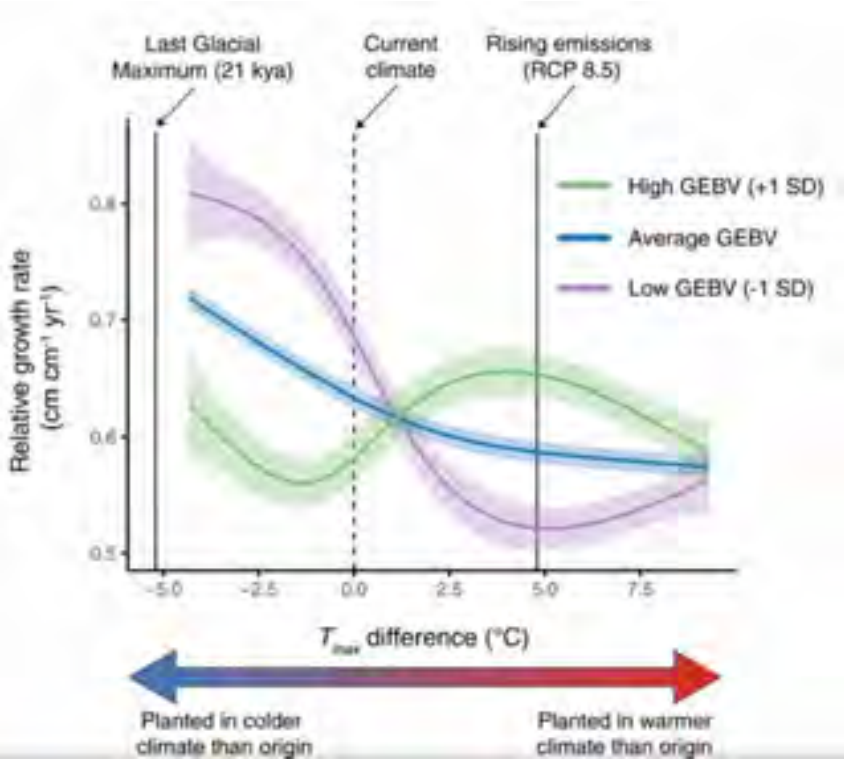
→ Progeny of maternal trees with high GEBVs were **preadaptated to warmer temperatures**.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

2. Genomic estimated breeding values (GEBVs) indicate the **genetic value** of each maternal tree (i.e. maternal genotype of each seedling) and were estimated as the **average progeny performance** (i.e. the predicted adjusted progeny relative growth rates) when grown in **warmer temperatures**.



Predictions:

→ Progeny of maternal trees with high GEBVs were **preadaptated to warmer temperatures**.

→ At a 4.8 °C increase in temperature (predicted under RCP-8.5 by 2070 to 2099), progeny of maternal trees with high GEBVs have 11% and 25% higher relative growth rates than the progeny of maternal trees with average GEBVs or low GEBVs, respectively.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

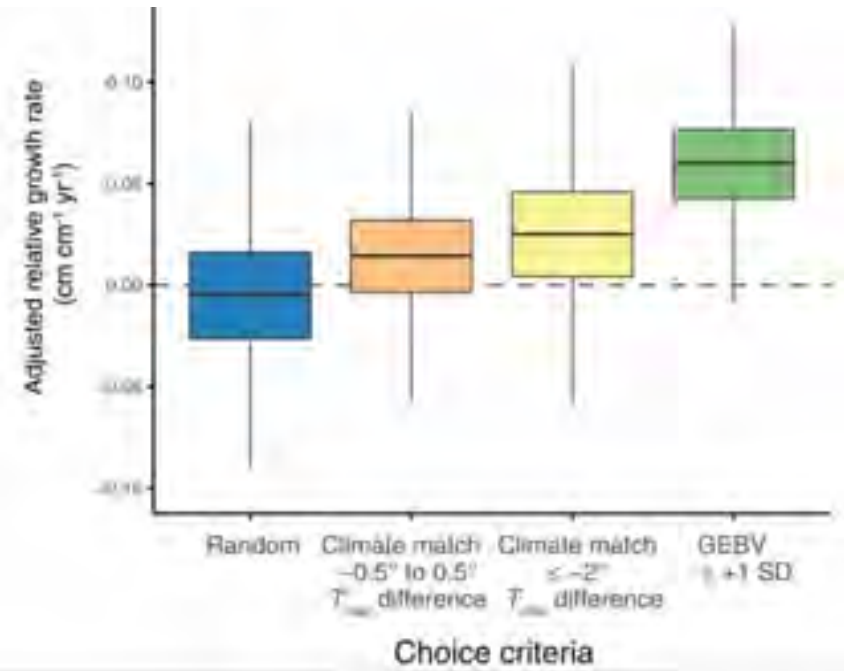
3. Progeny growth rates under **warmer temperatures** were predicted for four hypothetical sets of seedlings used as **seed source**.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

3. Progeny growth rates under **warmer temperatures** were predicted for four hypothetical sets of seedlings used as **seed source**.



Predictions:

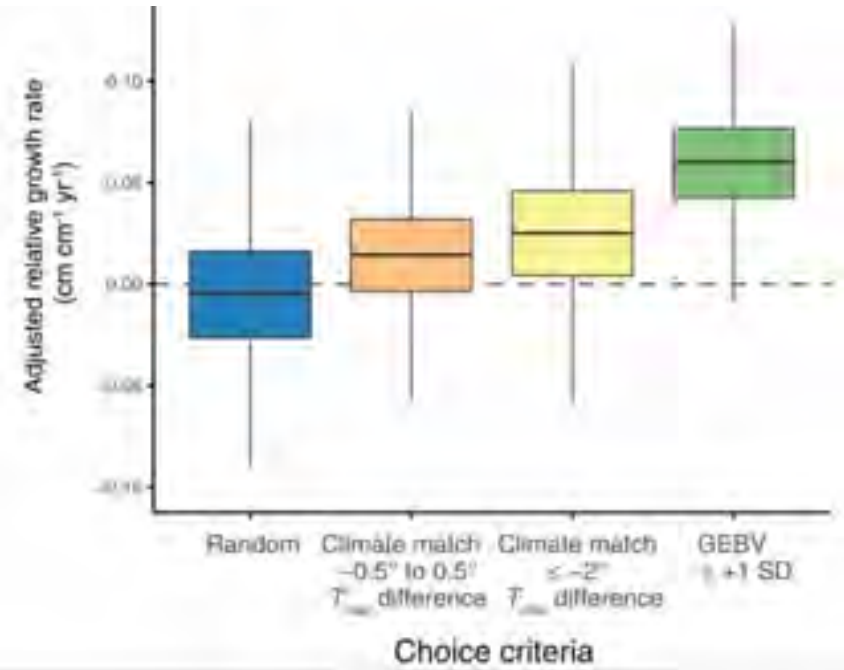
→ Choosing individuals with the highest GEBVs as seed sources leads to substantial increases in predicted progeny growth rates under warmer temperatures.

Predicting phenotypes based on genotypes in natural populations

Browne *et al.* 2019

Goal: evaluate the potential for a *genome-informed* assisted gene flow strategy to mitigate the negative impact of rising temperatures on growth rates in valley oak.

3. Progeny growth rates under **warmer temperatures** were predicted for four hypothetical sets of seedlings used as **seed source**.



Predictions:

→ Choosing individuals with the highest GEBVs as seed sources leads to substantial increases in predicted progeny growth rates under warmer temperatures.

Conclusion:

→ This study shows the potential of genome-informed assisted gene flow to mitigate the negative effects of future increases in temperature by selecting for genotypes predicted to have relatively high productivity in future climate scenarios.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al*. 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

For four **seedling traits** measured in common gardens, they compared the **proportion of phenotypic variance explained** by:

- the **geographic location** of the populations
- the **climate-of-origin** of the populations
- several sets of **genomic markers**:
  - the full SNP array
  - candidate SNPs identified with genotype–phenotype associations (GPAs)
  - candidate SNPs identified with genotype–environment associations (GEAs)
  - candidate SNPs identified with GPAs and GEAs.
  - control set of SNPs.

Predicting phenotypes based on genotypes in natural populations

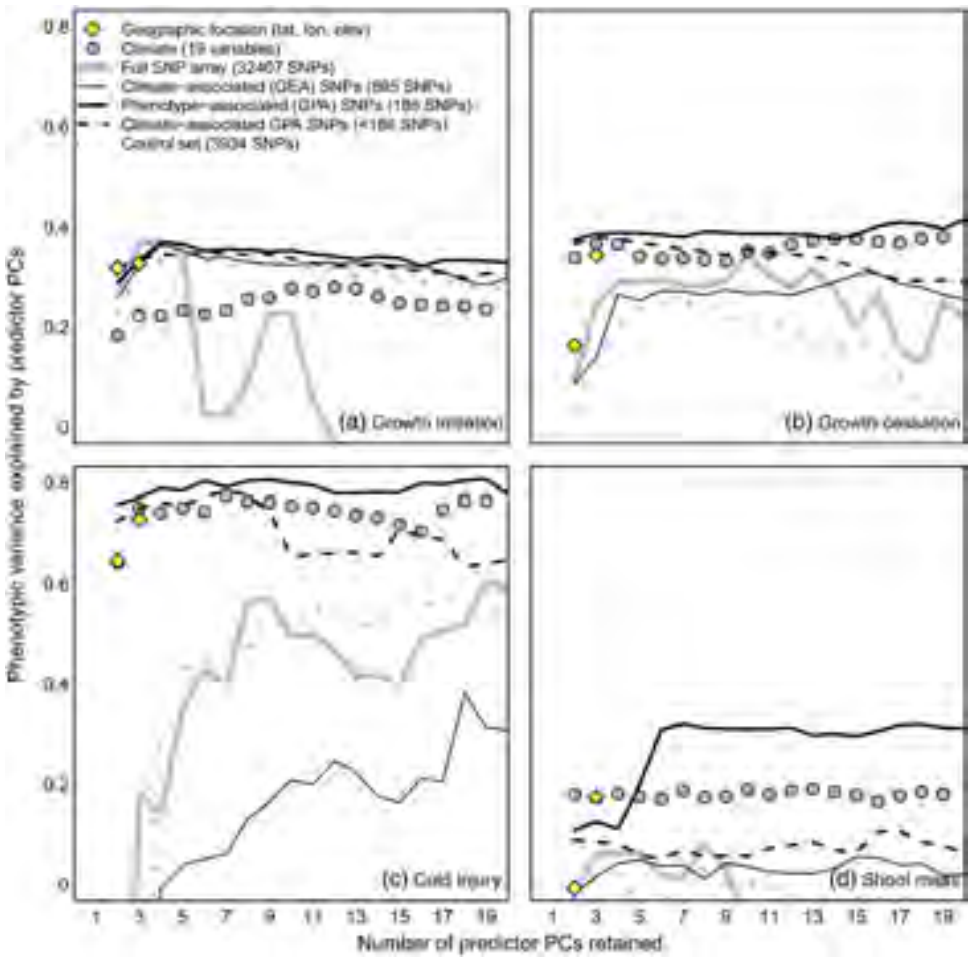
Mahony *et al*. 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.



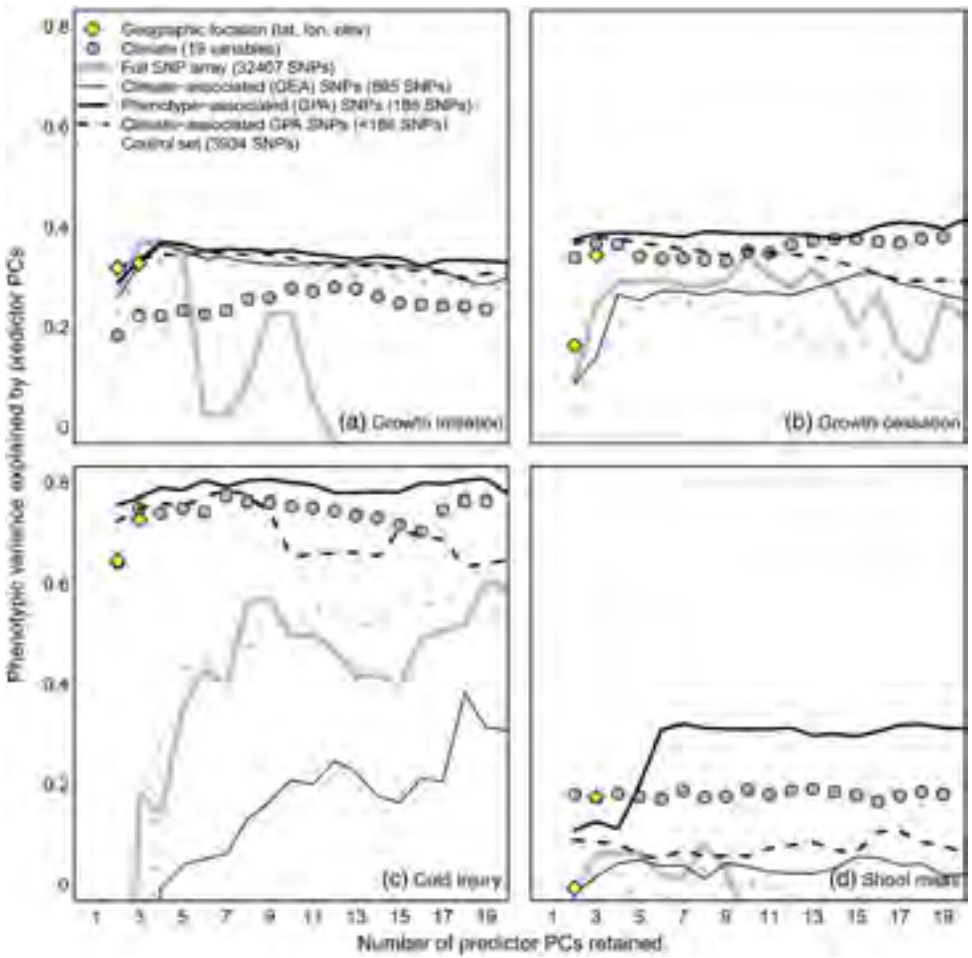
Results:

The **predictive power** of climate variables, geography, and genotypes varied greatly among seedling traits.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.



Results:

The **predictive power** of climate variables, geography, and genotypes varied greatly among seedling traits.

Can be explained by:

→ link between phenotype and fitness

→ degree of polygenicity

→ epistatic and pleiotropic interactions

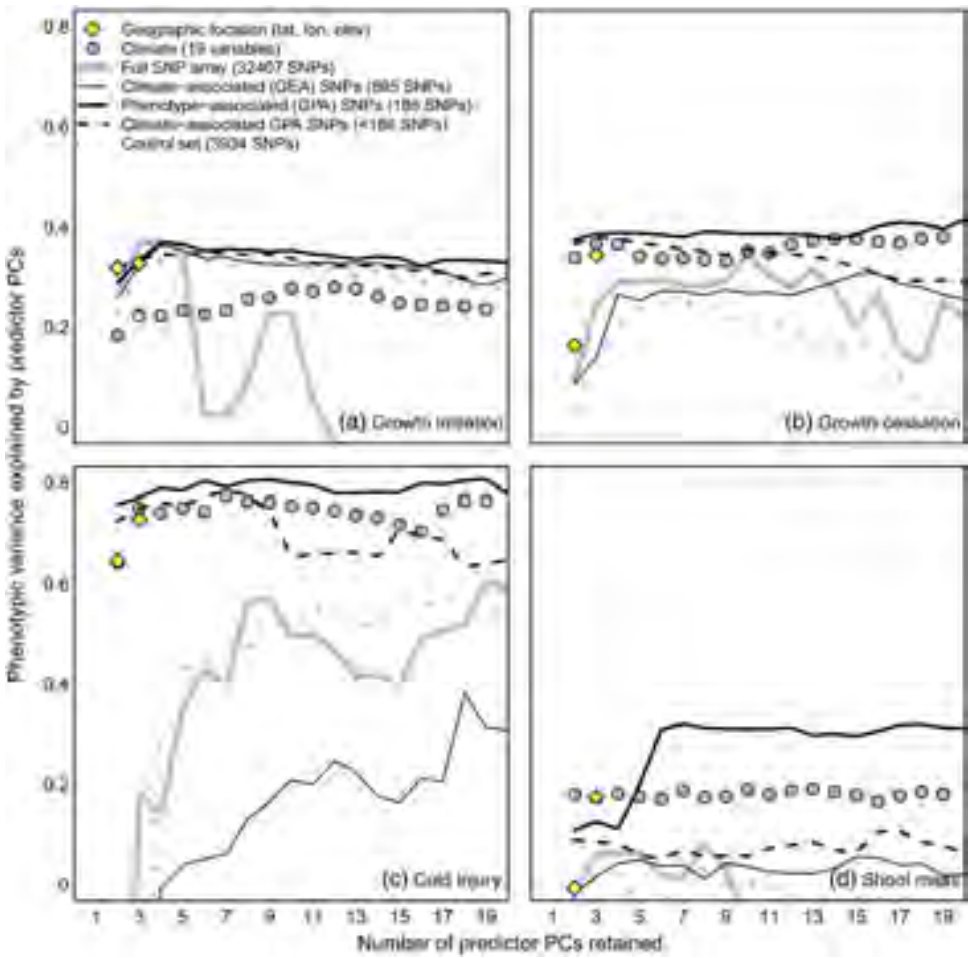
→ relative importance of natural selection vs demographic history

→ method shortcomings

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

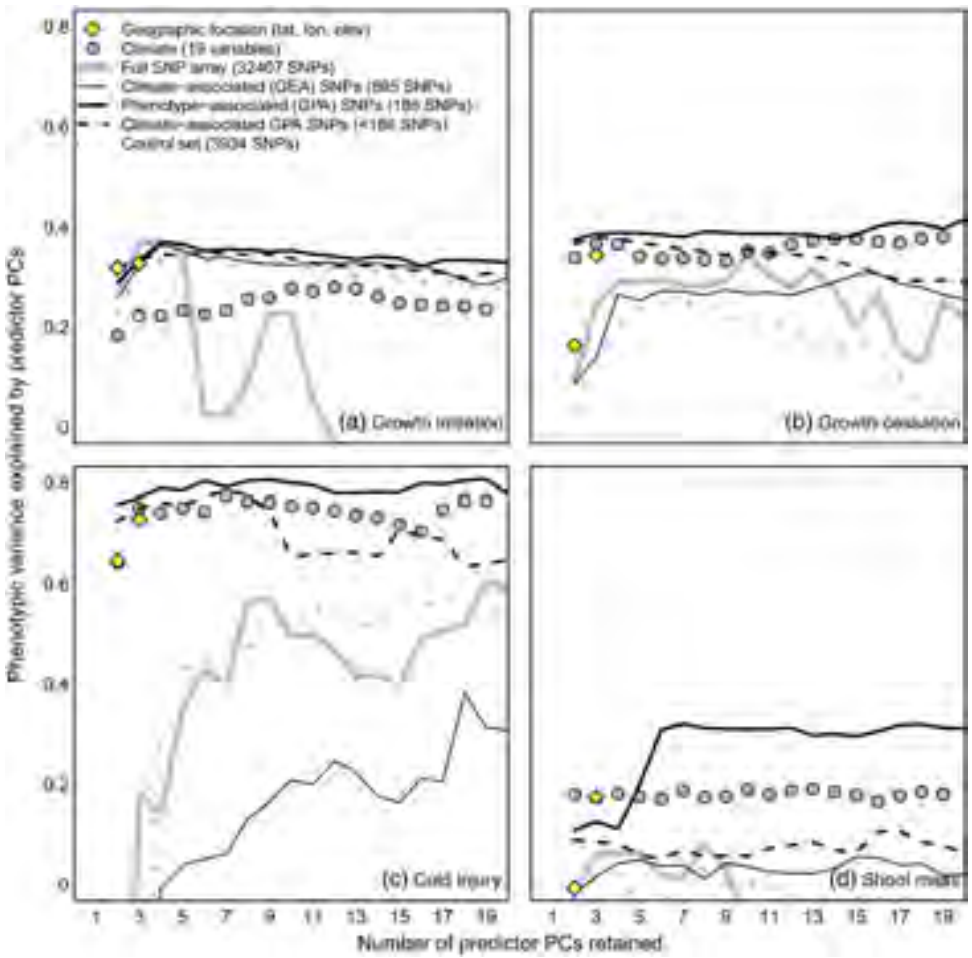


Results:

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.



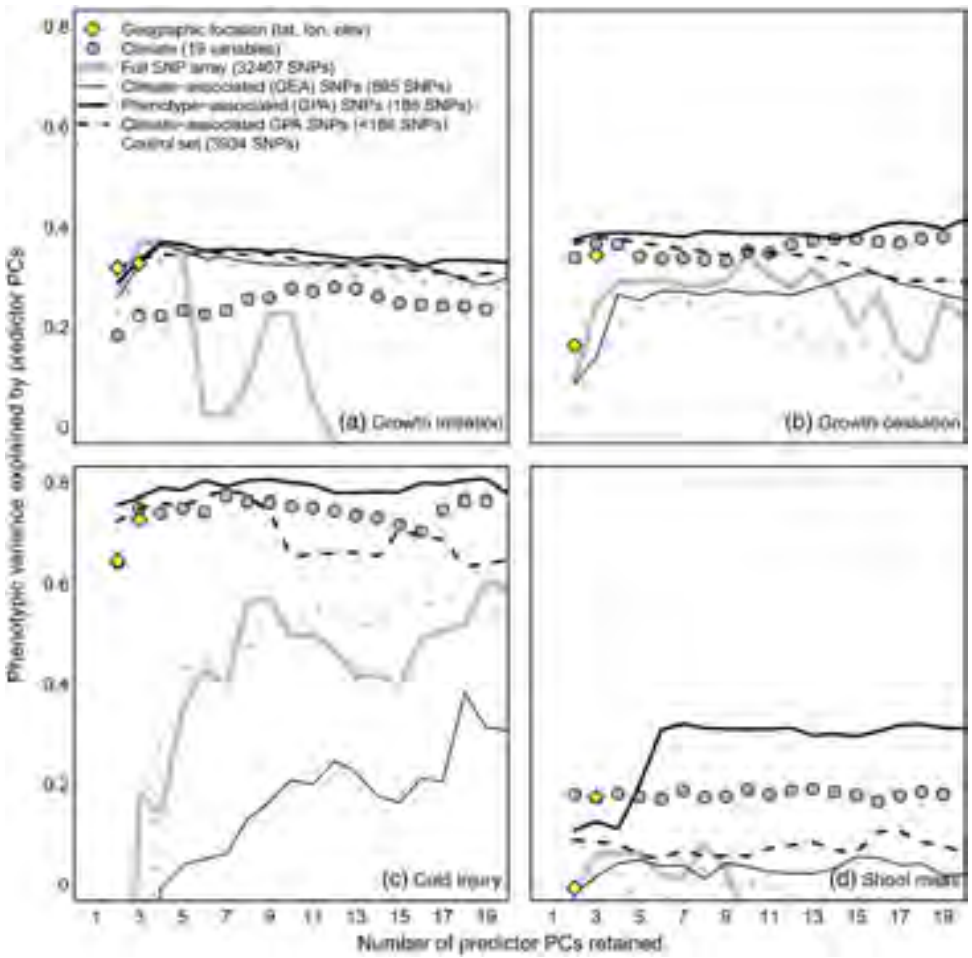
Results:

GPA SNPs are consistently the **best set of markers** for explaining variation in the studied seedling traits, while **GEA SNPs** are not good predictors of the phenotypes.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.



Results:

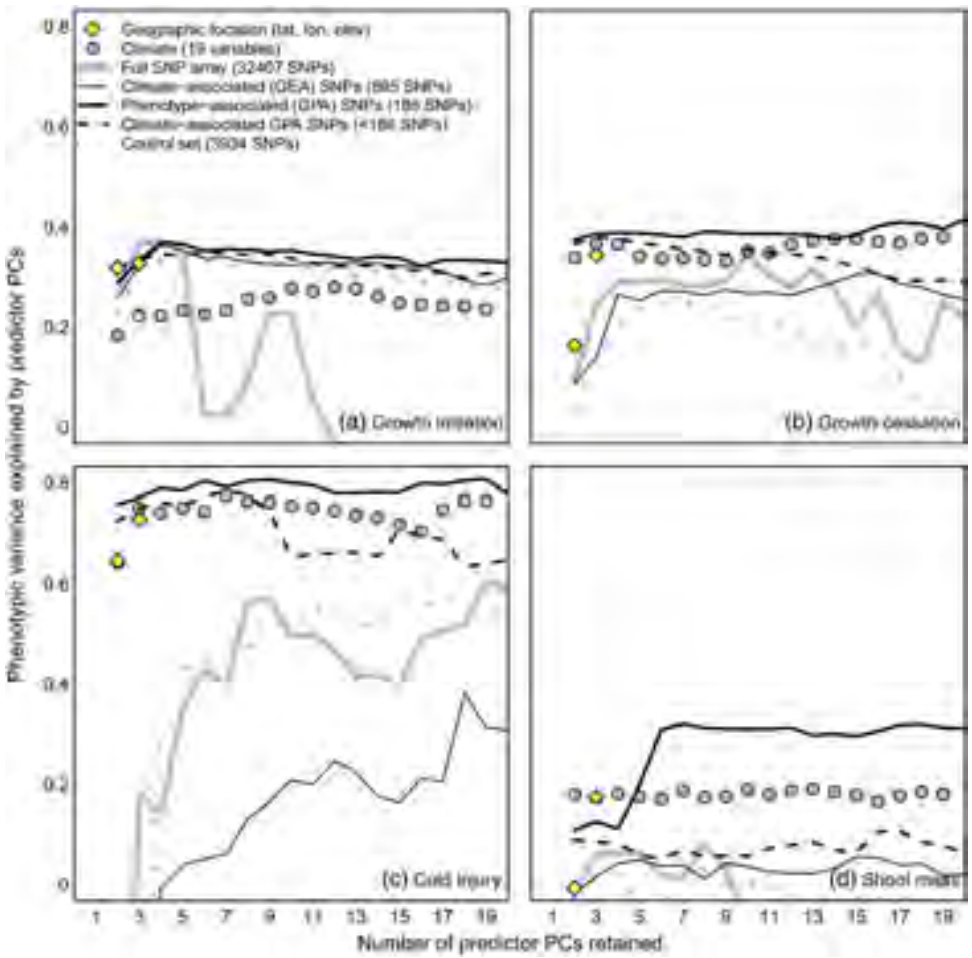
GPA SNPs are consistently the **best set of markers** for explaining variation in the studied seedling traits, while **GEA SNPs** are not good predictors of the phenotypes.

Climate and geography were also good predictors of the phenotypes.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.



Results:

GPA SNPs are consistently the **best set of markers** for explaining variation in the studied seedling traits, while **GEA SNPs** are not good predictors of the phenotypes.

Climate and geography were also good predictors of the phenotypes.

GPA SNPs may play an effective role in designing assisted gene flow strategies.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

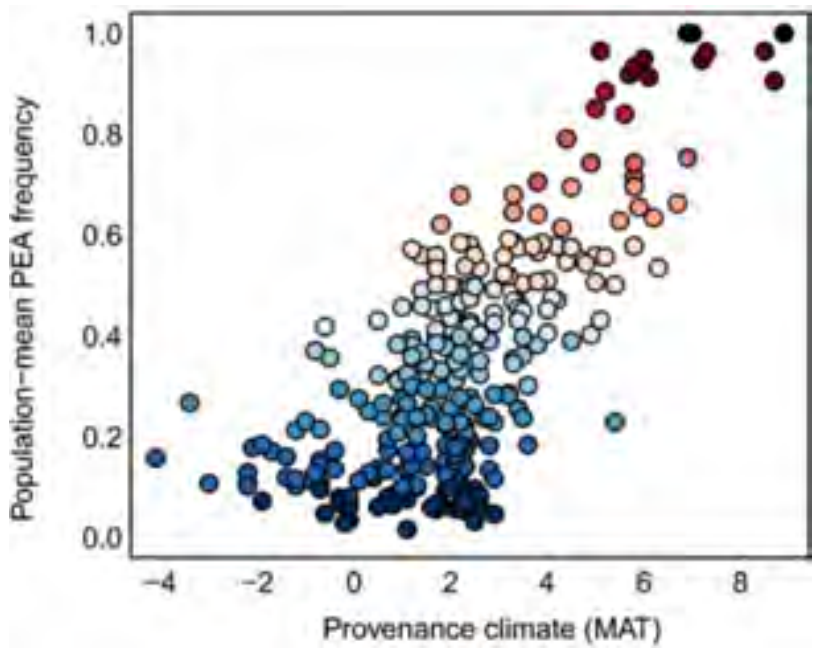
- Counts of **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the trait of interest.

Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

- Counts of **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the trait of interest.
- To characterize the spatial scale of the genetic clines associated with the traits of interest (here, autumn cold injury).

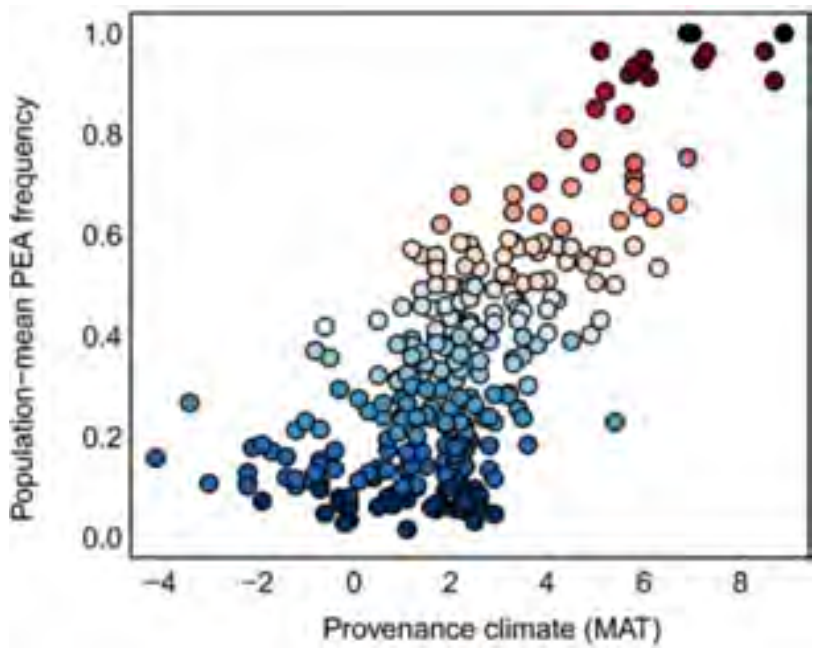


Predicting phenotypes based on genotypes in natural populations

Mahony *et al.* 2020

Goal: evaluate genomic data, relative to phenotypic and climatic data, as a basis for assisted gene flow and genetic conservation in lodgepole pine.

- Counts of **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the trait of interest.
- To characterize the spatial scale of the genetic clines associated with the traits of interest (here, autumn cold injury).



→ **No evidence for highly localized genetic clines** at scales that would constrain local seed transfer to scales finer than those indicated by phenotypic data or necessitate geographically small genetic conservation units.

Predicting phenotypes based on genotypes in natural populations

MacLachlan *et al.* 2021

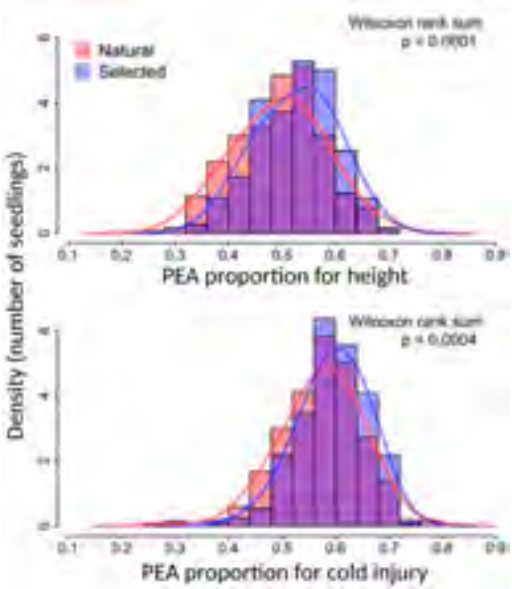
Goal: evaluate the effect of artificial selection for faster growth on climatically adaptive genes and traits in lodgepole pine.

Predicting phenotypes based on genotypes in natural populations

MacLachlan *et al.* 2021

Goal: evaluate the effect of artificial selection for faster growth on climatically adaptive genes and traits in lodgepole pine.

$\rightarrow$ Identify **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the traits of interest (growth, cold injury, growth initiation and growth cessation) in natural populations.



