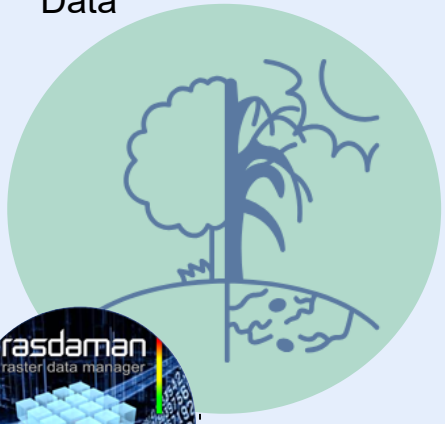
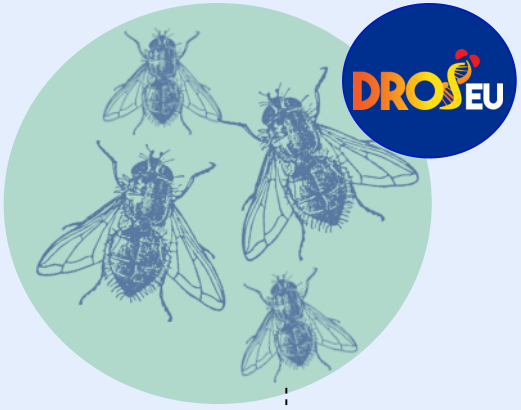
FairiCube UseCase3: Drosophila Genetics

Geospatial Data / Environmental Data



Genetic Data
for *Drosophila melanogaster*



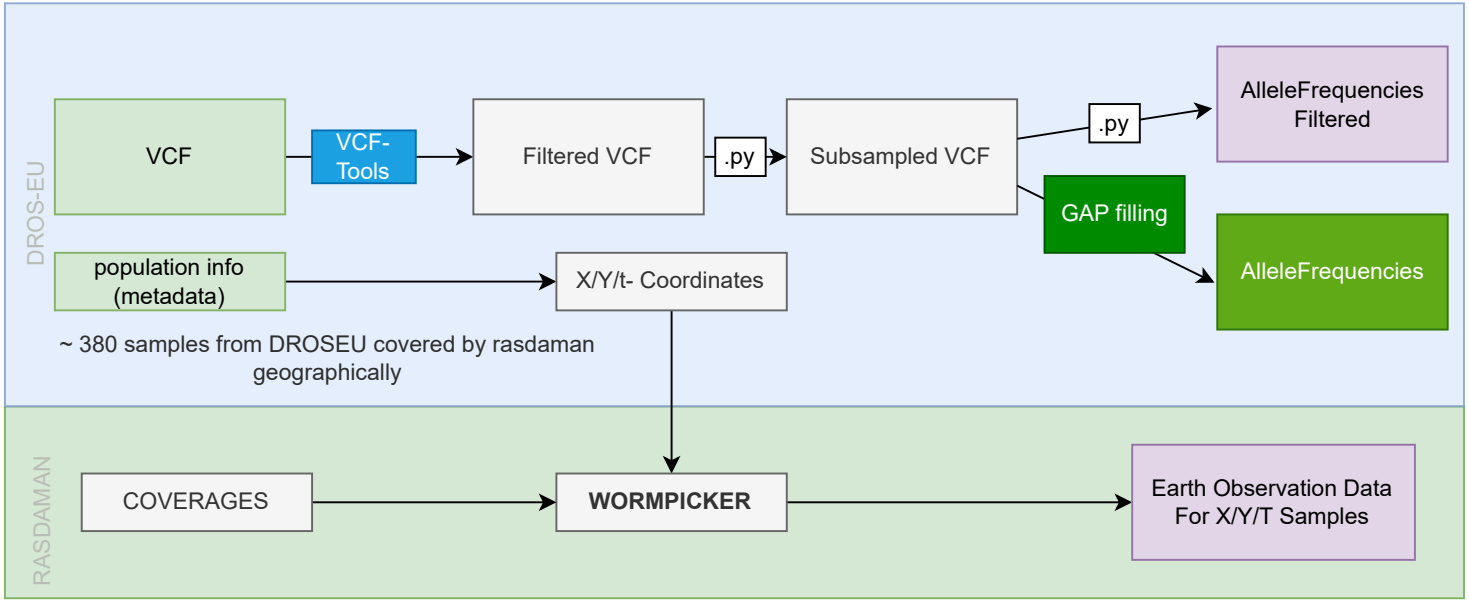
NHM
"Landscape Genomics Pipeline"

