slendr: a framework for spatio-temporal population genomic simulations on geographic landscapes

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

We present a new R package, slendr (slendr.net), designed for reproducible, declarative, and visually-focused encoding of complex spatio-temporal population models on real and abstract geographic landscapes. slendr uses a tailor-made SLiM script (messerlab.org/slim) as a simulation back end bundled with the package, and saves spatially-annotated tree sequences as outputs from its models. In addition to spatial models, slendr offers a new way to simulate traditional, random-mating demographic models—either with SLiM or using an alternative back end implemented with msprime. With its R-idiomatic interface to the tree sequence analysis library tskit (tskit.dev), slendr opens up the possibility of efficient, reproducible population genetic simulations and analyses entirely in R. In this poster we demonstrate the features of the R package on several complete examples.

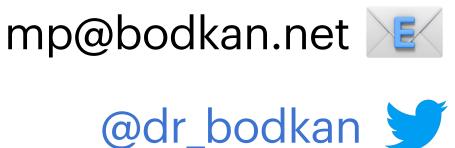
website: slendr.net

ancestry in "x1" and "x2"

tutorials: slendr.net/articles

preprint: biorxiv.org/content/10.1101/2022.03.20.485041v1

interactive examples from this poster: github.com/bodkan/probgen2022



Overview of a typical slendr workflow Model definition, simulation, and data analysis steps are all parts of a single reproducible R script,

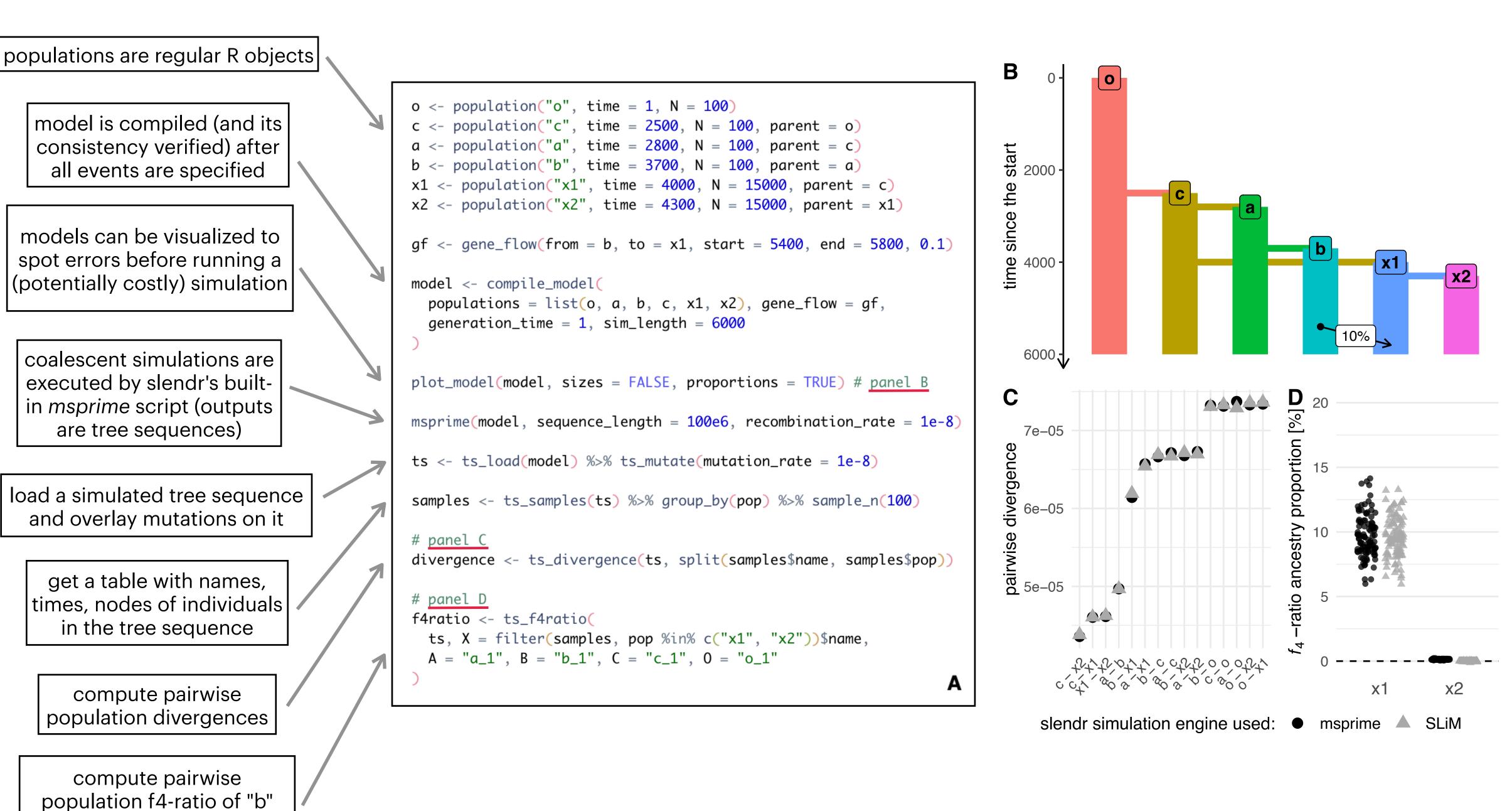
without having to write code in multiple languages or convert data between file formats.



