Insights into the environmental pressures driving adaptation in *Drosophila melanogaster*

María Bogaerts Márquez

6th October 2020

Introduction

Barcelona, Spain
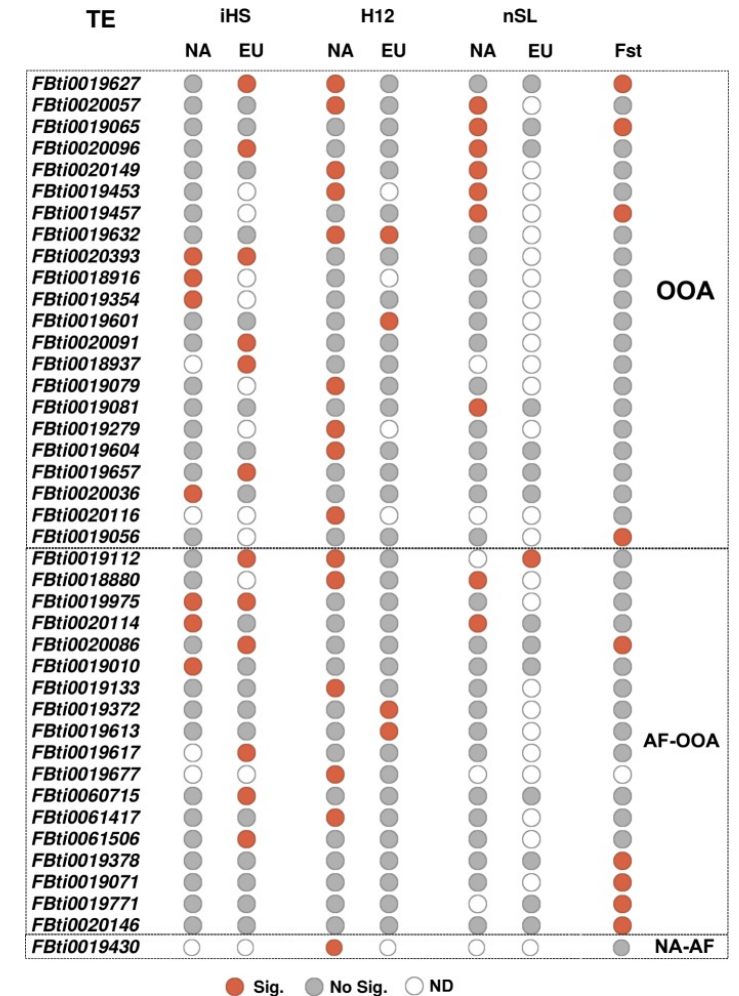

González Lab — Evolutionary and Functional Genomics

- How organisms adapt to their environments

- The role of Transposable Elements in adaptation

- Combining *omics* approaches with molecular and phenotypic analysis

- Mainly in *Drosophila melanogaster* but also in *Anopheles* and humans

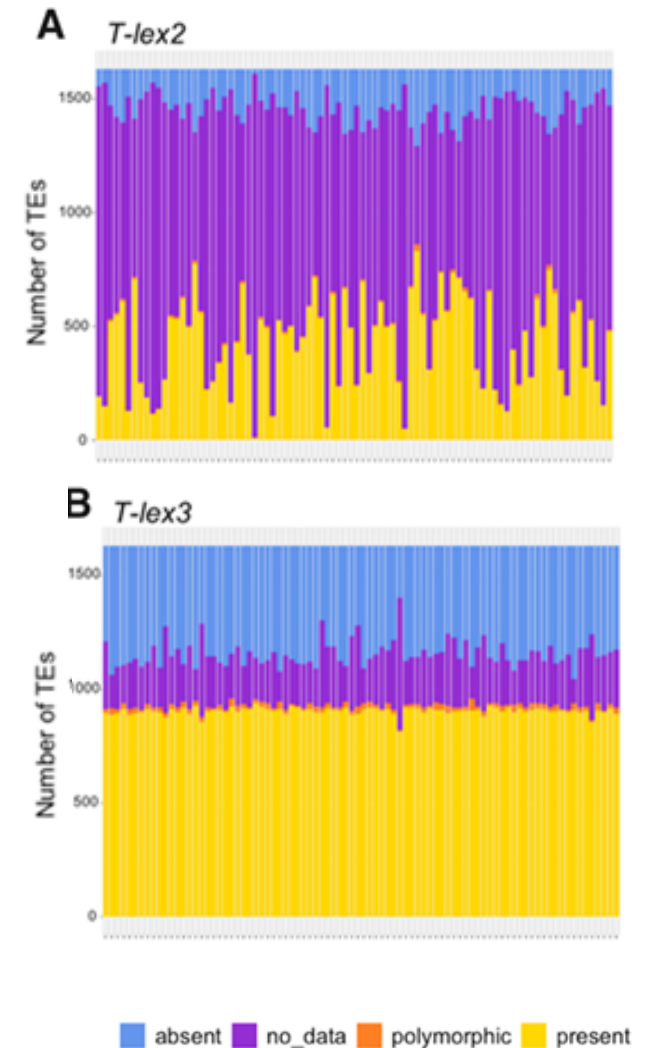INSTITUT de BIOLOGIA EVOLUTIVA — IBE — CSIC — upf.

