Individual-based simulations of genome evolution with ancestry: the GENOMEADMIXR R package

Thijs Janzen[1,2*], Fernando Diaz[3]

[1] Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 111039700 CC Groningen, The Netherlands

[2] Carl von Ossietzky University, Carl-von-Ossietzky-Str. 9–11, 26111, Oldenburg, Germany

[3] Department of Entomology, University of Arizona, Tucson, AZ, United States of America

[*] t.janzen@rug.nl

Running headline: the GENOMEADMIXR R package

ABSTRACT

1. Hybridization between populations or species results in a mosaic of the two parental genomes. Genome admixture has received increasing attention for its implications in speciation, human evolution, Evolve and Resequencing (E&R) and genetic mapping. However, a thorough understanding of how local ancestry changes after admixture, and how selection affects patterns of local ancestry remains elusive. The complexity of these questions limits analytical treatment, but these scenarios are specifically suitable for simulation. Currently, there is no simulation framework that models whole-genome evolution of local ancestry following admixture.

2. Here, we present the R package GenomeAdmixR, which uses an individual-based model to simulate genomic patterns of local ancestry following admixture forward in time. GenomeAdmixR provides user-friendly functions to set up and analyze simulations under evolutionary scenarios with selection, linkage and migration.

3. We show the flexible functionality of the GenomeAdmixR workflow by demonstrating 1) how to design an E&R simulation using GenomeAdmixR and 2) how to use GenomeAdmixR to verify analytical expectations following from the theory of junctions.

4. GenomeAdmixR provides a mechanistic approach to explore expected genome responses to realistic admixture scenarios. With this package, we aim to aid researchers in testing specific hypotheses based on empirical findings involving admixing populations.

Keywords: Genome evolution, admixture, ancestry, individual-based modeling, Evolution and Resequencing.

43 **INTRODUCTION**

44 Genetic exchange has long been recognized as an important driver of
45 genetic diversity, from the recombination of conspecific genomes and the
46 evolution of speciation with migration (Abbott *et al.* 2013) to introgressive
47 hybridization (Janzen *et al.* 2018; Lavretsky *et al.* 2019) and horizontal gene
48 transfer (Keeling & Palmer 2008). With recent advances in sequencing
49 technologies, the view of genetic exchange has now moved from information at
50 specific loci (*e.g.* alleles, genes) to the level of whole genome admixture (Leitwein
51 *et al.* 2018; Lavretsky *et al.* 2019; Chafin & Douglas 2020). The focus is now on
52 understanding how major drivers of evolution (*i.e.* recombination, selection and
53 migration) shape the linear connection of loci along the genome (*i.e.* synteny and
54 linkage) and how these processes evolve under particular scenarios of genome
55 divergence. Understanding these dynamics has helped to elucidate the impact of
56 migration on human evolution (Hellenthal *et al.* 2014; Payseur & Rieseberg
57 2016), the impact of hybridization on speciation (Schumer *et al.* 2014b, 2018),
58 and helped as well in designing experiments relying on admixture, for instance
59 Evolve and Resequencing (E&R) experiments (Franssen *et al.* 2017; Barghi &
60 Schlötterer 2019; Otte & Schlötterer 2020).

61 Recombination is one of the main drivers shaping macro-genomic
62 patterns after admixture, where contiguous blocks of ancestry are broken down
63 into smaller blocks over time (Fisher 1954, 1959). As time progresses, genomes
64 transform into macro-genomic mosaics consisting of many small contiguous
65 ancestry blocks. Our understanding of how these mosaics form is relatively
66 robust, but analytical treatments are only known for boundary cases, such as
67 admixture between two distant populations, or for restricted backcrossing

68 (Fisher 1954, 1959; Stam 1980; Macleod *et al.* 2005; Janzen *et al.* 2018;
69 Lavretsky *et al.* 2019).

70 Linkage patterns can also be generated by selection through genetic
71 hitchhiking or selection of epistatic interactions, which can alter genome-wide
72 patterns of genetic variation, population dynamics and evolvability (Gerrish *et al.*
73 2007; Arnold & Kunte 2017; Zhou *et al.* 2017). Thus, macro-genomic patterns
74 can be driven both by recombination and selection alike. However, a
75 mathematical treatment of the interaction between selection, recombination and
76 admixture is currently lacking. In contrast, individual-based simulations are very
77 suitable to explore these interactions, for instance using packages such as
78 `MimicrEE2` (Vlachos & Kofler 2018) (written in Java) and `SLiM` (Haller &
79 Messer 2017, 2018) (written in C++). Unfortunately, these do not explicitly
80 track local ancestry, and information on haplotypes is often only indirectly
81 available or requires extensive post-processing. `forqs` (Kessner & Novembre
82 2014) and `SELAM` (Corbett-Detig & Jones 2016) (both written in C++) remedy
83 this issue and are specifically focused on tracking local ancestry and inheritance
84 of ancestry blocks, however these lack ease of use (requiring for instance local
85 compilation of C++ code, and configuration of input files) and are not easily
86 applied cross-platform. Here, we present GENOMEADMIXR, a software package
87 written in the R language. The R language is readily used within biology and can
88 easily be used cross-platform. GENOMEADMIXR provides routines to simulate
89 admixture in a host of scenarios (Figure 1), and includes a wide range of routines
90 available to analyze and visualize simulation results.