run ex1.R in your browser on Binder: github.com/bodkan/probgen2022

Traditional, non-spatial demographic models slendr provides a new way to specify demographic models (population splits, population size

changes, gene-flow events) using a straightforward, declarative interface entirely in R. Models can be simulated by built-in simulation engines written in SLiM and msprime.



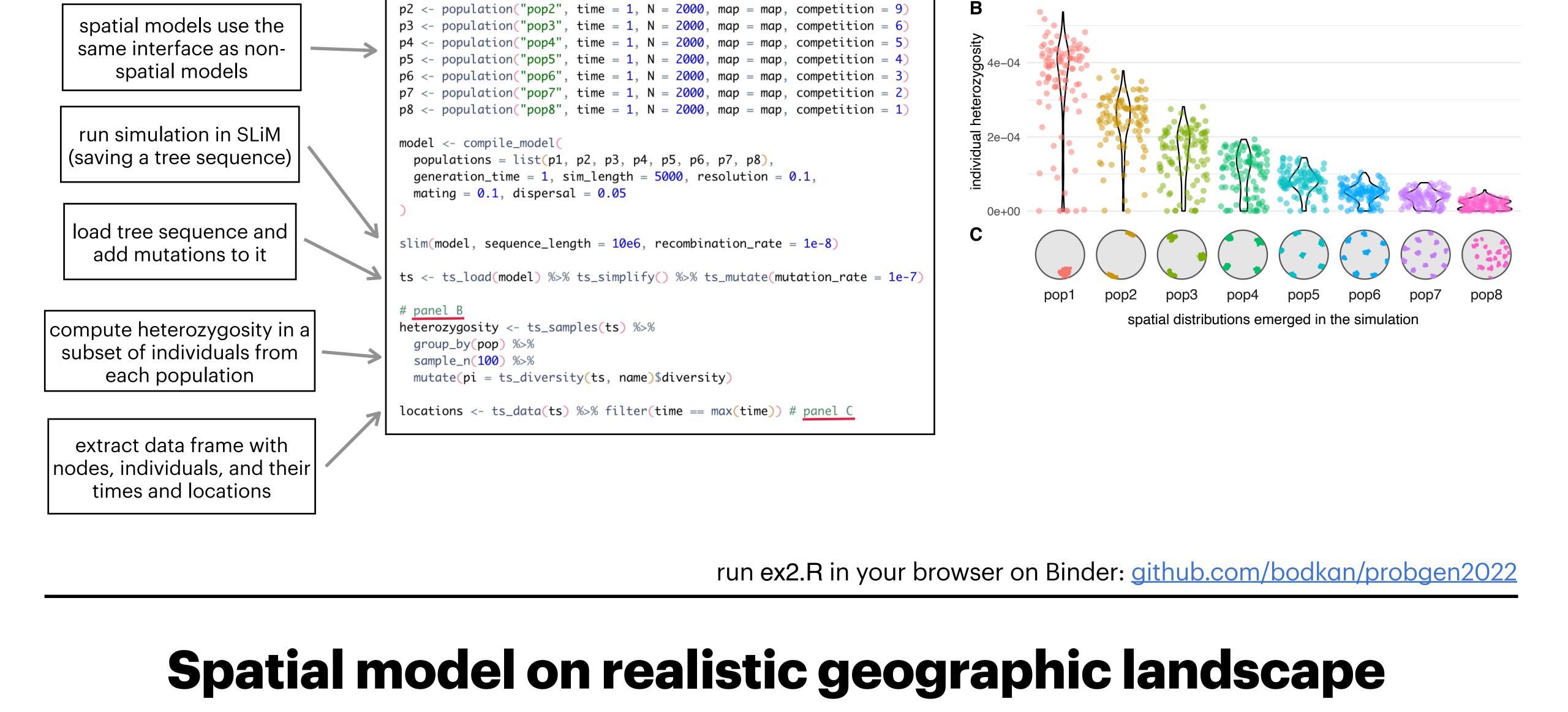
If the user defines a simulation world map (in this example an abstract, featureless map), the model can be simulated with a built-in SLiM back end script.