Previous work

- Rech GE **et al.** 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet.*, 15(2), e1007900. doi:10.1371/journal.pgen.1007900

- **Bogaerts-Márquez M** *et al.* 2019. T-lex3: an accurate tool to genotype and estimate population frequencies of transposable elements using the latest short-read whole genome sequencing data. *Bioinformatics*. doi:10.1093/bioinformatics/btz727

- Kapun M **et al.** 2020. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol. Biol. Evol.* doi:10.1093/molbev/msaa120
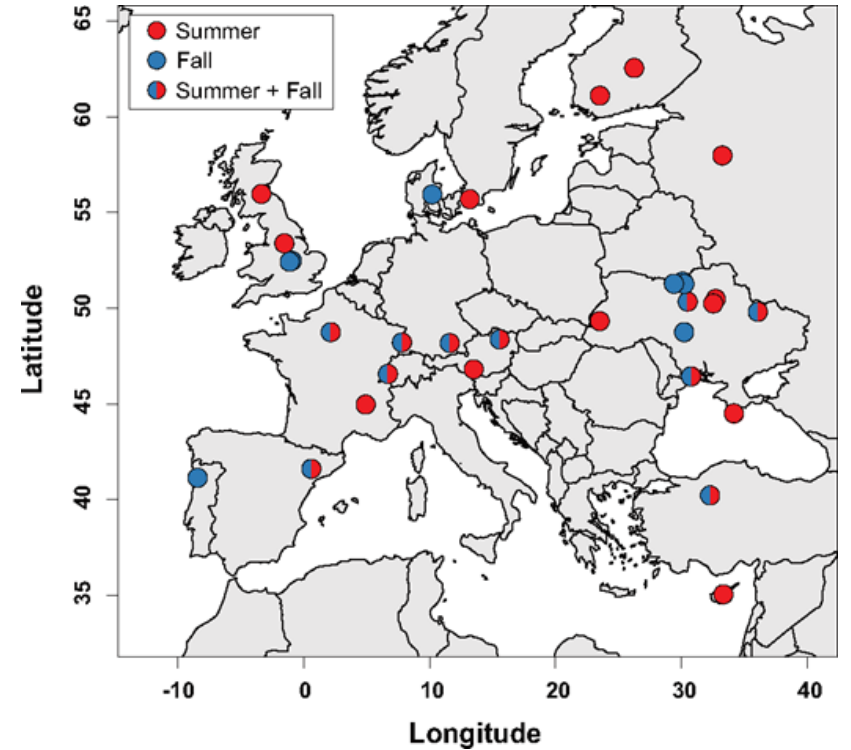
Previous work

• Rech GE *et al.* 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet.*, 15(2), e1007900. doi:10.1371/journal.pgen.1007900

• **Bogaerts-Márquez M** *et al.* 2019. T-lex3: an accurate tool to genotype and estimate population frequencies of transposable elements using the latest short-read whole genome sequencing data. *Bioinformatics*. doi:10.1093/bioinformatics/btz727

• Kapun M *et al.* 2020. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol. Biol. Evol.* doi:10.1093/molbev/msaa120

Previous work

• Rech GE *et al.* 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila. PLoS Genet.*, 15(2), e1007900. doi:10.1371/journal.pgen.1007900

• **Bogaerts-Márquez M** *et al.* 2019. T-lex3: an accurate tool to genotype and estimate population frequencies of transposable elements using the latest short-read whole genome sequencing data. *Bioinformatics.* doi:10.1093/bioinformatics/btz727

• Kapun M *et al.* 2020. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol. Biol. Evol.* doi:10.1093/molbev/msaa120

