

Summer school on Software and Statistical Methods for Population Genomics

19-23 Sep 2022 Aussois (France)

Genomic offset environmental distance and loss of fitness

Clément Gain & Olivier François



Outline

- Evaluating associations between environment and genetic variation
- Theory of genomic offset: Geometry of the ecological niche
- A new measure with dual interpretation
- Results from a simulation study and a common garden experiment
- A short tutorial in R (LEA package)

Genomic offset (GO)

- Goal: Predict the mal-adaptation of (sedentary) organisms to abrupt ecological changes
- GO measures are based on genomic and environmental data



Evaluating the effect of environmental variables on genetic variation

- Demographic factors can confound the effect of natural selection on genetic variation
- Latent factor mixed models estimate confounding factors (**U**) and environmental effect sizes (**B**) simultaneously

$$\mathbf{Y} = \mathbf{X}_{\text{env}} \mathbf{B}^T + \mathbf{U}_{\text{latent}} \mathbf{V}^T + \mathbf{E}_{\text{noise}}$$

Y = centered allele frequency

V = factor loadings

(Caye et al. MBE 2019)



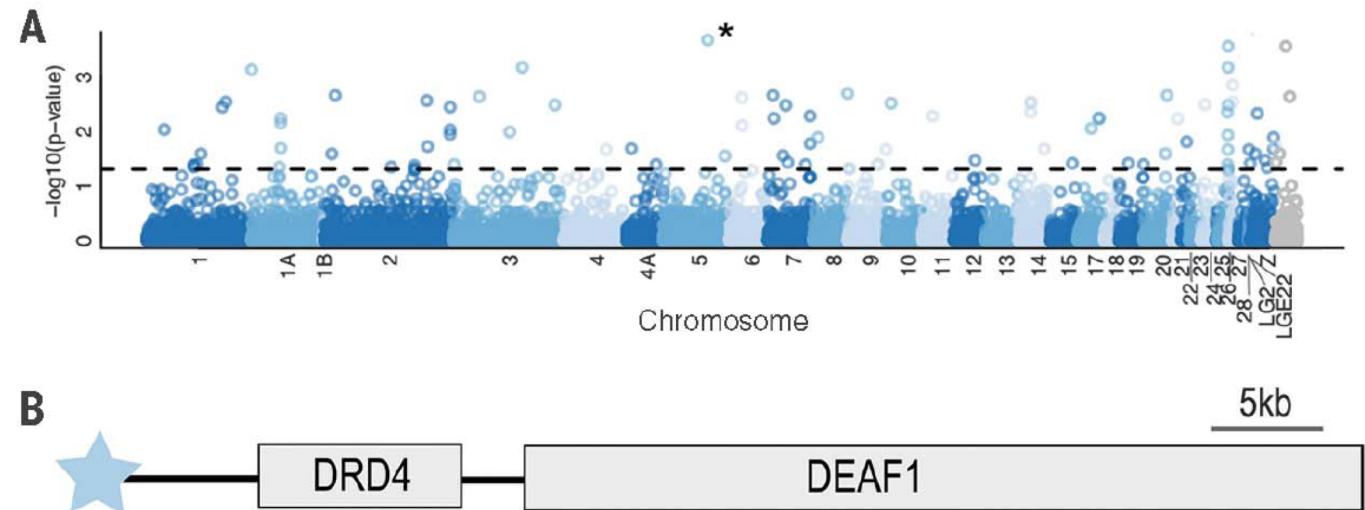
GEA study

RESEARCH

ECOLOGICAL GENOMICS

Genomic signals of selection predict climate-driven population declines in a migratory bird

Rachael A. Bay,^{1,2,*} Ryan J. Harrigan,¹ Vinh Le Underwood,¹ H. Lisle Gibbs,³
Thomas B. Smith,^{1,4} Kristen Ruegg^{1,5}



- LFMMs are most often used in hypothesis testing approaches that seek to identify significant levels of associations between genotype and environment.

Geometry of the ecological niche

- Grinnell (1917) and Hutchinson (1957) defined the ecological niche as the sum of the habitat conditions that allow organisms to survive and reproduce.
- Definition of GO as a squared distance in the ecological niche

$$G^2(\mathbf{x}, \mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*) \mathbb{E}[\mathbf{B}\mathbf{B}^T] (\mathbf{x} - \mathbf{x}^*)^T$$

$\mathbb{E}[\mathbf{B}\mathbf{B}^T]$ = covariance matrix of effect sizes estimated in the LFMM.

(Gain et al. online very soon)

Genetic gap: $G^2(\mathbf{x}, \mathbf{x}^*)$

G^2 solves some issues with GO approaches (Rellstab et al. 2021).

- models neutral population structure
- models correlated environmental effects
- determines the relative importance of environmental variables and of combinations of those variables
- accounts for the polygenic architecture of adaptive traits. The G^2 measure integrates over environmental effect sizes.