91

MATERIALS and METHODS

Description

GENOMEADMIXR is written in the R programming language (Team 2020). Similar to SliM (Haller & Messer 2017, 2018) and SELAM (Corbett-Detig & Jones 2016), GENOMEADMIXR simulates a population forward in time using a Wright-Fisher model with non-overlapping generations and a constant population size. GENOMEADMIXR simulated diploid individuals that sexually reproduce. For computational tractability, only one pair of chromosomes is simulated an all individuals are assumed to be hermaphroditic. Recombination is modeled in the same way as in for instance SLiM and SELAM (Corbett-Detig & Jones 2016; Haller & Messer 2018), i.e. as a Poisson process, with the number of crossovers Poisson distributed with the size in Morgan of the chromosome as rate parameter. The location of crossovers is drawn from a uniform distribution, without interference. The simulation tracks the locations of changes in local ancestry (junctions) (like SELAM) but does not explicitly model nucleotides (as for instance SLiM).

Usage

A development version can be installed from Github using the "devtools" package:

```
> install.packages("devtools")
> devtools::install_github("thijsjanzen/GenomeAdmixR")
```

Alternatively, soon a version will be available from CRAN:

```
> install.packages("GenomeAdmixR")
```

The core of the GENOMEADMIXR package is formed by the simulation function `simulate_admixture` (Table 1). To simulate a simple admixture scenario where two distinct populations hybridize to form a population, which continues to admix for 100 generations, we write:

```
> pop <- simulate_admixture(pop_size = 1000, number_of_founders = 2, total_runtime =
100)
```

Molecular markers and statistic estimations

To simulate uncertainty in tracking local ancestry, in line with molecular methods (Schumer *et al.* 2014a; Dennenmoser *et al.* 2019; Lavretsky *et al.* 2019), GENOMEADMIXR includes functionality to track local ancestry at molecular markers located along the genome. Resulting local ancestry information is returned from the function in long format, which facilitates downstream analysis. The number of used ancestral genomes will therefore represent the alternative alleles at marker position. Consequently, using two ancestral genomes would simulate a biallelic model.

Continuing along the selected scenario, but tracking 1000 markers along the genome, one would get:

```
> pop <- simulate_admixture(pop_size = 1000, number_of_founders = 2, total_runtime =
100, markers = seq(0, 1, length.out = 1000))
```

Based on these markers, a set of genetic diversity statistics can be calculated, including allele frequency (`calculate_allele_frequencies`, Table 1), heterozygosity (`calculate_heterozygosity`, Table 1), linkage

disequilibrium (`calculate_ld`, Table 1), Tajima's D (`calculate_tajima_d`, Table 1) and F_{ST} (`calculate_fst`, Table 1), as well as changes of these over generations for better assessing significance of simulated evolution. Caution should be taken when more than four distinct ancestries are included in the simulation, because most genetic diversity statistics assume a maximum nucleotide number of four. For a single population we can estimate Linkage Disequilibrium (LD), average heterozygosity and tajima's D:

```
> ld <- calculate_ld(pop)
> het <- calculate_heterozygosity(pop)
> td <- calculate_tajima_d(pop)
```

Visualization

`Simulate_admixture` by default returns the frequencies of all ancestors at the markers specified at the start and the end of the simulation. Allele frequency changes can be visualized using `plot_difference_frequencies_` (Table 1, Figure 2) or `plot_start_end` (Table 1, Figure 2) and using `plot_frequencies` (Table 1, Figure 2) the final frequencies are plotted. The functions `plot_over_time` (Table 1, Figure 2) and `plot_joyplot_frequencies` (Table 1, Figure 2) provide functionality to show allele frequency changes over time. All plotting functions return ggplot2 (Wickham 2009) objects, which can be further customized by the user.

Migration

The function `simulate_admixture_migration` evolves two independent populations in parallel, given a fraction of exchange of parental individuals per generation.

```
> migr_pop <- simulate_admixture_migration(total_runtime = 1000, pop_size = c(100, 100), migration_rate = 0.01)
```

Then, we can analyze genetic divergence across these two populations using the F_{ST} statistic:

```
> calculate_fst(migr_pop$population_1, migr_pop$population_2)
```

There is functionality included to stop the simulation once a threshold of genetic differentiation (*i.e.* F_{ST}) between the two populations is reached.