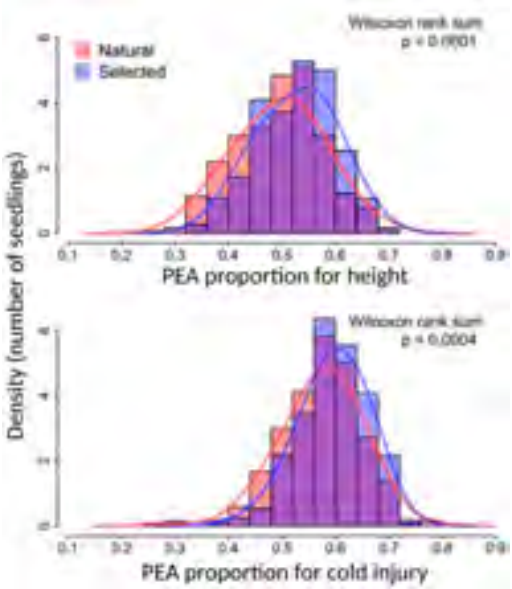
- Substantial shifts in PEA proportions and frequencies were observed across many loci after two generations of selective breeding for height.

Predicting phenotypes based on genotypes in natural populations

MacLachlan *et al.* 2021

Goal: evaluate the effect of artificial selection for faster growth on climatically adaptive genes and traits in lodgepole pine.

$\rightarrow$ Identify **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the traits of interest (growth, cold injury, growth initiation and growth cessation) in natural populations.



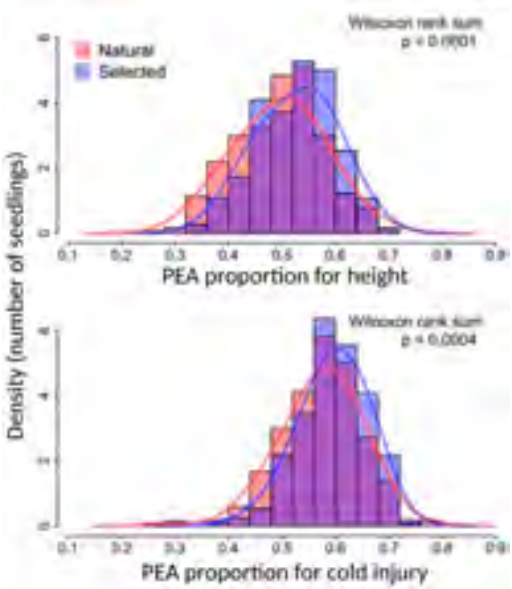
- Substantial shifts in PEA proportions and frequencies were observed across many loci after two generations of selective breeding for height.
- Genetic overlap among traits.

Predicting phenotypes based on genotypes in natural populations

MacLachlan *et al.* 2021

Goal: evaluate the effect of artificial selection for faster growth on climatically adaptive genes and traits in lodgepole pine.

$\rightarrow$ Identify **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the traits of interest (growth, cold injury, growth initiation and growth cessation) in natural populations.



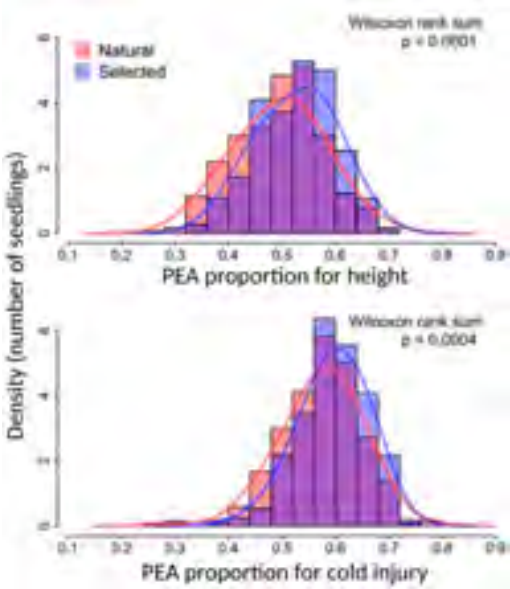
- Substantial shifts in PEA proportions and frequencies were observed across many loci after two generations of selective breeding for height.
- Genetic overlap among traits.
- Genetic correlations: alleles most strongly associated with greater height associated with greater cold injury and delayed phenology.

Predicting phenotypes based on genotypes in natural populations

MacLachlan *et al.* 2021

Goal: evaluate the effect of artificial selection for faster growth on climatically adaptive genes and traits in lodgepole pine.

→ Identify **positive-effect alleles** (PEAs), i.e. alleles that are positively associated with the traits of interest (growth, cold injury, growth initiation and growth cessation) in natural populations.



- Substantial shifts in PEA proportions and frequencies were observed across many loci after two generations of selective breeding for height.
- Genetic overlap among traits.
- Genetic correlations: alleles most strongly associated with greater height associated with greater cold injury and delayed phenology.

→ By affecting trait-associated genetic variation involved in climate adaptation, artificial selection for tree growth may compromise future adaption in a changing climate and assisted gene flow strategies.

Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

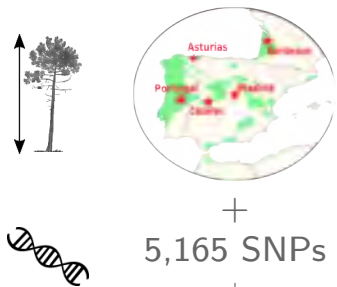
→ Calculating counts of height-associated positive-effect alleles (PEAs)

Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Calculating counts of height-associated positive-effect alleles (PEAs)

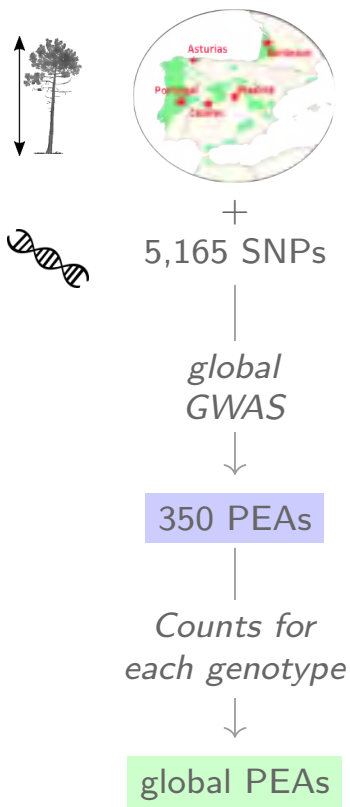


Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Calculating counts of height-associated positive-effect alleles (PEAs)

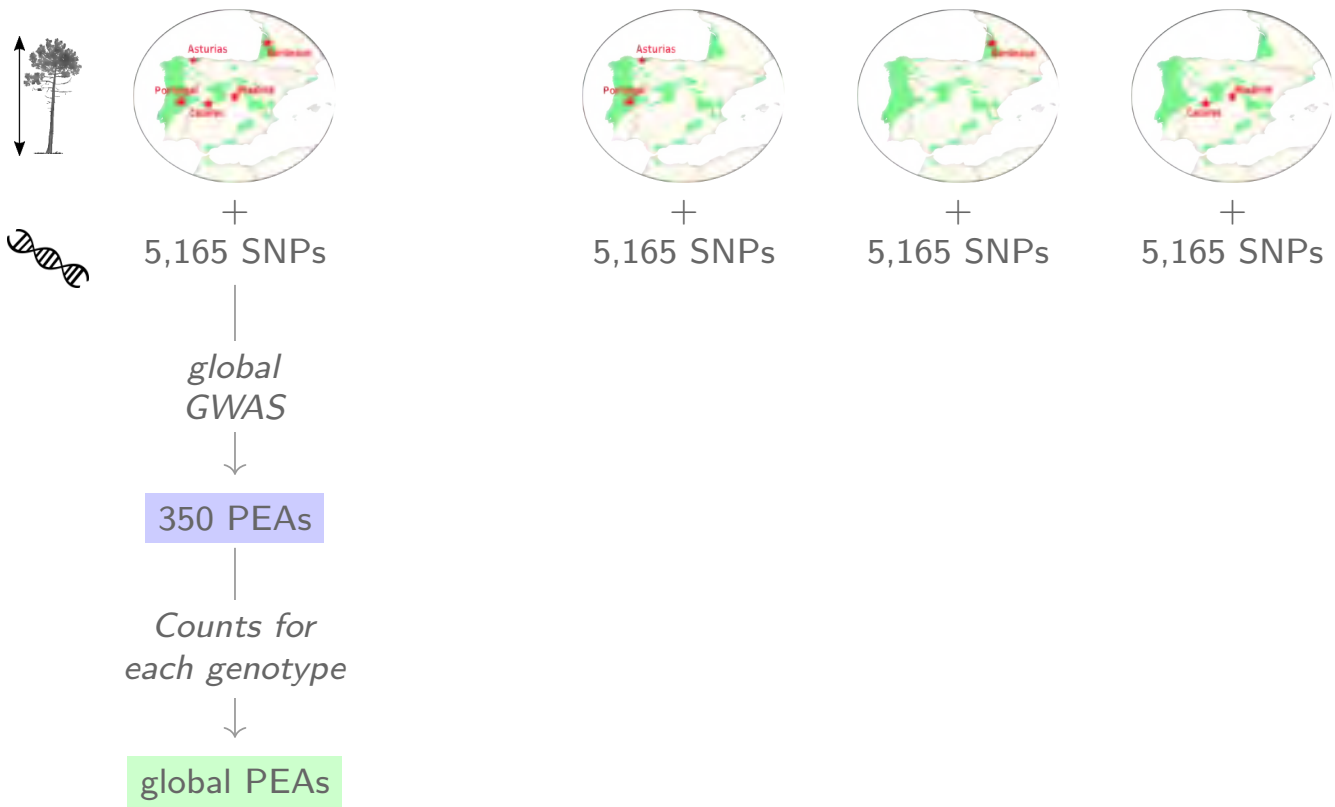


Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Calculating counts of height-associated positive-effect alleles (PEAs)

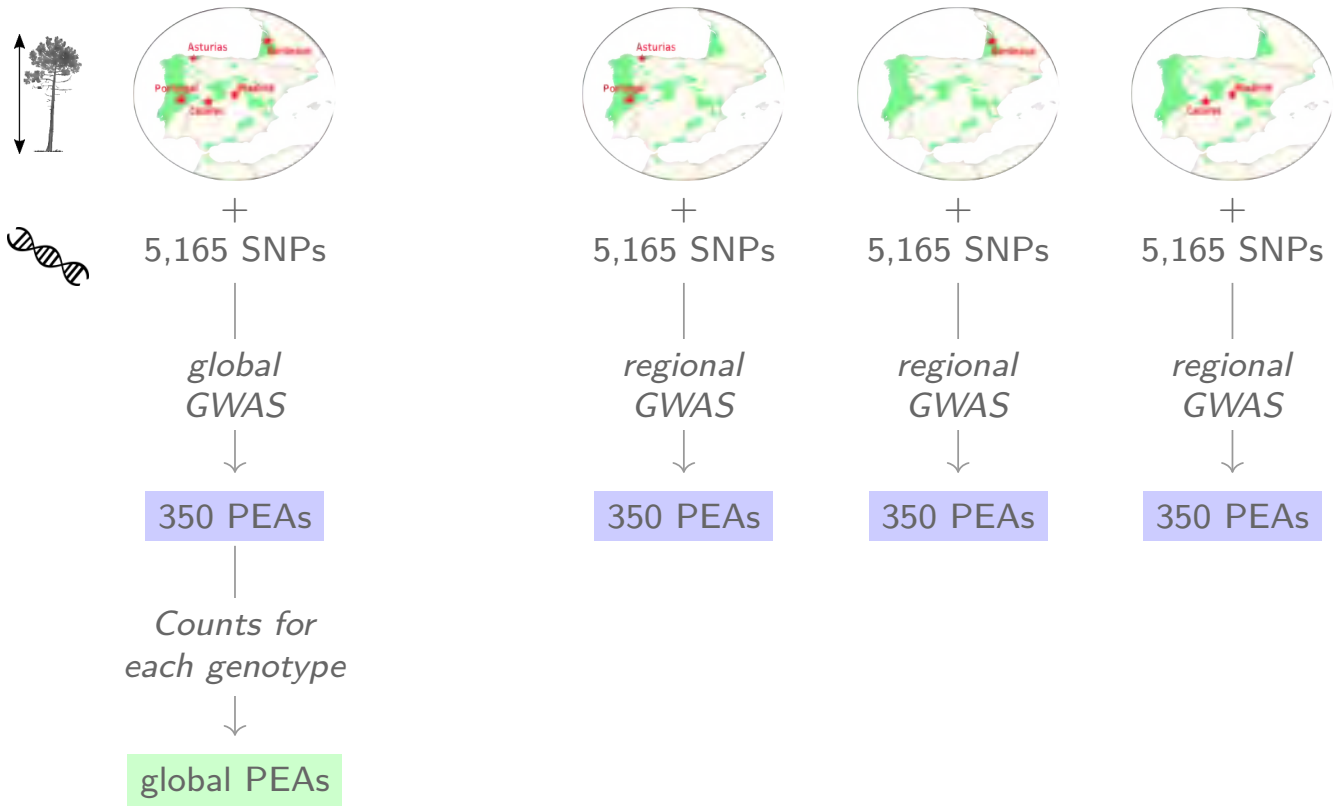


Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Calculating counts of height-associated positive-effect alleles (PEAs)

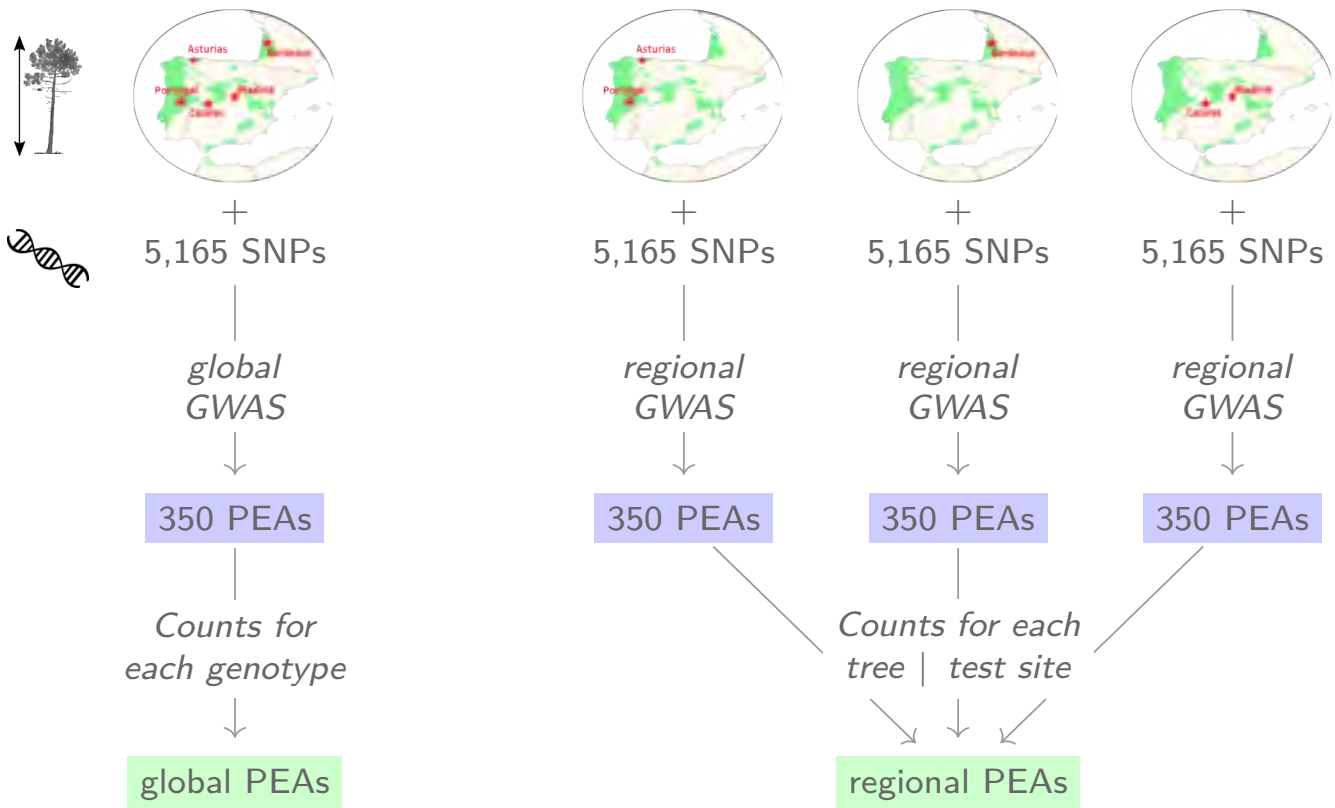


Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Calculating counts of height-associated positive-effect alleles (PEAs)



Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

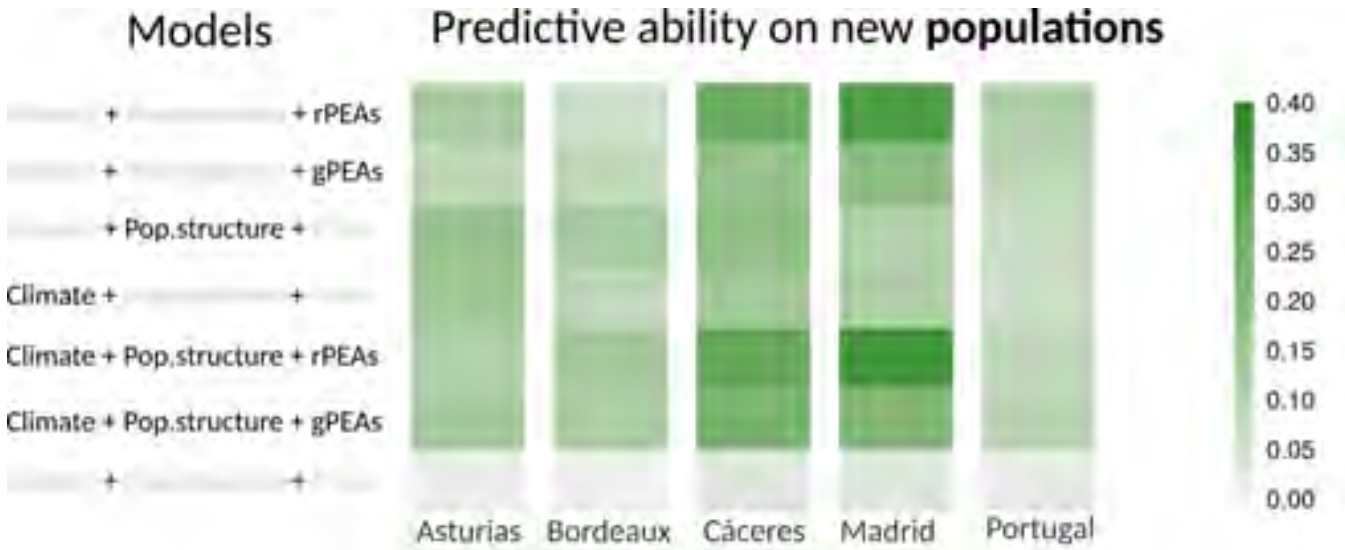
$\rightarrow$ Comparing the predictive ability of different models on new populations.

Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Comparing the predictive ability of different models on new populations.

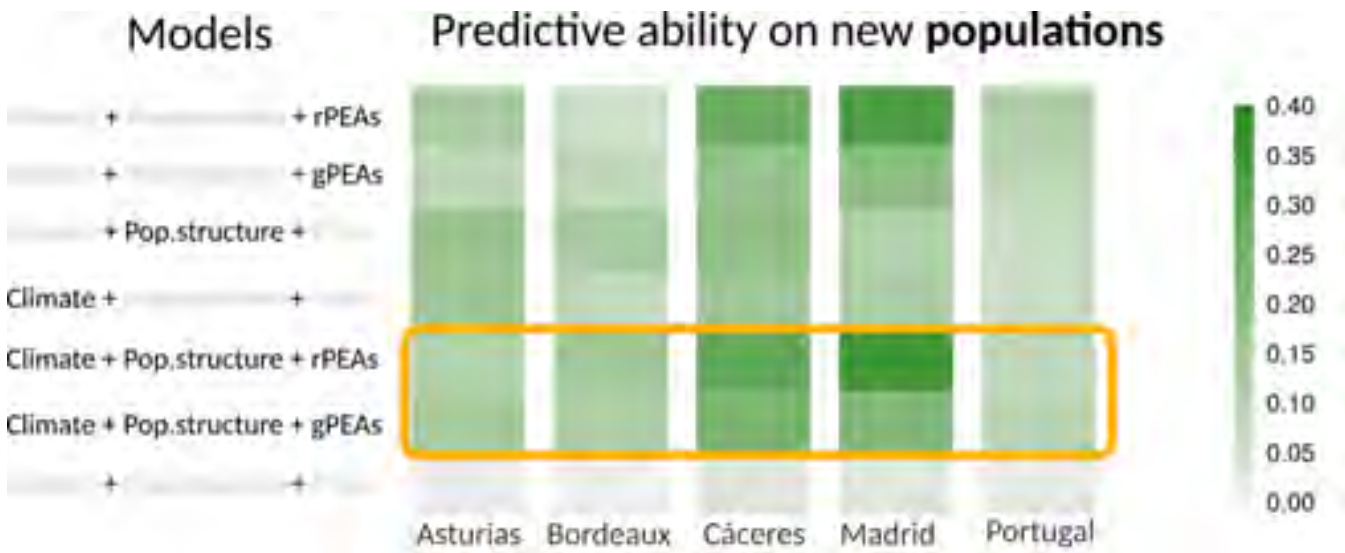


Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Comparing the predictive ability of different models on new populations.



Models **combining climatic and genomic information** had the best out-of-sample predictive ability.

Predicting phenotypes based on genotypes in natural populations

Archambeau *et al.* 2022

Goal: predicting height growth of populations with no phenotypic information in maritime pine.

→ Comparing the predictive ability of different models on new populations.

Models **combining climatic and genomic information** had the best out-of-sample predictive ability.

Better predictive ability of **regionally-selected PEAs (rPEAs)** than globally-selected PEAs (gPEAs).

Predicting phenotypes based on genotypes in natural populations

So, **GEBVs** or **PEA counts**, which one to use in **natural populations**?

Genomic estimated breeding values (GEBVs)

Counts of alleles associated with the phenotype **weighed** by their phenotypic effects

Equivalent to polygenic risk scores (PRS)

✓ Likely more precise within genetic groups (i.e. family, population)

✗ Less precise for highly polygenic traits?

✓ More precise for olygogenic traits?

✗ Low transferability across genetic groups (inconsistent allelic effects)

Positive-effect allele (PEA) counts

Counts of alleles positively associated with the phenotype

✗ Less precise within genetic groups

✗ Less precise for olygogenic traits

✓ More precise for highly polygenic traits?

✓ Higher transferability across genetic groups? (remains to be tested)

✓ More robust to small stochastic SNP sampling effects (whether a large effect alleles is included or not)

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches

- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change

- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Predicting phenotypes from gene-environment associations

Lasky *et al.* 2015

A huge dataset

— 1943 georeferenced sorghum landraces genotyped at 404,627 SNPs
— Phenotype measurements from experimental drought stress and aluminium toxicity.

Predicting phenotypes from gene-environment associations

Lasky *et al.* 2015

A huge dataset

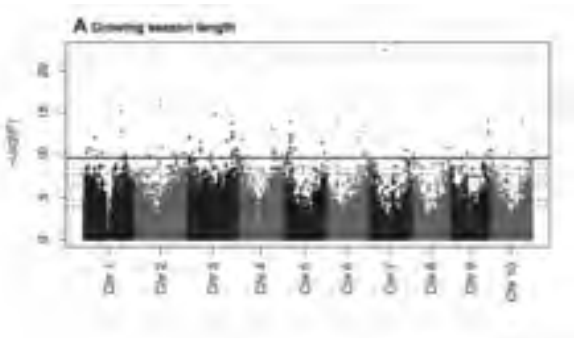
– 1943 georeferenced sorghum landraces genotyped at 404,627 SNPs
– Phenotype measurements from experimental drought stress and aluminium toxicity.

Goals

– Predicting **average phenotypes** across relatively benign and stressful environments.
– Predicting **phenotypic change** from relatively benign to stressful environments (i.e. **G×E interaction**).

Approaches to predict the genomic breeding values

– **Marker-assisted selection** (MAS) approach:



- Identifying SNPs with the **strongest environmental associations**.
- Filtering the SNPs with low LD.
- Predicting phenotypes using the **summed SNP-environment associations**.

Predicting phenotypes from gene-environment associations

Lasky *et al.* 2015

A huge dataset

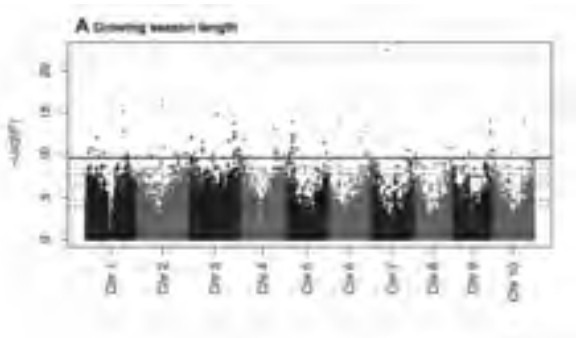
– 1943 georeferenced sorghum landraces genotyped at 404,627 SNPs
– Phenotype measurements from experimental drought stress and aluminium toxicity.

Goals

– Predicting **average phenotypes** across relatively benign and stressful environments.
– Predicting **phenotypic change** from relatively benign to stressful environments (i.e. **G×E interaction**).

Approaches to predict the genomic breeding values

– **Marker-assisted selection** (MAS) approach:



- Identifying SNPs with the **strongest environmental associations**.
- Filtering the SNPs with low LD.
- Predicting phenotypes using the **summed SNP-environment associations**.

– **Genomic selection** (GS) approach:

Predicting phenotypes based on **environmental associations** with estimated **genome-wide kinship**.

Predicting phenotypes from gene-environment associations

Lasky *et al.* 2015

A huge dataset

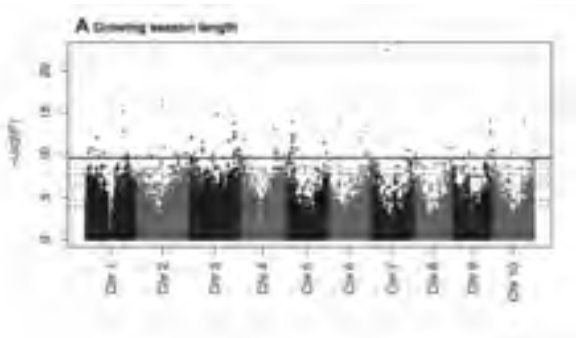
– 1943 georeferenced sorghum landraces genotyped at 404,627 SNPs
– Phenotype measurements from experimental drought stress and aluminium toxicity.

Goals

– Predicting **average phenotypes** across relatively benign and stressful environments.
– Predicting **phenotypic change** from relatively benign to stressful environments (i.e. **G×E interaction**).

Approaches to predict the genomic breeding values

– **Marker-assisted selection** (MAS) approach:



- Identifying SNPs with the **strongest environmental associations**.
- Filtering the SNPs with low LD.
- Predicting phenotypes using the **summed SNP-environment associations**.

– **Genomic selection** (GS) approach:

Predicting phenotypes based on **environmental associations** with estimated **genome-wide kinship**.

– **Hybrid** approach combining MAS and GS approaches.

Predicting phenotypes from gene-environment associations

Lasky *et al.* 2015

→ Environment-associated SNPs are good predictors of **average phenotypes**.

e.g. per-plant biomass best predicted by SNPs with the strongest association with aridity in Austin, United States.

Predicting phenotypes from gene-environment associations

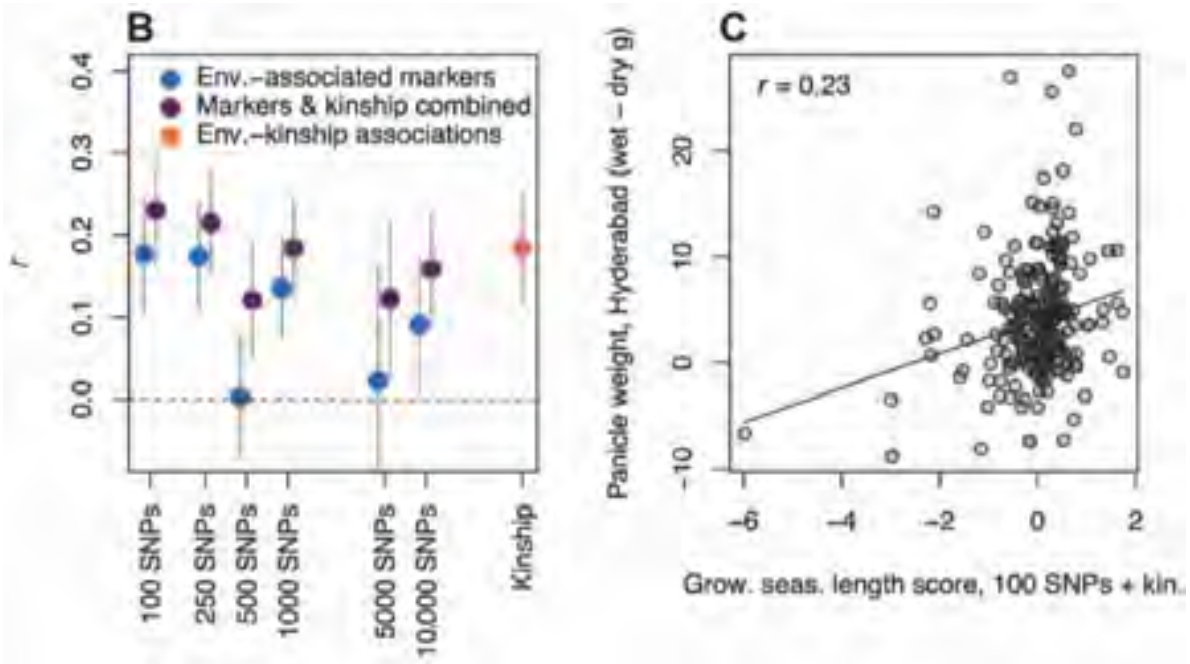
Lasky *et al.* 2015

→ Environment-associated SNPs are good predictors of **average phenotypes**.

e.g. per-plant biomass best predicted by SNPs with the strongest association with aridity in Austin, United States.

→ Environment-associated SNPs are good predictors of **genotype-by-environment interactions**.

e.g. reductions in panicle height under drought conditions best predicted by SNPs with the strongest associations with growing season length.



Predicting phenotypes from gene-environment associations

Predicting the **environmental range** of the genotypes

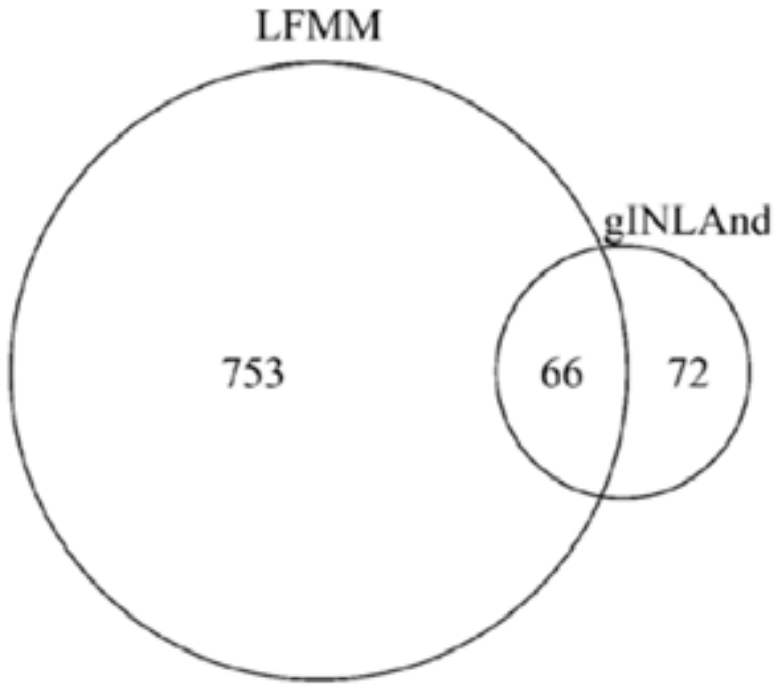
→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

Predicting phenotypes from gene-environment associations

Predicting the **environmental range** of the genotypes

→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

1. 66 SNPs were significantly associated with aridity in two gene-environment association (GEA) methods.