Implemented in LEA 3.9.5 as `LEA::genetic.gap()`

(Gain et al. online some day)

Population genetic interpretations of the genetic gap

- Dual interpretation as a squared distance in genetic space

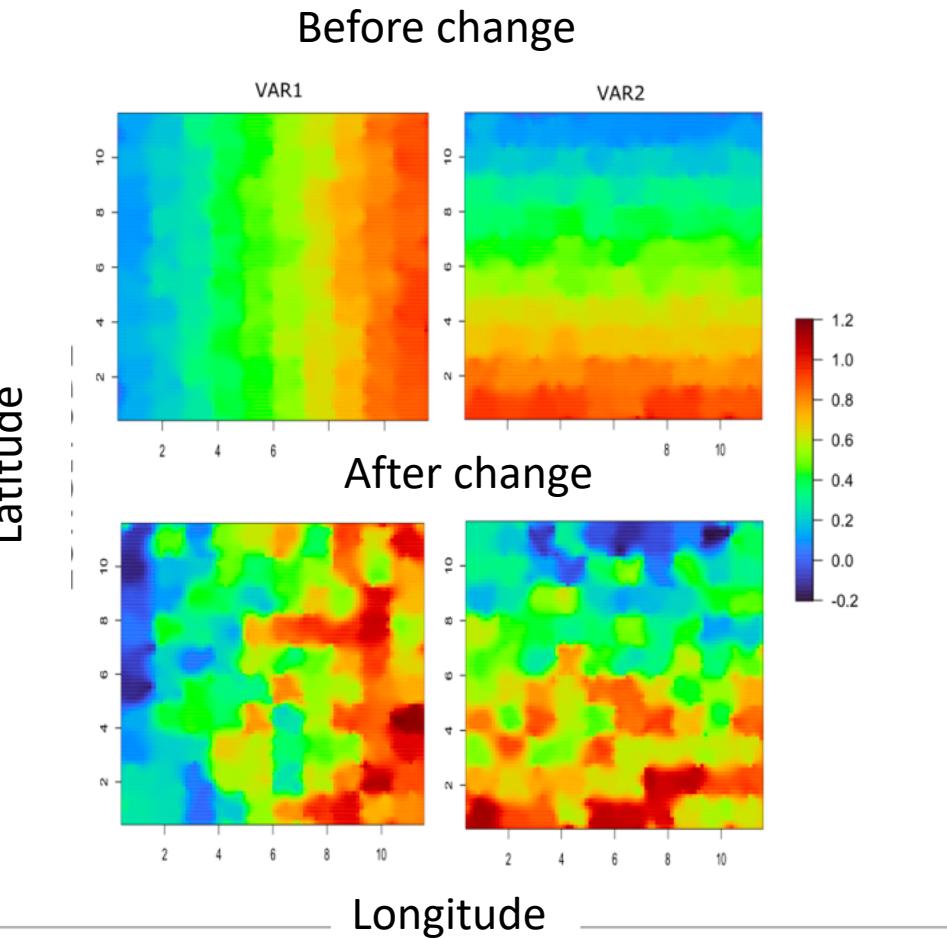
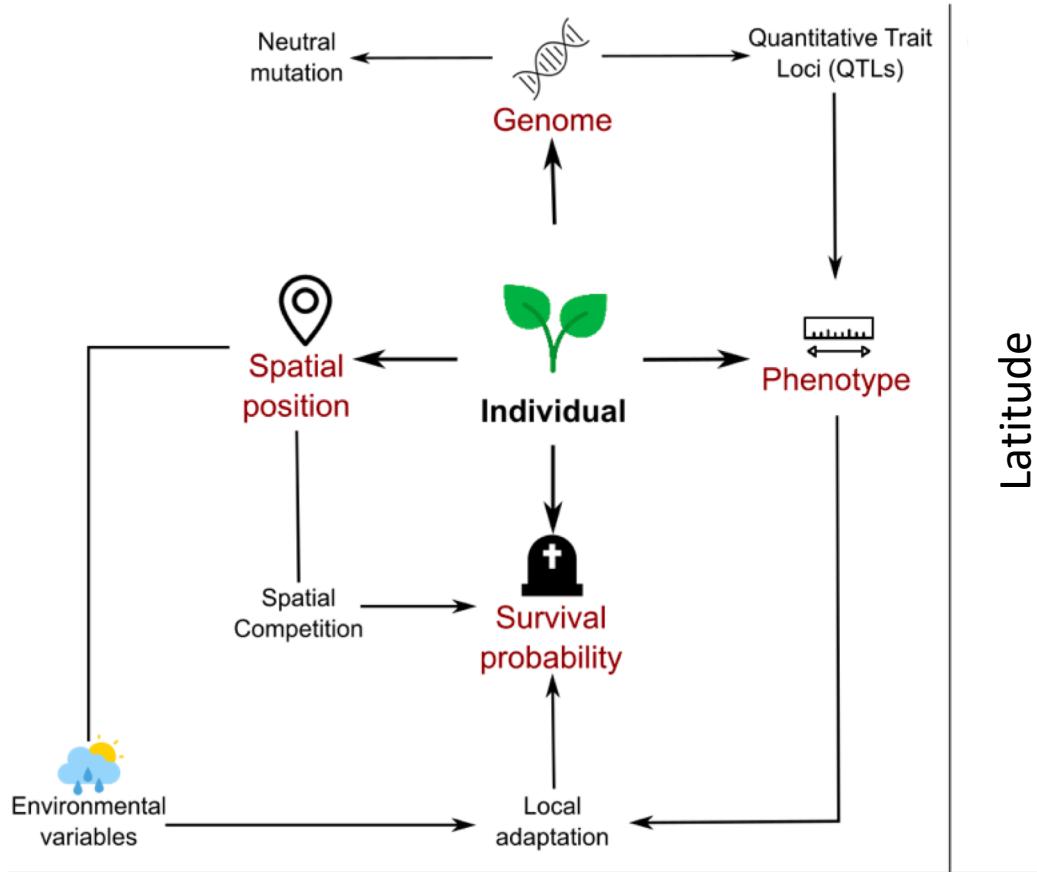
$$G^2(\mathbf{x}, \mathbf{x}^*) = \mathbb{E}[(f(\mathbf{x}) - f(\mathbf{x}^*))^2]$$