uc3-drosophola-genetics/projects/LandscapeGenomicsPipeline at main · FAIRiCUBE/uc3-drosophola-genetics (github.com)

So far pipeline is adjusted to work with only a small subset due to computational efforts (10/737 potential samples, 10.000 Positions).
Shellscript / Python / R

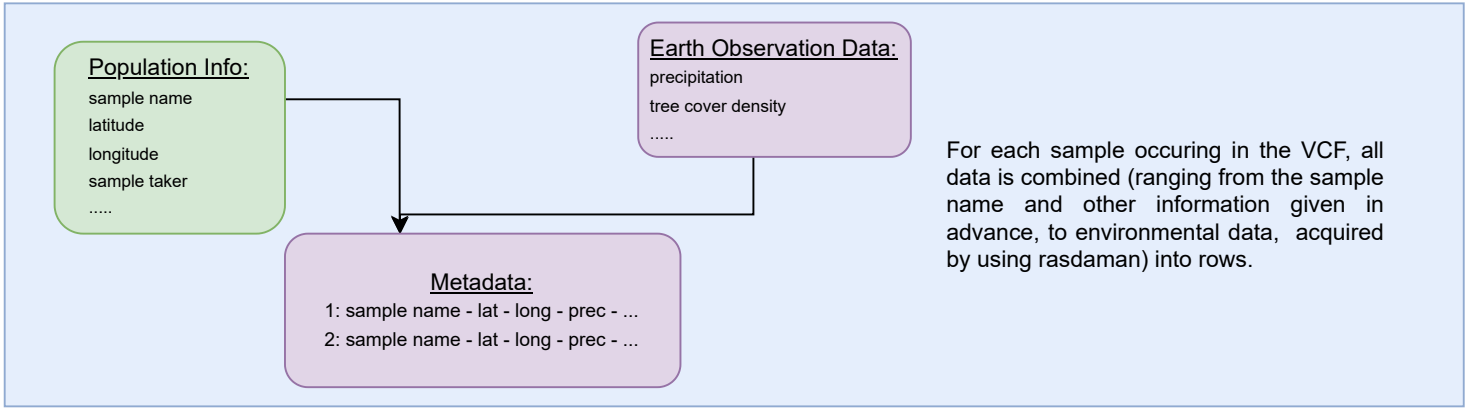
Step 1: Obtain and Filter Data

DrosEU: Download data as VCF, Subsample, convert to allele frequency table
Rasdaman: Install Rasdaman Software, get environmental data for spatial points covered in sampling



Step 2: Intersect / Combine Data

Match environmental data with metadata table (samplename, ID, population size,...)