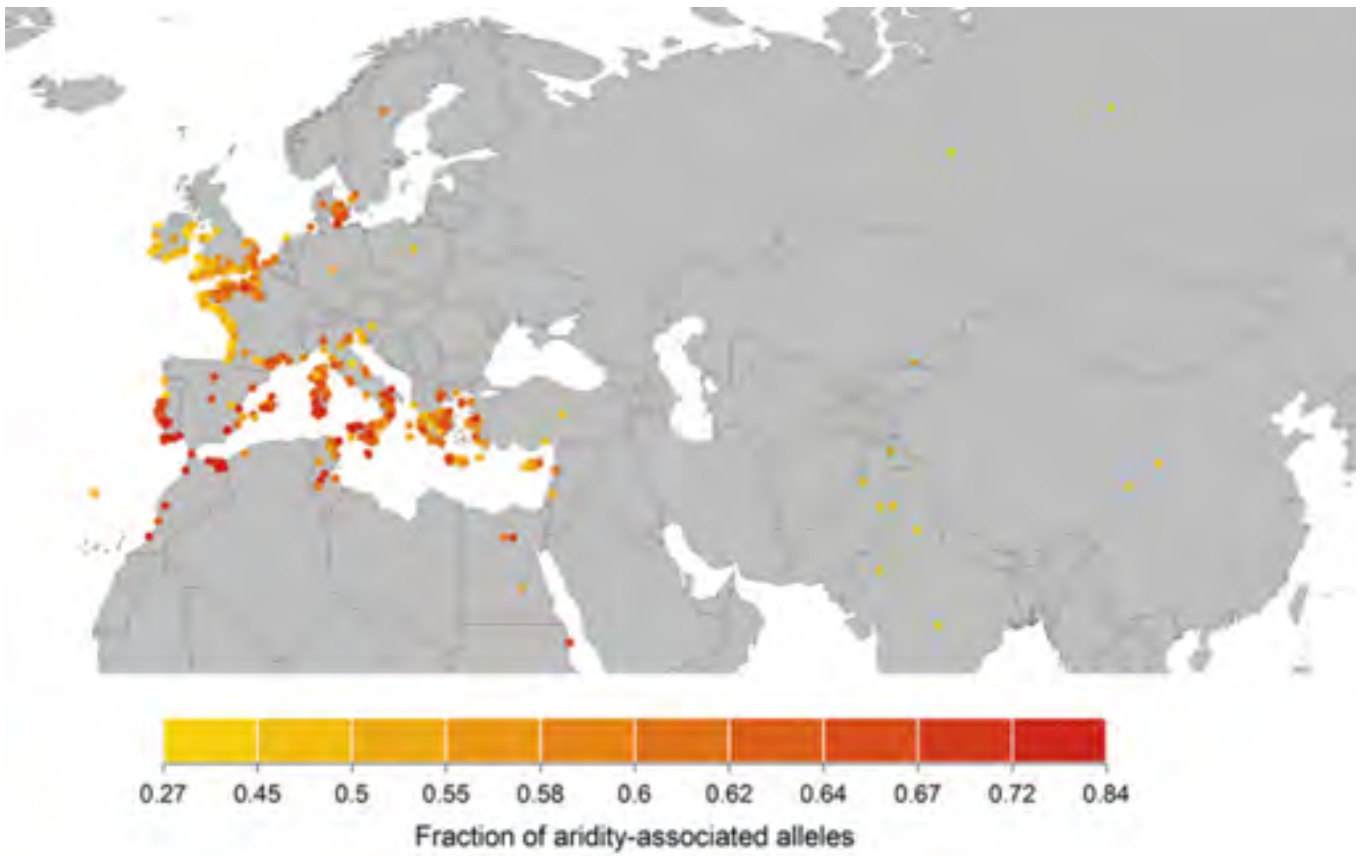


Predicting phenotypes from gene-environment associations

Predicting the **environmental range** of the genotypes

→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

2. The fraction of SNPs positively associated with aridity was calculated for each individual.

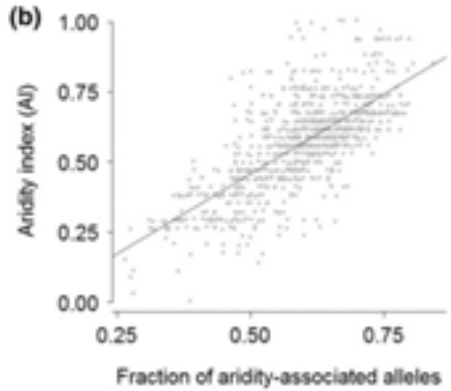


Predicting phenotypes from gene-environment associations

Predicting the **environmental range** of the genotypes

→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

3. The association between the aridity index and the fraction of aridity-associated alleles was estimated in the training dataset,

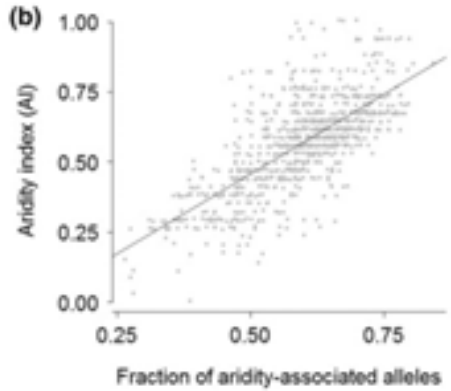


Predicting phenotypes from gene-environment associations

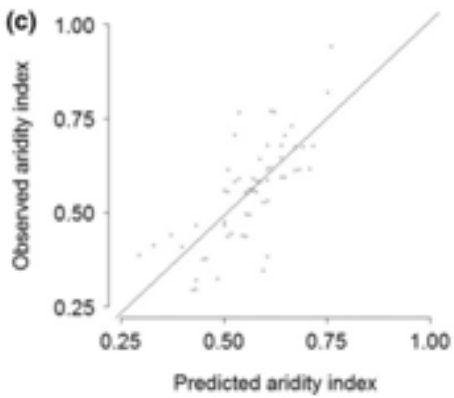
Predicting the **environmental range** of the genotypes

→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

3. The association between the aridity index and the fraction of aridity-associated alleles was estimated in the training dataset,



4. and validated in an independent dataset.



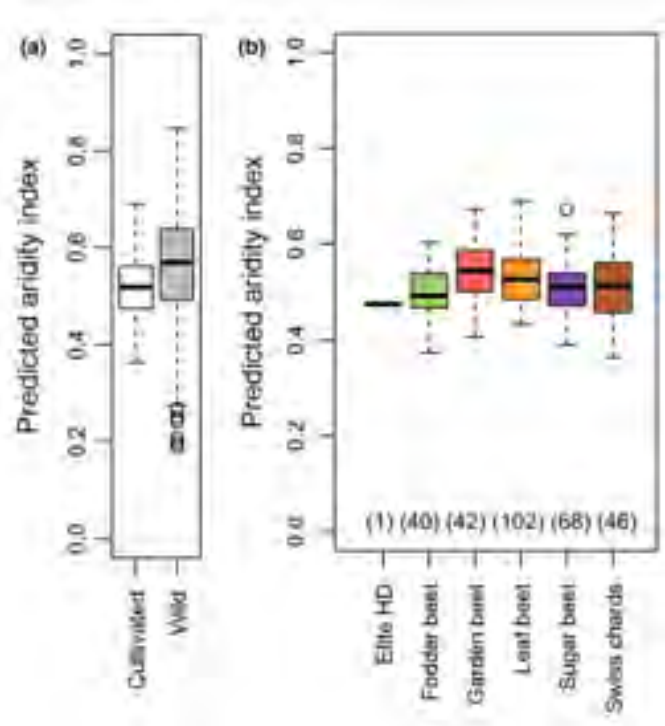
Predicting phenotypes from gene-environment associations

Predicting the **environmental range** of the genotypes

→ A case study on aridity tolerance in sugar beet (Manel *et al.* 2018)

5. This approach can be used to predict the **environmental range** of individuals for which only the genotype is known.

e.g. the aridity range of cultivated beets



Predicting phenotypes from gene-environment associations

Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

1. Estimating a **genomic-relationship matrix** (GRM) based on **average allelic correlation** with more than 200,000 SNPs.

Predicting phenotypes from gene-environment associations

Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

1. Estimating a **genomic-relationship matrix** (GRM) based on **average allelic correlation** with more than 200,000 SNPs.

2. Fitting an **animal model** with the **climate-of-origin** of each accession (minimum winter temperature) used as phenotype.

Predicting phenotypes from gene-environment associations

Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

1. Estimating a **genomic-relationship matrix** (GRM) based on **average allelic correlation** with more than 200,000 SNPs.

2. Fitting an **animal model** with the **climate-of-origin** of each accession (minimum winter temperature) used as phenotype.

3. Predicting **breeding values**, which can be interpreted as *latent cold-adapted phenotypes* based on the assumption that populations are locally adapted along climatic gradients.

Predicting phenotypes from gene-environment associations

Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

1. Estimating a **genomic-relationship matrix** (GRM) based on **average allelic correlation** with more than 200,000 SNPs.

2. Fitting an **animal model** with the **climate-of-origin** of each accession (minimum winter temperature) used as phenotype.

3. Predicting **breeding values**, which can be interpreted as *latent cold-adapted phenotypes* based on the assumption that populations are locally adapted along climatic gradients.

→ This approach **does not require** experiments or phenotypic measurements: predictions are based solely on the **relationship** between **GRM** and **native environments**.

Predicting phenotypes from gene-environment associations

Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

4. Validating the predictions with experimental data (individual relative fitness of 157 ecotypes) from four common gardens.

Predicting phenotypes from gene-environment associations

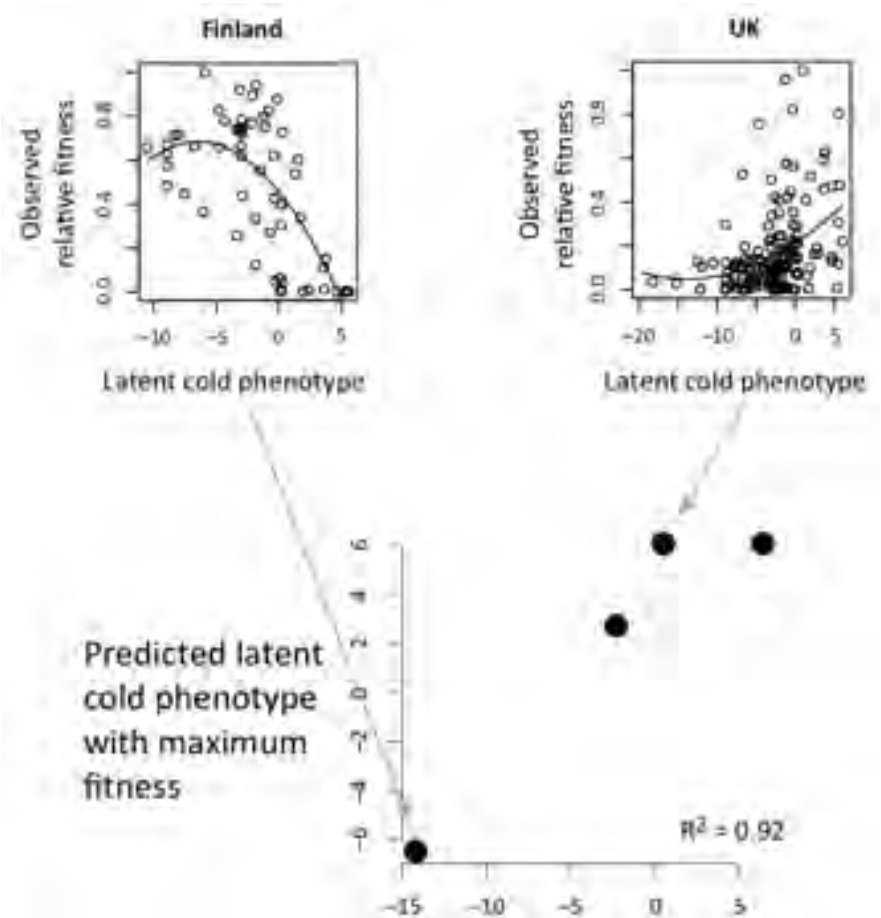
Predicting **locally adapted genotypes** for a given site

→ A case study in *Arabidopsis thaliana* (Gienapp *et al.* 2017)

4. Validating the predictions with experimental data (individual relative fitness of 157 ecotypes) from four common gardens.

- Identifying the **expected most locally adapted phenotype** in each common garden based on the association between the observed relative fitness and the predicted latent cold-adapted phenotypes.
- Predicted latent cold-adapted phenotypes in the four common gardens were **strongly associated** with temperature conditions in the common gardens.

→ This approach can be used to **predict locally adapted genotypes** in a **given environment**, which is particularly useful to guide conservation and restoration strategies.



Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches
- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change
- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches
- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change
- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Exposure: captures **changes in climatic suitability** as predicted by the ENMs and the **extent of environmental dissimilarity** between present and future conditions.

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Exposure: captures **changes in climatic suitability** as predicted by the ENMs and the **extent of environmental dissimilarity** between present and future conditions.

Sensitivity: based on the frequency of alleles associated with warmer and drier climatic conditions (**adaptive sensitivity**) and the levels of neutral genetic diversity (**neutral sensitivity**).

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Exposure: captures **changes in climatic suitability** as predicted by the ENMs and the **extent of environmental dissimilarity** between present and future conditions.

Sensitivity: based on the frequency of alleles associated with warmer and drier climatic conditions (**adaptive sensitivity**) and the levels of neutral genetic diversity (**neutral sensitivity**).

Range shift potential: captures the **degree of connectivity** to **other populations** under future conditions or to **areas** predicted to be **climatically suitable**.

Razgour *et al.* 2017: an innovative integrated framework

Goal: identify populations under **threat** from **climate change** in a bat species with limited dispersal abilities.

Population	Exposure	Sensitivity (adaptive; neutral)	Range shift	Risk level
Lisboa	1	− (−; 0)	+	Low
Bizkaia	1	+ (++; −)	+	Low
Granada	2	− (−; 0)	−	Medium
Girona	3	0 (+; −)	−	Medium
Valladolid	3	− (−; −)	−	Medium-high
Albacete	4	− (−; −)	−	Medium-high
Valencia	4	+ (++; −)	−	High

Implications: Assigned levels of threat can help prioritize and inform conservation action under climate change, such as:

- **rescuing high-risk populations** (through translocation of the entire population or of individuals with relevant adaptive variation into the population)
- **increasing landscape connectivity** to facilitate range shifts and the spread of adaptive genetic variation to reduce threats to medium- and medium–high-risk populations.

Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.

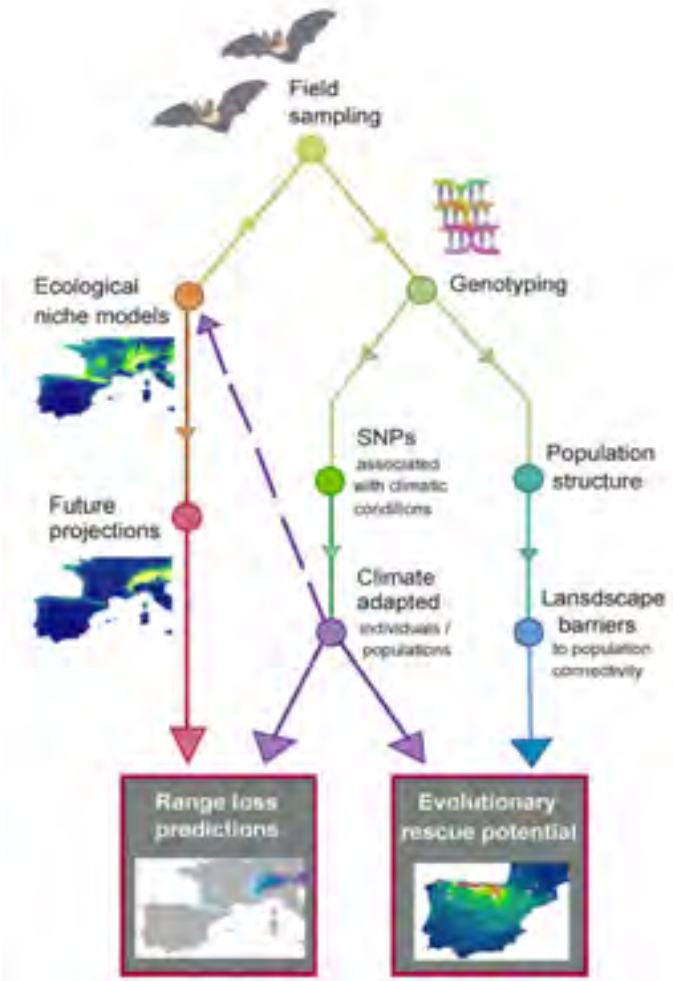
Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.

1. Identifying local climatic adaptations in wild populations by using genotype–environment association (GEA) analysis.

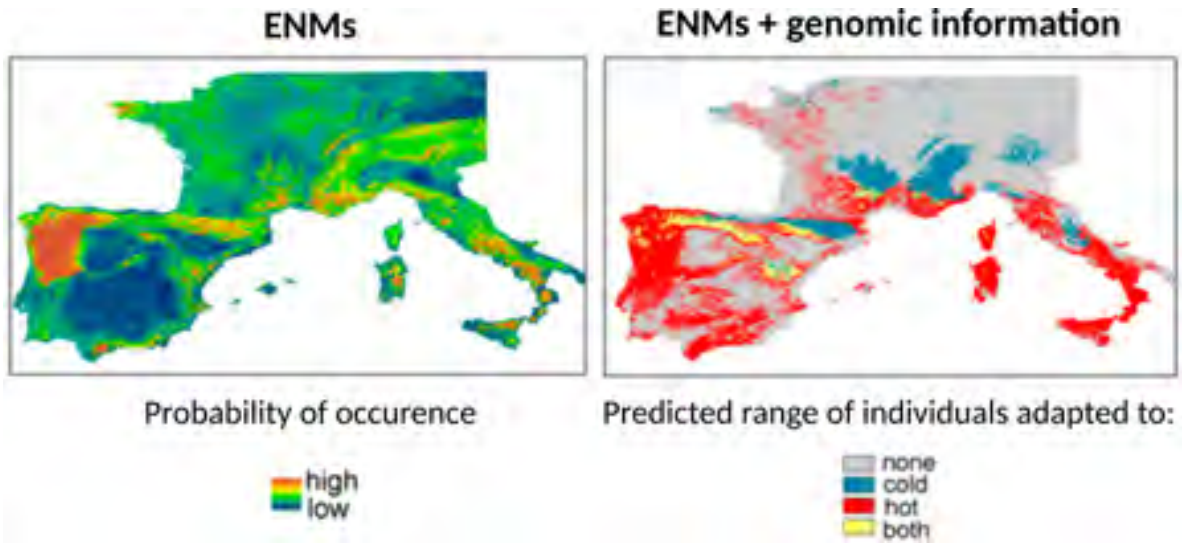
2. Incorporating this information into:

- forecasts of **range changes** under future climate.
- assessment of population persistence through the spread of climate-adaptive genetic variation (**evolutionary rescue potential**).



Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

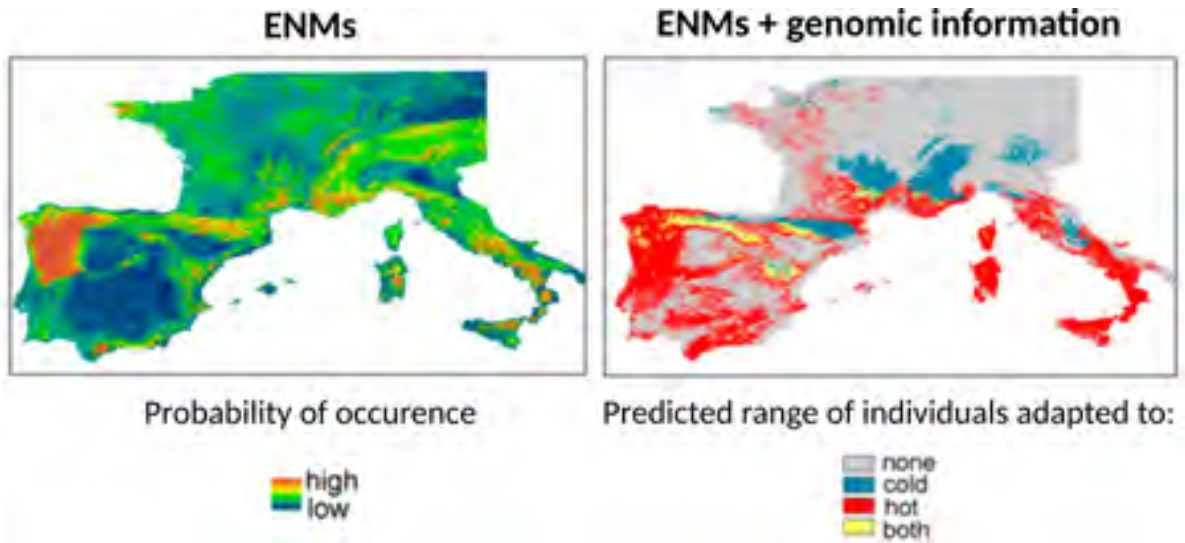
Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.



→ Hot–dry genotypes are predicted to expand their range at the expense of cold–wet genotypes.

Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.



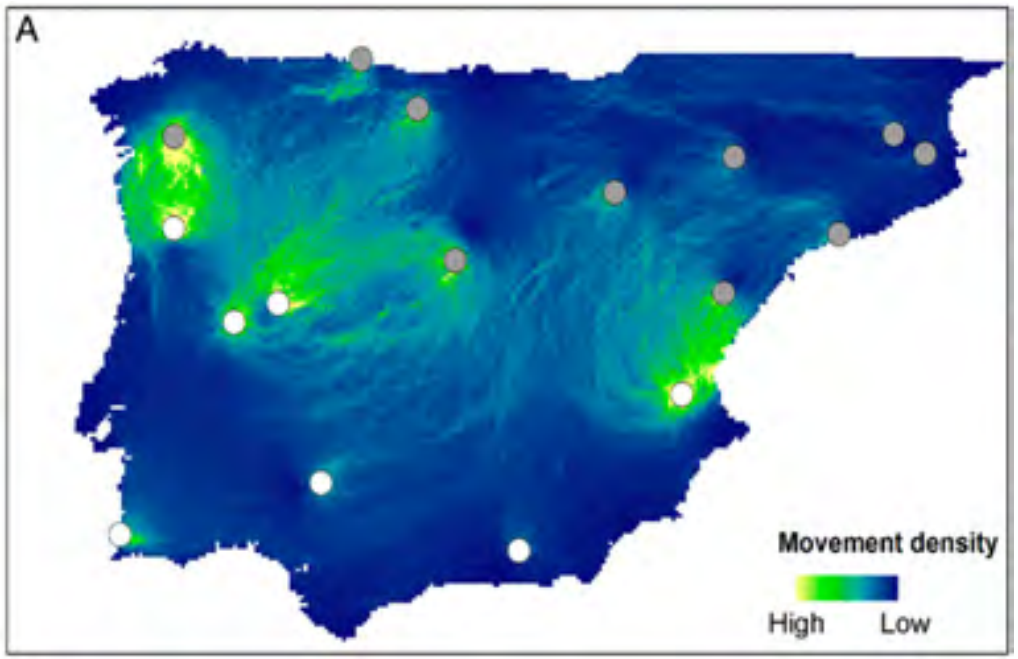
→ Hot–dry genotypes are predicted to expand their range at the expense of cold–wet genotypes.

→ Considering local climatic adaptations in ecological niche models (ENMs):

- reduces future range loss projections.
- **increases** the predicted potential for interspecific competition, through increased range overlap.

Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

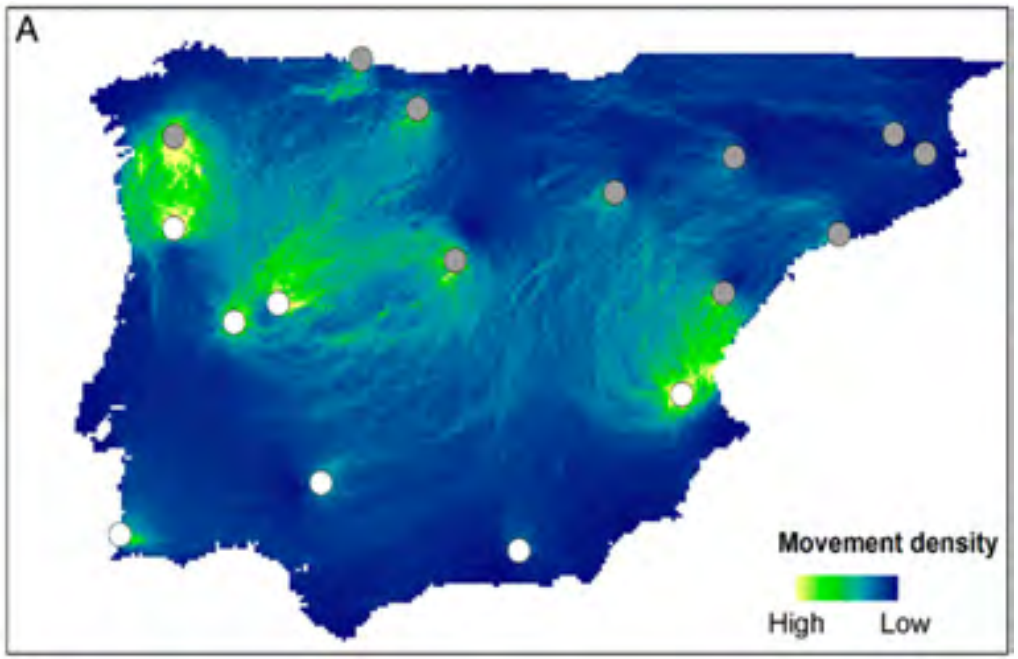
Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.



Evolutionary rescue potential under future climate change = predicted density of movement from populations or individuals adapted to hot–dry conditions (white circles) to those adapted to cold–wet conditions (gray circles) based on the effect of slope and tree/forest cover (landscape resistance) on movement.

Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.

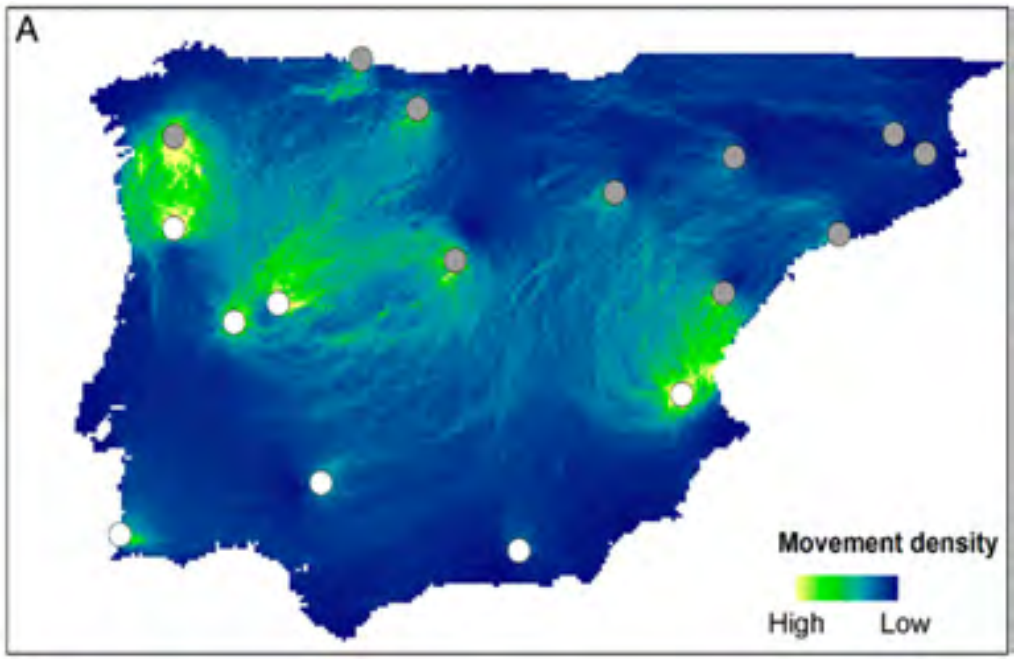


Evolutionary rescue potential under future climate change = predicted density of movement from populations or individuals adapted to hot–dry conditions (white circles) to those adapted to cold–wet conditions (gray circles) based on the effect of slope and tree/forest cover (landscape resistance) on movement.

→ In one species, **landscape barriers to movement** may **limit** the ability of individuals adapted to **hot–dry** conditions to reach areas that will become **climatically unsuitable** for **cold–wet-adapted** individuals, even though these areas will become suitable for hot–dry genotypes.

Razgour *et al.* 2019: incorporating directly climate-adaptive genetic variation into ENMs

Goal: assess the **vulnerability** of populations to **climate change** in two cryptic bat species with limited long-distance dispersal abilities.



Evolutionary rescue potential under future climate change = predicted density of movement from populations or individuals adapted to hot–dry conditions (white circles) to those adapted to cold–wet conditions (gray circles) based on the effect of slope and tree/forest cover (landscape resistance) on movement.

→ In one species, **landscape barriers to movement** may **limit** the ability of individuals adapted to **hot–dry** conditions to reach areas that will become **climatically unsuitable** for **cold–wet-adapted** individuals, even though these areas will become suitable for hot–dry genotypes.

→ In the other species, evolutionary rescue **may not be necessary** as most areas will remain suitable to cold-adapted individuals.

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches
- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change
- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

The genomic offset framework

Landmark papers

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology letters*, *18*(1), 1-16.

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., ... & Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. *Molecular Ecology*, *25*(23), 5907-5924.

Important reviews

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*(1).

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, *14*(5), 1202-1212.