$f(\mathbf{x})$ = allele frequency predicted by the LFMM for variable \mathbf{x}

- G^2 corresponds to Patterson's F_2 and to Nei's D_{ST} ($= F_{ST} \times H_T$)
- Under Gaussian stabilizing selection, G^2 measures the loss of fitness.

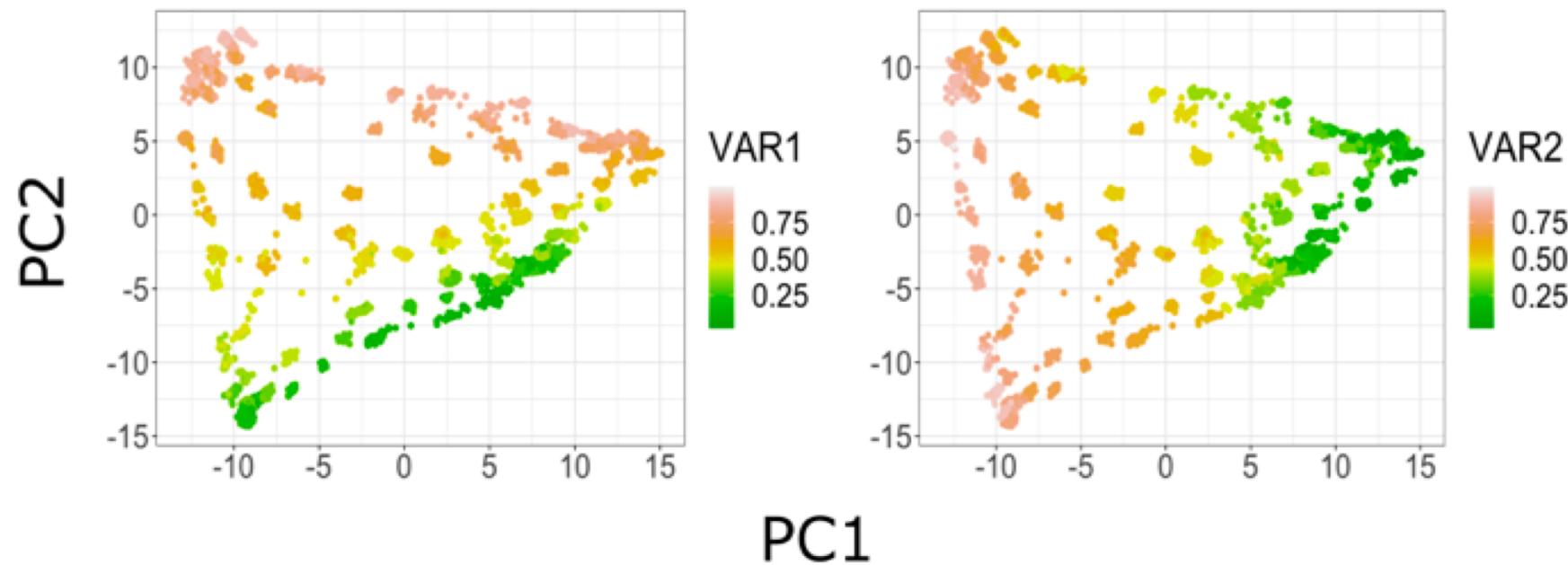
(Gain et al. not yet online)

Simulation study (using SLIM 3)