Selection

In the absence of selection, allelic frequencies at haplotype sequences of ancestral types remain constant, barring drift effects due to finite population size. `simulate_admixture` provides additional options to impose fitness benefits upon individuals that contain a marker under selection (where selection is interpreted as favouring alleles from a specific ancestor). We follow conventional fitness notation (Crow & Kimura 1970) and denote [w_{aa}, w_{Aa}, w_{AA}] = [*1, 1+hs, 1+s*], where *w* indicates fitness, *s* indicates the selective benefit of allele *A*, and *h* indicates the degree of dominance. Parents of offspring in the next

generation are drawn from the current parental population proportional to fitness and the probability of selecting individual parent j as one of the parents of an offspring in the next generation is given by:

$$P(j) = \frac{w_j}{\sum_{i=1}^{N} w_i}$$

The selection matrix [w_{aa}, w_{Aa}, w_{AA}] can be provided by the user, allowing for the implementation of a wide range of potential selective benefits upon receiving a marker, including overdominance and epistasis. Furthermore, the user is not restricted to only providing selection on a single marker, and can expand the selection matrix by adding extra rows for each marker. If multiple loci are under selection, the fitness of an individual is the product of the finesses at each locus. An example of adding selection to the previously selected hybridization scenario is:

```
> s_matrix <- as.matrix(t(c(0.5, 1, 1.1, 1.5, 0)))
> pop <- simulate_admixture(pop_size = 1000, number_of_founders = 2, total_runtime =
100, markers = seq(0, 1, length.out = 1000), select_matrix = s_matrix)
```

Demonstration of the package

To demonstrate the functionality of GENOMEADMIXR, we firstly show a series of simulations to compare alternative strategies in Evolution & Resequencing (*E&R*) experiments. Secondly, we show how simulations of GENOMEADMIXR match analytical predictions in junction theory.

E&R experiment

219 One of the goals in *E&R* experiments is to identify underlying genetic loci driving
220 responses to selection (Burke & Rose 2009; Schlötterer *et al.* 2015). Increased
221 genetic variability in the starting population improves the chances of capturing
222 additional candidate loci, but LD when sampling diverging populations can mask
223 the changes at individual positions (Kessner & Novembre 2015). We compare
224 two different sampling methods to increase genetic variability. *Model 1* starts
225 with a well-mixed population formed from 100 isofemale lines sampled from a
226 single (wild) population. *Model 2* aims to maximize genetic variability by
227 sampling two isofemale lines from different and diverging populations. Using
228 GENOMEADMIXR, we simulate the entire E&R pipeline and explore the impact of
229 both sampling schemes. Figure S1 summarizes the entire simulation setup and R
230 code to replicate our findings can be found in the supplementary material. Figure
231 3 (for a single simulation) indicates that the resolution to detect the marker
232 under selection in the experiment was much higher in *Model 1.* Also, *Model 1*
233 seems to experience less genetic hitchhiking compared to *Model 2*. We
234 performed 10 independent simulations with the same results, and replicates can
235 be used for further analysis and statistical tests. Overall, the obtained results
236 indicate that *Model 1* is for this pipeline the preferred approach (see for a more
237 detailed analysis the Supplementary Material).
238
239 *Theory of junctions*
240 From the extended theory of junctions (Janzen *et al.* 2018) we know that the
241 expected number of junctions (where a junction delineates the end of one
242 contiguous stretch of genomic content from the same ancestor, and the start of
243 another), depends on the initial heterozygosity H_0. Janzen's extended theory of

244 junctions focuses specifically on the scenario with two ancestors, whereas
245 GENOMEADMIXR allows for an arbitrary number of ancestors. We simulate the
246 accumulation of junctions using GENOMEADMIXR for two population sizes (100
247 and 1000 individuals) and vary the number of ancestors *n* in [2, 4, 8]. We find
248 that when we average the number of accumulated junctions over 1000
249 replicates, the average number of junctions closely follows that of our analytical
250 expectation (Figure 4).
251
252 **DISCUSSION**
253 We have presented here a new R package that uses individual-based
254 simulations to study patterns in local ancestry after admixture. We have shown
255 how GENOMEADMIXR can be used within an experimental evolution framework
256 and how results obtained using GENOMEADMIXR are in line with theoretical
257 expectations.
258 We expect GENOMEADMIXR to be applicable in many fields. Firstly, we
259 expect GENOMEADMIXR to allow for improved ease of use of individual-based
260 simulations to study patterns in the formation of contiguous tracts of ancestry,
261 and their breakdown due to recombination. The study of the breakdown of
262 ancestry blocks has recently attracted renewed interest, with studies expanding
263 the theory of junctions (the delineations between two ancestry tracts) towards
264 potential marker effects (Janzen *et al.* 2018), studies utilizing patterns in the
265 distribution of ancestry tracts to infer the onset of hybridization (Ungerer *et al.*
266 1998; Buerkle & Rieseberg 2008; Pool & Nielsen 2009; Gravel 2012; Liang &
267 Nielsen 2014; Corbett-Detig & Nielsen 2017) and the inclusion of ancestral tract
268 information in the analysis of human history (Hellenthal *et al.* 2014; Payseur &

269 Rieseberg 2016). `GenomeAdmixR` complements these analyses by providing a
270 framework to verify findings through simulation.