The genomic offset framework

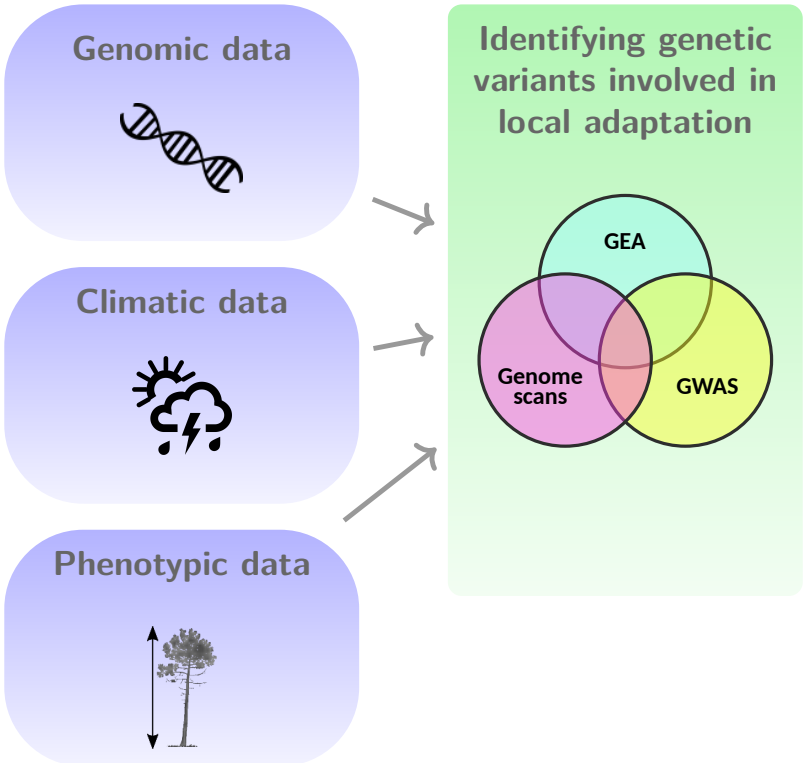
The genomic offset framework

Genomic data

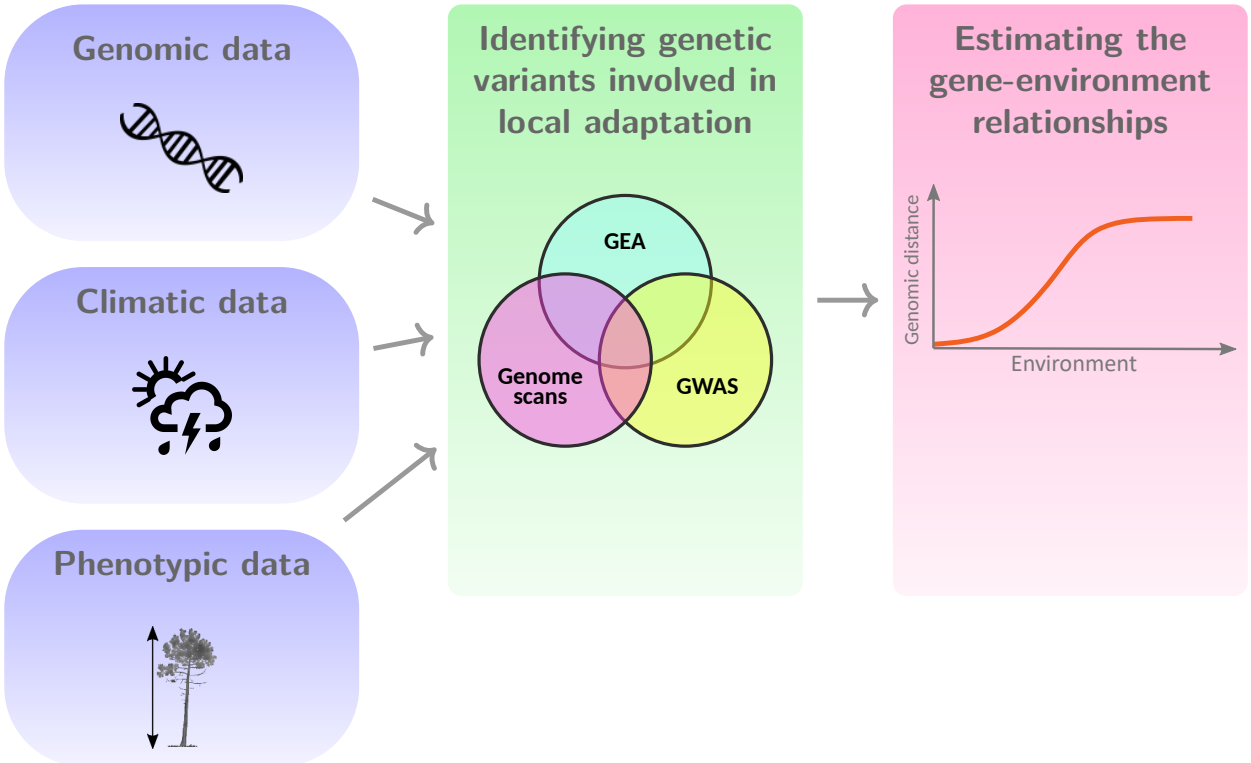
Climatic data

Phenotypic data

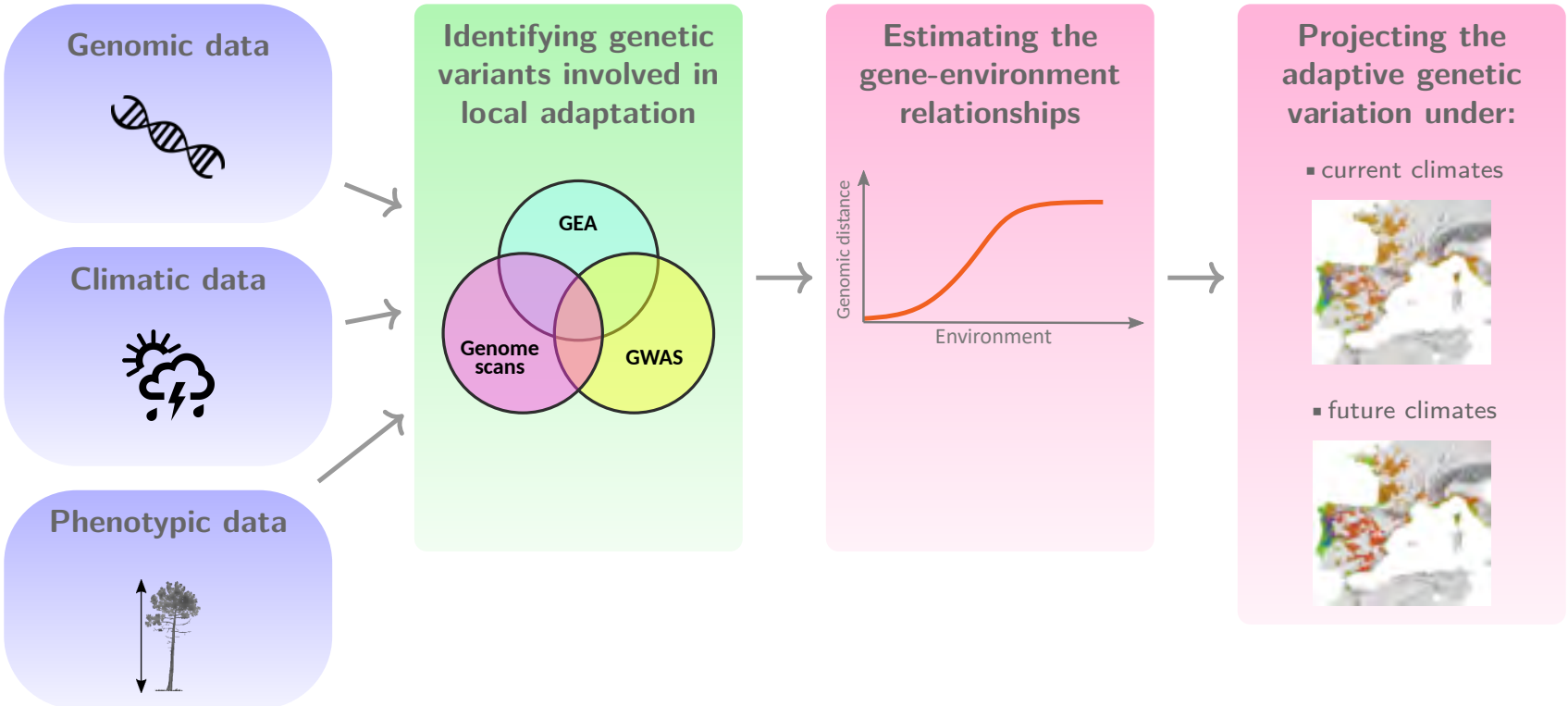
The genomic offset framework



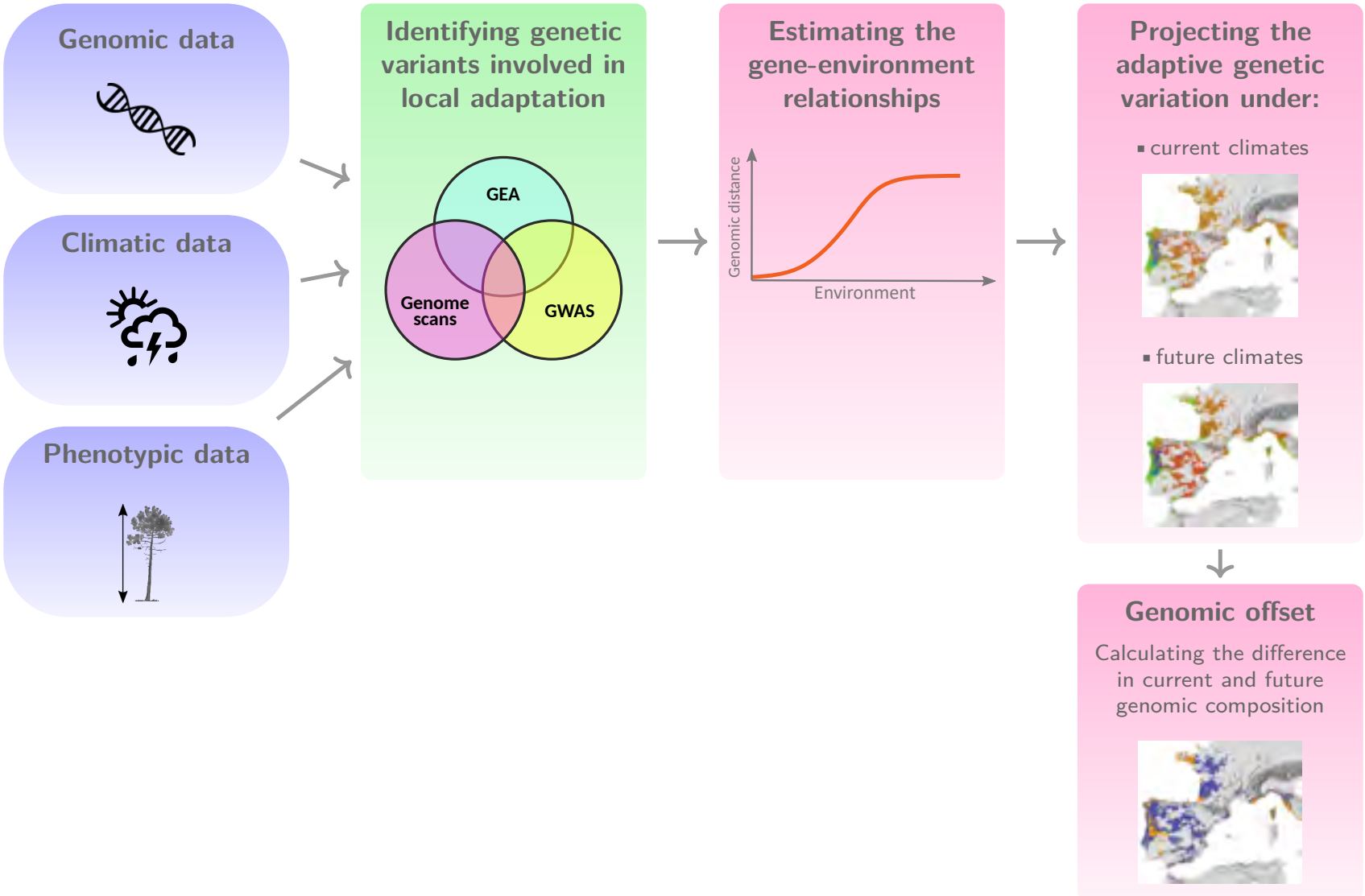
The genomic offset framework



The genomic offset framework



The genomic offset framework



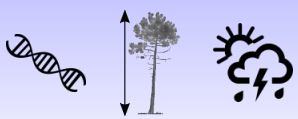
The genomic offset framework

Genomic data

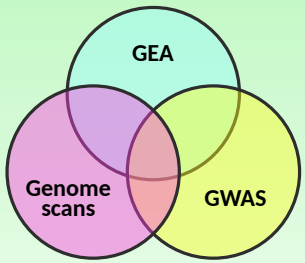
Climatic data

Phenotypic data

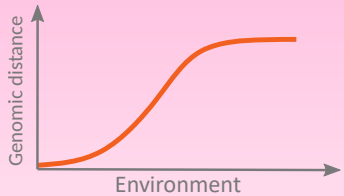
Independent data



Identifying genetic variants involved in local adaptation



Estimating the gene-environment relationships



Projecting the adaptive genetic variation under:

- current climates

- future climates

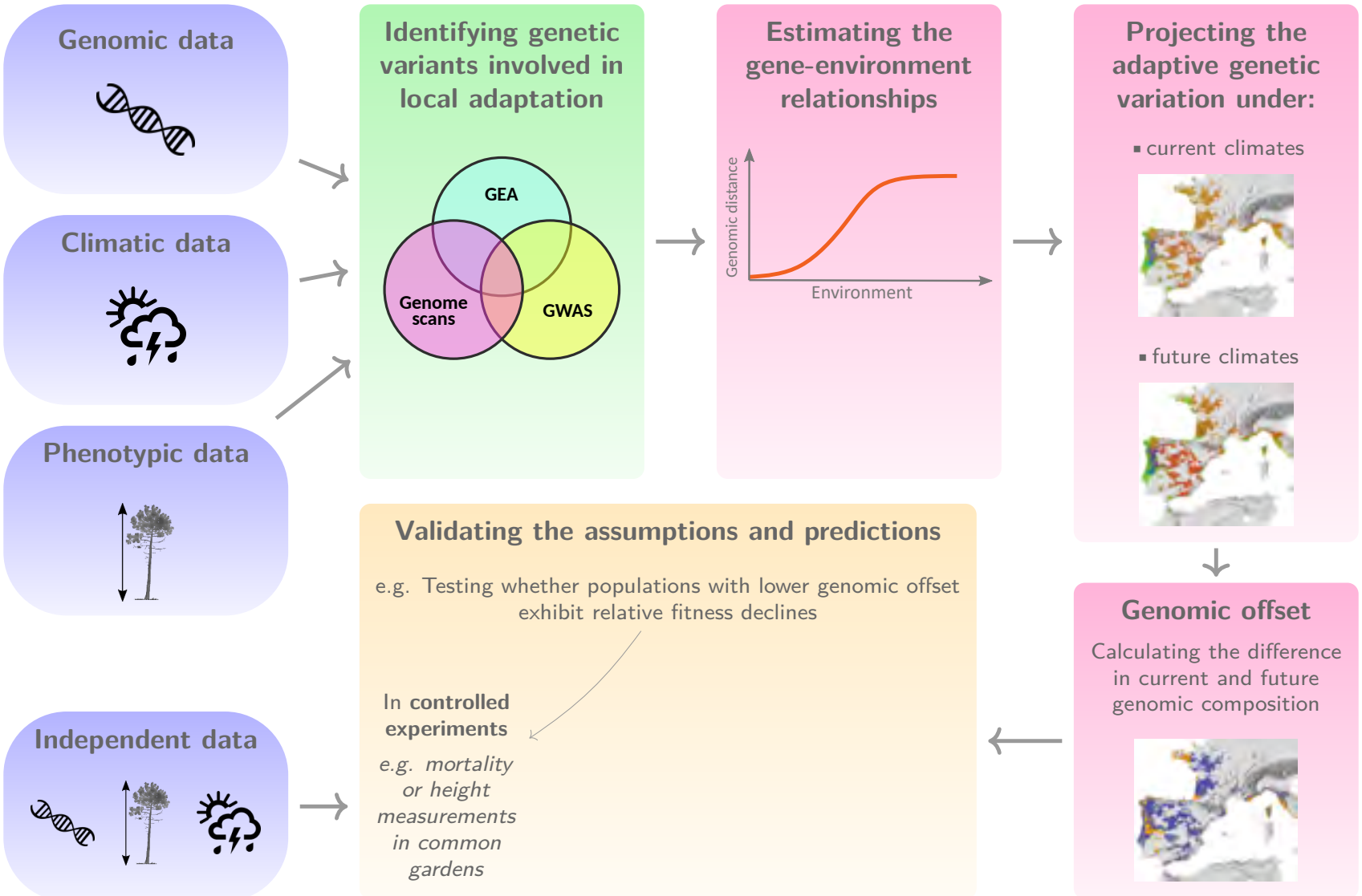
Genomic offset

Calculating the difference in current and future genomic composition

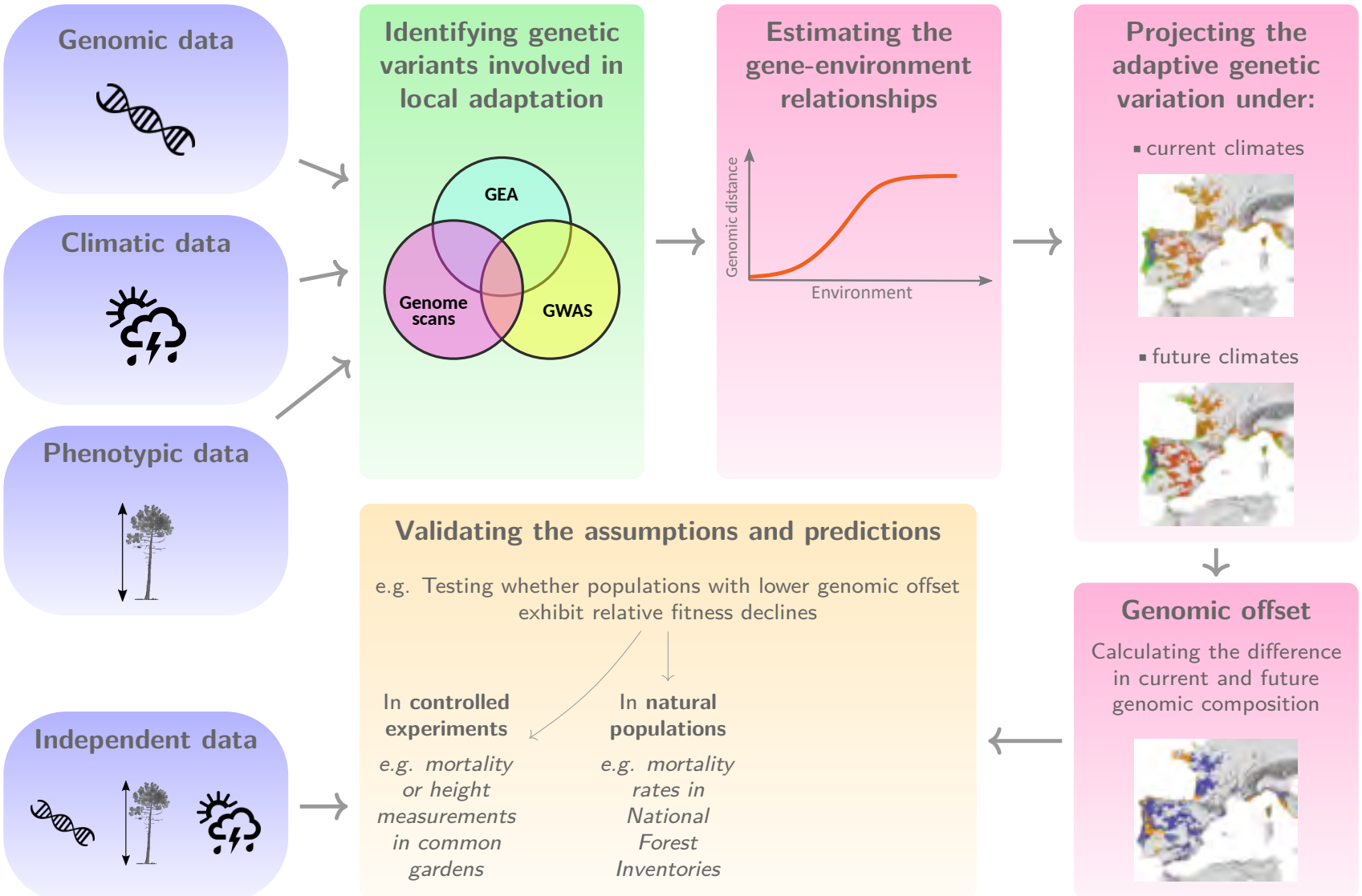
Validating the assumptions and predictions

e.g. Testing whether populations with lower genomic offset exhibit relative fitness declines

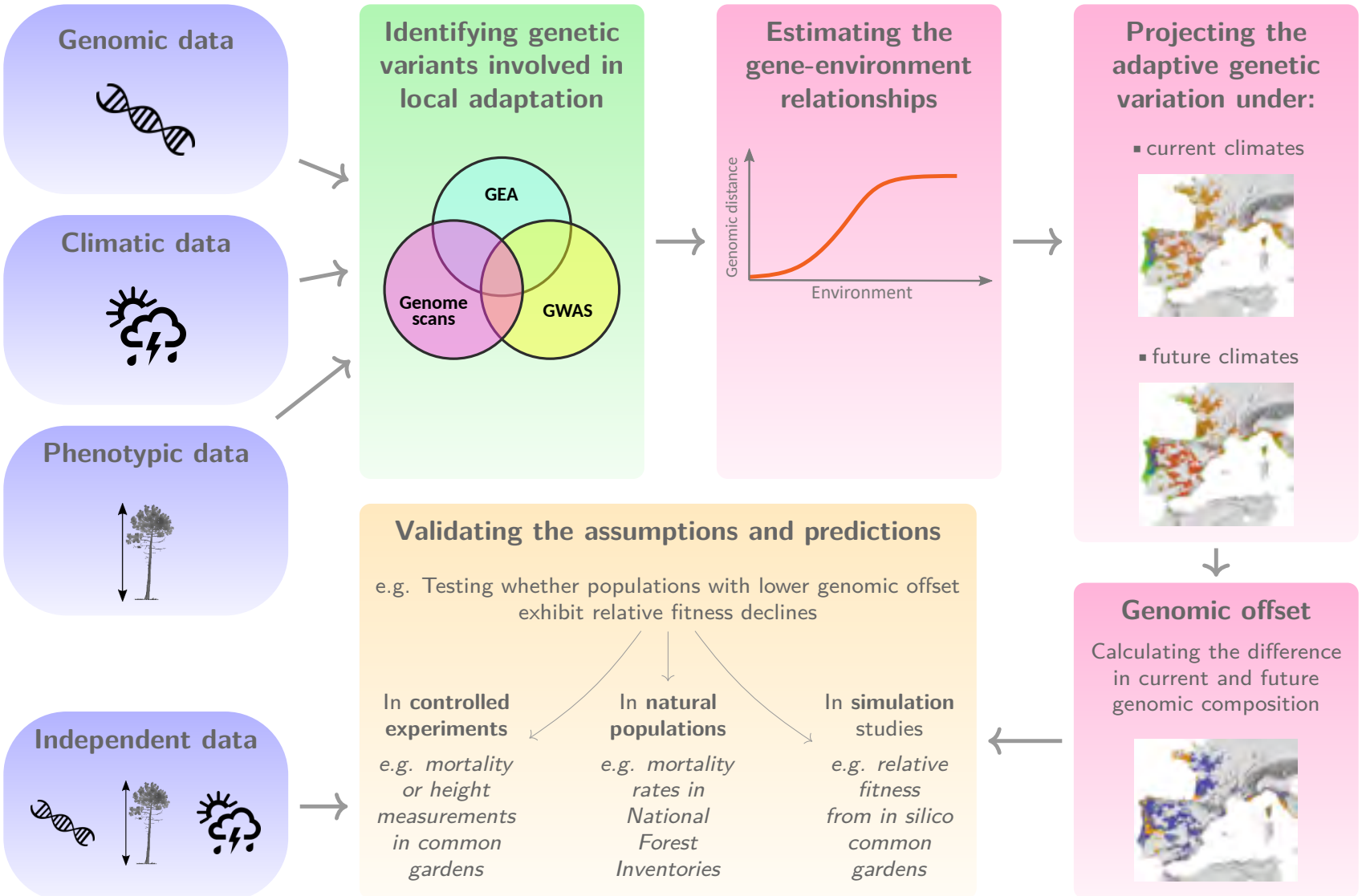
The genomic offset framework



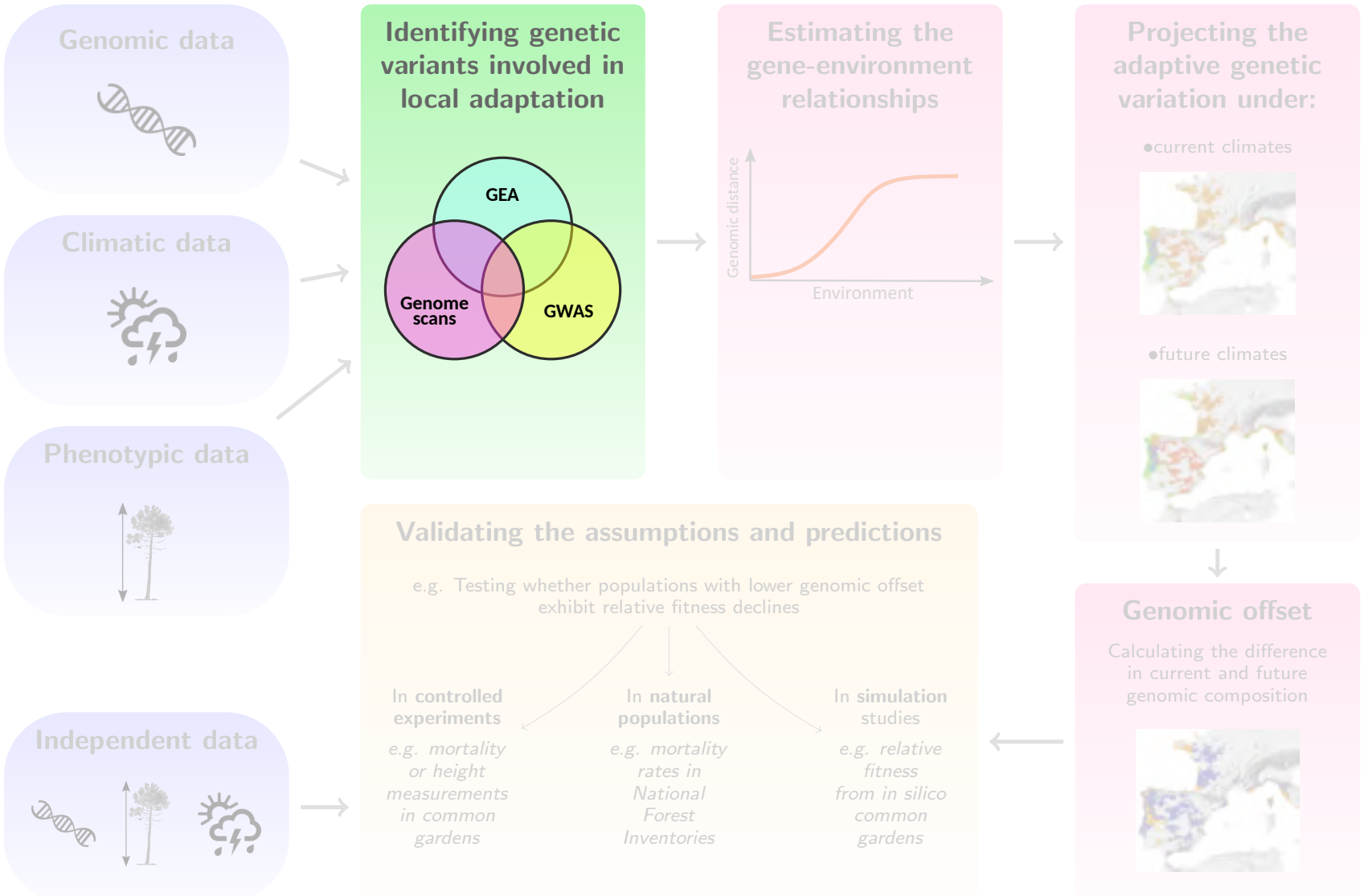
The genomic offset framework



The genomic offset framework



Discovery phase



Identifying genetic variants involved in local adaptation

Genome scans (outlier tests)

Identification of loci **deviating** from the **neutral distribution** of genetic variation.

F_{ST}-based methods: Arlequin and Bayescan.

Method based on selective sweeps: SweepFinder2.

Other methods: SelEstim.

Fitzpatrick and Keller (2015), Pina-Martins *et al.* (2019)

Identifying genetic variants involved in local adaptation

Genome scans (outlier tests)

Identification of loci **deviating** from the **neutral distribution** of genetic variation.

F_{ST}-based methods: Arlequin and Bayescan.

Method based on selective sweeps: SweepFinder2.

Other methods: SelEstim.

Fitzpatrick and Keller (2015), Pina-Martins *et al.* (2019)

Gene-environment association (GEA) methods

Identification of loci whose frequencies are associated with **climatic gradients**.

Single-locus methods: Bayenv, Bayescenv, LFMM, gINLAnd, Samβada

Multi-loci methods: RDA

Fitzpatrick and Keller (2015), Ingvarsson and Bernhardsson (2018), Ruegg *et al.* (2018), Bay *et al.* (2018), Rochat *et al.* (2021), Carvalho *et al.* 2019

Identifying genetic variants involved in local adaptation

Genome scans (outlier tests)

Identification of loci **deviating** from the **neutral distribution** of genetic variation.

F_{ST}-based methods: Arlequin and Bayescan.

Method based on selective sweeps: SweepFinder2.

Other methods: SelEstim.

Fitzpatrick and Keller (2015), Pina-Martins *et al.* (2019)

Gene-environment association (GEA) methods

Identification of loci whose frequencies are associated with **climatic gradients**.

Single-locus methods: Bayenv, Bayescenv, LFMM, gINLAnd, Samβada

Multi-loci methods: RDA

Fitzpatrick and Keller (2015), Ingvarsson and Bernhardsson (2018), Ruegg *et al.* (2018), Bay *et al.* (2018), Rochat *et al.* (2021), Carvalho *et al.* 2019

Genome-wide association studies (GWAS)

Identification of loci whose frequencies are associated with **adaptive traits**.

Single-locus methods: LFMM, BSLMM-GEMMA

Multi-loci methods: piMASS, RDA

Fitzpatrick and Keller (2015), Exposito-Alonso *et al.* (2019), Carvalho *et al.* 2019

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.
- ✗ Phenotypes can be difficult to measure: trait expression at specific ontogenic stages or in specific environments (genotype-by-environment interaction).

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.
- ✗ Phenotypes can be difficult to measure: trait expression at specific ontogenic stages or in specific environments (genotype-by-environment interaction).

→ GEAs and genome scans: a focus on **selection**-driven changes in **allele frequency** along **climatic**/**spatial** gradients.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.
- ✗ Phenotypes can be difficult to measure: trait expression at specific ontogenic stages or in specific environments (genotype-by-environment interaction).

→ GEAs and genome scans: a focus on **selection**-driven changes in **allele frequency** along **climatic**/**spatial** gradients.

- ✓ No need to choose which phenotype to measure and in which developmental stages or environments.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.
- ✗ Phenotypes can be difficult to measure: trait expression at specific ontogenic stages or in specific environments (genotype-by-environment interaction).

→ GEAs and genome scans: a focus on **selection**-driven changes in **allele frequency** along **climatic**/**spatial** gradients.

- ✓ No need to choose which phenotype to measure and in which developmental stages or environments.
- ✓ Can be particularly relevant for **long-lived species** or species with complex life histories or when phenotypes involved in climate adaptation are unknown.

Identifying genetic variants involved in local adaptation

Which method to use?

Capblancq *et al.* 2020

→ GWAS: a focus on **climate-adaptive phenotypes**.

- ✓ Phenotypes are the direct target of selection.
- ✓ Means of incorporating *a priori* knowledge of traits most important to climate adaptation in the species considered.
- ✗ Phenotypes can be difficult to measure: trait expression at specific ontogenic stages or in specific environments (genotype-by-environment interaction).

→ GEAs and genome scans: a focus on **selection**-driven changes in **allele frequency** along **climatic**/**spatial** gradients.

- ✓ No need to choose which phenotype to measure and in which developmental stages or environments.
- ✓ Can be particularly relevant for **long-lived species** or species with complex life histories or when phenotypes involved in climate adaptation are unknown.
- ✗ Choosing a set of climate variables and appropriate spatial and temporal scales to capture the most likely drivers of selection is challenging.

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

→ Reduces **false positives** but at the expense of **true positives**.

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

→ Reduces **false positives** but at the expense of **true positives**.

→ Biases detections towards **strong selective sweeps**, as most methods are unlikely to detect **weak signals of selection** (e.g. recent selection or selection on standing genetic variation).

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

→ Reduces **false positives** but at the expense of **true positives**.

→ Biases detections towards **strong selective sweeps**, as most methods are unlikely to detect **weak signals of selection** (e.g. recent selection or selection on standing genetic variation).

→ Limits detections to the least powerful (or most conservative) method.

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

→ Reduces **false positives** but at the expense of **true positives**.

→ Biases detections towards **strong selective sweeps**, as most methods are unlikely to detect **weak signals of selection** (e.g. recent selection or selection on standing genetic variation).

→ Limits detections to the least powerful (or most conservative) method.

✓ Can be relevant in **exploratory analyses** or when the **selective gradients** are **unknown**.

Identifying genetic variants involved in local adaptation

Should we select overlapping detections across different methods?

Lotterhos and Witlock 2015, Forester *et al.* 2018, Ahrens *et al.* 2021

→ Reduces **false positives** but at the expense of **true positives**.

→ Biases detections towards **strong selective sweeps**, as most methods are unlikely to detect **weak signals of selection** (e.g. recent selection or selection on standing genetic variation).

→ Limits detections to the least powerful (or most conservative) method.

✓ Can be relevant in **exploratory analyses** or when the **selective gradients** are **unknown**.

✗ Too conservative when **selection is well understood** and uninformative predictors are not used, and when the goal is to detect **weakly selected loci**.

Identifying genetic variants involved in local adaptation

Multiple confounding factors...

- **Demographic history**

Excoffier *et al.* 2009, Bonhomme *et al.* 2010, Lotterhos and Witlock 2014, Ahrens *et al.* 2018.

- **Population genetic structure** reflecting gene flow levels and population isolation

Vasemägi 2006, Le Corre and Kremer 2012.

- Gene surfing

Bierne *et al.* 2011, Ahrens *et al.* 2018.

- Background selection

Bierne *et al.* 2011, Lotterhos and Witlock 2015.