Spatial model on an abstract spatial landscape

map \leftarrow world(xrange = c(0, 10), yrange = c(0, 10), define a circular world map landscape = region(center = c(5, 5), radius = 5))

p1 <- population("pop1", time = 1, N = 2000, map = map, competition = 0)

 $p2 \leftarrow population("pop2", time = 1, N = 2000, map = map, competition = 9)$

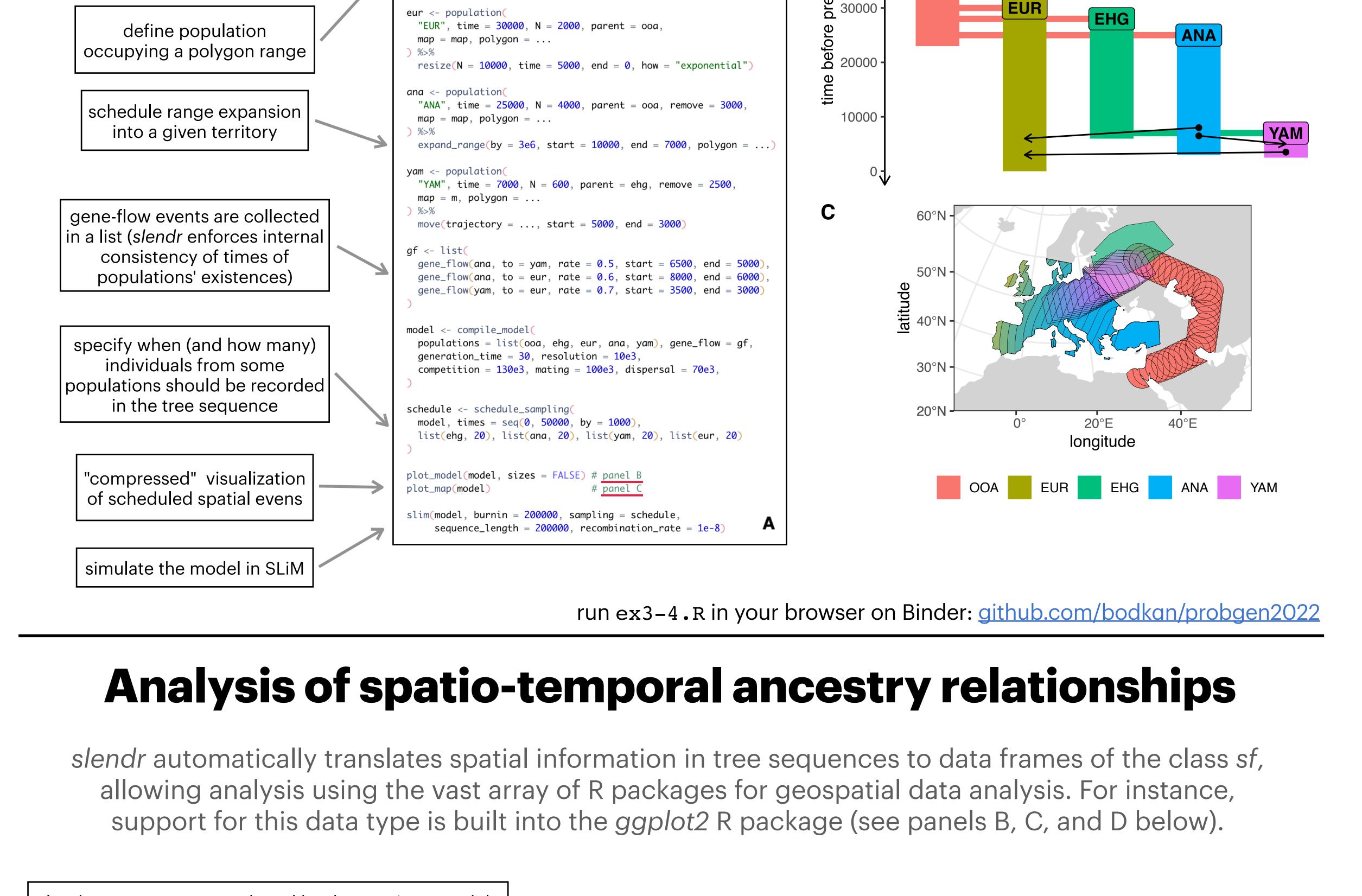


slendr allows scheduling of large-scale population migrations and range expansions using a set of dedicated functions, without the need for handling spatial geometric operations. These events can occur on abstract landscapes but can be also defined on realistic regions on Earth (such as in this example).

download geographic spatial <- world(xrange = c(-15, 60), yrange = c(20, 65), crs = 3035) features of West Eurasia

ooa <- population(

"00A", time = 50000, N = 500, remove = 23000**B** 50000 map = map, center = c(33, 30), radius = 400e3OOA define a circular population, move(trajectory = ..., start = 50000, end = 40000)schedule a movement ehg <- population(</pre> 40000 "EHG", time = 28000, N = 1000, parent = ooa, remove = 6000, along a given trajectory present map = map, polygon = ...



load tree sequence produced by the previous model (and make a smaller version of it through simplification)