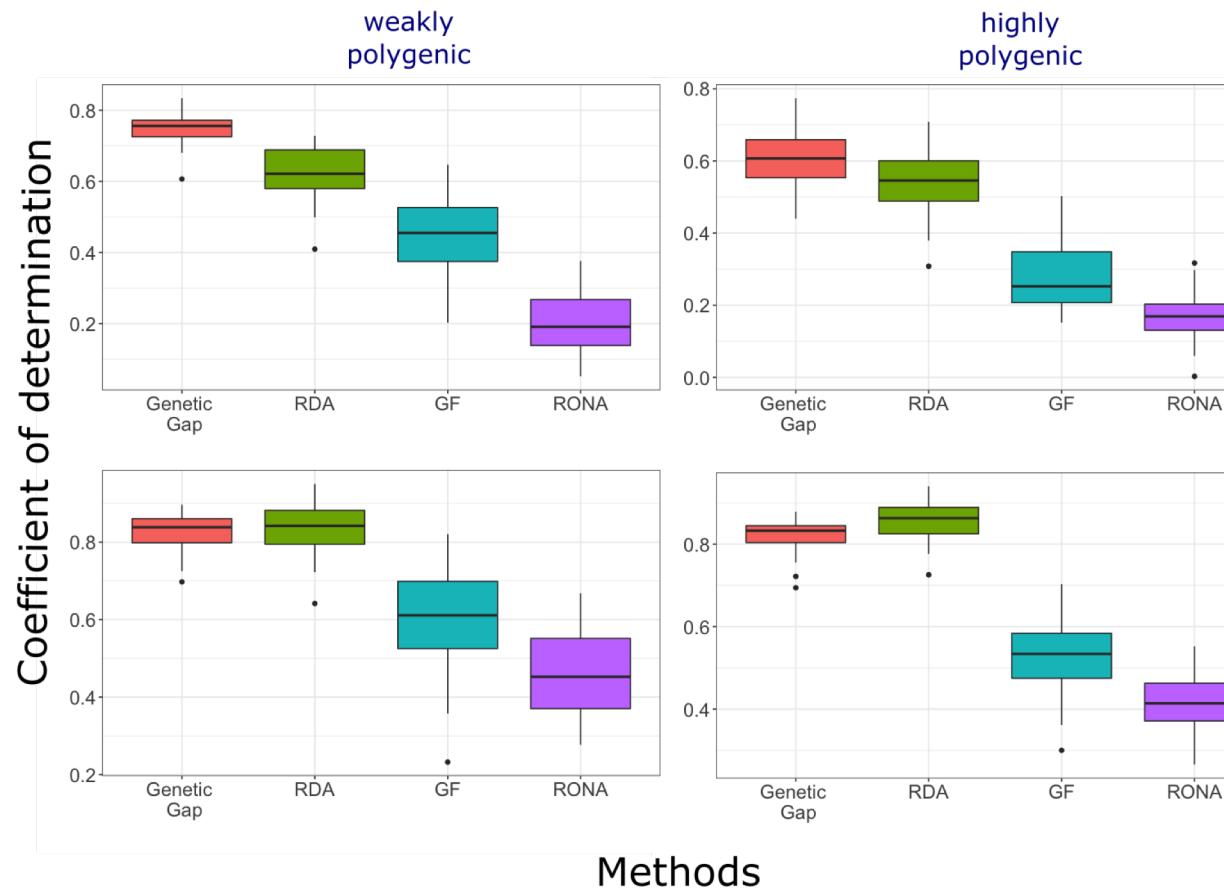


A look at population structure (PCA)