271 Secondly, we expect GENOMEADMIXR to be an important tool in
272 investigating population genomics and designing *E&R* experiments, where
273 GENOMEADMIXR can be used to test whether the setup of an experiment is
274 sufficient to detect signatures of selection on a background of existing LD. The
275 functionality of the package allows simulating the effect of *i)* previous population
276 dynamics that occurred before sampling (*natural populations*), *ii)* initial
277 parameters when obtaining the starting population for selection (*mixed*
278 *populations*) as well as *iii)* variation in genomic and *selection* parameters. Here,
279 we demonstrate how to use GENOMEADMIXR to estimate the level of resolution
280 obtained when combining inbred lines from different populations in the starting
281 population used in an E&R experiment (Parts *et al.* 2011; Koide *et al.* 2012). We
282 find that although this model narrowed down the haplotype blocks around the
283 selected maker substantially, this still lacks resolution when compared with
284 traditional sampling (Kofler & Schlötterer 2014; Schlötterer *et al.* 2015) from a
285 single population (even after 30 admixing generations). However, the obtained
286 resolution seems deep enough to investigate questions involving variation
287 between evolutionarily independent entities if these are not possible to address
288 by sampling a single population. Haplotype block dynamics have always been
289 important for understanding *E&R* experiments, but have recently been used to
290 explain massive genomic regions responding to selection in these studies
291 (Orozco-Terwengel *et al.* 2012; Turner & Miller 2012; Tobler *et al.* 2013),
292 limiting the ability to differentiate causal variant from hitchhiking regions.
293 Recent methods focus on haplotype block estimations and analyses (Franssen *et*

294 *al.* 2017; Burny *et al.* 2020; Otte & Schlötterer 2020) so much so that the
295 paradigm is shifting from analysis of single nucleotide polymorphisms to
296 selected haplotype blocks (Barghi & Schlötterer 2019).

297 Thirdly, we expect GENOMEADMIXR function as a useful teaching tool,
298 where it can be used to demonstrate the impact of hybridization on linkage
299 patterns, the subsequent interaction between selection and drift and more
300 generally to provide an easy toolkit to explore the interaction between
301 recombination and hybridization.

302 GENOMEADMIXR shares many similarities with it's predecessors SliM
303 (Haller & Messer 2017, 2018) and SELAM (Corbett-Detig & Jones 2016), both
304 powerful admixture simulation programs. However, SLiM is mainly focused on
305 more advanced population demographic scenarios and is not directly suited for
306 tracking ancestry. In contrast, SELAM does track local ancestry, but requires
307 local compilation of C++ code (which can be difficult) and uses complicated input
308 tables to parameterize the simulations. GENOMEADMIXR aims to alleviate both
309 these issues, firstly GENOMEADMIXR is specifically focused on tracking ancestry,
310 and it's admixture, over time. Secondly, because GENOMEADMIXR is an R package,
311 parameterization can be done through the use of the R scripting language, which
312 facilitates sharing of code between platforms, replication of approaches and ease
313 of use.

314 In conclusion, we believe that GENOMEADMIXR has a bright future ahead,
315 with a myriad of potential implementations across different fields of population
316 genetics and molecular ecology.

317

318 **AUTHOR CONTRIBUTIONS**

TJ and FD conceived the model, TJ developed the R code, FD and TJ jointly wrote and revised the manuscript.

DATA AVAILABILITY

A development version of the package can be found at https://github.com/thijsjanzen/GenomeAdmixR, and a stable version will soon be available on CRAN (https://cran.r-project.org/package=GenomeAdmixR).

ACKNOWLEDGEMENTS

We thank Omer Markovitch, Pratik R. Gupte and Cyrus Mallon for helpful comments on an earlier version of the manuscript. The authors indicate that they have no conflicts of interest.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, a., Buerkle, C. a., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, a., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, a. G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, a. W., Parisod, C., Pfennig, K., Rice, a. M., Ritchie, M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf, J.B.W. & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Arnold, M.L. & Kunte, K. (2017). Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation. *Trends in Ecology & Evolution*, **32**, 601–611.

Barghi, N. & Schlötterer, C. (2019). Shifting the paradigm in evolve and resequence studies: From analysis of single nucleotide polymorphisms to selected haplotype blocks. *Molecular Ecology*, **28**, 521–524.