Previous work
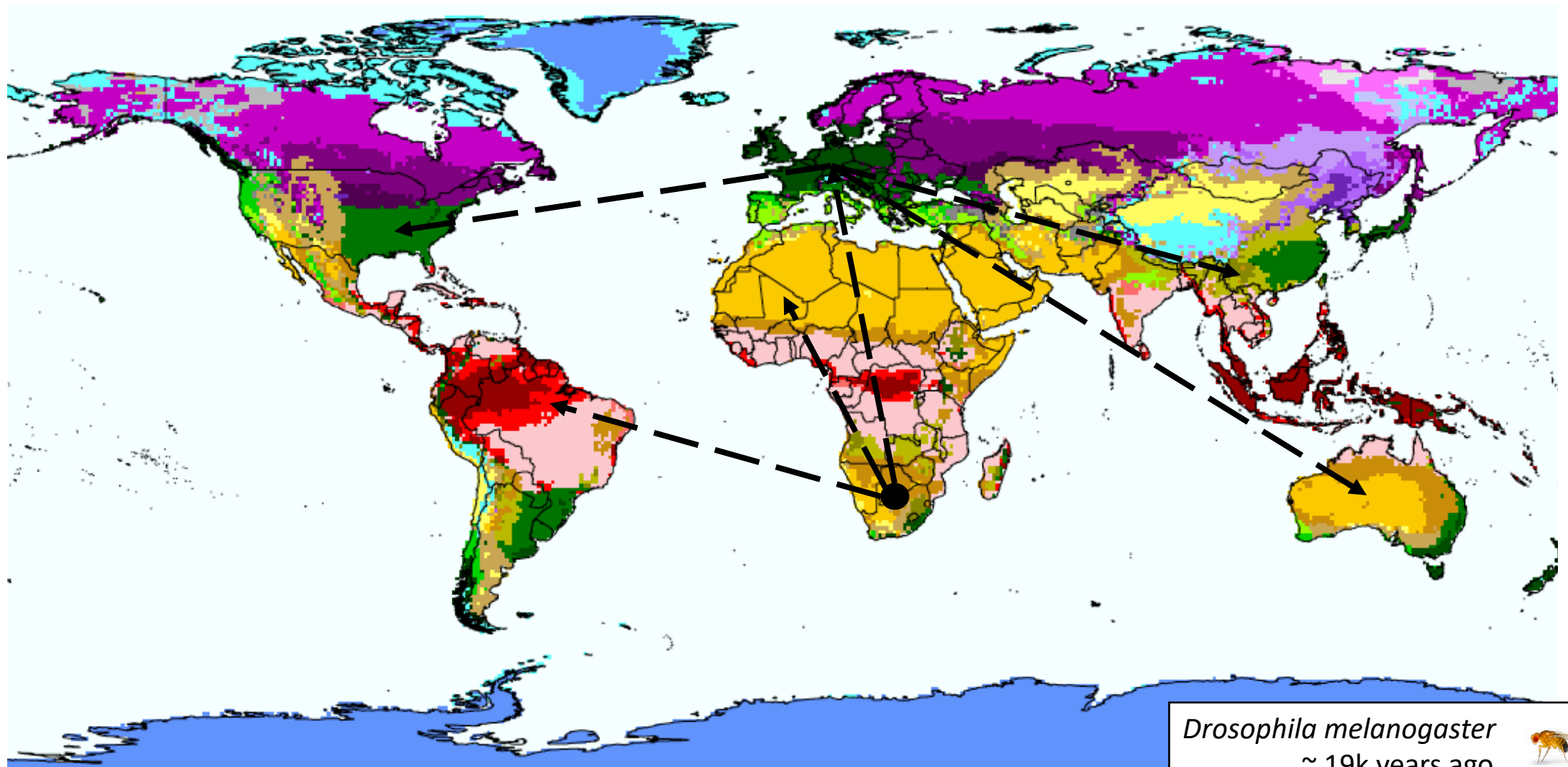
Selective Sweeps in DrosEU 2014 populations

- Pool-hmm (Boitard et al. 2013)
  - Identifies candidate sweep regions via distortions in the allele frequency spectrum

- 30/48 samples → 19 populations

- Well supported sweeps: *wapl*, HDAC6, Hen1, CR18217, *mgl*, *phantom, Cyp18a1* and *Cy6g1*

- 64 genes in sweeps detected in 19 populations
  - 52 were located in the 10% of regions with the lowest values of Tajima's *D*
  - 43 were located in regions with reduced Tajima's *D* (lowest 10%) in African populations
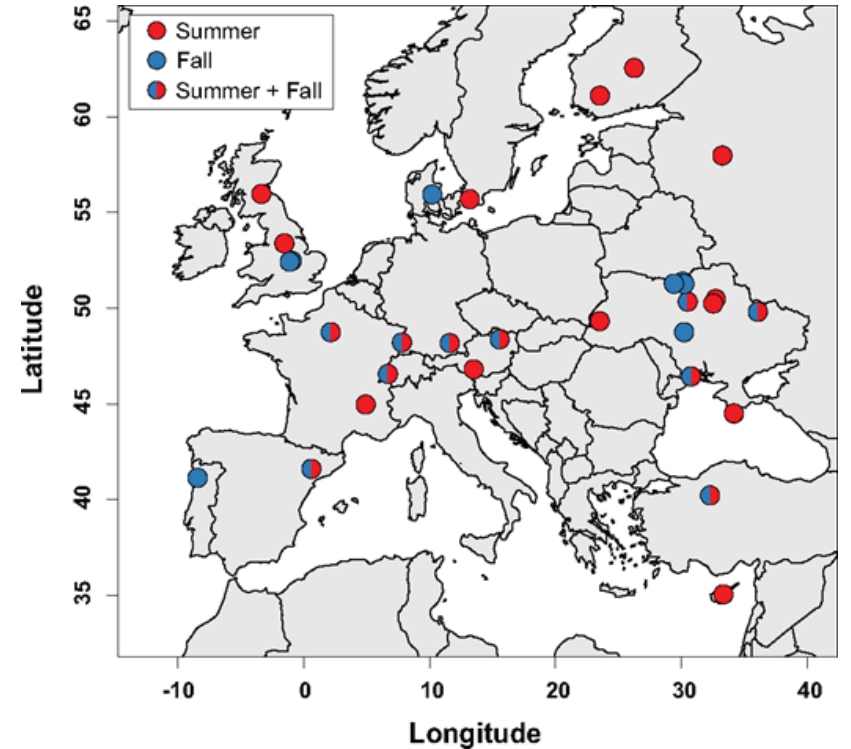
Insights into the environmental pressures driving adaptation in *Drosophila melanogaster*