- Endogenous genetic barriers

Bierne *et al.* 2011, de Mita *et al.* 2013

- Genomic variation in the recombination rate

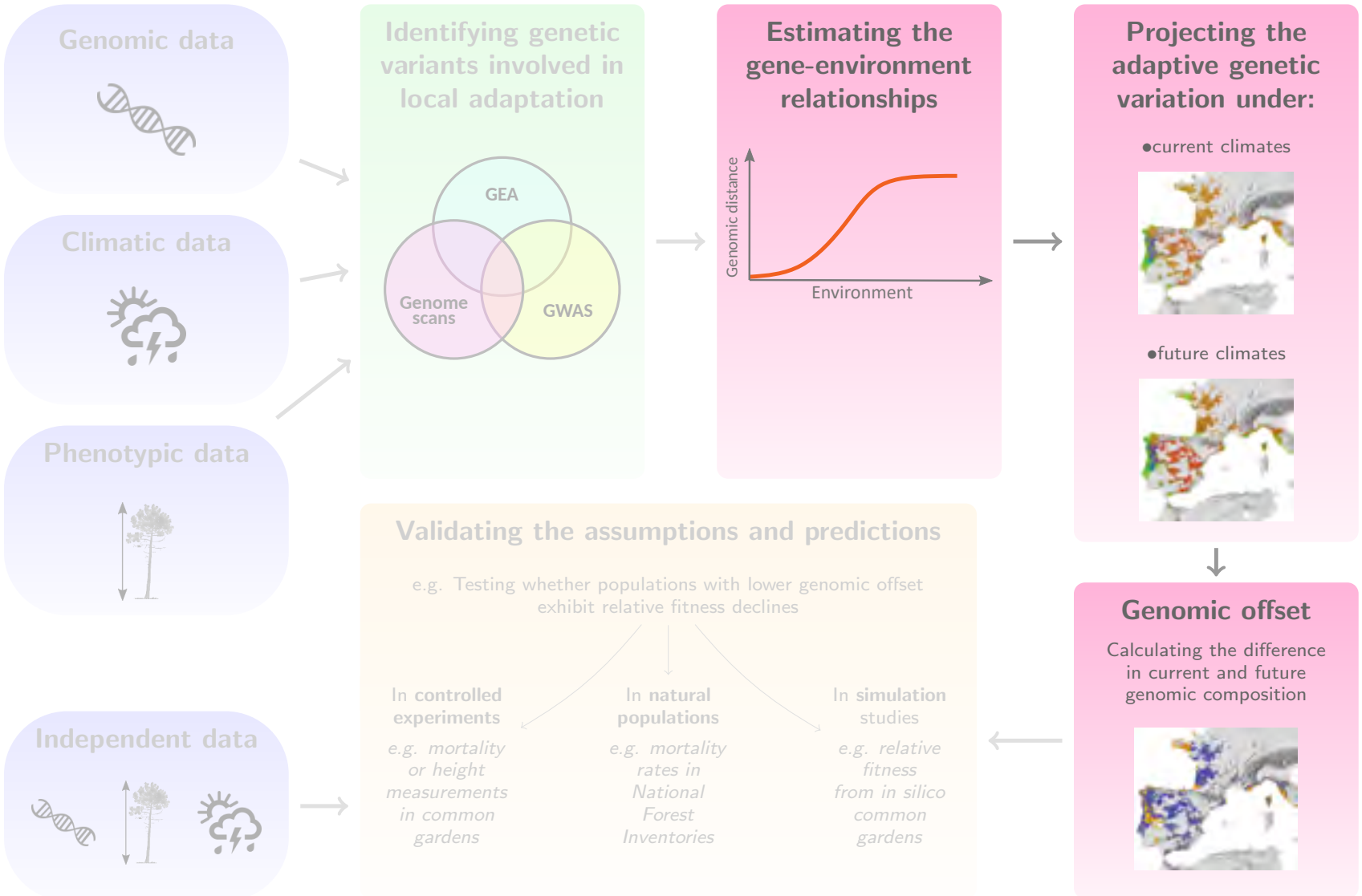
Noor and Bennett 2009, Roesti *et al.* 2012, Cruickshank and Hahn 2014.

- Genetic hitchhiking

Oleksyk *et al.* 2010, Pritchard and Di Rienzo 2010, Tiffin and Ross-Ibarra 2014.

...

Prediction phase



Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

  Generalized Dissimilarity Modelling (GDM, Ferrier *et al.* 2007)

  $\longrightarrow$ nonlinear extension of permutational matrix regression.

Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

  Generalized Dissimilarity Modelling (GDM, Ferrier *et al.* 2007)

  $\longrightarrow$ nonlinear extension of permutational matrix regression.

  Gradient Forest (GF, Ellis *et al.* 2012)

  $\longrightarrow$ extension of the nonparametric machine-learning regression tree approach known as *random forests*.

Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

  Generalized Dissimilarity Modelling (GDM, Ferrier *et al.* 2007)

  $\longrightarrow$ nonlinear extension of permutational matrix regression.

  Gradient Forest (GF, Ellis *et al.* 2012)

  $\longrightarrow$ extension of the nonparametric machine-learning regression tree approach known as *random forests*.

- Conceptual leap to spatial modelling of genomic variation in **Fitzpatrick and Keller (2015)**.

Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

  Generalized Dissimilarity Modelling (GDM, Ferrier *et al.* 2007)

  $\longrightarrow$ nonlinear extension of permutational matrix regression.

  Gradient Forest (GF, Ellis *et al.* 2012)

  $\longrightarrow$ extension of the nonparametric machine-learning regression tree approach known as *random forests*.

- Conceptual leap to spatial modelling of genomic variation in **Fitzpatrick and Keller (2015)**.

- **Nonlinear multivariate** approaches that model the **turnover in allele frequencies** along **environmental gradients**.

Predicting the genomic offset with GDM and GF

- Approaches from community-level modelling of species assemblages:

  Generalized Dissimilarity Modelling (GDM, Ferrier *et al.* 2007)

  $\longrightarrow$ nonlinear extension of permutational matrix regression.

  Gradient Forest (GF, Ellis *et al.* 2012)

  $\longrightarrow$ extension of the nonparametric machine-learning regression tree approach known as *random forests*.

- Conceptual leap to spatial modelling of genomic variation in **Fitzpatrick and Keller (2015)**.

- **Nonlinear multivariate** approaches that model the **turnover in allele frequencies** along **environmental gradients**.

- **Increasingly used** in **forest trees**, as particularly well suited to long-lived and sessile organisms.

  e.g. Fitzpatrick *et al.* 2021, Gougherty *et al.* 2020, Keller *et al.* 2018, Lu *et al.* 2019, Supple *et al.* 2018, Vanhove *et al.* 2021.

Predicting the genomic offset with GDM and GF

Modelling the turnover in allele frequency along environmental gradients.

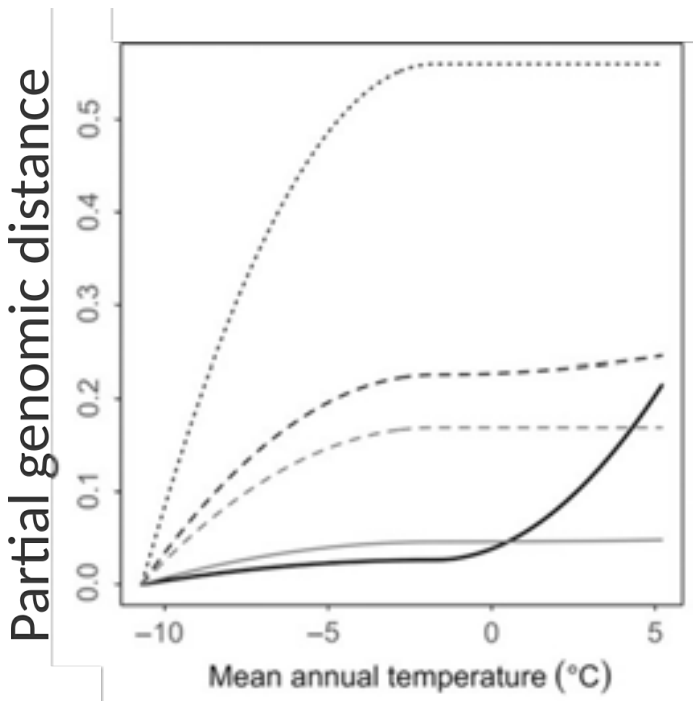
Predicting the genomic offset with GDM and GF

Modelling the turnover in allele frequency along environmental gradients.

Generalized Dissimilarity Modelling

Models **pairwise genetic dissimilarity** among populations as a function of **pairwise population differences** in **environmental** and **geographic** variables.

→ **I-spline turnover functions**



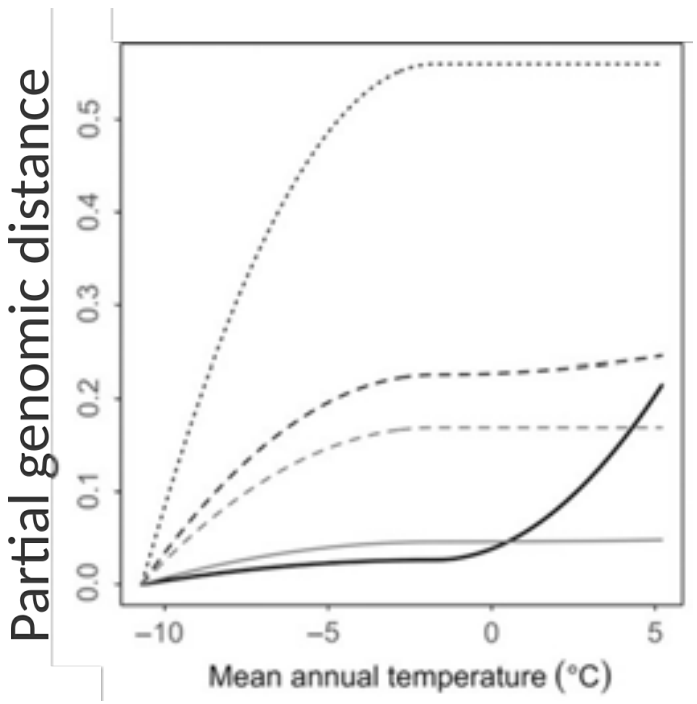
Predicting the genomic offset with GDM and GF

Modelling the turnover in allele frequency along environmental gradients.

Generalized Dissimilarity Modelling

Models **pairwise genetic dissimilarity** among populations as a function of **pairwise population differences** in **environmental** and **geographic** variables.

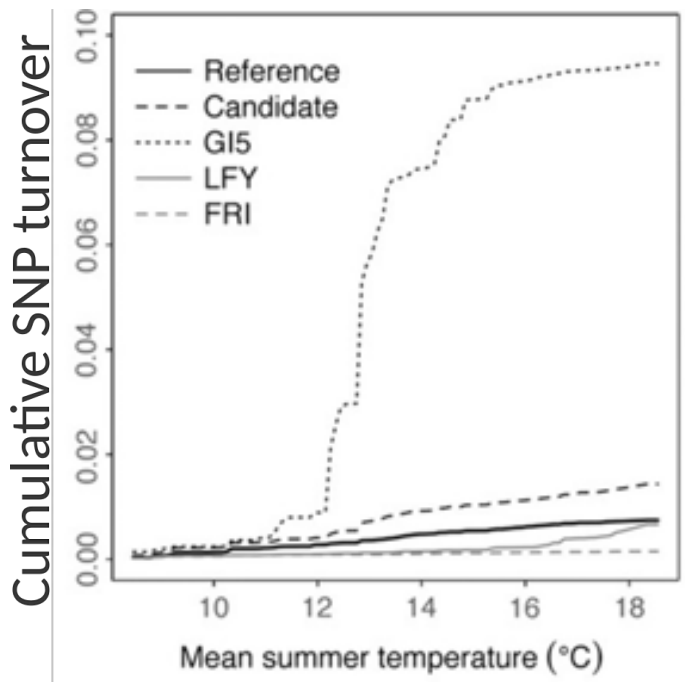
→ **I-spline turnover functions**



Gradient Forests

Fits an ensemble of **decision trees** that determine how well **partitions** distributed at numerous **split values** along each gradient explain **changes in allele frequencies** on either side of a split.

→ **Cumulative importance turnover functions**



Predicting the genomic offset with GDM and GF

Projecting the genomic composition under current and future climates

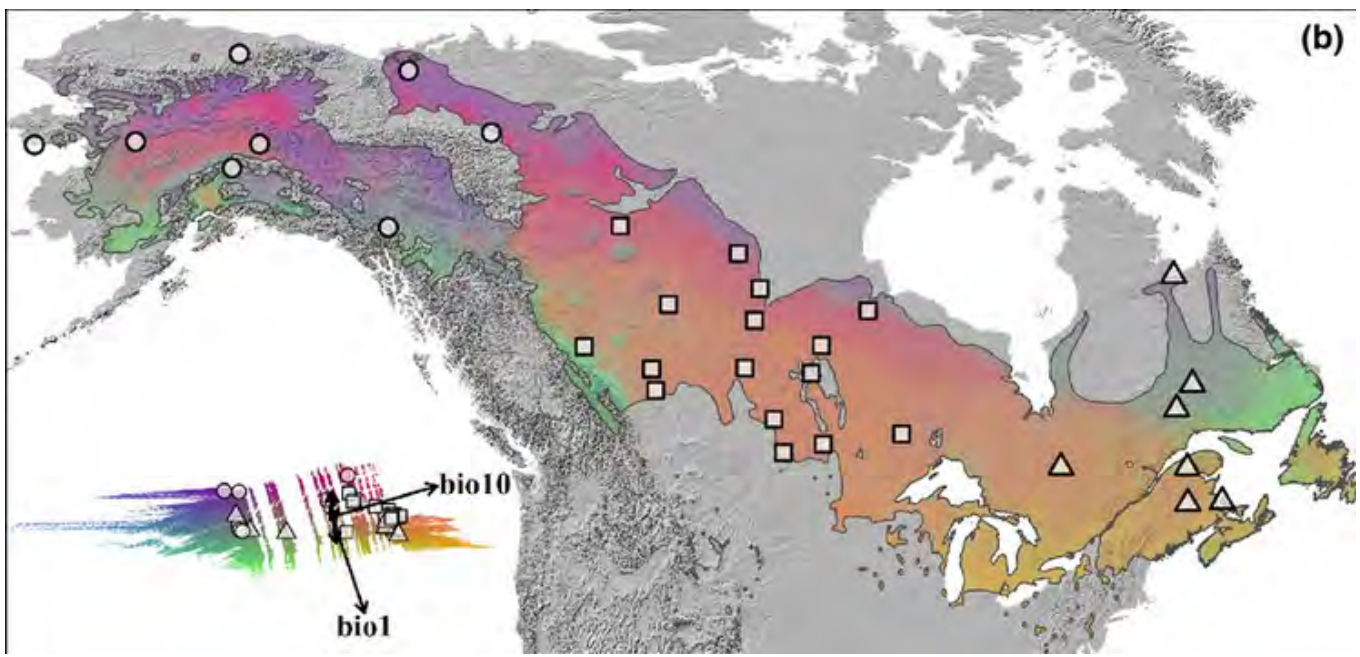
The GDM and GF turnover functions are used to **transform** the current and future values of the environmental covariates into **genetic importance values**, thus which allows to project the **current** and **future genomic composition** across the area of interest (e.g the species range).

Predicting the genomic offset with GDM and GF

Projecting the genomic composition under current and future climates

The GDM and GF turnover functions are used to **transform** the current and future values of the environmental covariates into **genetic importance values**, thus which allows to project the **current** and **future genomic composition** across the area of interest (e.g the species range).

To aid visualisation, a principal component analysis (PCA) is used to reduce the dimensionality of the transformed environmental covariates and a RGB colour palette is assigned to the first three PCs, with resulting **similar colour** for **similar expected patterns of genomic composition**.



From Fitzpatrick and Keller (2015)

Predicting the genomic offset with GDM and GF

Predicting the genomic offset across the landscape

Predicting the genomic offset with GDM and GF

Predicting the genomic offset across the landscape

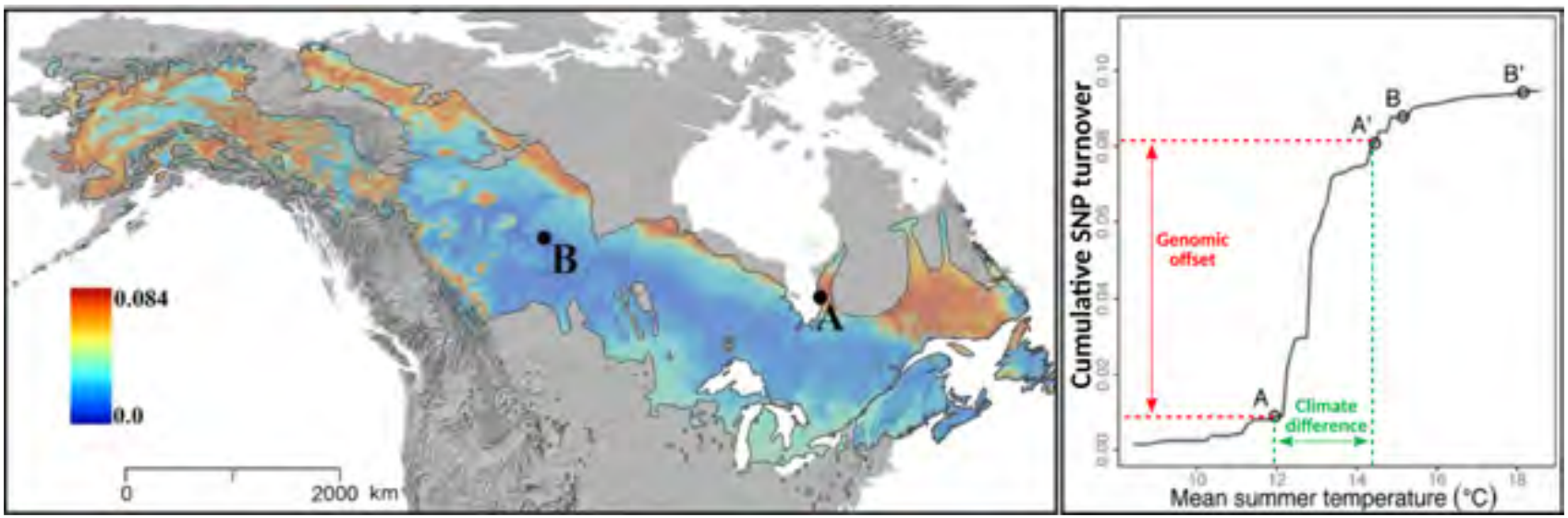
The **genomic offset** is calculated as the Euclidean **distance** between the **current** and **future** genetic importance values.

Predicting the genomic offset with GDM and GF

Predicting the genomic offset across the landscape

The **genomic offset** is calculated as the Euclidean **distance** between the **current** and **future** genetic importance values.

It is expected to indicate how much the **genomic composition** would have to **change** to **maintain** the **gene-environment associations** observed under **current** environmental conditions.



From Fitzpatrick and Keller (2015)

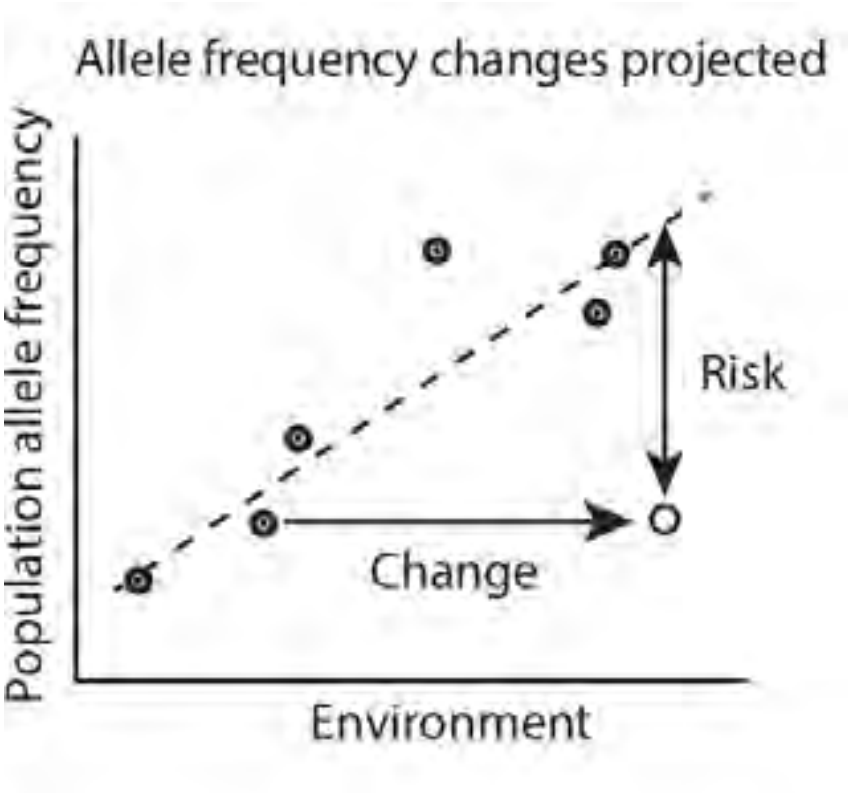
Predicting the genomic offset with linear regression (RONA)

Predicting the genomic offset with linear regression (RONA)

Rellstab *et al.* 2016.

Predicting the genomic offset with linear regression (RONA)

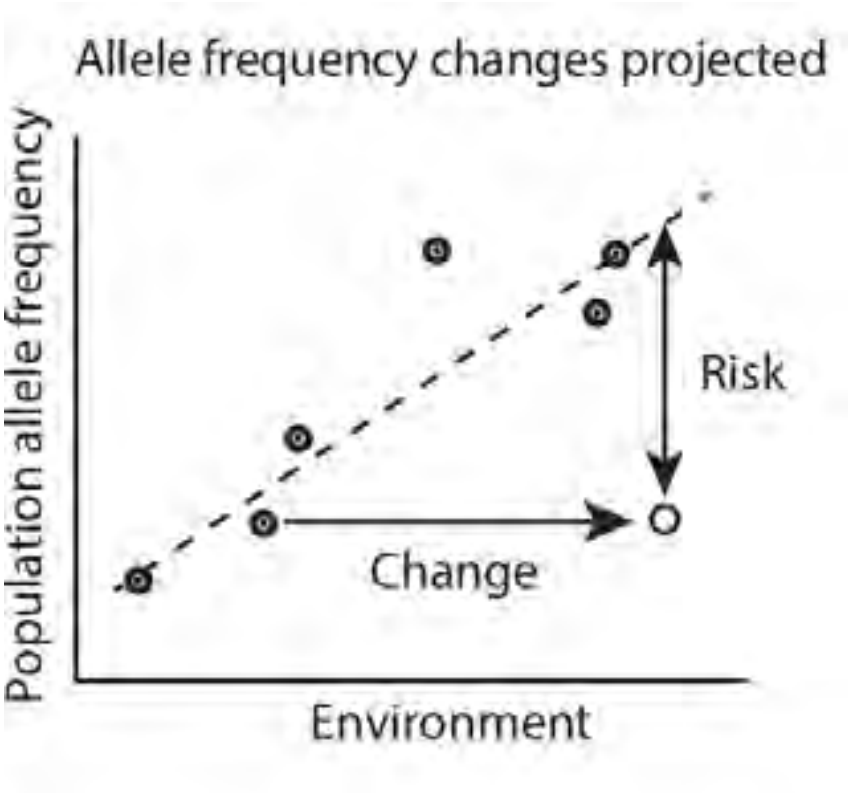
Rellstab *et al.* 2016.



Model the **linear association** between the frequencies of previously identified candidate SNPs and the environment.

Predicting the genomic offset with linear regression (RONA)

Rellstab *et al.* 2016.

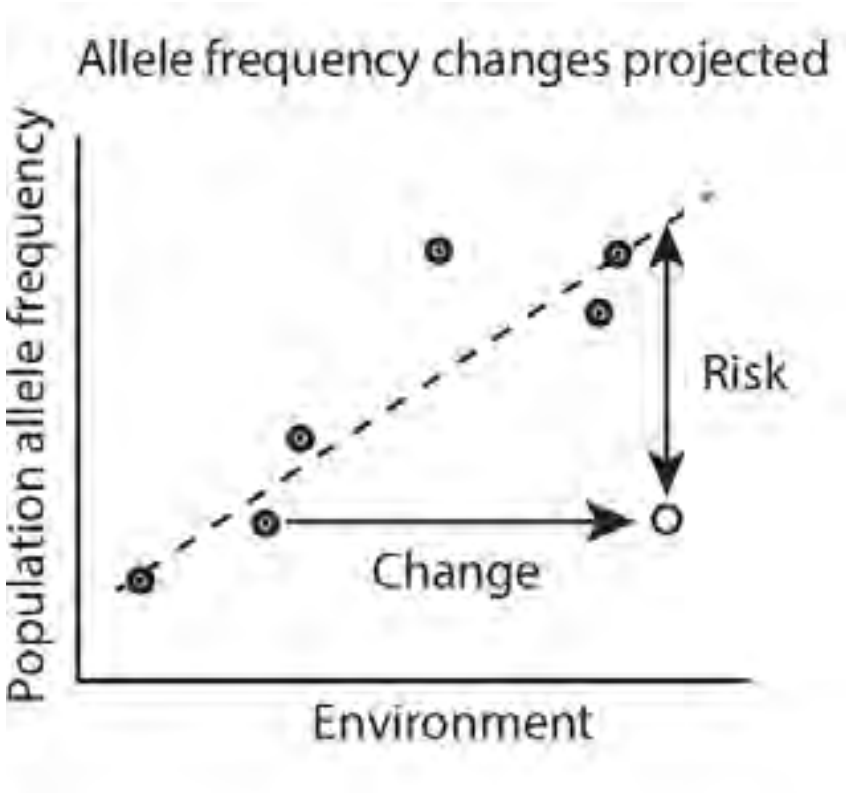


Model the **linear association** between the frequencies of previously identified candidate SNPs and the environment.

Estimated at the level of a **single-locus**.

Predicting the genomic offset with linear regression (RONA)

Rellstab *et al.* 2016.



Model the **linear association** between the frequencies of previously identified candidate SNPs and the environment.

Estimated at the level of a **single-locus**.

Risk of nonadaptedness (RONA) = average difference across the selected loci between the observed allele frequencies under current climates and the expected allele frequencies under future climates.

Predicting the genomic offset with Redundancy analysis (RDA)

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Genotype-environment associations

RDA is a **constrained ordination** that can be used to model **linear** relationships among **multiple environment predictors** and **multilocus genetic variation** (either individual genotypes or population allele frequencies) .

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Genotype-environment associations

RDA is a **constrained ordination** that can be used to model **linear** relationships among **multiple environment predictors** and **multilocus genetic variation** (either individual genotypes or population allele frequencies) .

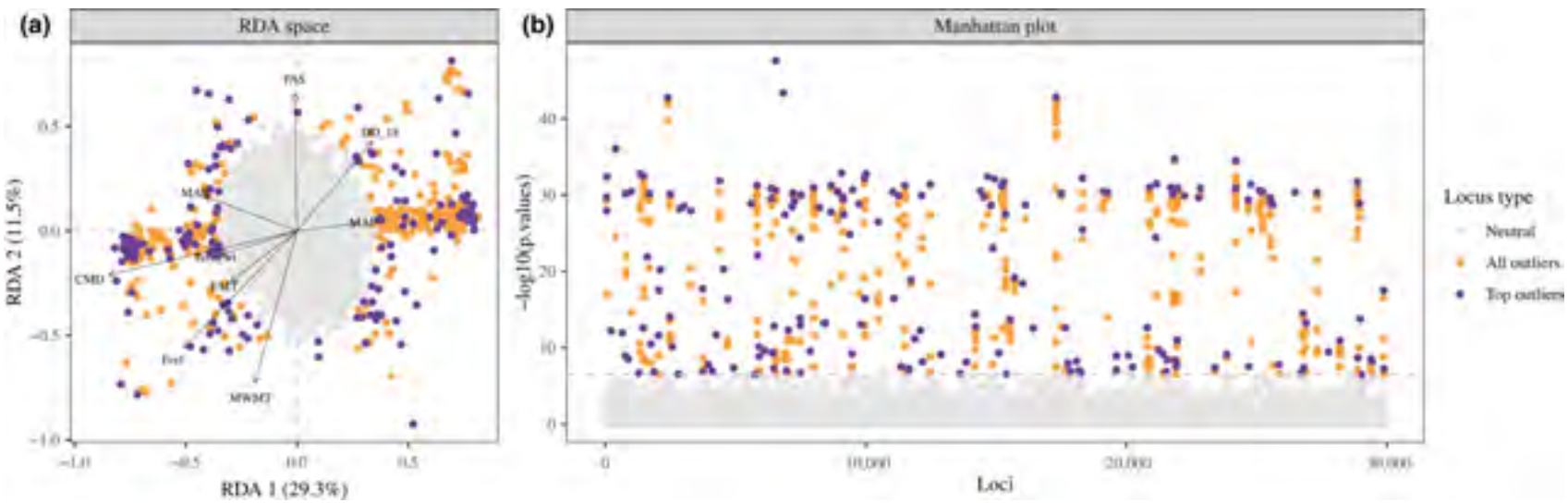
One of the **best-performing multivariate GEA methods** to date and **highly robust** to **sampling design** (Capblancq *et al*. 2018, Forester *et al*. 2018).

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Genotype-environment associations

RDA is a **constrained ordination** that can be used to model **linear** relationships among **multiple environment predictors** and **multilocus genetic variation** (either individual genotypes or population allele frequencies) .

One of the **best-performing multivariate GEA methods** to date and **highly robust** to **sampling design** (Capblancq *et al.* 2018, Forester *et al.* 2018).



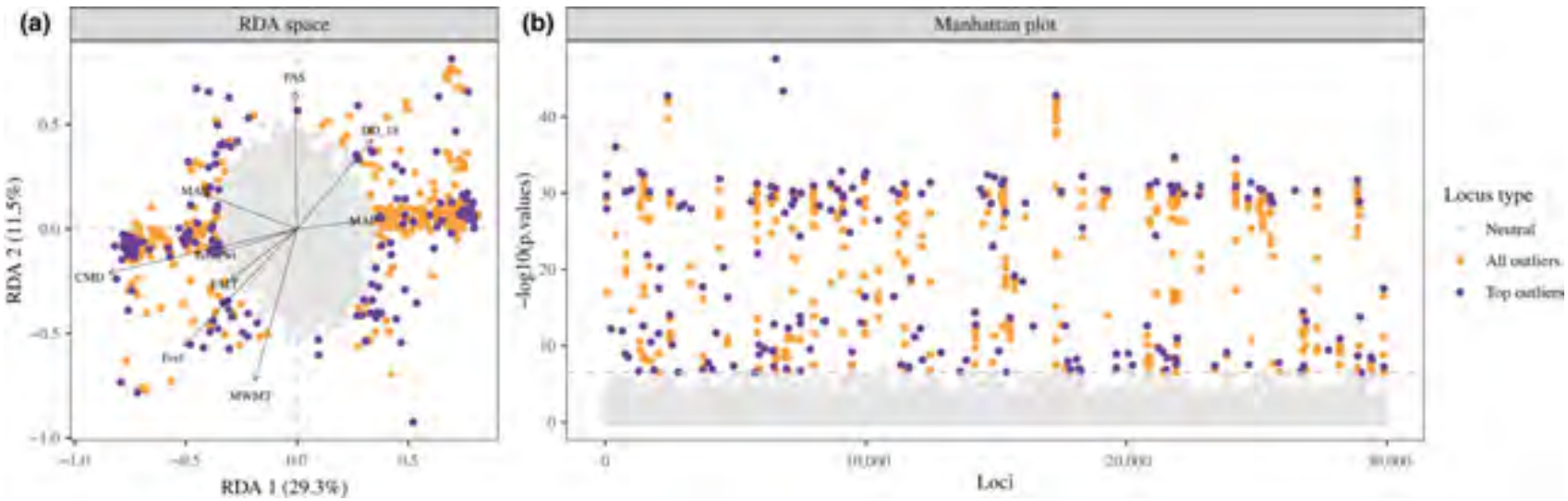
→ The **constrained ordination axes** consist of covarying sets of genotypes that are correlated with the multivariate environment.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Genotype-environment associations

RDA is a **constrained ordination** that can be used to model **linear** relationships among **multiple environment predictors** and **multilocus genetic variation** (either individual genotypes or population allele frequencies) .

One of the **best-performing multivariate GEA methods** to date and **highly robust** to **sampling design** (Capblancq *et al.* 2018, Forester *et al.* 2018).



→ The **constrained ordination axes** consist of covarying sets of genotypes that are correlated with the multivariate environment.

→ The **loci putatively under selection** are the markers showing **extreme loadings** (i.e. score) along one or multiple axes.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Spatial projections of adaptive gradient(s)

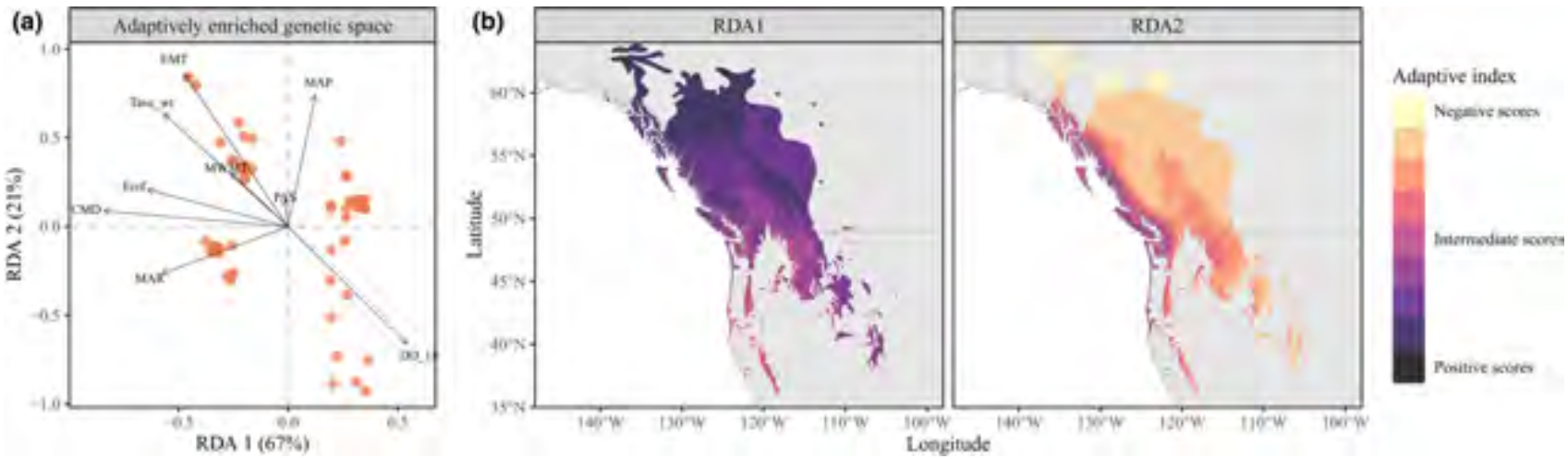
The **genotype–environment associations** (more precisely the scores of the environmental variables along the RDA axes) can be used to predict an **index of adaptive genetic composition** for any **environment** across the landscape.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Spatial projections of adaptive gradient(s)

The **genotype–environment associations** (more precisely the scores of the environmental variables along the RDA axes) can be used to predict an **index of adaptive genetic composition** for any **environment** across the landscape.