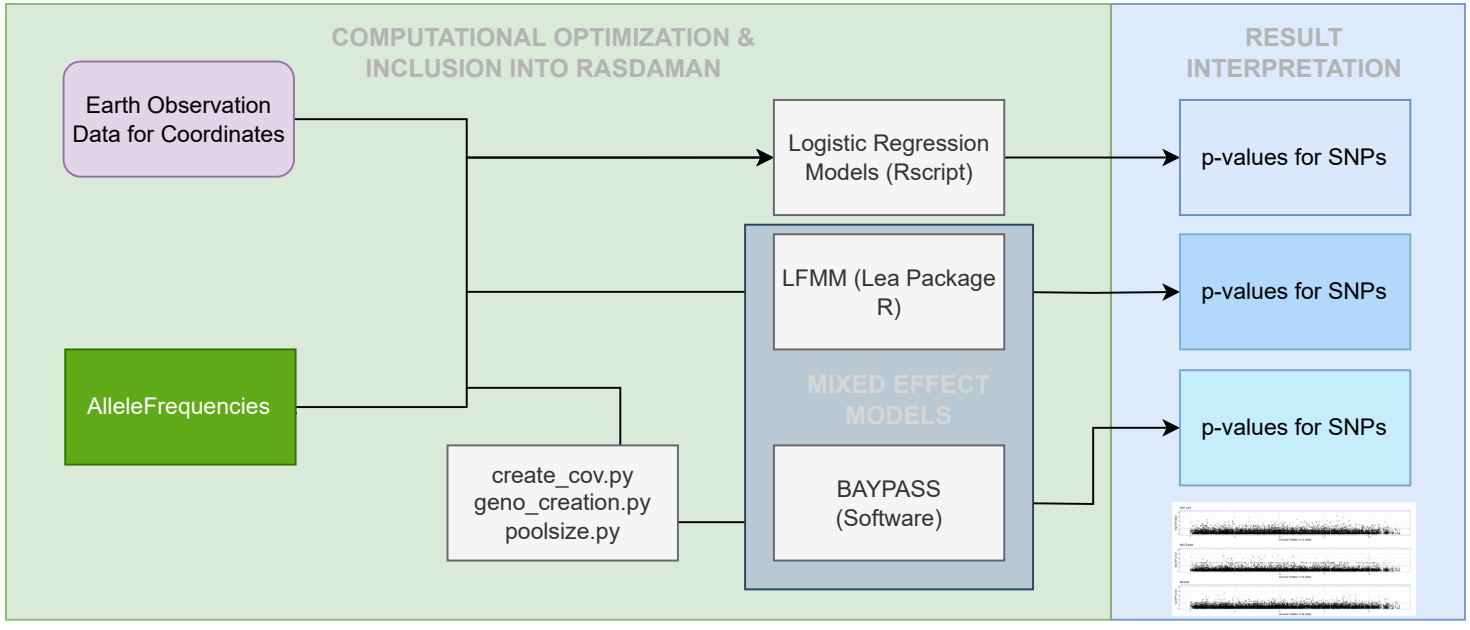
For each sample occuring in the VCF, all data is combined (ranging from the sample name and other information given in advance, to environmental data, acquired by using rasdaman) into rows.

Step 3: Apply Statistical Methods / Hypothesis Testing

Test for correlation of polymorphism-frequency with variable:
- Linear Regression, Logistic Regression, Latent Factor mixed Models

Usage of specific programs written for (population) genetic analyses:
- LEA, BAYPASS,...

--> Computationally heavy for multiple 100 samples, pipeline so fa rnot fully adjusted to computationally intense number of samples



LFMM: include a strategy to correcly estimate number of latent factors (strucutre)

BAYPASS: Calculation of p-values not clear yet,filter??? Inlcude more models?
Way of comparing (not visually) the results?? Overlapping ?

SPATIAL DISTRIBUTION OF GENETIC VARIATION

DROS-EU

Allele Frequency

npstat

Genetic Diversity

population info (metadata)

X/Y/t- Coordinates

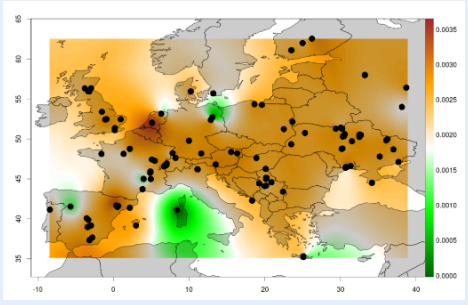
RASDAMAN

desired layers

WORMPICKER

FAIRiCUBE

Regression to predict which environmental variable influences genetic diversity



THE INFLUENCE OF ENVIRONMENTAL VARIATION ON SPECIES DISTRIBUTION

DROS-EU

TAXODROS (taxodros.uzh.ch)

GBIF

DrosDist.r

X/Y/T - Coordinates

RASDAMAN

desired layers

WORMPICKER

FAIRiCUBE

species distribution modelling