First part

Drosophila melanogaster
~ 19k years ago

Koppen-Geiger climate distribution: *Rubel & Kottek 2010*

- *Drosophila melanogaster*
- Males
- Pool-sequenced samples

North America (Machado et al. 2019)
- 11 samples from 11 locations in North American East coast (collection 2003-2014)
- ~ 2.4 M SNPs
- 344 Referenced Transposable Elements

Europe (Kapun et al. 2020)
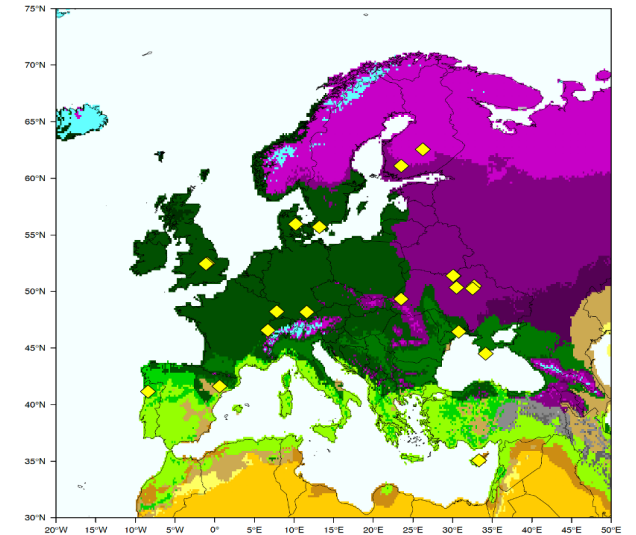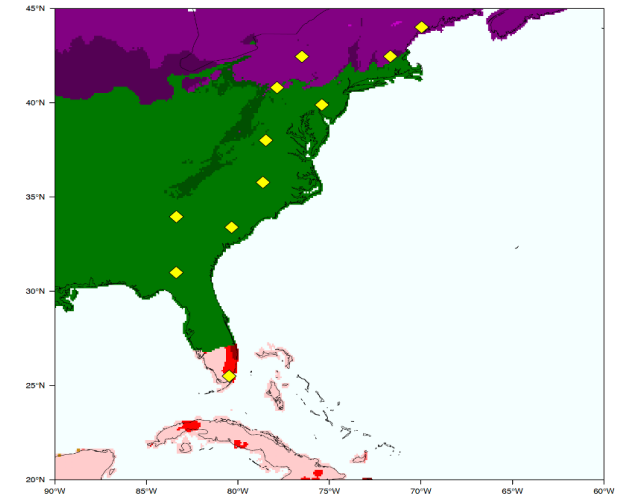- 26 samples from 20 locations in Europe (collection 2014)

Europe:
- 20 samples
- ~ 3 M SNPs
- 302 TEs

Europe Summer:
- 14 samples
- ~ 2.8 M SNPs
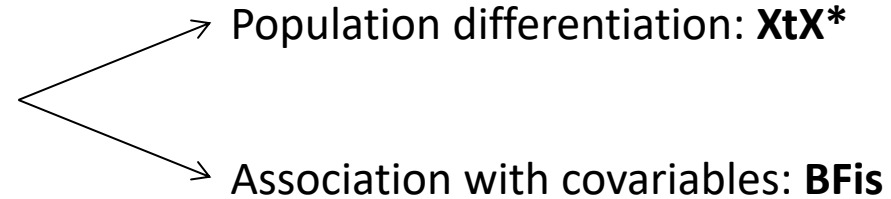- 264 TEs

Europe Fall:
- 10 samples
- ~ 2.8 M SNPs
- 276 TEs

The Method:

BayPass
(Gautier 2015)