PCA of the genomic data colored by environment



Simulation results

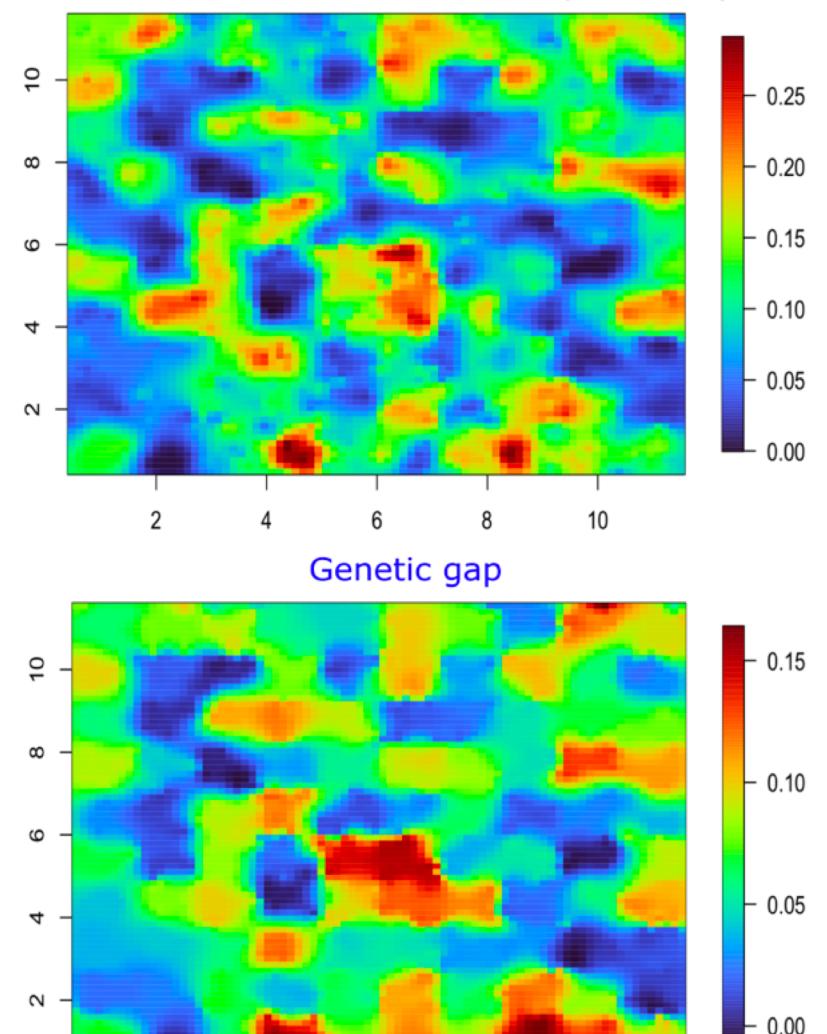


Confounding due
to population
structure

HIGH

LOW

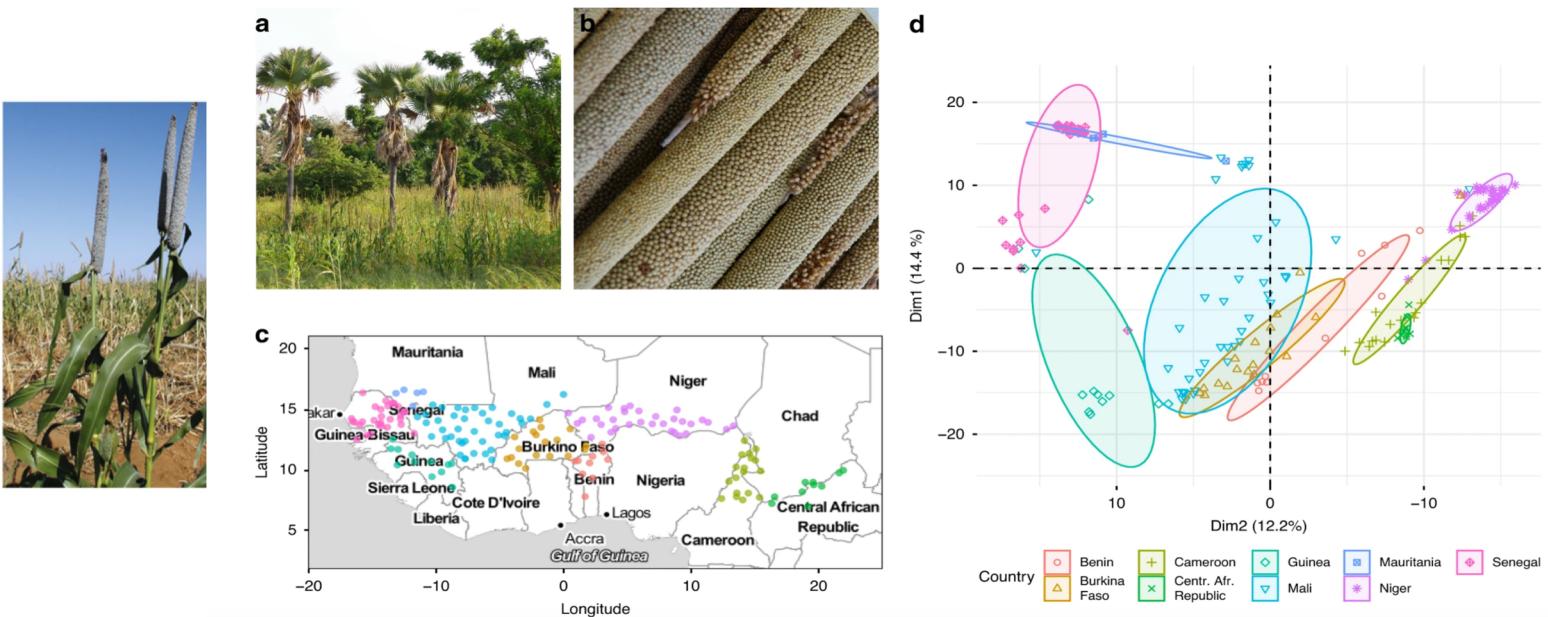
Absolute decrease in survival probability