Buerkle, C.A. & Rieseberg, L.H. (2008). The rate of genome stabilization in homoploid hybrid species. *Evolution*, **62**, 266–275.

Burke, M.K. & Rose, M.R. (2009). Experimental evolution with Drosophila. *Am J Physiol Regul Integr Comp Physiol*.

Burny, C., Nolte, V., Nouhaud, P., Dolezal, M., Schlötterer, C. & Baer, C. (2020). Secondary Evolve and Resequencing: An Experimental Confirmation of Putative Selection Targets without Phenotyping. *Genome Biology and Evolution*, **12**, 151–159.

Chafin, T.K. & Douglas, M.R. (2020). Genome-wide local ancestries discriminate homoploid hybrid speciation from secondary introgression in the red wolf (Canidae : Canis rufus). 1–49.

Corbett-Detig, R. & Jones, M. (2016). SELAM: Simulation of epistasis and local adaptation during admixture with mate choice. *Bioinformatics*, **32**, 3035–3037.

Corbett-Detig, R. & Nielsen, R. (2017). A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genetics*, **13**, 1–40.

Crow, J.F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper & Row, New York. Retrieved August 25, 2020, from https://www.cabdirect.org/cabdirect/abstract/19710105376

Dennenmoser, S., Schatz, M.C., Sedlazeck, F.J., Zytnicki, M., Nolte, A.W. & Altmüller, J. (2019). Genome-wide patterns of transposon proliferation in an evolutionary young hybrid fish. *Molecular Ecology*, 0–2.

Fisher, R.A. (1954). A fuller theory of "Junctions" in inbreeding. *Heredity*, **8**, 187–197.

Fisher, R.A. (1959). An algebraically exact examination of junction formation and transmission in parent-offspring inbreeding. *Heredity*, **13**, 179–186.

Franssen, S.U., Barton, N.H. & Schlotterer, C. (2017). Reconstruction of haplotype-blocks selected during experimental evolution. *Molecular Biology and Evolution*, **34**, 174–184.

Gerrish, P.J., Colato, A., Perelson, A.S. & Sniegowski, P.D. (2007). Complete genetic linkage can subvert natural selection. *PNAS*, **104**, 6266–6271.

Gravel, S. (2012). Population genetics models of local ancestry. *Genetics*, **191**, 607–619.

Haller, B.C. & Messer, P.W. (2017). SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, **34**, 230–240.

Haller, B.C. & Messer, P.W. (2018). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, **36**, 632–637.

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D. & Myers, S. (2014). A Genetic Atlas of Human. *Science*, **343**, 747–751.

Janzen, T., Nolte, A.W. & Traulsen, A. (2018). The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. *Evolution*, **72**, 735–750. Retrieved from http://doi.wiley.com/10.1111/evo.13436

Keeling, P.J. & Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, **9**, 605–618.

Kessner, D. & Novembre, J. (2014). Forqs: Forward-in-time simulation of recombination, quantitative traits and selection. *Bioinformatics*, **30**, 576–577.

Kessner, D. & Novembre, J. (2015). Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. *Genetics*, **199**, 991–1005.

Kofler, R. & Schlötterer, C. (2014). A guide for the design of evolve and resequencing studies. *Molecular Biology and Evolution*, **31**, 474–483.

Koide, T., Goto, T. & Takano-Shimizu, T. (2012). Genomic mixing to elucidate the genetic system of complex traits. *Experimental Animals*, **61**, 503–509.

Lavretsky, P., Janzen, T. & McCracken, K.G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, **9**, 3470–3490.

Leitwein, M., Gagnaire, P.A., Desmarais, E., Berrebi, P. & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, **27**, 3466–3483.

Liang, M. & Nielsen, R. (2014). The lengths of admixture tracts. *Genetics*, **197**, 953–967.

Macleod, A.K., Haley, C.S., Woolliams, J.A. & Stam, P. (2005). Marker densities and the mapping of ancestral junctions. *Genetical Research*, **85**, 69–79.

Orozco-Terwengel, P., Kapun, M., Nolte, V., Kofler, R., Flatt, T. & Schlãtterer, C. (2012). Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular Ecology*, **21**, 4931–4941.

Otte, K.A. & Schlötterer, C. (2020). Detecting selected haplotype blocks in Evolve and Resequence experiments. *Molecular Ecology Resources*, 1–37.

Parts, L., Cubillos, F.A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S.J., Molin, M., Zia, A., Simpson, J.T., Quail, M.A., Moses, A., Louis, E.J., Durbin, R. & Liti, G. (2011). Revealing the genetic structure of a trait by sequencing a population under selection. 1131–1138.

Payseur, B.A. & Rieseberg, L.H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Pool, J.E. & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, **181**, 711–719.

Schlötterer, C., Kofler, R., Versace, E., Tobler, R. & Franssen, S.U. (2015). Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity*, **114**, 431–440.