Covariance matrix → Population differentiation: **XtX***

Covariance matrix → Association with covariables: **BFis**

- **Long-term**: average 1970-2000

WorldClim

19 Standard Bioclimatic Variables:

Temperature
Rainfall

- **Year-specific**: one year before collection date

ECMWF

UNITED STATES NAVAL OBSERVATORY

HM Nautical Almanac Office

Temperature
Rainfall
Evaporation
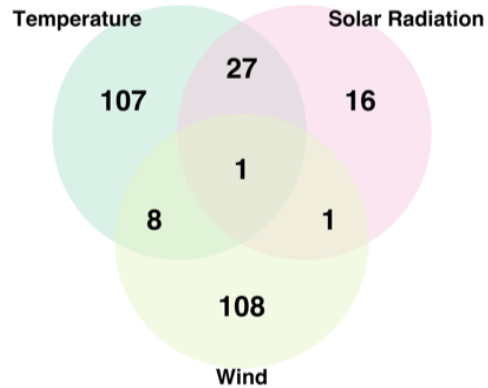**Solar Radiation**
Wind
Soil
Daylight hours

Population differentiation: XtX*

- Significance threshold: **top 0.05%**
  - SNPs in gene body and in regulatory regions

- Genes known to play a role in adaptation: *sgg*, *cpo*, *Ace* and *mth*

- *In(2L)t* enriched in **Europe** and **Europe Fall** and *In(3R)P* enriched in **North America**

- GO enrichment main clusters: development, signaling, morphogenesis

- 55 candidate genes significantly overlapping between North America and Europe (SuperExactTest < 0.05)
  - 19/55 not reported in other clinal studies
  - GO enriched clusters: regulation, signaling and response to stimulus and development
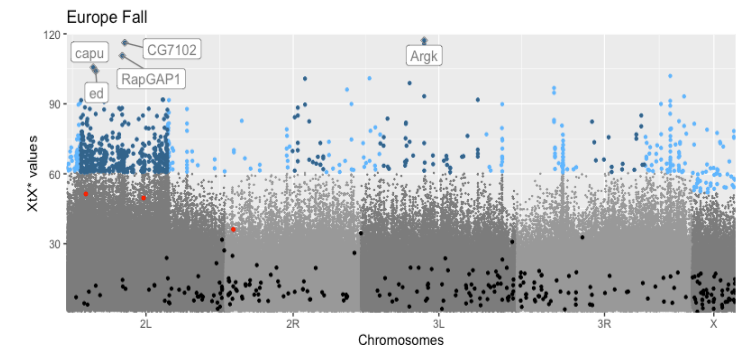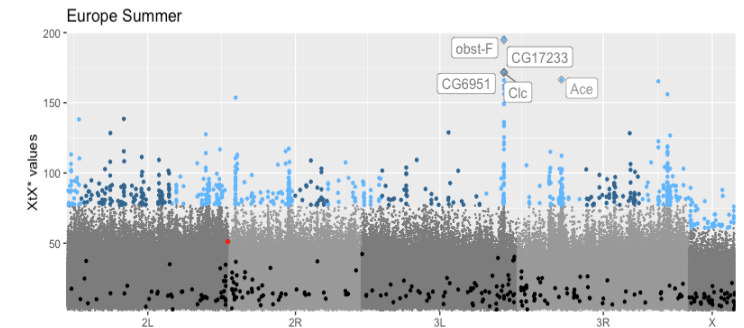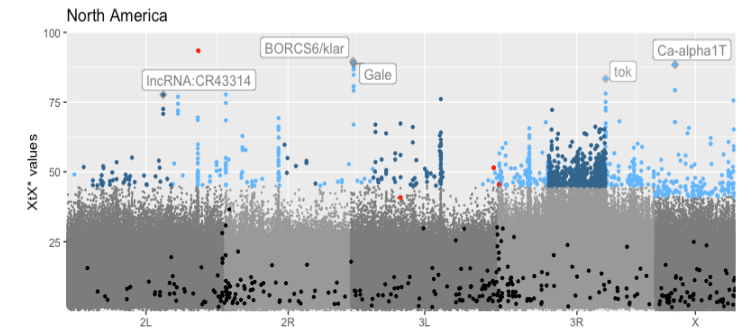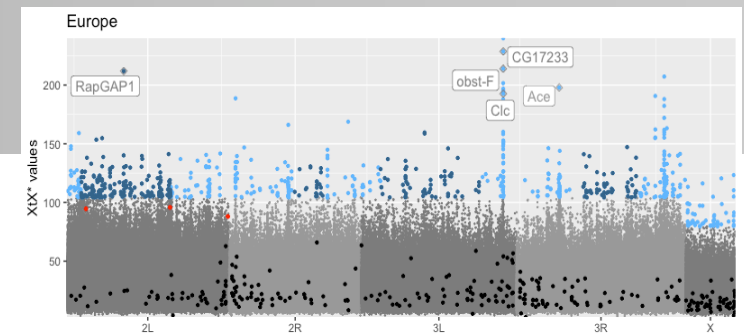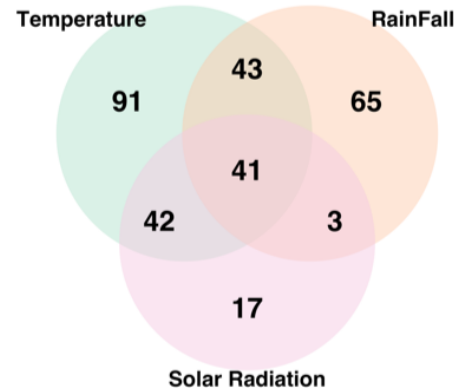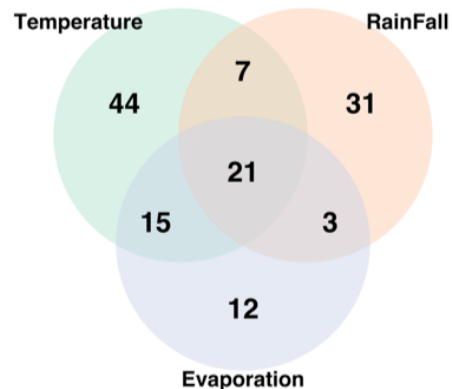
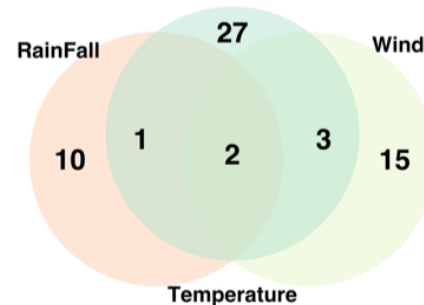# Association with environmental variables: BFis