convert the 10th tree of the tskit tree sequence into ts <- ts_load(model) the phylogenetic format of the ape R package ts_small <- ts_simplify(ts, c("EUR_599", "ANA_322", "EHG_7", "EUR_578", "EUR_501", "YAM_30")) tree <- ts_phylo(ts_small, i = 10) # panel B nodes <- ts_data(tree)</pre> # panel C extract tables of spatial locations branches <- ts_branches(tree)</pre> # panel C of each node and edge in the tree

ancestors <- ts_ancestors(ts, "EUR_599") # panel D</pre>

collect the locations (and times) of all ancestral nodes of a given individual across the entire tree sequence D B (EUR_599) EUR_599 (chromosome 1, node 10) EUR_599 (chromosome 2, node 11) 60°N (EUR_578) 8 (EUR_578) 50°N latitude N°05 6 (EUR_501) (ANA_322) (EHG_7) 0 (EUR_599) 30°N 7 (EUR_501)

2 (ANA_322)

0 (EHG_7)

40000

NIH award R01HG010774.

20000

time before present [years ago]

run ex3-4.R in your browser on Binder: github.com/bodkan/probgen2022

40°E

20°E

0°

longitude

40°E

20°E

1. Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Denmark, 2. Department of Computational Biology, Cornell University, Ithaca, NY, USA,

supported by a Lundbeck Foundation grant (R302-2018-2155) and a Novo Nordisk Foundation grant (NNF18SA0035006) given to the GeoGenetics Centre. PR was supported by

20°N

30°E

longitude

10°E

20°E

40°E

50°E

3. Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA /// FR was supported by a Villum Young Investigator Grant (project no. 00025300). MP was