Schumer, M., Cui, R., Powell, D.L., Dresner, R., Rosenthal, G.G. & Andolfatto, P. (2014a). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife*, **2014**, 1–21.

Schumer, M., Rosenthal, G.G. & Andolfatto, P. (2014b). How common is homoploid hybrid speciation? *Evolution*, **68**, 1553–1560.

Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, C., Sankararaman, S., Andolfatto, P., Rosenthal, G.G. & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, **360**, 656–660.

Stam, P. (1980). The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetical Research*, **35**, 131.

Team, R.C. (2020). R: A Language and Environment for Statistical Computing.

Tobler, R., Franssen, S.U., Kofler, R., Orozco-terwengel, P., Nolte, V., Hermisson, J. & Schlotterer, C. (2013). Massive Habitat-Specific Genomic Response in D . melanogaster Populations during Experimental Evolution in Hot and Cold Environments. *Molecular Biology and Evolution*.

Turner, T.L. & Miller, P.M. (2012). Investigating natural variation in drosophila courtship song by the evolve and resequence approach. *Genetics*, **191**, 633–642.

Ungerer, M.C., Baird, S.J., Pan, J. & Rieseberg, L.H. (1998). Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 11757–11762.

Vlachos, C. & Kofler, R. (2018). MimicrEE2: Genome-wide forward simulations of Evolve and Resequencing studies. *PLoS Computational Biology*, **14**, 1–10. Retrieved from http://dx.doi.org/10.1371/journal.pcbi.1006413

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis - Hadley Wickham - Google Books*. Retrieved September 9, 2020, from https://books.google.nl/books?hl=nl&lr=&id=XgFkDAAAQBAJ&oi=fnd&pg=PR8&dq=ggplot2&ots=so58bP5WbN&sig=ThS6gEgxaK9XADL_HeG5gDU04Pc#v=onepage&q=ggplot2&f=false

Zhou, Y., Qiu, H. & Xu, S. (2017). Modeling Continuous Admixture Using Admixture-Induced Linkage Disequilibrium. *Scientific Reports*, **7**, 1–10.

Figures

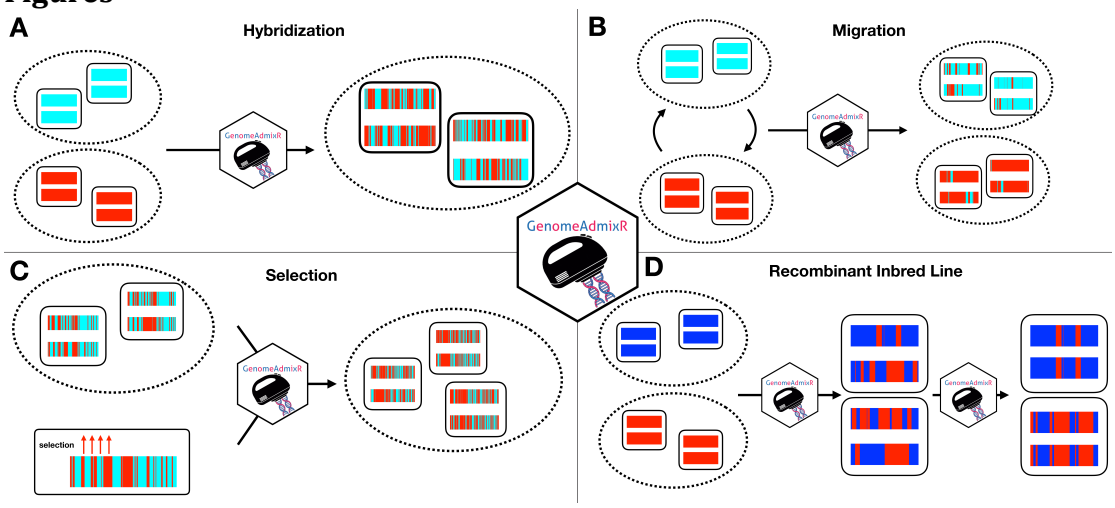

Figure 1. Overview of applications of the GENOMEADMIXR package. The GENOMEADMIXR package can be used to **A**) simulate admixture of genomes (rectangular, colored) over time, passed on between individuals (rounded squares) when two populations (dotted lines) meet. **B**) Simulate ongoing genetic exchange under migration, **C**) explore the impact of selection for alleles from the red ancestor in the first section of the genome and **D**) Simulate the formation of Recombinant Inbred Lines, where continued inbreeding generates fully homozygous.