This index provides an estimate of **adaptive genetic similarity or difference** of all pixels on the landscape as a function of the values of the **environmental predictors** at that location.

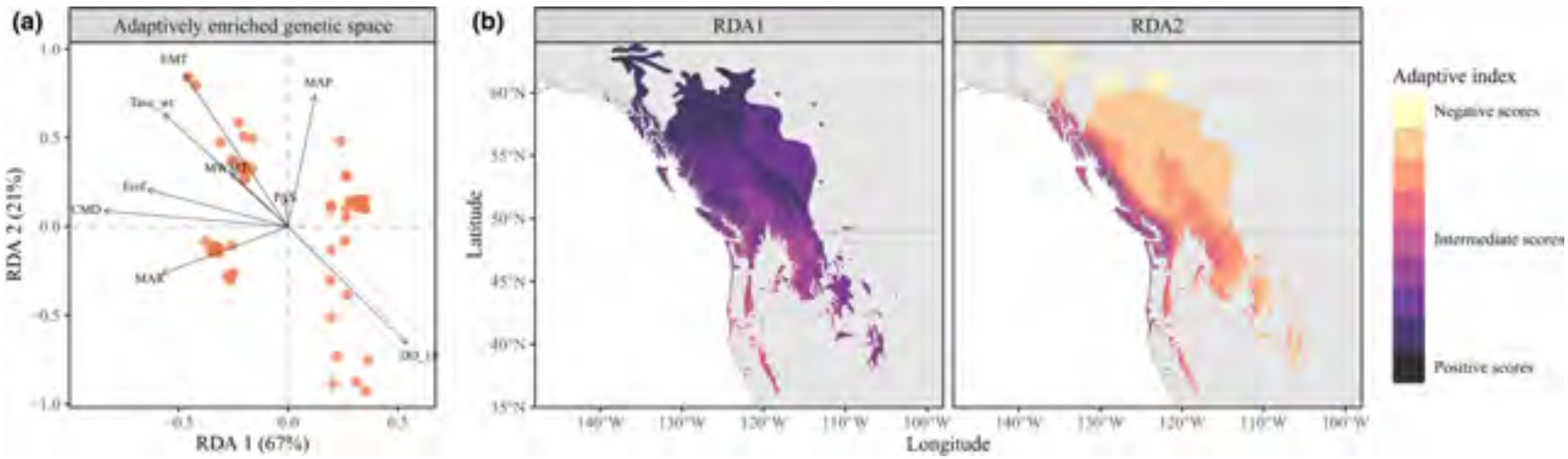


Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Spatial projections of adaptive gradient(s)

The **genotype–environment associations** (more precisely the scores of the environmental variables along the RDA axes) can be used to predict an **index of adaptive genetic composition** for any **environment** across the landscape.

This index provides an estimate of **adaptive genetic similarity or difference** of all pixels on the landscape as a function of the values of the **environmental predictors** at that location.



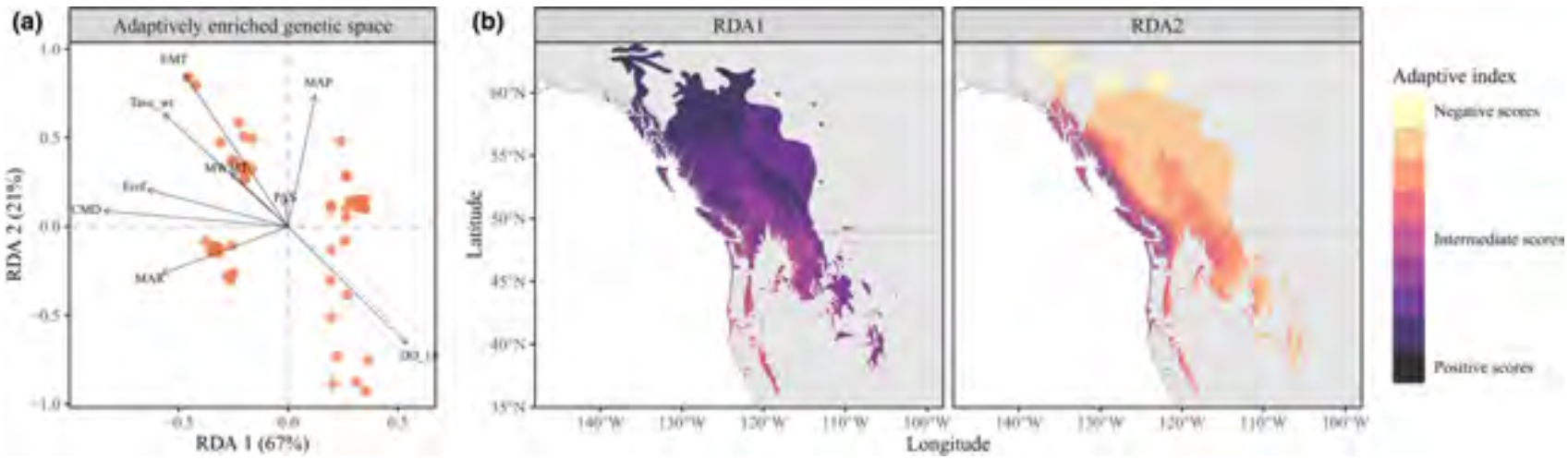
→ The '**adaptively enriched genetic space**' shows how the genetic markers correlate with the different climatic variables as well as how the climatic variables correlate with each other.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Spatial projections of adaptive gradient(s)

The **genotype–environment associations** (more precisely the scores of the environmental variables along the RDA axes) can be used to predict an **index of adaptive genetic composition** for any **environment** across the landscape.

This index provides an estimate of **adaptive genetic similarity or difference** of all pixels on the landscape as a function of the values of the **environmental predictors** at that location.



→ The '**adaptively enriched genetic space**' shows how the genetic markers correlate with the different climatic variables as well as how the climatic variables correlate with each other.

→ In this example, the **spatial projection** of the **adaptive gradient** across the landscape shows how RDA1 index contrasts southern/low elevation areas from northern/high elevation areas, and how RDA2 index contrasts coastal/southern areas from inland/mountainous areas.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Predicting the genomic offset

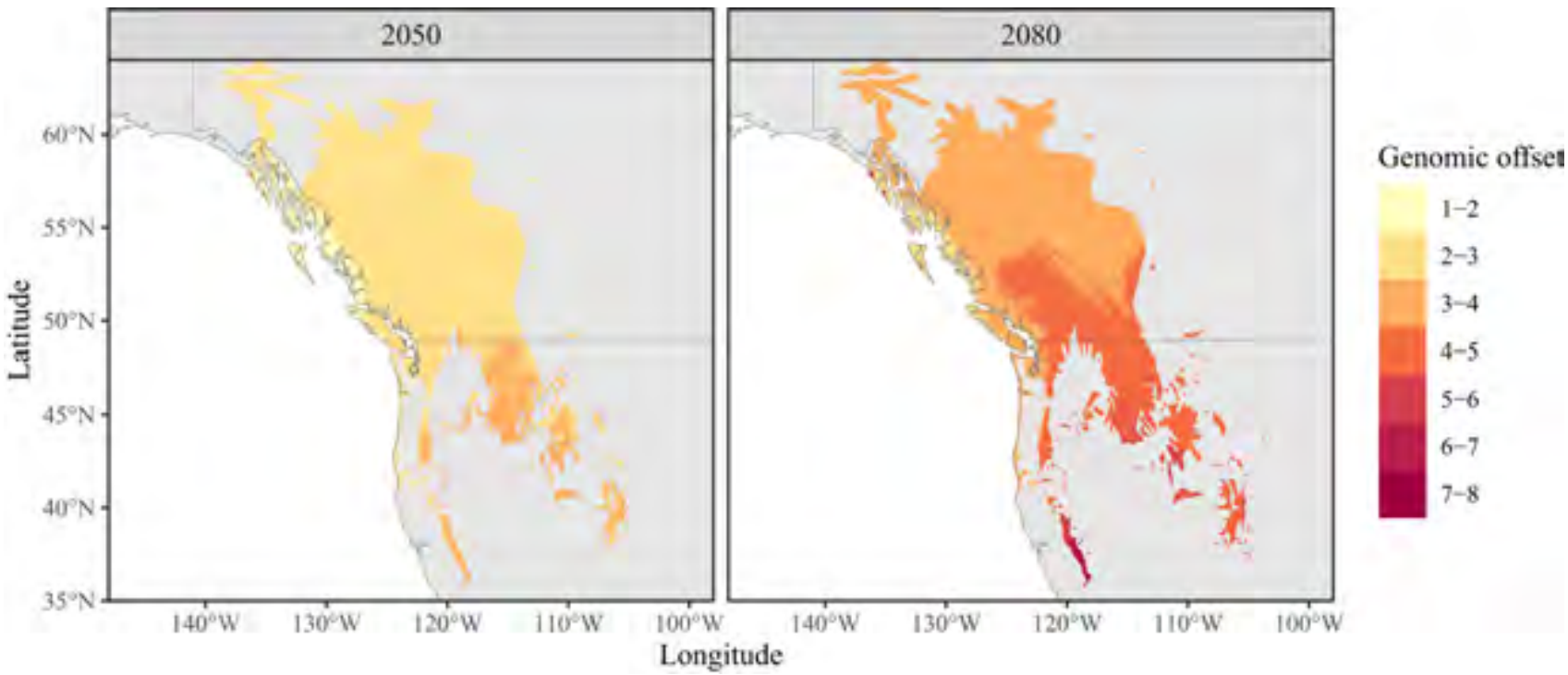
The **genotype–environment associations** can be used to predict the **optimal adaptive index** of all pixels on the landscape using both current and future environmental conditions.

Predicting the genomic offset with Redundancy analysis (RDA)

Capblancq and Forester 2021. Predicting the genomic offset

The **genotype–environment associations** can be used to predict the **optimal adaptive index** of all pixels on the landscape using both current and future environmental conditions.

The **genomic offset** is calculated as the difference between the two predictions, and provides an estimate of the **shift in adaptive index** that would be required to track climate change.



Predicting the genomic offset with multivariate logistic regressions

Rochat *et al.* 2021.

Predicting the genomic offset with multivariate logistic regressions

Rochat *et al.* 2021.

- Predict the frequency of adaptive genotypes called **SPatial Areas of Genotypes Probabilities** (SPAG) based on **logistic genotype-environment associations**.

Predicting the genomic offset with multivariate logistic regressions

Rochat *et al.* 2021.

- Predict the frequency of adaptive genotypes called **SPatial Areas of Genotypes Probabilities** (SPAG) based on **logistic genotype-environment associations**.
- Allows for the combination of adaptive genotypes in different additive/epistatic relationships.

Predicting the genomic offset with multivariate logistic regressions

Rochat *et al.* 2021.

- Predict the frequency of adaptive genotypes called **SPatial Areas of Genotypes Probabilities** (SPAG) based on **logistic genotype-environment associations**.
- Allows for the combination of adaptive genotypes in different additive/epistatic relationships.
- The genomic offset is measured as the difference between current and future SPAGs.

Predicting the genomic offset with multivariate logistic regressions

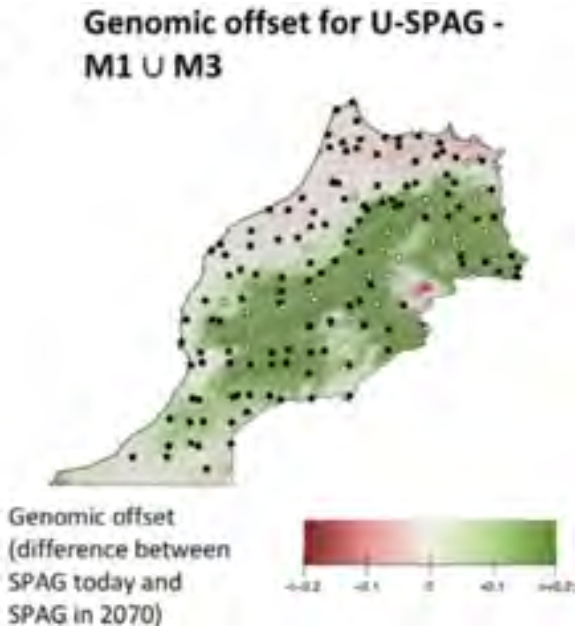
Rochat *et al.* 2021.

- Predict the frequency of adaptive genotypes called **SPatial Areas of Genotypes Probabilities** (SPAG) based on **logistic genotype-environment associations**.
- Allows for the combination of adaptive genotypes in different additive/epistatic relationships.
- The genomic offset is measured as the difference between current and future SPAGs.
- Strong advantage: a **cross-validation** step to assess the accuracy of the evolutionary potential predictions under the different intergenic relationships.

Predicting the genomic offset with multivariate logistic regressions

Rochat *et al.* 2021.

- Predict the frequency of adaptive genotypes called **SPatial Areas of Genotypes Probabilities** (SPAG) based on **logistic genotype-environment associations**.
- Allows for the combination of adaptive genotypes in different additive/epistatic relationships.
- The genomic offset is measured as the difference between current and future SPAGs.
- Strong advantage: a **cross-validation** step to assess the accuracy of the evolutionary potential predictions under the different intergenic relationships.



Union SPAGs (U-SPAGs) = probabilities of finding at least one of the adaptive single-locus genotypes of interest.

→ **epistatic situation** where the presence of an adaptive single-locus genotype is sufficient for securing the selective advantage, while additional adaptive genotypes do not confer additional advantages.

Predicting the genomic offset with latent factor mixed models (LFMM)

Gain and François *et al.* 2021.

Predicting the genomic offset with latent factor mixed models (LFMM)

Gain and François *et al.* 2021.

• The **genotype–environment associations** are estimated with LFMMs, which optimally separate **neutral genetic variation**—modelled in the **latent factors** and based on previously computed ancestry coefficients—from **adaptive genetic variation** modelled in the **effect sizes** of environmental covariates.

Predicting the genomic offset with latent factor mixed models (LFMM)

Gain and François *et al.* 2021.

- The **genotype–environment associations** are estimated with LFMMs, which optimally separate **neutral genetic variation**—modelled in the **latent factors** and based on previously computed ancestry coefficients—from **adaptive genetic variation** modelled in the **effect sizes** of environmental covariates.
- Assumes that the effects of the environment are potentially weak but **highly polygenic**.

Predicting the genomic offset with latent factor mixed models (LFMM)

Gain and François *et al.* 2021.

- The **genotype–environment associations** are estimated with LFMMs, which optimally separate **neutral genetic variation**—modelled in the **latent factors** and based on previously computed ancestry coefficients—from **adaptive genetic variation** modelled in the **effect sizes** of environmental covariates.
- Assumes that the effects of the environment are potentially weak but **highly polygenic**.
- Uses **effect sizes** instead of significance values.

Predicting the genomic offset with latent factor mixed models (LFMM)

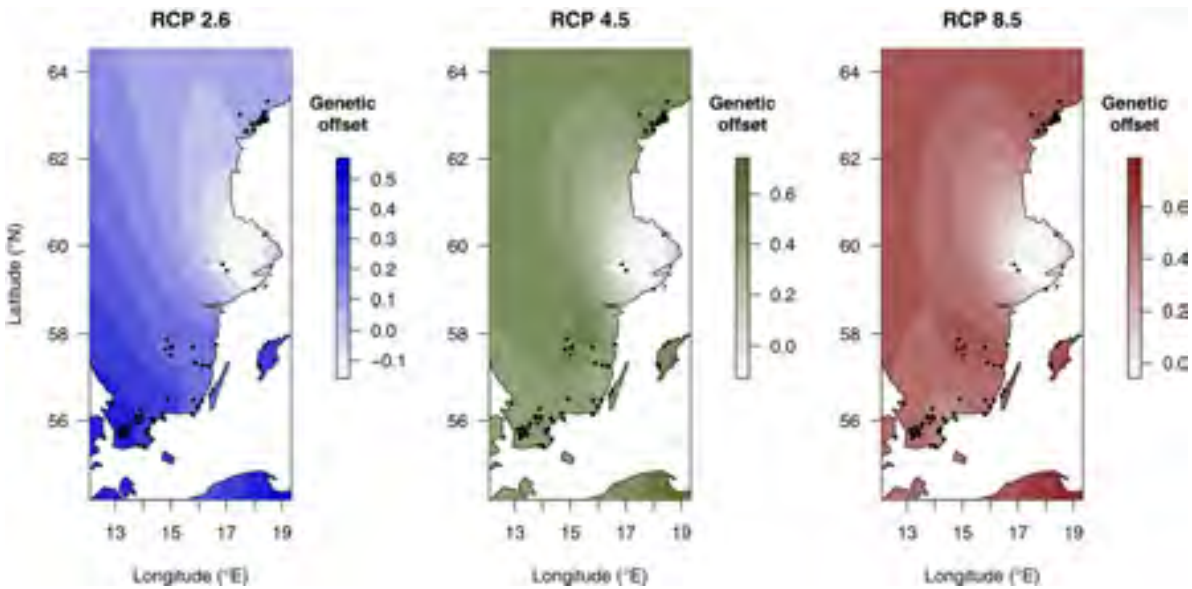
Gain and François *et al.* 2021.

- The **genotype–environment associations** are estimated with LFMMs, which optimally separate **neutral genetic variation**—modelled in the **latent factors** and based on previously computed ancestry coefficients—from **adaptive genetic variation** modelled in the **effect sizes** of environmental covariates.
- Assumes that the effects of the environment are potentially weak but **highly polygenic**.
- Uses **effect sizes** instead of significance values.
- Leverages the entire set of genotyped loci to predict genomic variation under projected conditions.

Predicting the genomic offset with latent factor mixed models (LFMM)

Gain and François *et al.* 2021.

- The **genotype–environment associations** are estimated with LFMMs, which optimally separate **neutral genetic variation**—modelled in the **latent factors** and based on previously computed ancestry coefficients—from **adaptive genetic variation** modelled in the **effect sizes** of environmental covariates.
- Assumes that the effects of the environment are potentially weak but **highly polygenic**.
- Uses **effect sizes** instead of significance values.
- Leverages the entire set of genotyped loci to predict genomic variation under projected conditions.



Predicting the genomic offset

To sum-up:

$\rightarrow$ Genomic offset approaches are in their infancy of development and applications.

$\rightarrow$ A growing number of methods, each with different pros and cons.

Accounting for population structure

Accounting for population structure

- Generalized Dissimilarity Modelling

Using any measure of geographic separation between sites as a predictor.

Accounting for population structure

- **Generalized Dissimilarity Modelling**

Using any measure of geographic separation between sites as a predictor.

- **Gradient Forests**

Cannot directly accommodate spatial effects.

Moran's eigenvector map (MEM) variables have been used to indirectly capture the influence of spatial processes and unmeasured environmental variation, thus assuming that population structure is somehow correlated with geographic position, which might not be the case in species with complex demography and highly heterogeneous habitats.

Accounting for population structure

- **Generalized Dissimilarity Modelling**

Using any measure of geographic separation between sites as a predictor.

- **Gradient Forests**

Cannot directly accommodate spatial effects.

Moran's eigenvector map (MEM) variables have been used to indirectly capture the influence of spatial processes and unmeasured environmental variation, thus assuming that population structure is somehow correlated with geographic position, which might not be the case in species with complex demography and highly heterogeneous habitats.

- **RONA**

No direct correction for population structure.

Can be indirectly accounted for by using transformed or residual allele frequencies after correcting for population structure.

Accounting for population structure

- **Generalized Dissimilarity Modelling**

Using any measure of geographic separation between sites as a predictor.

- **Gradient Forests**

Cannot directly accommodate spatial effects.

Moran's eigenvector map (MEM) variables have been used to indirectly capture the influence of spatial processes and unmeasured environmental variation, thus assuming that population structure is somehow correlated with geographic position, which might not be the case in species with complex demography and highly heterogeneous habitats.

- **RONA**

No direct correction for population structure.

Can be indirectly accounted for by using transformed or residual allele frequencies after correcting for population structure.

- **Redundancy Analysis**

Population structure can be included as a conditioning variable adjusted in a *partial RDA*, using ancestry coefficients, PC axes or spatial eigenvectors (e.g. Moran's eigenvector maps).

Accounting for population structure

- **Generalized Dissimilarity Modelling**

Using any measure of geographic separation between sites as a predictor.

- **Gradient Forests**

Cannot directly accommodate spatial effects.

Moran's eigenvector map (MEM) variables have been used to indirectly capture the influence of spatial processes and unmeasured environmental variation, thus assuming that population structure is somehow correlated with geographic position, which might not be the case in species with complex demography and highly heterogeneous habitats.

- **RONA**

No direct correction for population structure.

Can be indirectly accounted for by using transformed or residual allele frequencies after correcting for population structure.

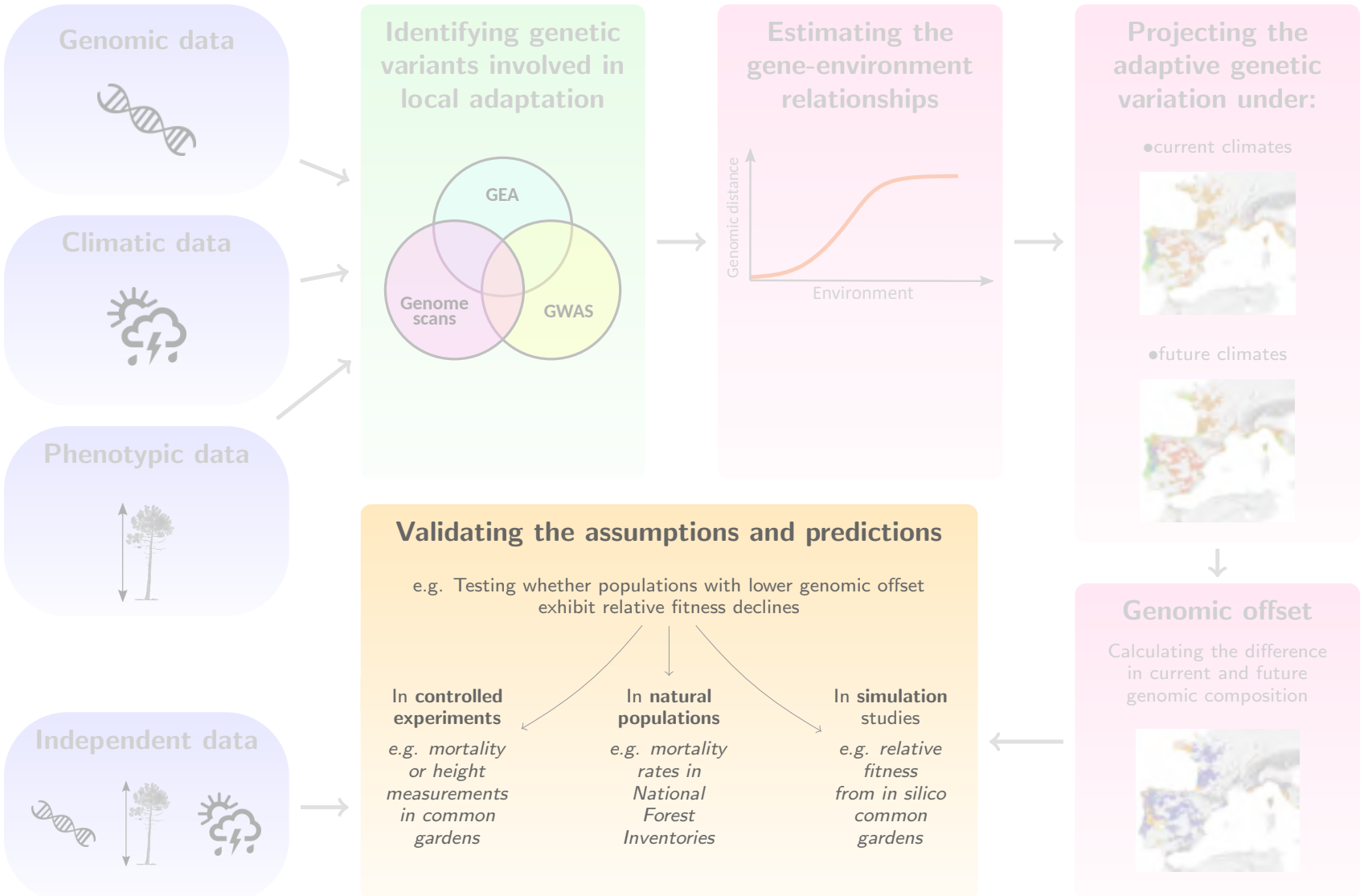
- **Redundancy Analysis**

Population structure can be included as a conditioning variable adjusted in a *partial RDA*, using ancestry coefficients, PC axes or spatial eigenvectors (e.g. Moran's eigenvector maps).

- **Latent Factor Mixed Models**

Directly incorporates neutral population genetic structure through the use of latent factors.

Validation phase



Validation phase

An approach that relies on key assumptions in need of validation:

Capblancq *et al.* 2020, Rellstab *et al.* 2021

Validation phase

An approach that relies on key assumptions in need of validation:

Capblancq *et al.* 2020, Rellstab *et al.* 2021

- The alleles involved in local adaptation have been correctly identified.

→ a common concern to studies using GWAS, GEA methods or genome scans to detect **candidate genes** (Ioannidis *et al.* 2009, Oetting *et al.* 2018)

Validation phase

An approach that relies on key assumptions in need of validation:

Capblancq *et al.* 2020, Rellstab *et al.* 2021

- The alleles involved in local adaptation have been correctly identified.

→ a common concern to studies using GWAS, GEA methods or genome scans to detect **candidate genes** (Ioannidis *et al.* 2009, Oetting *et al.* 2018)

- The estimated gene-environment associations reflect the adaptive optima between the available standing genetic variation and the current climates.

→ Implicit in all GEA methods whereas most natural populations are likely not at equilibrium (Hereford 2009, Leimu and Fischer 2008, Browne *et al.* 2019).

→ Numerous reasons (Brady *et al.* 2019), e.g. gene flow from other populations, insufficient genetic variation, genetic drift, rapid environmental shifts, dispersal lag.

Validation phase

An approach that relies on key assumptions in need of validation:

Capblancq *et al.* 2020, Rellstab *et al.* 2021

- The alleles involved in local adaptation have been correctly identified.

→ a common concern to studies using GWAS, GEA methods or genome scans to detect **candidate genes** (Ioannidis *et al.* 2009, Oetting *et al.* 2018)

- The estimated gene-environment associations reflect the adaptive optima between the available standing genetic variation and the current climates.

→ Implicit in all GEA methods whereas most natural populations are likely not at equilibrium (Hereford 2009, Leimu and Fischer 2008, Browne *et al.* 2019).

→ Numerous reasons (Brady *et al.* 2019), e.g. gene flow from other populations, insufficient genetic variation, genetic drift, rapid environmental shifts, dispersal lag.

- The **magnitude** of the genomic offset **is proportional** to the expected decrease in **relative fitness**.

Validation phase

An approach that relies on key assumptions in need of validation:

Capblancq *et al.* 2020, Rellstab *et al.* 2021

- The alleles involved in local adaptation have been correctly identified.

→ a common concern to studies using GWAS, GEA methods or genome scans to detect **candidate genes** (Ioannidis *et al.* 2009, Oetting *et al.* 2018)

- The estimated gene-environment associations reflect the adaptive optima between the available standing genetic variation and the current climates.

→ Implicit in all GEA methods whereas most natural populations are likely not at equilibrium (Hereford 2009, Leimu and Fischer 2008, Browne *et al.* 2019).

→ Numerous reasons (Brady *et al.* 2019), e.g. gene flow from other populations, insufficient genetic variation, genetic drift, rapid environmental shifts, dispersal lag.

- The **magnitude** of the genomic offset **is proportional** to the expected decrease in **relative fitness**.

- The current **spatial** associations between allele frequencies and climatic conditions can be used to infer the potential adaption lags induced by a change in climate (i.e. disruption of genotype–climate associations) across **time**.

→ a inherent assumption of **space-for-time substitution approaches**.

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

From statistical/correlational approaches

less costly but with a lower degree of evidence

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

From statistical/correlational approaches

less costly but with a lower degree of evidence

- verifying that the candidate genes identified with GWAS, GEA methods or genome scans can be detected in **independent samples**.

e.g. Bay *et al.* 2018.

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

From statistical/correlational approaches

less costly but with a lower degree of evidence

- verifying that the candidate genes identified with GWAS, GEA methods or genome scans can be detected in **independent samples**.

e.g. Bay *et al.* 2018.

- verifying that the **climatic variables** with the strongest associations in GEA methods are also **associated with fitness** in experiments (e.g. common gardens).

e.g. Aitken *et al.* 2008, De Kort *et al.* 2014, Mahony *et al.* 2020, Rehfeldt *et al.* 2002, Steane *et al.* 2014.

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

From statistical/correlational approaches

less costly but with a lower degree of evidence

- verifying that the candidate genes identified with GWAS, GEA methods or genome scans can be detected in **independent samples**.

e.g. Bay *et al.* 2018.

- verifying that the **climatic variables** with the strongest associations in GEA methods are also **associated with fitness** in experiments (e.g. common gardens).

e.g. Aitken *et al.* 2008, De Kort *et al.* 2014, Mahony *et al.* 2020, Rehfeldt *et al.* 2002, Steane *et al.* 2014.

From more direct experimental support

highly costly but with a **higher degree of evidence**

Validation phase

Have the alleles involved in local adaptation been correctly identified?

Strengthen the robustness of candidate gene identification by accumulating independent evidence:

Capblancq *et al.* 2020

From statistical/correlational approaches

less costly but with a lower degree of evidence

- verifying that the candidate genes identified with GWAS, GEA methods or genome scans can be detected in **independent samples**.

e.g. Bay *et al.* 2018.

- verifying that the **climatic variables** with the strongest associations in GEA methods are also **associated with fitness** in experiments (e.g. common gardens).

e.g. Aitken *et al.* 2008, De Kort *et al.* 2014, Mahony *et al.* 2020, Rehfeldt *et al.* 2002, Steane *et al.* 2014.

From more direct experimental support

highly costly but with a **higher degree of evidence**

- verifying the **phenotypic effects** of the candidate genes under **controlled conditions** or via **functional validation**, e.g. through genetic transformation into specific mutant backgrounds, gene editing of putatively causal sites, or by knocking back gene expression.

e.g. Exposito-Alonso et al. 2019, Yoder et al. 2014, Curtin et al. 2017, Monroe et al. 2018, Rohde et al. 2018

Validation phase

Are populations at their adaptive equilibrium?

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

But for **long-lived** species such as forest trees, populations may often **not be** at their **phenotypic optimum**.

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

But for **long-lived** species such as forest trees, populations may often **not be** at their **phenotypic optimum**.

- Boreal, temperate and Mediterranean populations at the **leading edge** may **grow** and **survive more** in **warmer** conditions, at least in the short term.

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

But for **long-lived** species such as forest trees, populations may often **not be** at their **phenotypic optimum**.

• Boreal, temperate and Mediterranean populations at the **leading edge** may **grow** and **survive more** in **warmer** conditions, at least in the short term.

→ may be caused by **dispersal constraints** and **insufficient time** for populations to reach equilibrium since the last ice age (i.e. post-glacial expansion still ongoing).

e.g. Fréjaville *et al.* 2020, García-Valdés *et al.* 2013, Pedlar and McKenney 2017, Rehfeldt *et al.* 2018, 2003, 2002, Savolainen *et al.* 2007

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

But for **long-lived** species such as forest trees, populations may often **not be** at their **phenotypic optimum**.

• Boreal, temperate and Mediterranean populations at the **leading edge** may **grow** and **survive more** in **warmer** conditions, at least in the short term.

→ may be caused by **dispersal constraints** and **insufficient time** for populations to reach equilibrium since the last ice age (i.e. post-glacial expansion still ongoing).

e.g. Fréjaville *et al.* 2020, García-Valdés *et al.* 2013, Pedlar and McKenney 2017, Rehfeldt *et al.* 2018, 2003, 2002, Savolainen *et al.* 2007

• **Adaptation lags** to temperature observed in valley oak, with fastest growth rates occurring at cooler temperatures than populations are currently experiencing.

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with **reciprocal common garden experiments**.

But for **long-lived** species such as forest trees, populations may often **not be** at their **phenotypic optimum**.

• Boreal, temperate and Mediterranean populations at the **leading edge** may **grow** and **survive more** in **warmer** conditions, at least in the short term.

