

A framework for simulating spatio-temporal population genomic data on a real geographic landscape

Martin Petr¹, Benjamin C. Haller², Peter L. Ralph³, Fernando Racimo¹

Most evolutionary processes in nature have an implicit spatial dimension. However, this spatial component of evolution is often neglected, partly because the tools to study complex spatio-temporal population dynamics have been lacking. Here we present a new R package for programmatic and visual definition of spatio-temporal population models which generates spatially-explicit data with the forward population genetics software SLiM (<https://messerlab.org/slim>) as a simulation back end.

Simulation workflow

1. spatial context

Populations can occupy a real landscape with a coordinate reference system (CRS) appropriate for the location on Earth's surface (here West Eurasia). By default, all coordinates are specified in degrees longitude and latitude and internally transformed into the specified CRS, but the `world` function can also define a completely abstract two-dimensional space.

2. population dynamics

The function `population` defines the spatial boundary of a population. Functions such as `move` and `expand` encode the spatio-temporal dynamics of populations and internally perform transformations of polygons in a two-dimensional space.

3. geneflow events

Genetic exchange between populations is specified using a dedicated `geneflow` function that designates the timing of the exchange and makes sure that the populations overlap in space (if this is necessary).

4. model compilation

Up to this point, population boundaries have been stored in a fully vectorised form. The `compile` function converts this vectorised encoding to a series of raster maps (i.e. two-dimensional matrices) and saves them to disk in a format understood by SLiM (step 6).

5. model visualization

We provide an R *shiny* (<https://shiny.rstudio.com>) browser app (on the right) spawned by the `explore` function, which allows one to step through the spatio-temporal model and inspect it interactively before it is executed in SLiM (step 6).

6. execution in SLiM

A built-in template SLiM script which loads the compiled spatio-temporal model (step 4) is executed by the `slim` function (on the right). Here we save the location of every individual who ever lived in a 100 Mb tree-sequence file for further analysis (calculating summary statistics, genotype calling etc.). Users can also provide custom snippets of SLiM code to be run on top of the spatio-temporal demographic model.

Complete simulation of Holocene human history:

```
library(sleendr)

map <- world(xrange = c(-15, 60), yrange = c(20, 65),
  landscape = "natureearth", crs = "EPSG:3035")

# geographic regions used later to define population boundaries
europe <- region(map = map, polygon = list(
  c(-8, 35), c(10, 38), c(20, 35), c(25, 35),
  c(32, 45), c(20, 58), c(-5, 60), c(-15, 50))
)
anatolia <- region(map = map, polygon = list(
  c(28, 35), c(40, 35), c(42, 40),
  c(30, 43), c(27, 40), c(25, 38))
)

ehg <- population( # Eastern hunter-gatherers
  "EHG", time = 10000, N = 4000, map = map, remove = 7000,
  polygon = list(c(26, 55), c(38, 53), c(48, 53), c(60, 53),
    c(60, 60), c(48, 63), c(38, 63), c(26, 60))
)

ana <- population( # Anatolian farmers
  name = "ANA", time = 10000, N = 8000, remove = 4000,
  map = map, polygon = anatolia
) %>%
  expand(by = 2500e3, start = 10000, end = 7000,
    snapshots = 10, polygon = join(europe, anatolia))

eur <- population( # European population
  name = "EUR", time = 10000, N = 5000,
  map = map, polygon = europe
)

yam <- population( # Yamnaya steppe population
  name = "YAM", time = 7000, N = 3000, remove = 3000,
  parent = ehg, polygon = list(
    c(26, 50), c(38, 49), c(48, 50),
    c(48, 56), c(38, 59), c(26, 56)
  )
) %>%
  move(trajecory = c(15, 50), start = 5000, end = 3000,
    snapshots = 8)

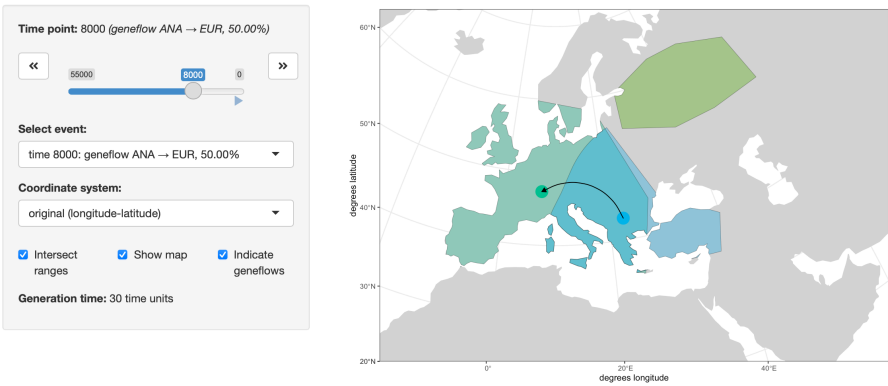
geneflows <- list(
  geneflow(ana, yam, start = 6500, end = 6000, rate = 0.3),
  geneflow(ana, eur, start = 8000, end = 6000, rate = 0.5),
  geneflow(yam, eur, start = 4000, end = 3000, rate = 0.75)
)

model <- compile(
  populations = list(ehg, ana, yam, eur), geneflows = geneflows,
  generation_time = 30, resolution = 10e3,
  competition_dist = 200e3, mate_dist = 200e3,
  offspring_dist = 100e3, dir = "/tmp/test-model"
)

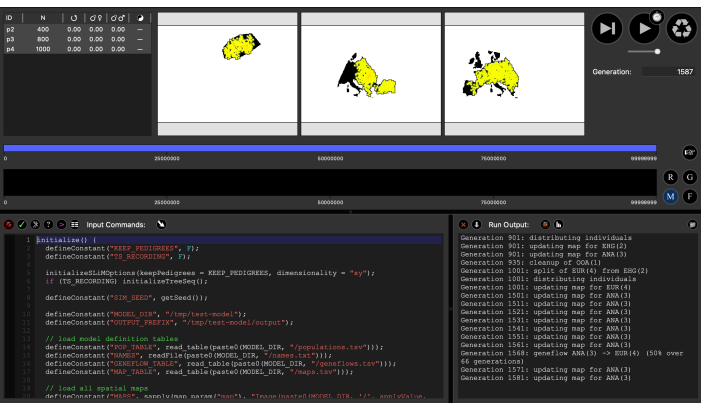
explore(model)

slim(
  model, seq_length = 100e6, recomb_rate = 1e-8,
  save_locations = TRUE, ts_recording = TRUE, method = "gui"
)
```

R shiny-based explorer of spatio-temporal models:



Spatio-temporal simulation running in SLiMgui:



Code, tutorial and an interactive demo: github.com/bodkan/smbe2021

¹ Lundbeck GeoGenetics Centre, Globe Institute, University of Copenhagen | ² Department of Computational Biology, Cornell University, Ithaca, New York | ³ Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon

This project is supported by Lundbeck Foundation, Novo Nordisk Foundation and Villum Foundation.