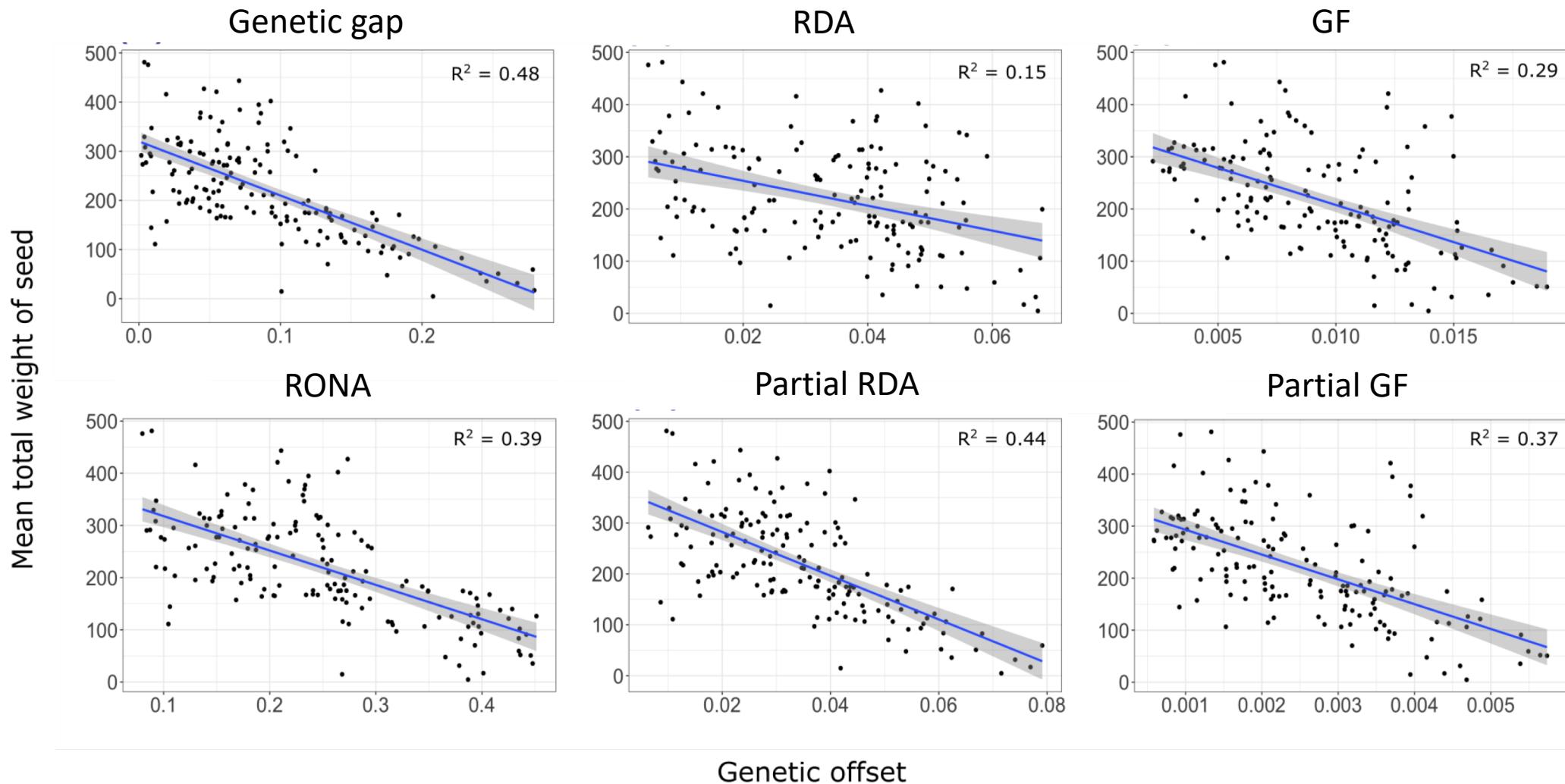
RDA: Capblancq et al. 2021, GF: Fitzpatrick & Keller 2015, RONA: Rellstab et al. 2016

Common garden experiment

- Pearl millet is one of the most nutritious low input staple cereal cultivated in the arid and low fertility soils of Africa and Asia.
- Rohné et al. (2020) grew 123 landraces in a common garden in Sadoré (Niger) and measured their yield



Genomic offset analysis



A short tutorial in R (source: GitHub repo SSMPG2022)

```
# To install LEA, remove the comment character in the next command line  
# devtools::install_github("bcm-uga/LEA")  
library(LEA)
```

Data: 200 diploid individuals genotyped at 510 SNPs and 4 environmental variables.

About 30% of all loci are influenced by environment.