→ may be caused by **dispersal constraints** and **insufficient time** for populations to reach equilibrium since the last ice age (i.e. post-glacial expansion still ongoing).

e.g. Fréjaville *et al.* 2020, García-Valdés *et al.* 2013, Pedlar and McKenney 2017, Rehfeldt *et al.* 2018, 2003, 2002, Savolainen *et al.* 2007

• **Adaptation lags** to temperature observed in valley oak, with fastest growth rates occurring at cooler temperatures than populations are currently experiencing.

→ may be caused by the **rapid rate** of **current climate change** to which populations **are not adapting fast enough**.

Browne *et al.* 2019

Validation phase

Are populations at their adaptive equilibrium?

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with the genomic offset approach itself!

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with the genomic offset approach itself!

Borell *et al.* 2019

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with the genomic offset approach itself!

Borell *et al.* 2019

The **risk of non-adapteness** (RONA; equivalent to the genomic offset) was calculated for **specific populations**.

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with the genomic offset approach itself!

Borell *et al.* 2019

The **risk of non-adapteness** (RONA; equivalent to the genomic offset) was calculated for **specific populations**.

$\rightarrow$ measures the **mismatch** between their **current** genomic composition and the **required** genomic composition to be adapted to their current environment (based on the gene-environment associations measured in the other populations).

Validation phase

Are populations at their adaptive equilibrium?

Can be tested with the genomic offset approach itself!

Borell *et al.* 2019

The **risk of non-adapteness** (RONA; equivalent to the genomic offset) was calculated for **specific populations**.

$\rightarrow$ measures the **mismatch** between their **current** genomic composition and the **required** genomic composition to be adapted to their current environment (based on the gene-environment associations measured in the other populations).

$\rightarrow$ is proportional to the variation in allele frequency that would be required for the population to optimally adapt to its local environment.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Can be evaluated with:

- controlled experiments such as common gardens.
- observational data.
- *in silico*.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Can be evaluated with:

- controlled experiments such as common gardens.
- observational data.
- *in silico*.

Informs on model robustness in the face of:

- **demographic** and **stochastic processes** that might also affect the spatial variation in allele frequencies (e.g. expansion history, gene flow, genetic drift, mutation).
- **genetic architecture** of climate adaptation (polygenic trait architectures, G×E interactions, nonadditive genetic variance).

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

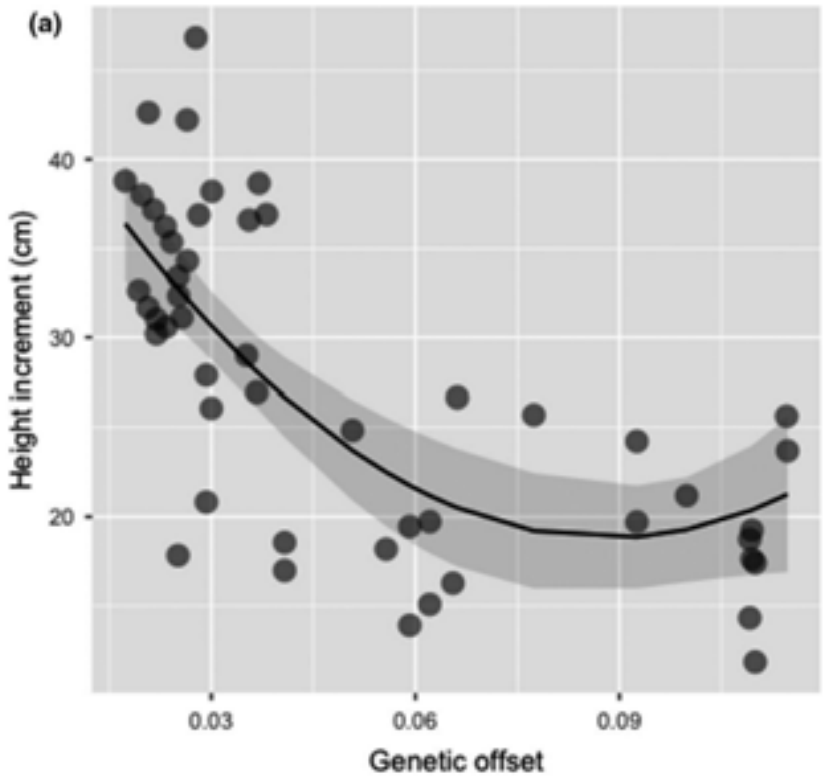
The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.



Fitzpatrick *et al.* 2021.

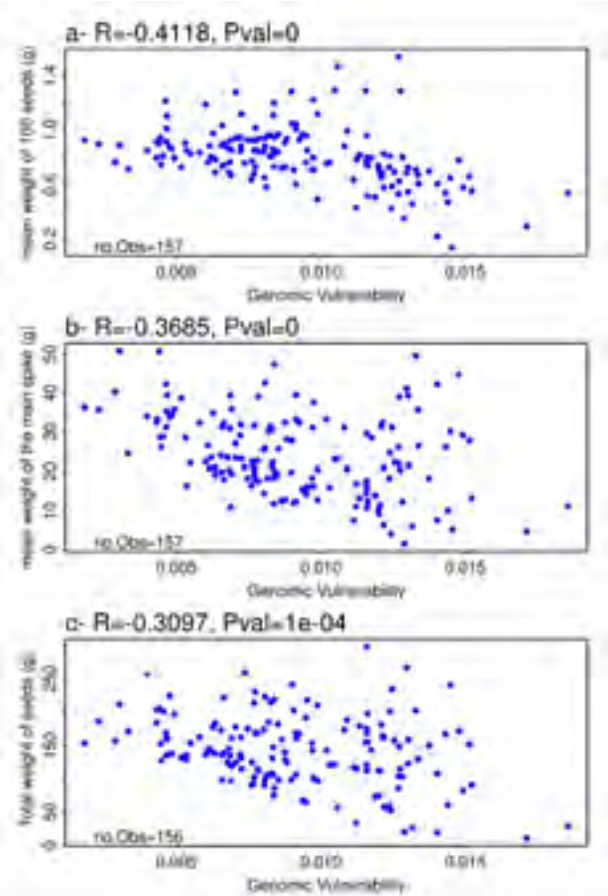
→ As expected, the predicted genomic offset was negatively associated with height growth increment measured in common gardens.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Validating genomic offset predictions in ***common gardens***

The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.



Rhoné *et al.* 2021.

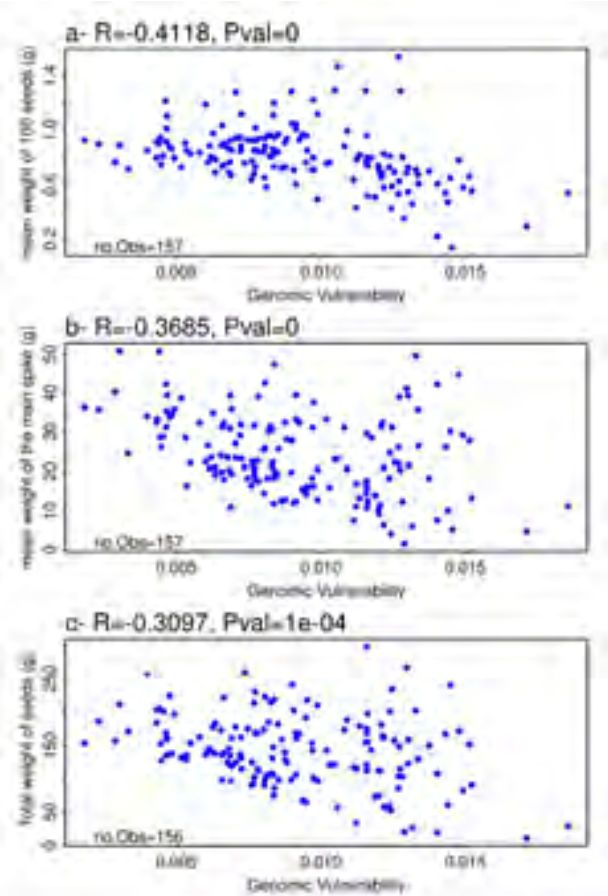
A study investigating the patterns of potential maladaptation in 173 pearl millet landraces of agronomic importance in West Africa.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.



Rhoné *et al.* 2021.

A study investigating the patterns of potential maladaptation in 173 pearl millet landraces of agronomic importance in West Africa.

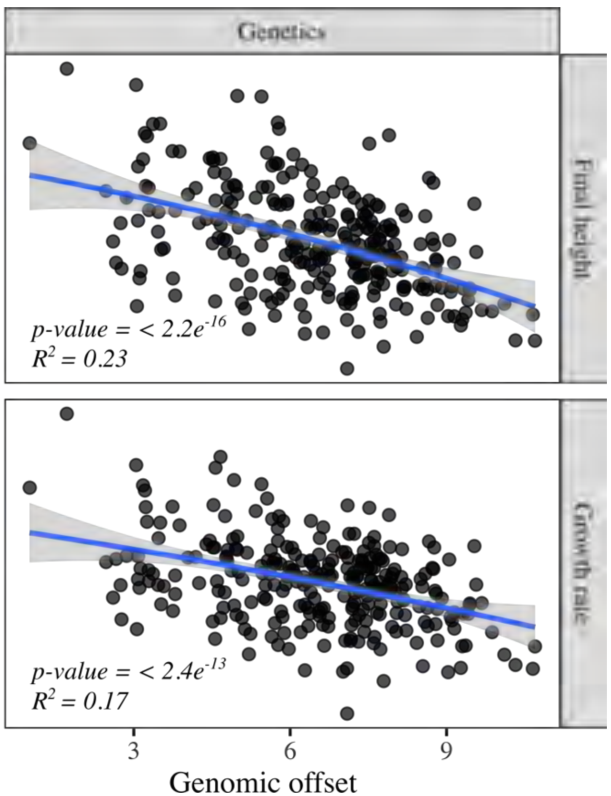
→ As expected, the predicted genomic offset is **negatively associated** with **yield-related traits** in a common garden experiment in Niger.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.



Capblancq and Forester 2021.

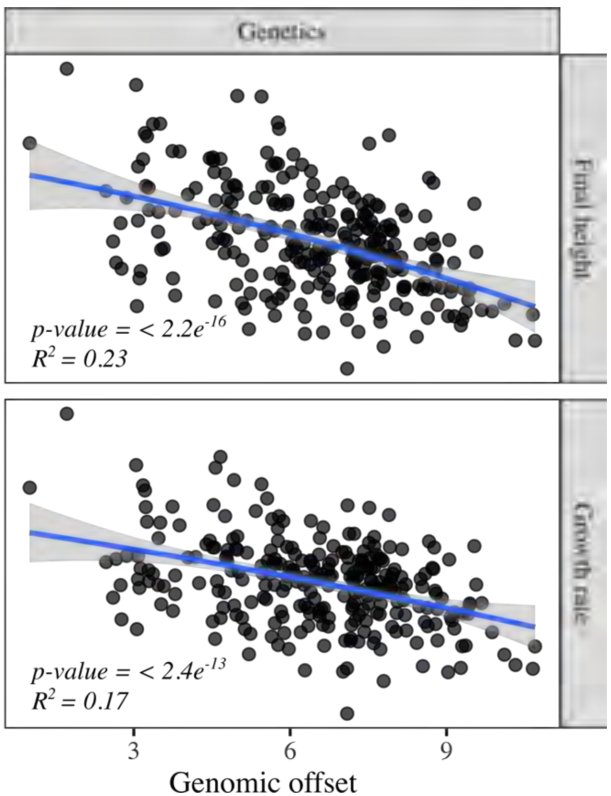
1,594 seedlings (from 281 provenances) of lodgepole pine planted in a common garden in Vancouver and genotyped for 29,941 SNPs (after filtering).

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **common gardens***

The turnover functions are used to predict the genomic offset of populations when transplanted from their environment of origin to the environment of the common garden.



Capblancq and Forester 2021.

1,594 seedlings (from 281 provenances) of lodgepole pine planted in a common garden in Vancouver and genotyped for 29,941 SNPs (after filtering).

→ Strong negative relationship between genomic offset and seedling fitness (i.e. height and growth rate) in the common garden.

Validation phase

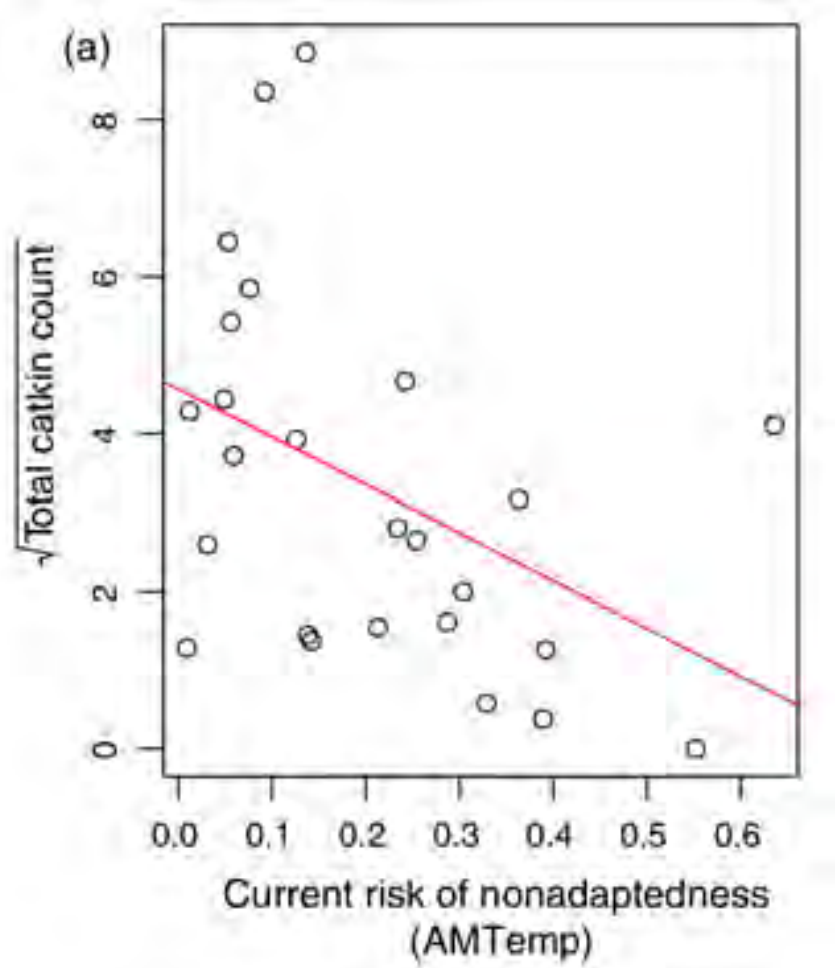
Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***



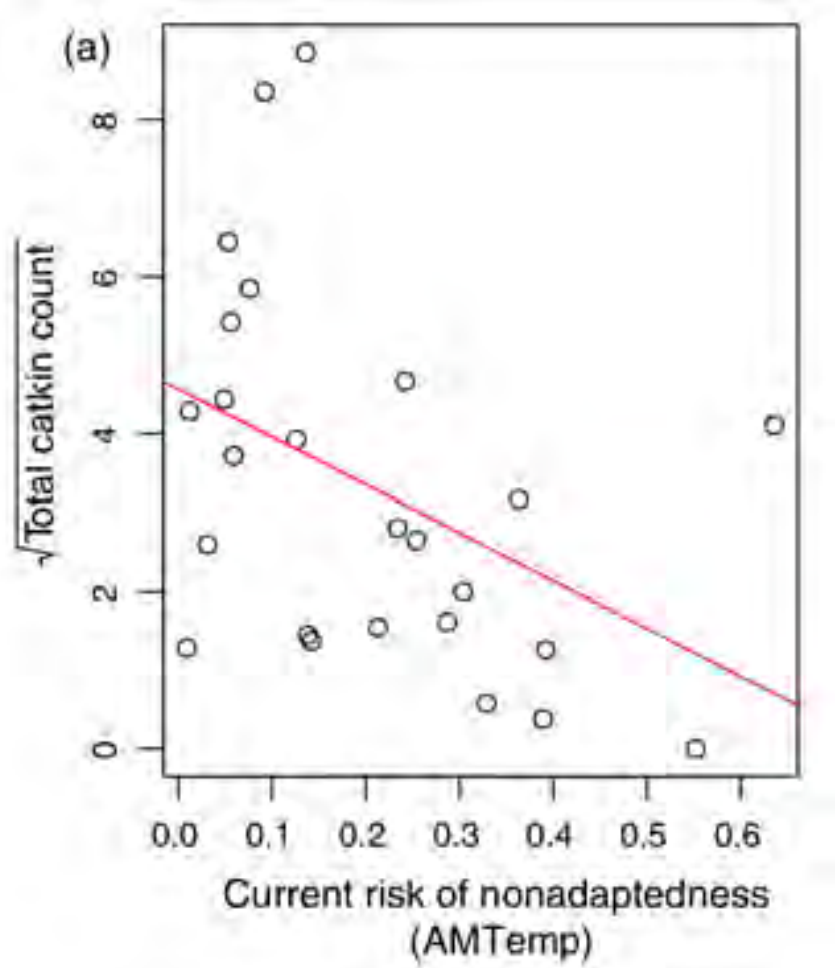
Borell *et al.* 2019.

Evaluating maladaptation in Scottish populations of the montane plant dwarf birch.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***



Borell *et al.* 2019.

Evaluating maladaptation in Scottish populations of the montane plant dwarf birch.

→ Populations predicted to be currently maladapted (i.e. with a mismatch between their allele frequencies at loci associated with annual mean temperature and their local environment) had reduced catkin production.

Comment: the possibility that reduced reproductive output could be an adaptive response to a poorer environment cannot be excluded, but this seems unlikely given the short timescales involving a handful of generations.

Validation phase

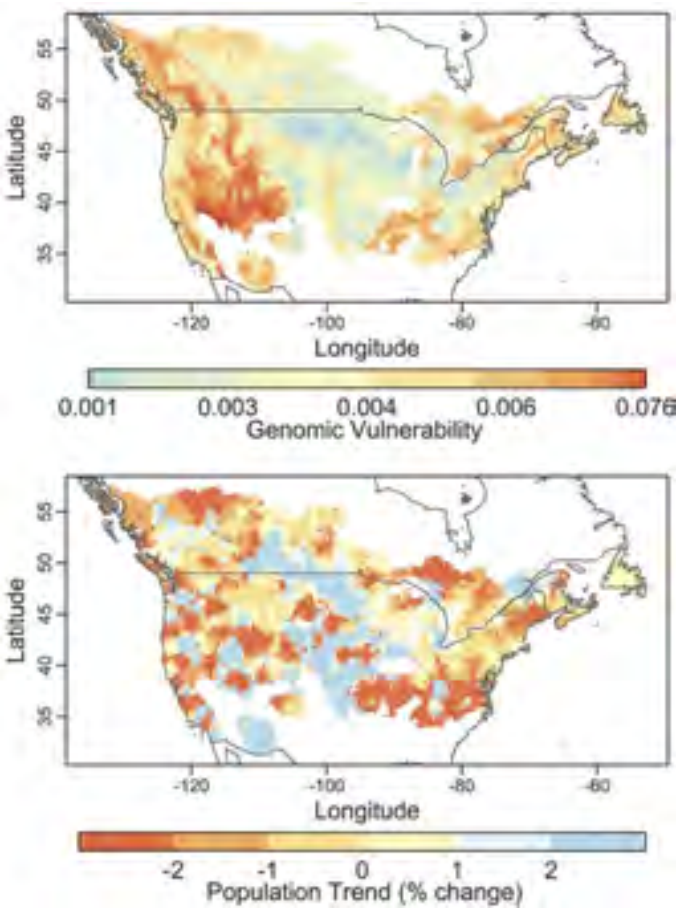
Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***



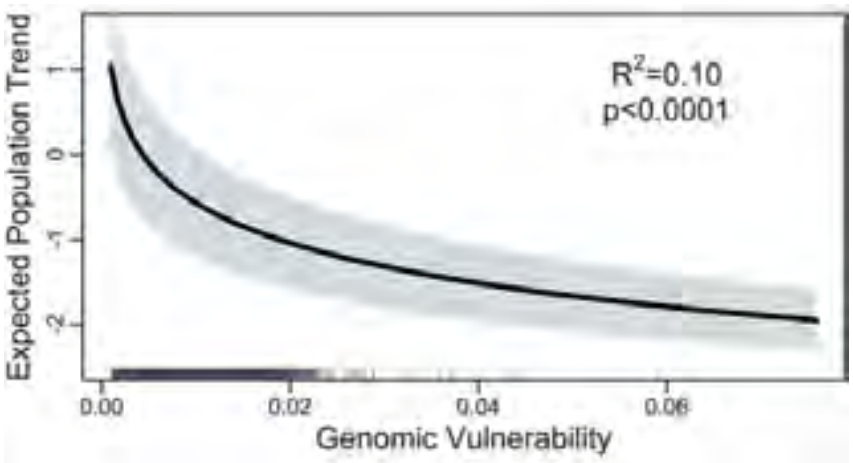
Bay *et al*. 2018.

Compared **genomic offset predictions** in a migratory bird species (yellow warbler) to spatial extrapolations of **population declines** over the past half century.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Validating genomic offset predictions in ***observational studies***



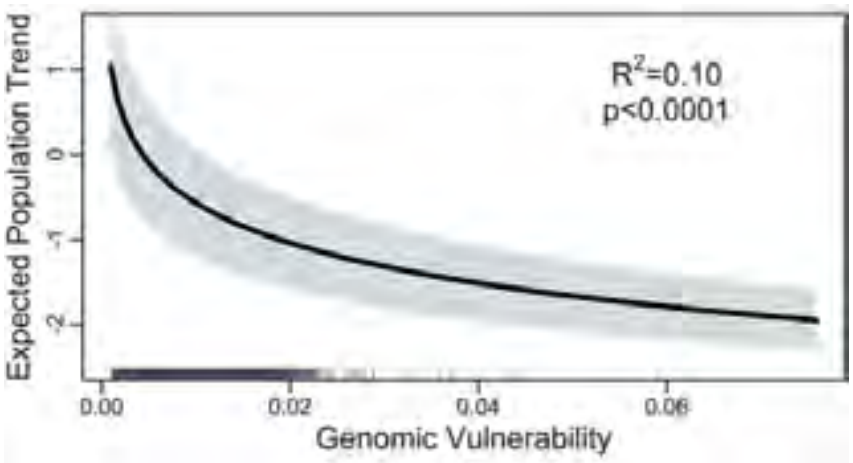
Bay *et al*. 2018.

→ Populations with highest genomic offset have experienced the largest population declines.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

Validating genomic offset predictions in ***observational studies***



Bay *et al.* 2018.

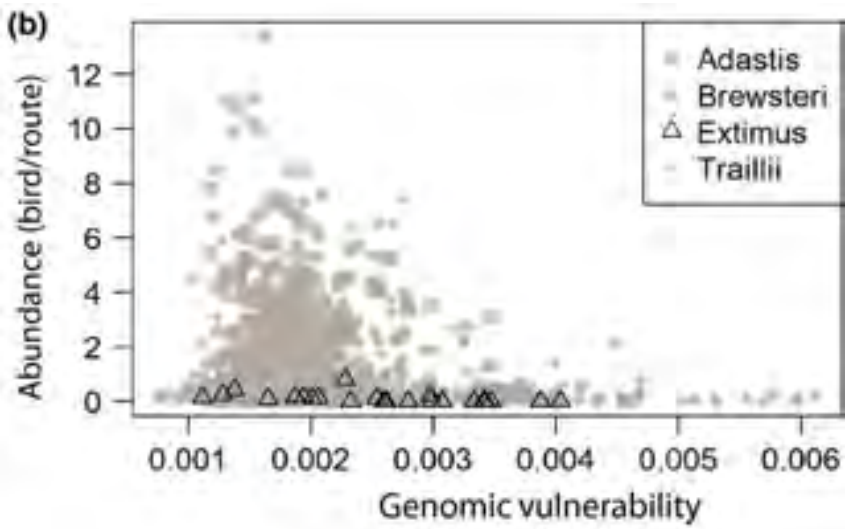
→ Populations with highest genomic offset have experienced the largest population declines.

Authors' interpretation: populations already in decline are most vulnerable to future climate change and a mismatch between genomic variation and climate may have already negatively affected populations.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***



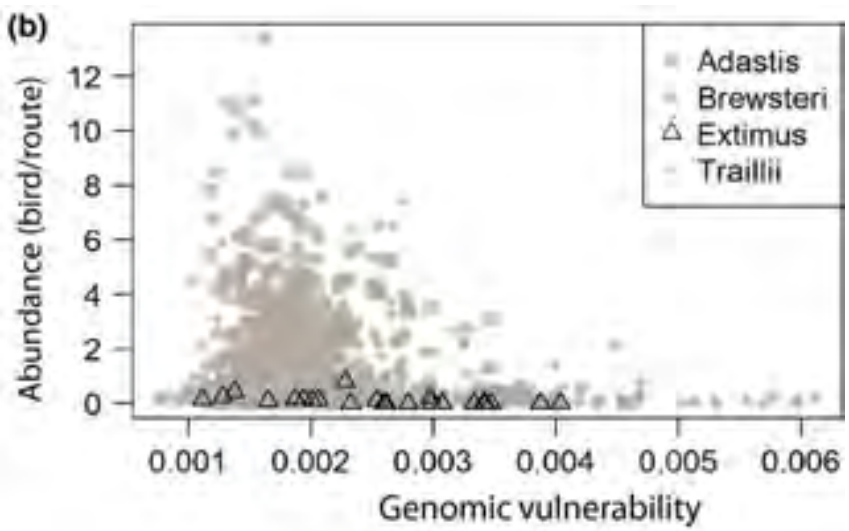
Ruegg *et al*. 2018.

Applying the genomic offset approach to 24 populations of willow flycatcher (an endangered desert passerine) using 105,000 SNPs from 219 individuals.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***



Ruegg *et al*. 2018.

Applying the genomic offset approach to 24 populations of willow flycatcher (an endangered desert passerine) using 105,000 SNPs from 219 individuals.

→ Genomic offset predictions were negatively associated with abundance, which was mostly due to the especially low abundance in sites with the highest genomic offset.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.

- the **historical** and **future climate shifts** were **assumed to be correlated** whereas it is probably not the case.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.
- the **historical** and **future climate shifts** were **assumed to be correlated** whereas it is probably not the case.
- the relationship between **historical climate** and **population trends** was not evaluated, whereas it appears as a necessary first step for predicting **future climate-driven population dynamics** based on historical associations.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.
- the **historical** and **future climate shifts** were **assumed to be correlated** whereas it is probably not the case.
- the relationship between **historical climate** and **population trends** was not evaluated, whereas it appears as a necessary first step for predicting **future climate-driven population dynamics** based on historical associations.
- the genomic offset predictions were based on **all SNPs** with $\mathcal{R}^2 > 0$ in the GF analysis ($\sim$8,000 SNPs), which therefore mostly capture **neutral genetic variation**.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.
- the **historical** and **future climate shifts** were **assumed to be correlated** whereas it is probably not the case.
- the relationship between **historical climate** and **population trends** was not evaluated, whereas it appears as a necessary first step for predicting **future climate-driven population dynamics** based on historical associations.
- the genomic offset predictions were based on **all SNPs** with $\mathcal{R}^2 > 0$ in the GF analysis ($\sim$8,000 SNPs), which therefore mostly capture **neutral genetic variation**.
- the **robustness** of the fitted relationship between population trends and genomic vulnerability may be questioned as **extrapolations** are treated as **observations**, which artificially inflates the sample size, could exacerbate spatial non-independence in the data and ultimately **increases the likelihood of finding a significant relationship**.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions in **observational studies***

However, according to **Fitzpatrick *et al.* 2018**, the statistical analyses and interpretation of the genomic offset proposed in Bay et al. 2018 have severe problems:

- this approach **cannot be considered** as a validation of the genomic offset concept (i.e. a metric reflecting decreased population fitness in face of climate change) because the genomic offset was estimated as a function of **current** climates and not as a function of **past** climates.
- the **historical** and **future climate shifts** were **assumed to be correlated** whereas it is probably not the case.
- the relationship between **historical climate** and **population trends** was not evaluated, whereas it appears as a necessary first step for predicting **future climate-driven population dynamics** based on historical associations.
- the genomic offset predictions were based on **all SNPs** with $\mathcal{R}^2 > 0$ in the GF analysis ($\sim$8,000 SNPs), which therefore mostly capture **neutral genetic variation**.
- the **robustness** of the fitted relationship between population trends and genomic vulnerability may be questioned as **extrapolations** are treated as **observations**, which artificially inflates the sample size, could exacerbate spatial non-independence in the data and ultimately **increases the likelihood of finding a significant relationship**.

$\rightarrow$ **Initiated a debate** on the conditions and assumptions underlying the genomic offset approach for reliable estimates of vulnerability to climate change.

Validation phase

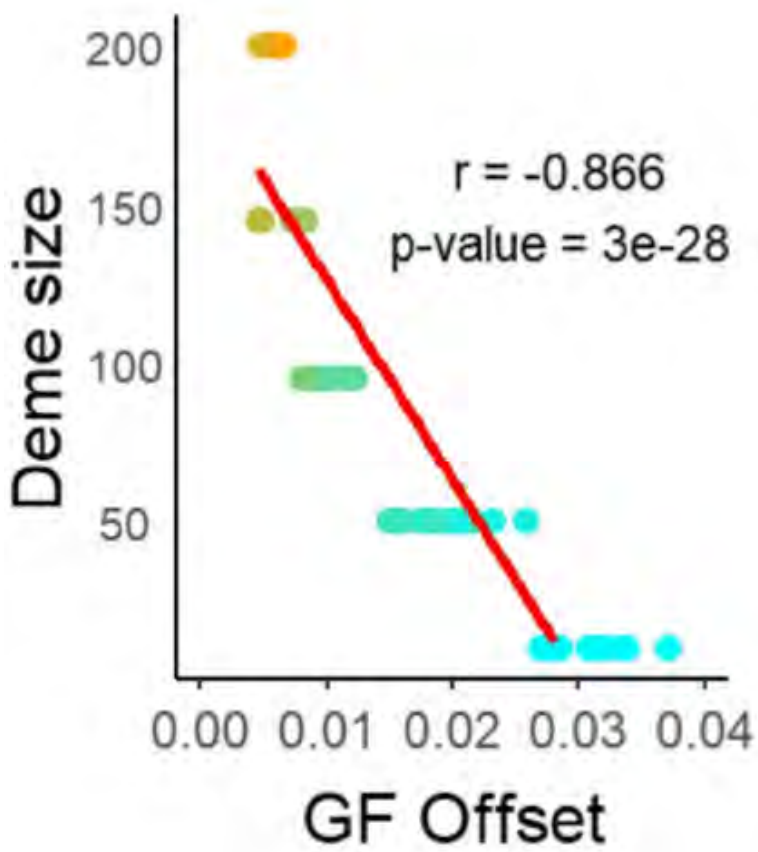
Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***



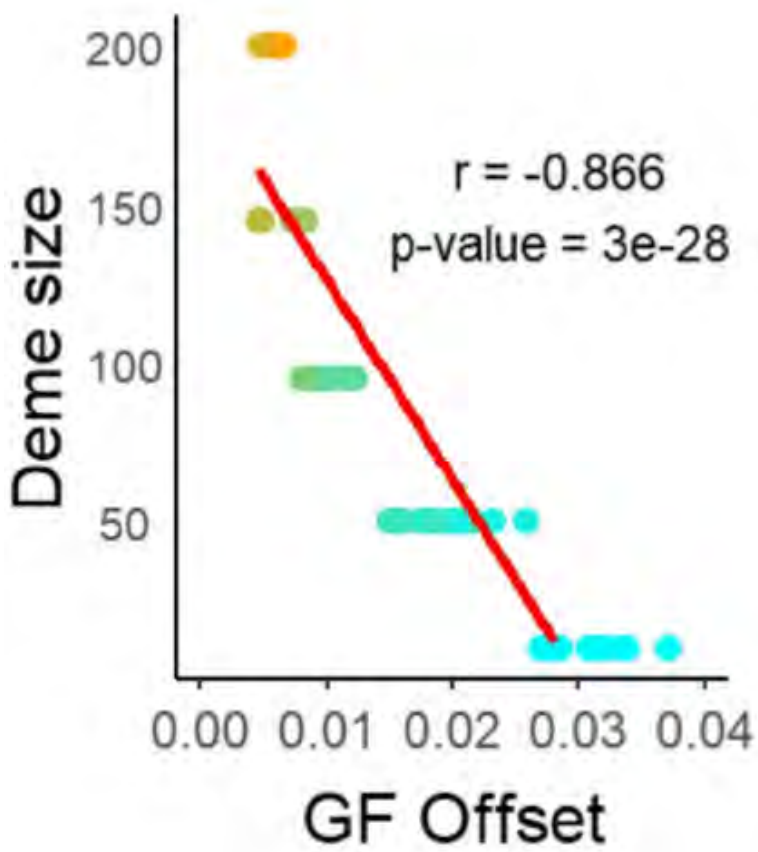
Láruson *et al.* 2021.