A) Europe

B) North America

C) Europe Summer

D) Europe Fall

- Significance threshold: Bfis > 30
  - SNPs in gene body and in regulatory regions

- 748 candidate genes among the four data sets
  - **Temperature: 400**
  - **Rainfall: 241**
  - Evaporation: 153
  - Solar Radiation: 153
  - **Wind: 226**
  - Soil: 4
  - Daylight hours: 73

- 32 candidate genes significantly overlapping between North America and Europe (SuperExactTest < 0.05)
  - 12/32 not reported in other clinal studies

- 289 genes not shown in population differenciation

Transposable elements

Population differenciation (XtX*)

Association with an environmental variable

Europe

5

3

6

North America

4

2

4
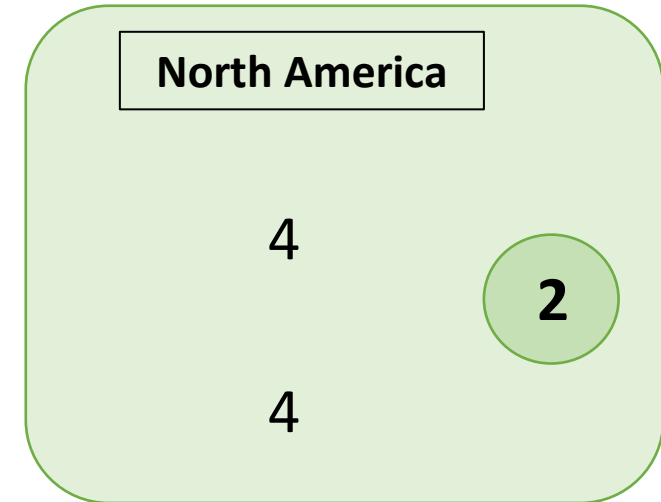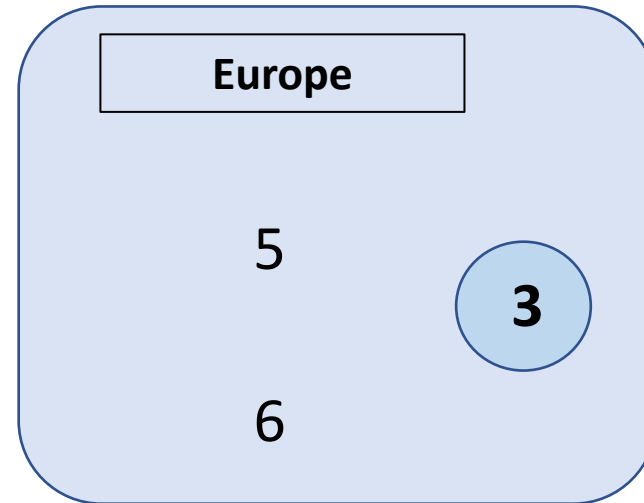
FBti0019112

- Previous signatures of selection: iHS, nLS, H12
- First intron *lilli* gene
- Population differentiation in **Europe** and **Europe Fall**
- Association with environmental variables in Europe: **Temperature**

FBti0061428

- Previous signatures of selection: no
- Upstream/downstream CG31809/CG6012
- Population differentiation in **North America**
- Association with environmental variables in North America: **Wind**

Take-home messages

- In addition to the already studied variables related to temperature and rainfall, **wind** could be a putative environmental pressure which plays a role in adaptation

- **Wind** variables showed **small overlap** with other variables

- **GEA analysis** are needed to find other adaptive signatures which are **not** explained by population differentiation patterns
  - Limited to the **variables** used and how they are involved in the population structure
  - Limited to the **strongest outliers**