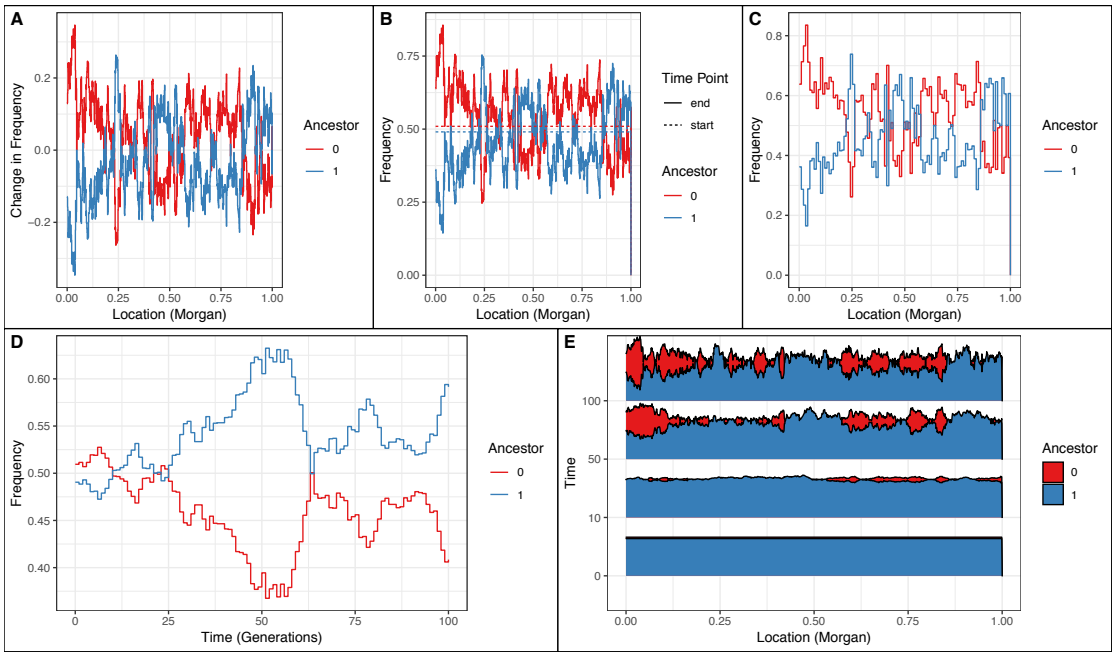

Figure 2. Example visualizations available in GenomeAdmixR. Shown are example plots after an hybridization scenario between two distinct ancestral populations, parameter values used are: population size = 10,000, total runtime = 100. 1000 markers evenly spaced in [0, 1] were used. **A**: `plot_difference_frequencies`, which visualizes the average change in frequency between the start and end of the simulation **B**: plot_start_end, which plots the average frequencies per ancestor, both at the start and the end of the simulation. **C**: plot_frequencies, which plots the average frequency of each ancestor across the genome **D**: plot_over_time, which plots the average frequency of a single marker (here a marker at 50 cM) in the population, over time. **E**: plot_joyplot_frequencies, which visualizes the average frequency over time.