→ Genomic offset at neutral loci **negatively associated** with **population size** varying along an environmental gradient.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***



Láruson *et al.* 2021.

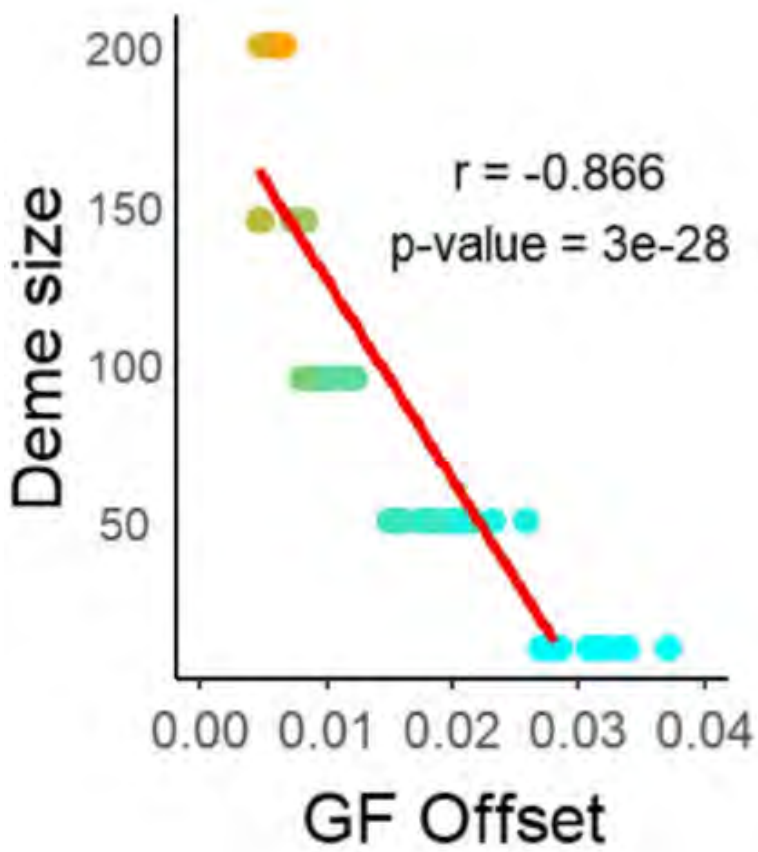
→ Genomic offset at neutral loci **negatively associated** with **population size** varying along an environmental gradient.

Likely due to **increased genetic drift** in **small populations**.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***



Láruson *et al.* 2021.

→ Genomic offset at neutral loci **negatively associated** with **population size** varying along an environmental gradient.

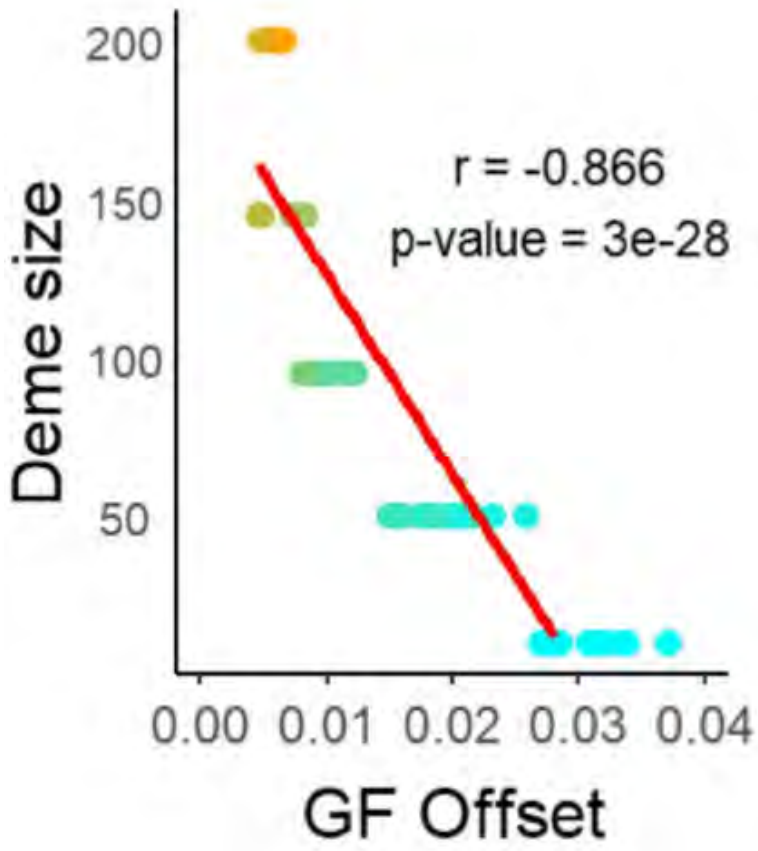
Likely due to **increased genetic drift** in **small populations**.

Implications: The negative associations between genomic offset and population size found in some studies (Bay et al. 2018, Ruegg et al. 2018) may originate from **neutral genetic drift** and not natural selection as assumed.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***



Láruson *et al.* 2021.

→ Genomic offset at neutral loci **negatively associated** with **population size** varying along an environmental gradient.

Likely due to **increased genetic drift** in **small populations**.

Implications: The negative associations between genomic offset and population size found in some studies (Bay et al. 2018, Ruegg et al. 2018) may originate from **neutral genetic drift** and not natural selection as assumed.

→ **Empirical studies** identifying an association between population size and genomic offset predictions **should not be considered** as validation of the genomic offset concept.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Láruson *et al.* 2021.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Láruson *et al.* 2021.

Evaluating the assumption that genomic offset (obtained from Gradient Forest) is inversely related to fitness by conducting *in silico* common garden experiments under monogenic and polygenic architectures.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Láruson *et al.* 2021.

Evaluating the assumption that genomic offset (obtained from Gradient Forest) is inversely related to fitness by conducting *in silico* common garden experiments under monogenic and polygenic architectures.

→ The genomic offset predicted well the relative fitness in common gardens under both single locus and polygenic architectures, and regardless of whether or not nonadapted loci and noncausal environments were included in the analysis.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Láruson *et al.* 2021.

Evaluating the assumption that genomic offset (obtained from Gradient Forest) is inversely related to fitness by conducting *in silico* common garden experiments under monogenic and polygenic architectures.

→ The genomic offset predicted well the relative fitness in common gardens under both single locus and polygenic architectures, and regardless of whether or not nonadapted loci and noncausal environments were included in the analysis.

However, **neutral demography** (population sizes and migration rates), **genomic architecture** and the nature of the adaptive environment can confound the relationship between genomic offset and fitness.

Validation phase

Do populations with **higher predicted genomic offset** show **lower relative fitness**?

*Validating genomic offset predictions with **simulations***

Láruson *et al.* 2021.

Evaluating the assumption that genomic offset (obtained from Gradient Forest) is inversely related to fitness by conducting *in silico* common garden experiments under monogenic and polygenic architectures.

→ The genomic offset predicted well the relative fitness in common gardens under both single locus and polygenic architectures, and regardless of whether or not nonadapted loci and noncausal environments were included in the analysis.

However, **neutral demography** (population sizes and migration rates), **genomic architecture** and the nature of the adaptive environment can confound the relationship between genomic offset and fitness.

Genomic offset predictions were also sensitive to **sampling schemes**.

Validation phase

Is the space-for-time substitution assumption valid?

- A first major source of uncertainty comes from how **current adaptive genetic variation** may recombine into **novel genotypic combinations**, in particular under **novel climates** (i.e. novel combinations of climatic variables).

Validation phase

Is the space-for-time substitution assumption valid?

- A first major source of uncertainty comes from how **current adaptive genetic variation** may recombine into **novel genotypic combinations**, in particular under **novel climates** (i.e. novel combinations of climatic variables).

Space-for-time substitution approaches are **correlative**, and therefore do not account for potential **plastic** or **adaptive responses** to climate change or other **evolutionary processes** shaping future patterns of **genetic variation** (e.g. gene flow).

Validation phase

Is the space-for-time substitution assumption valid?

- A first major source of uncertainty comes from how **current adaptive genetic variation** may recombine into **novel genotypic combinations**, in particular under **novel climates** (i.e. novel combinations of climatic variables).

Space-for-time substitution approaches are **correlative**, and therefore do not account for potential **plastic** or **adaptive responses** to climate change or other **evolutionary processes** shaping future patterns of **genetic variation** (e.g. gene flow).

→ How populations evolve over time can be investigated using temporal series such as 'evolve and re-sequence' experiments, biological archives, historical samples, herbarium specimens, different age cohorts of long-lived species (Rellstab *et al*. 2021).

Validation phase

Is the space-for-time substitution assumption valid?

• A first major source of uncertainty comes from how **current adaptive genetic variation** may recombine into **novel genotypic combinations**, in particular under **novel climates** (i.e. novel combinations of climatic variables).

Space-for-time substitution approaches are **correlative**, and therefore do not account for potential **plastic** or **adaptive responses** to climate change or other **evolutionary processes** shaping future patterns of **genetic variation** (e.g. gene flow).

→ How populations evolve over time can be investigated using temporal series such as 'evolve and re-sequence' experiments, biological archives, historical samples, herbarium specimens, different age cohorts of long-lived species (Rellstab *et al*. 2021).

• Another great deal of uncertainty around the future climate conditions that populations will have to face, as climate forecasts vary strongly among existing general circulation models (GCMs).

Validation phase

Is the space-for-time substitution assumption valid?

• A first major source of uncertainty comes from how **current adaptive genetic variation** may recombine into **novel genotypic combinations**, in particular under **novel climates** (i.e. novel combinations of climatic variables).

Space-for-time substitution approaches are **correlative**, and therefore do not account for potential **plastic** or **adaptive responses** to climate change or other **evolutionary processes** shaping future patterns of **genetic variation** (e.g. gene flow).

→ How populations evolve over time can be investigated using temporal series such as 'evolve and re-sequence' experiments, biological archives, historical samples, herbarium specimens, different age cohorts of long-lived species (Rellstab *et al*. 2021).

• Another great deal of uncertainty around the future climate conditions that populations will have to face, as climate forecasts vary strongly among existing general circulation models (GCMs).

→Therefore, **space-for-time substitution approaches** has to be used **with caution** and further work is needed to **integrate** these **uncertainties** into the **predictions**.

Validation phase

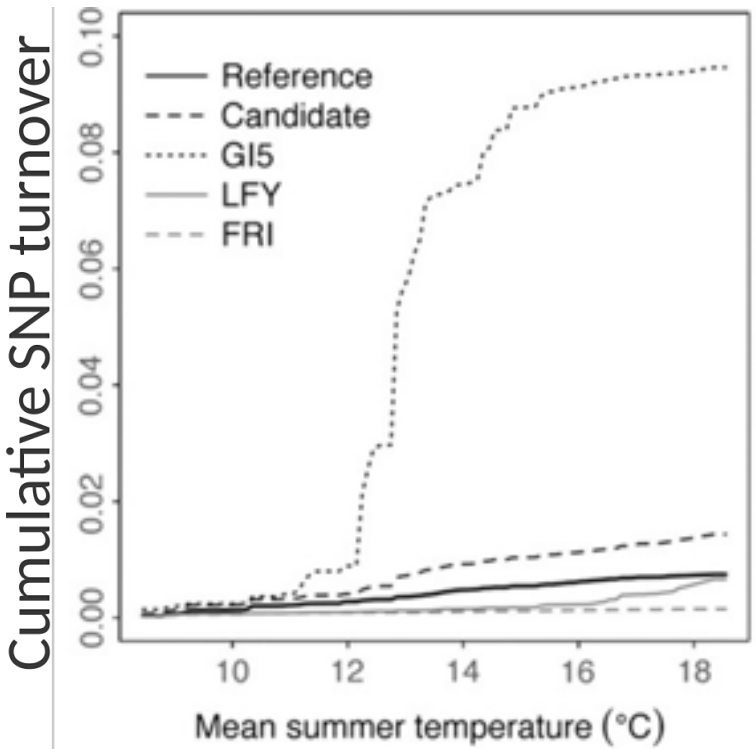
Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021



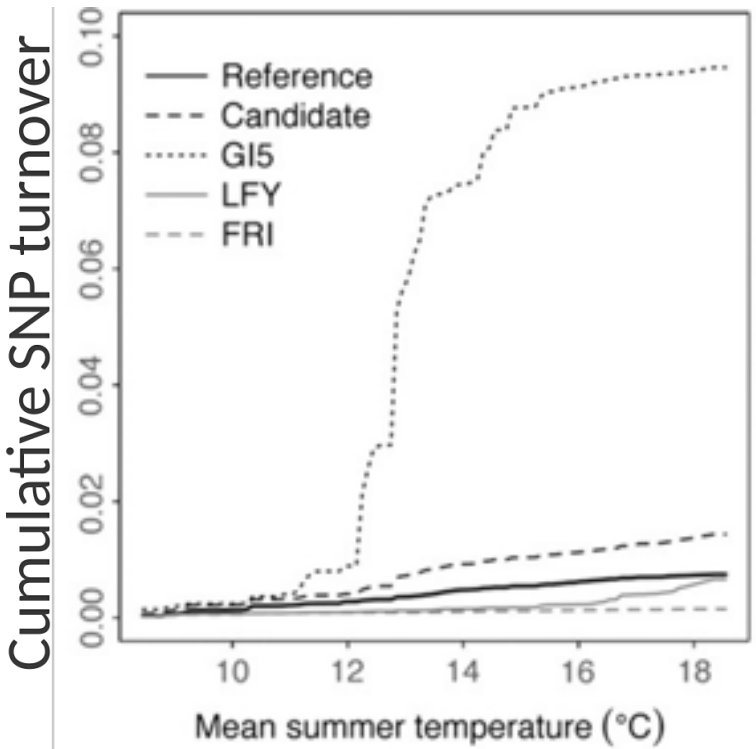
Proposed interpretations:

• 'The **maximum height** of each curve indicates the **total amount of turnover in allele frequencies** associated with that variable, and by extension, the **relative importance** of that variable in explaining changes in allele frequency, holding all other variables constant.'

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021



Proposed interpretations:

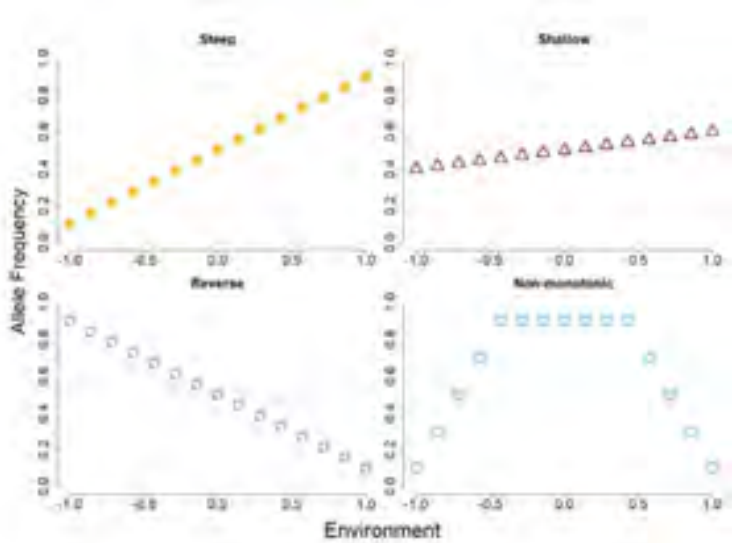
- 'The **maximum height** of each curve indicates the **total amount of turnover in allele frequencies** associated with that variable, and by extension, the **relative importance** of that variable in explaining changes in allele frequency, holding all other variables constant.'
- 'The **shape** of each function indicates how the **rate of change in allele frequencies** varies along the gradient.'

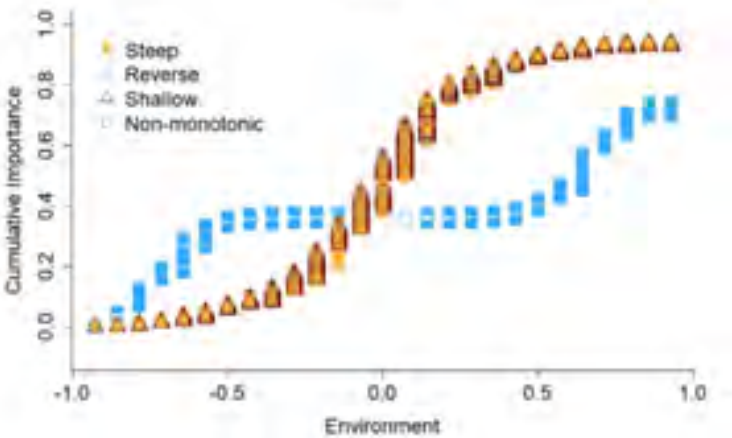
From Fitzpatrick and Keller (2015)

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021



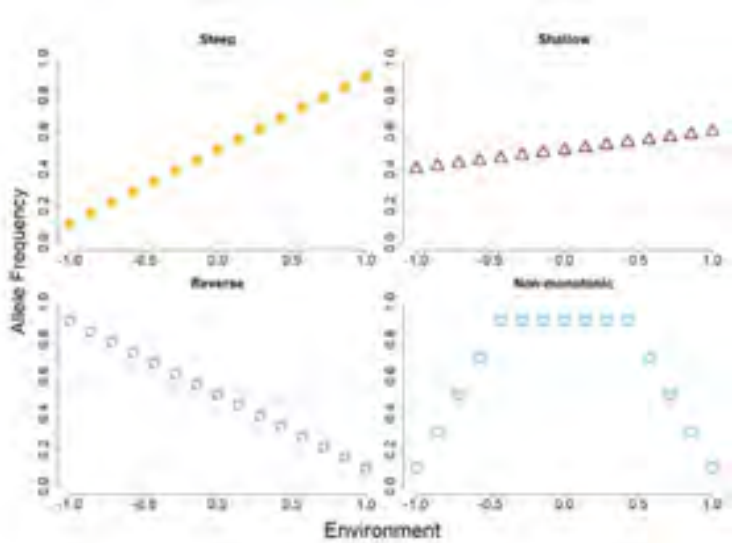


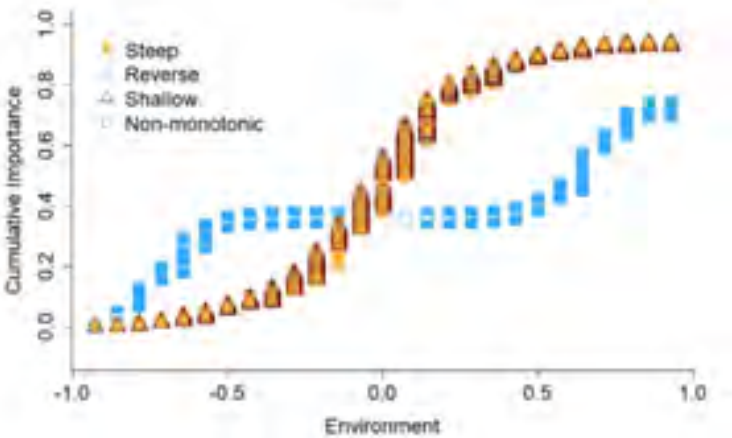
- For **linear** allele frequency clines, the **steepness** of the turnover function **does not reflect** the **rate** of allele frequency change.

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021



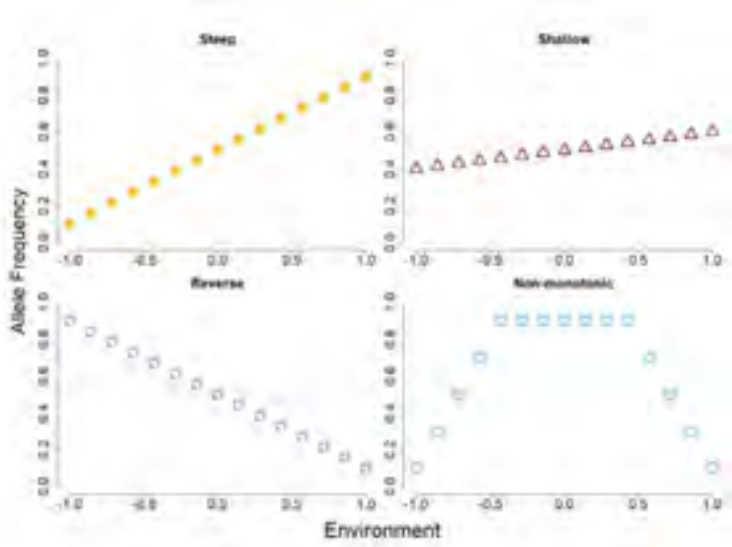


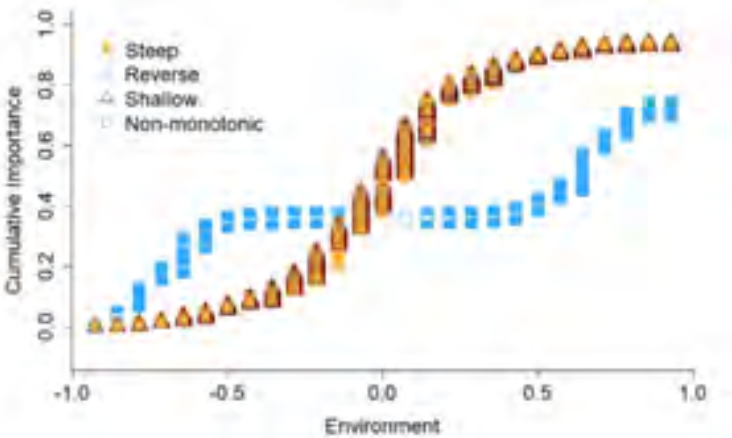
- For **linear** allele frequency clines, the **steepness** of the turnover function **does not reflect** the **rate** of allele frequency change.

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021





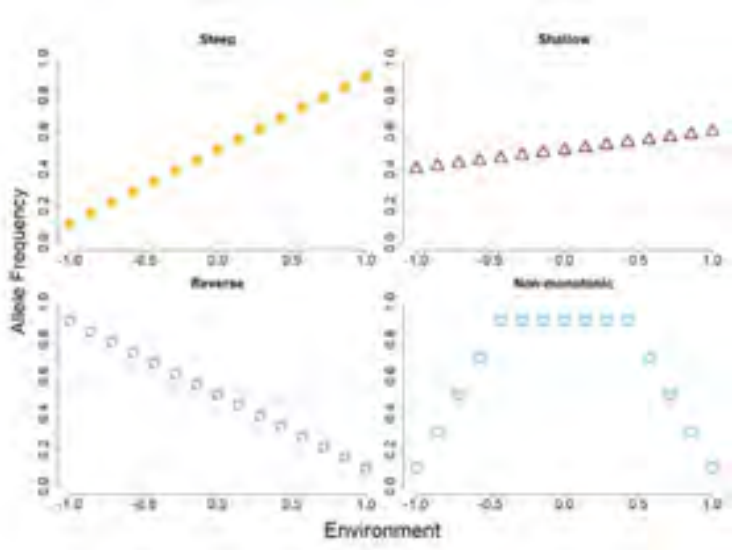
- For **linear** allele frequency clines, the **steepness** of the turnover function **does not reflect** the **rate** of allele frequency change.

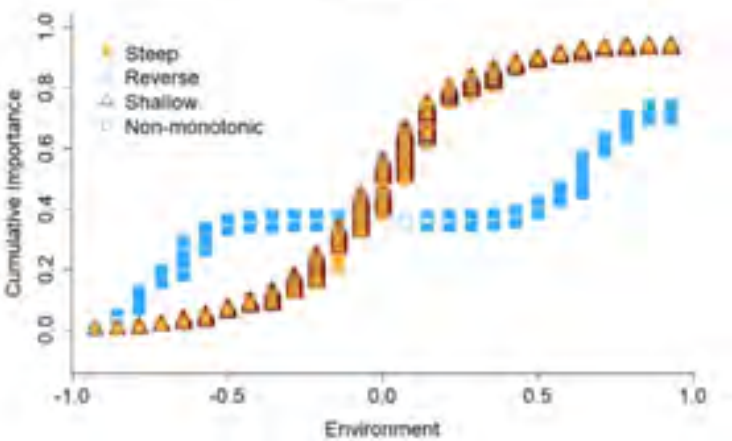
- The **maximum height** of each curve **does not indicate** the **total amount** of turnover in allele frequency along the gradient (i.e. the absolute difference in allele frequency along the gradient).

Validation phase

Evaluating some **interpretations** of the turnover functions ***in silico***.

Láruson *et al.* 2021





- For **linear** allele frequency clines, the **steepness** of the turnover function **does not reflect** the **rate** of allele frequency change.

- The **maximum height** of each curve **does not indicate** the **total amount** of turnover in allele frequency along the gradient (i.e. the absolute difference in allele frequency along the gradient).

Comment: In the nonmonotonic case, although the allele frequencies are the same at the environmental extremes (-1 and 1), a population would be predicted to have a nonzero genomic offset for an environmental shift from -1 to 1 for that locus.

Validation phase

More **empirical** and **simulation** studies are clearly needed to determine:

- the validity of the genomic offset predictions
- the limits of the interpretation of the genomic offset approach, i.e. in which cases and to what extent the genomic offset reflects future maladaptation.

.

Introduction: from species-level to genome-informed modelling

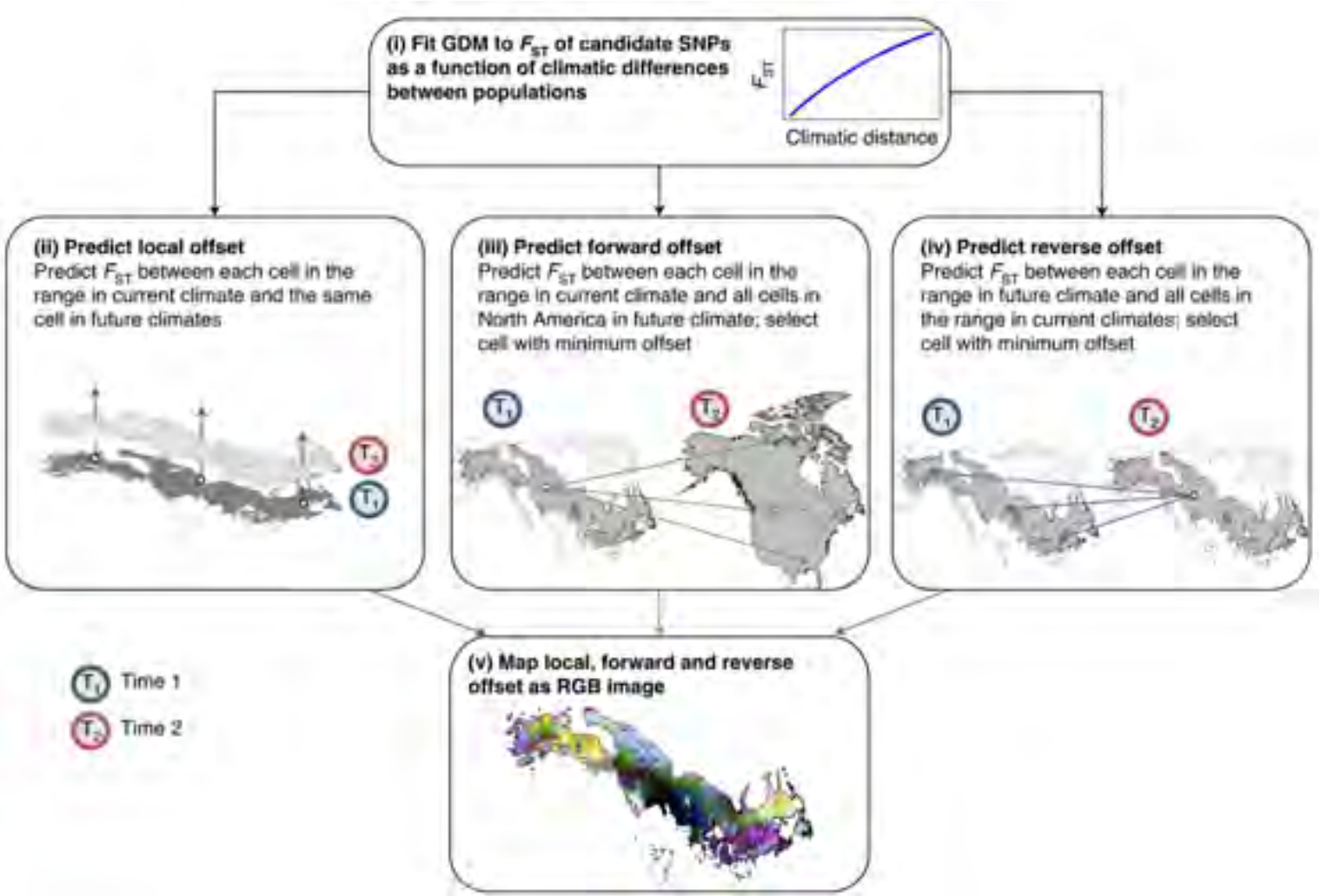
Predicting short-term performance with trait-based approaches
- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change
- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Extending the genomic offset framework

Incorporating migration distances and genotype-climate novelty

Gougherty *et al.* 2021.



Extending the genomic offset framework

Incorporating migration distances and genotype-climate novelty

Gougherty *et al.* 2021.

Local offset: may be the most relevant metric of vulnerability for **long-lived sessile organisms** with long generation times.

Forward offset: assumes that populations have **unlimited migration ability**.

$\rightarrow$ The current genotype-climate associations of populations with high forward offset may disappear from the landscape.

Reverse offset: captures the possibility that some populations in the current range may be preadatpated to the future climates at a particular location.

$\rightarrow$ Locations with high reverse offset may correspond to areas to which no populations will be adapted in the future, and in which we might expect some changes along existing gradients of genotype-climate associations.

Introduction: from species-level to genome-informed modelling

Predicting short-term performance with trait-based approaches
- Methods and limits in predicting phenotypes based on genotypes
- Predicting phenotypes based on genotypes in natural populations
- Predicting phenotypes from gene-environment associations

Predicting the adaptive potential of populations under climate change
- Incorporating intraspecific climatic adaptations into ENMs
- The genomic offset framework
- Extending the genomic offset framework
- Potential applications

Potential applications

Spatially explicit predictions of potential maladaptation to climate can help highlight the relative vulnerability of populations to climate change, which may serve to guide conservation policies.

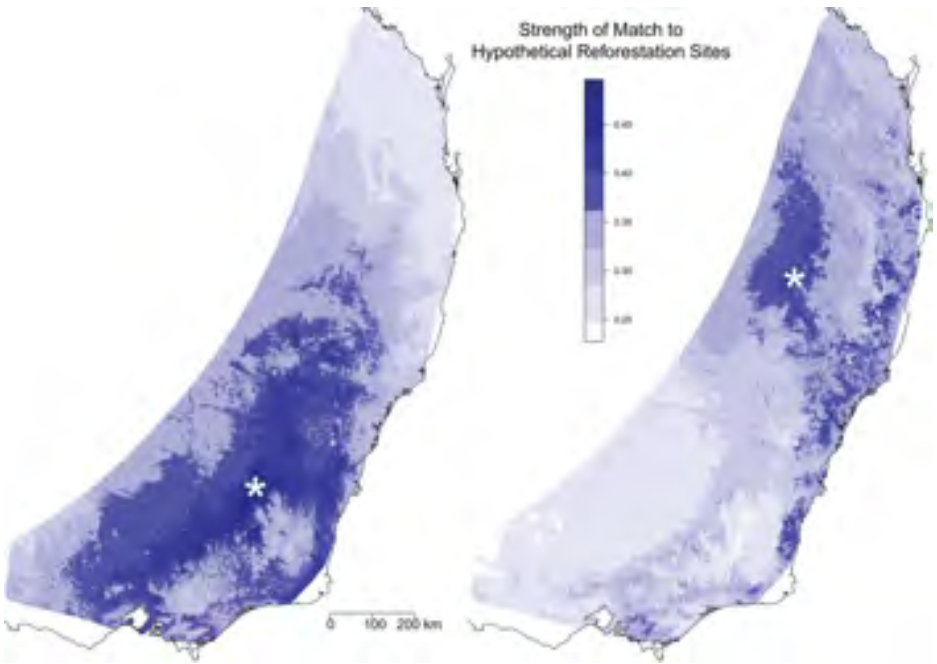
Potential applications

Providing recommendations for seed sourcing

Stupple *et al.* 2018.

Yellow box (*Eucalyptus melliodora*) is an emblematic Australian tree, essential to many native ecosystems, and **critically endangered**.

Can be difficult for reforestation practitioners to decide which seeds to plants in a given area as they have to select the ones with the highest chances of surviving now and in the future.



Genetic material from 275 *E. melliodora* trees at 36 different sites.

The genomic offset approach was used to match **hypothetical reforestation sites** (white asterisks) with the **seeds that will perform the best** in those sites under future climates.

Dark blue areas indicate **seed sourcing areas** predicted to **best match** future conditions at the hypothetical reforestation sites.