- **Transposable elements** which may play a role in adaptation
  - Lab experiments needed
  - Too limited because of the reference annotation: **non-reference TEs needed**

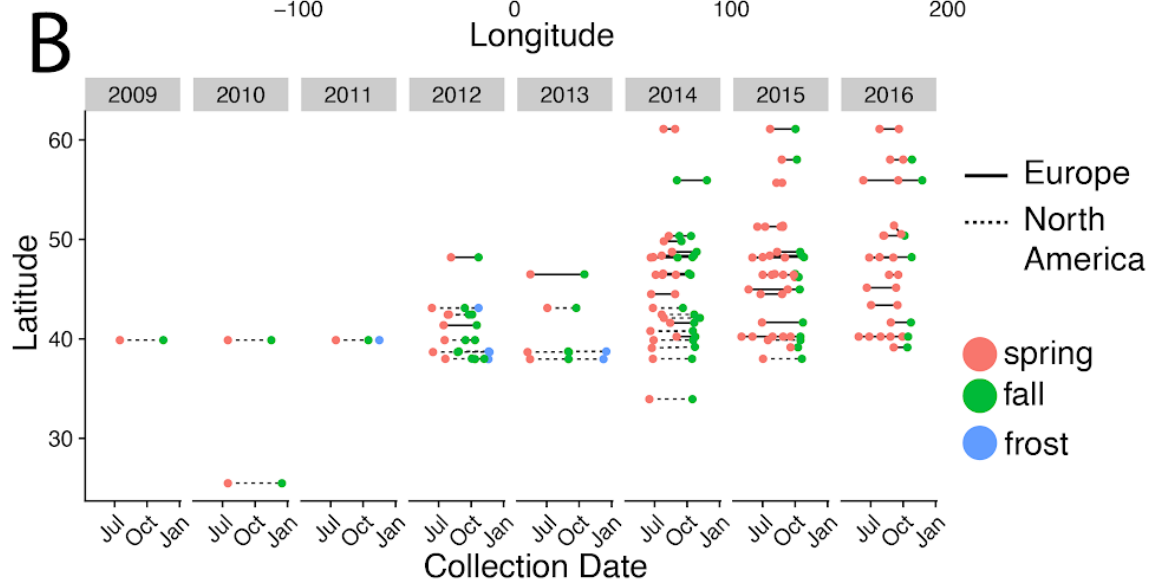# Insights into the environmental pressures driving adaptation in *Drosophila melanogaster*