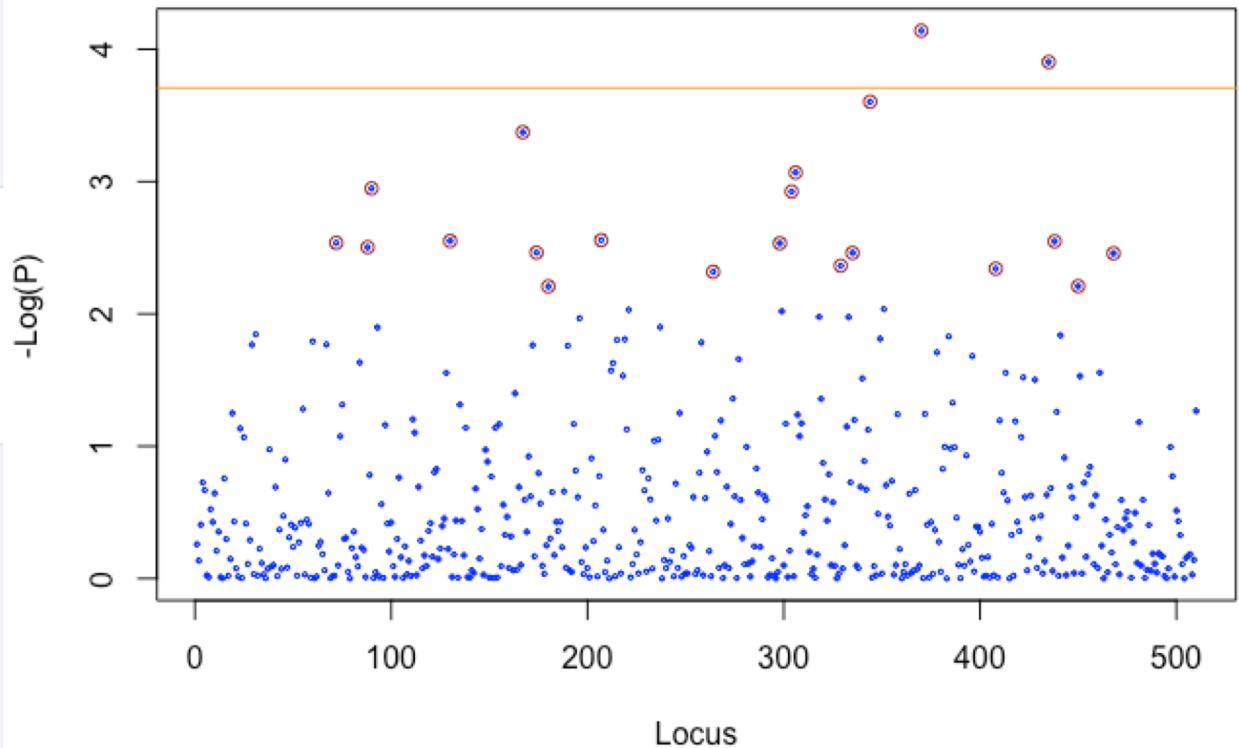
```
data("offset_example")  
# genotypes for 200 individuals  
Y <- offset_example$geno  
  
# Four environmental variables for 200 individuals  
X <- offset_example$env
```

GEA analysis

```
mod_lfmm2 <- lfmm2(input = Y,  
                     env = X,  
                     K = 3)
```

An LFMM with $K = 3$ latent factors is adjusted to the data

```
pv_lfmm2 = lfmm2.test(object = mod_lfmm2,  
                      input = Y,  
                      env = X,  
                      full = TRUE)$pvalues
```



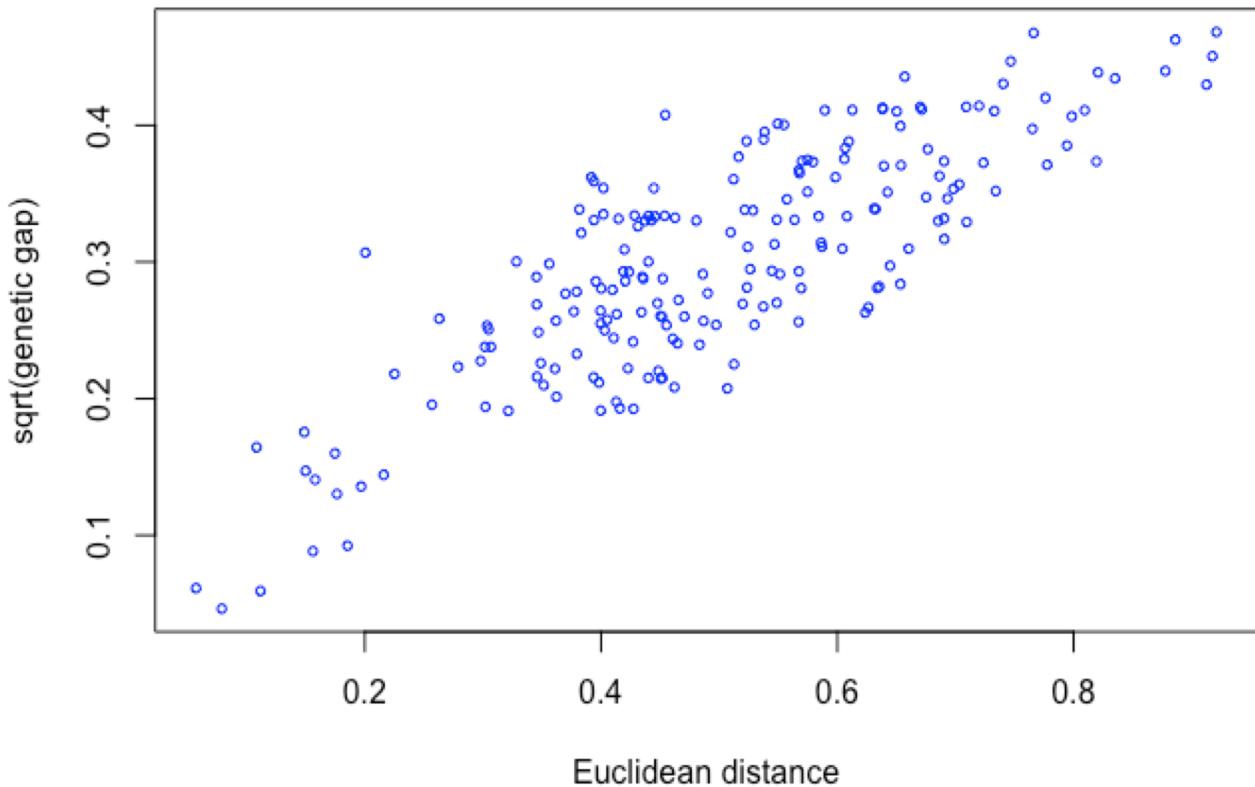
Quite disappointing: 90% of SNPs under selection remained undiscovered (FDR level = 20%)