Second part
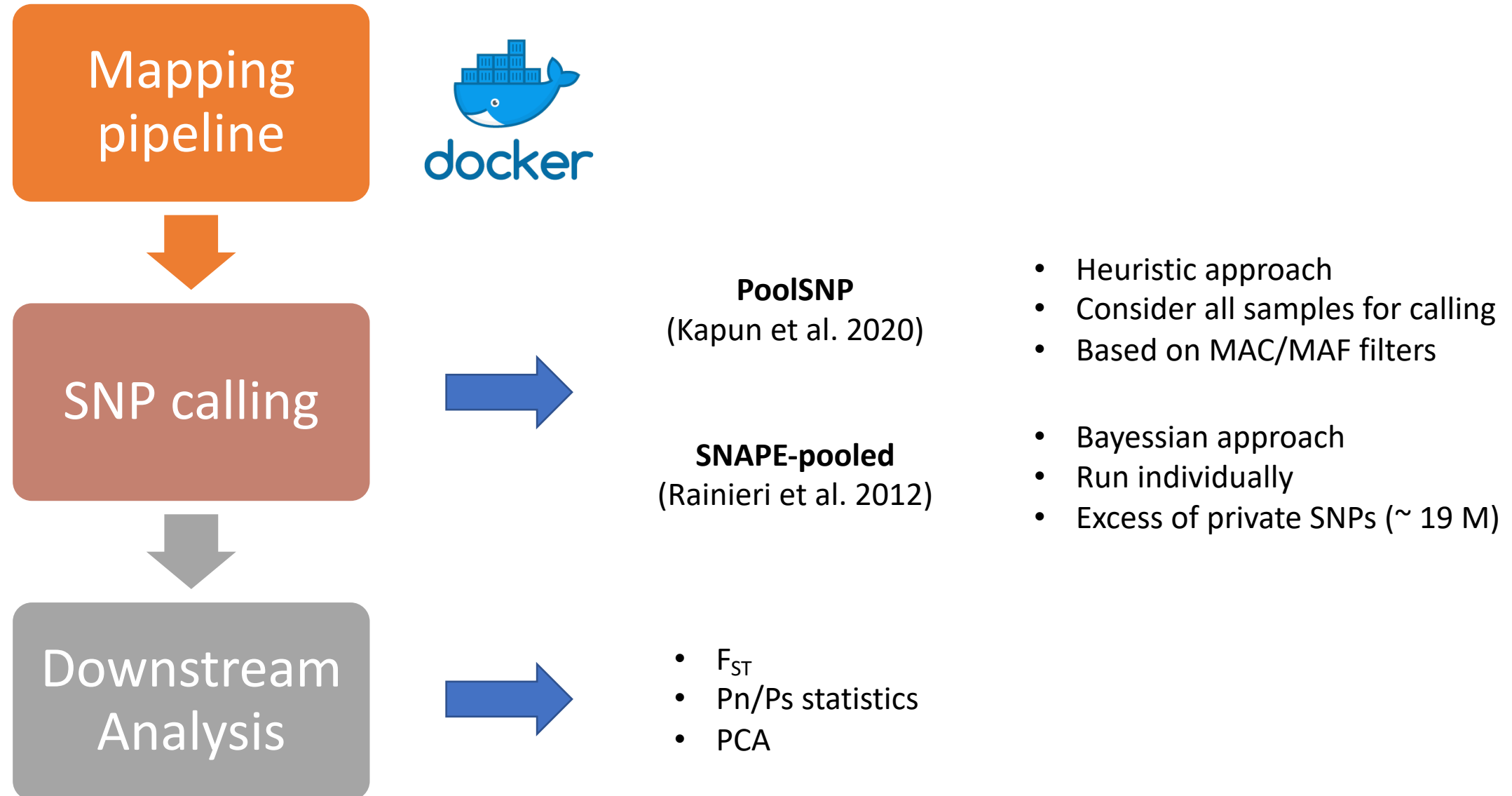
Drosophila Evolution in Space and Time
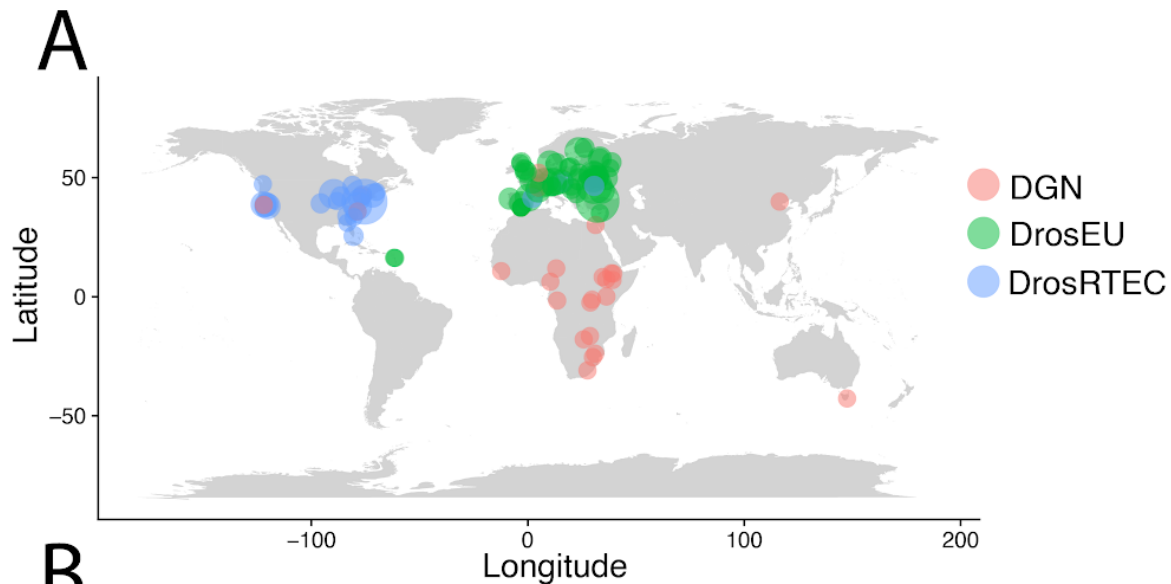
DEST data set:
Drosophila Evolution in Space and Time

- Pool-seq sequenced *Drosophila melanogaster*

- 169 European Populations (DrosEU)

- 77 North American Populations (Dros-RTEC)

- 26 *Drosophila melanogaster* Nexus (Lack et al. 2016): single-individual sequencing data from several ancestral African populations

Kapun et al. (in prep)

Drosophila Evolution in Space and Time

PoolSNP
(Kapun et al. 2020)

- Heuristic approach
- Consider all samples for calling
- Based on MAC/MAF filters

SNAPE-pooled
(Rainieri et al. 2012)

- Bayessian approach
- Run individually
- Excess of private SNPs (~ 19 M)

Mapping pipeline

SNP calling

Downstream Analysis

- F_{ST}
- Pn/Ps statistics
- PCA

Drosophila Evolution in Space and Time

o Study seasonality using *C2* BayPass contrast test in Summer/Fall samples

o Study changes in the same population in time (i.e. same population collected in three consecutive years)

o Study environmental association
  o Global adaptation
  o Continent-specific adaptation
  o Focusing in one specific variable

o Transposable elements
  o Reference transposable elements
  o Non-reference transposable elements
      o TEMP (Zhuang *et al.* 2014)
      o Tidal (Rahman *et al.* 2015)

Thank you