Genetic gap analysis

```
X_pred <- offset_example$env.pred
```

The genetic gap is computed at each sampled location

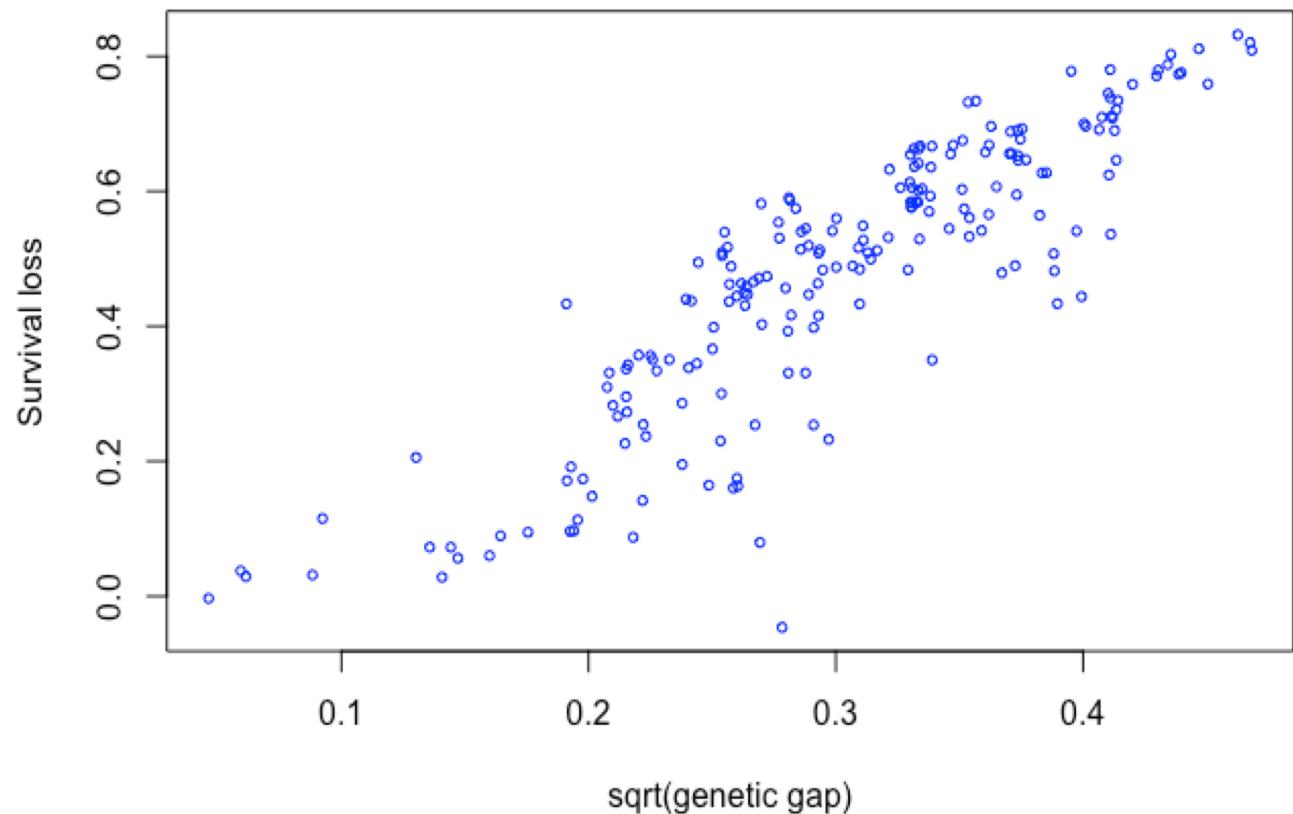
```
g_gap <- genetic.gap(input = Y,  
                      env = X,  
                      pred.env = X_pred,  
                      K = 3)
```



Prediction of fitness loss

$R^2(G, \text{survival loss}) = 76.75\%$

$R^2(\text{Euclidean dist.}, \text{survival loss}) = 55.51\%$



Conclusion

- A genetic offset based on a non-Euclidean geometry of the ecological niche
- Addresses several issues of genetic offsets
- Dual interpretation as a distance between environmental vectors and as a distance between predicted allele frequencies
- Improved prediction of fitness variation after translocation in modified environments
- Implemented in the function `genetic.gap` of LEA (runs fast)

Acknowledgments

- Flora Jay
- Isra Salazar
- Florence Forbes
- Bénédicte Rohné
- Philippe Cubry
- Yves Vigouroux
- ANR Genomic prediction of vulnerability of African crops to future climate
- ANR Predictive ecological genomics
- We have opportunities for PhD and postdoctoral fellowships: contact us.