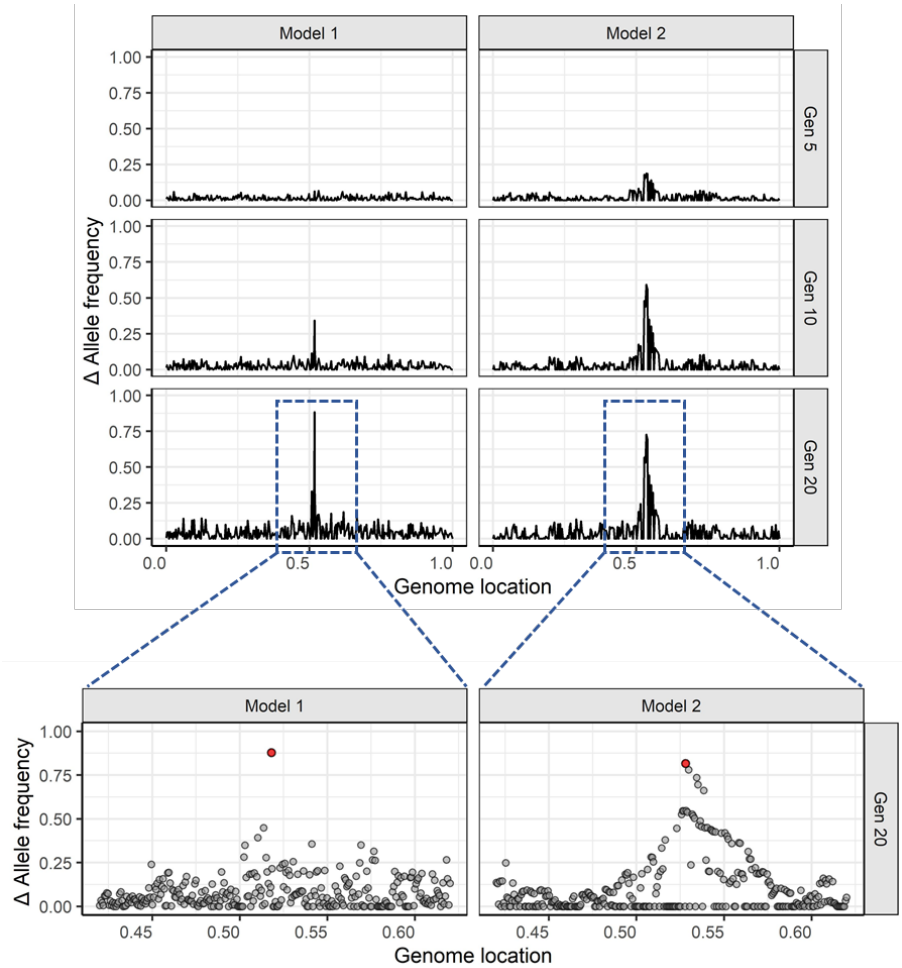

Figure 3. Summary of results obtained after simulating two scenarios for an *E&R* experiment using the GENOMEADMIXR package. Genome-wide Δ allele frequency after 5, 10 and 20 generations of selection are compared between *Model 1* and *Model 2*. The models differ in starting populations, with *Model* 1 only founded by 100 isofemales form a single population, and *Model* 2 founded by two isofemales form diverging populations. Selection was simulated for only one location (red point in the lower zoomed panel) and tracked in 300 biallelic markers, while the rest of the genome evolves neutrally.

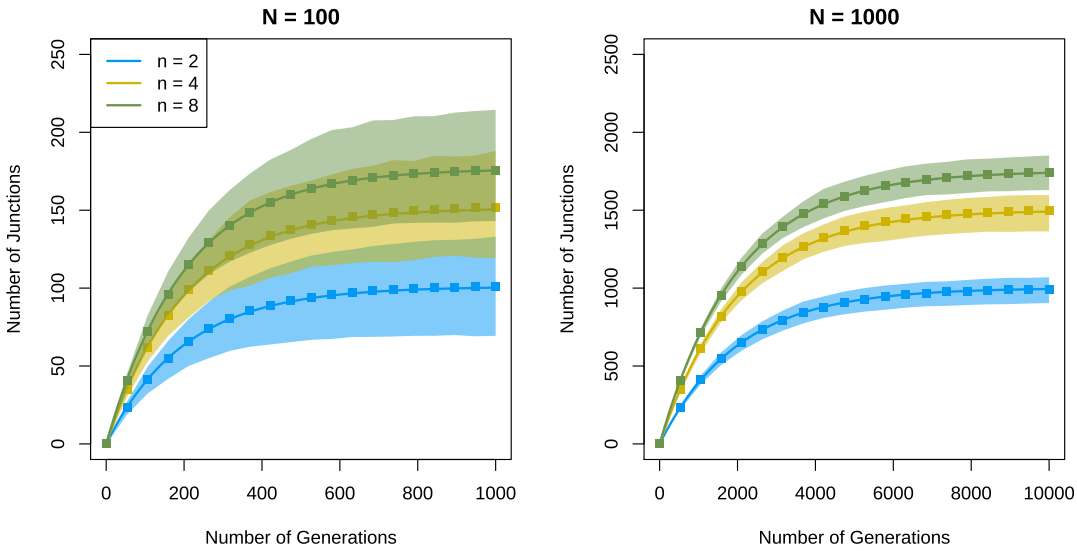

Figure 4. Accumulation of junctions for different numbers of ancestors (2, 4 and 8 unique ancestors), for a small population (N = 100) and a population of intermediate size (N = 1000). Square dots indicate the mean number of junctions observed across 1000 replicate simulations with the GENOMEADMIXR package. Shaded areas indicate the 95% Confidence Interval across 1000 replicates. The solid lines indicate the analytical prediction following equation 1. Mean simulation dynamics follow the analytical prediction very closely.

Table 1. Overview of available functions in the package

Simulation	Input	Output	Summary
`simulate_admixture`	- input population (if available) - population size - number of unique ancestors (if no input population was given) - initial frequency of each unique ancestor - number of generations to simulate - size of chromosome in Morgan - selection matrix (see text) - a vector with locations of molecular markers	- Admixed population - tracked allele frequencies through time	Forward simulate a population
`simulate_admixture_migration`	Same input as `simulate_admixture` except: - two input populations (if available) - population sizes for both populations - migration rate - possibility to stop at a critical Fst value	- Two admixed populations - tracked allele frequencies through time	Forward simulate two populations connected via a fixed rate of migration *m*
Statistical methods			
`calculate_heterozygosity`	- input population - vector with locations of markers	- population average heterozygosity	Calculates average heterozygosity
`calculate_fst`	- two input populations - number of individuals sampled to calculate the statistic - number of markers sampled to calculate the statistic	Fst value	Calculates the Weir & Cockerham Fst for a set of markers
`calculate_LD`	- input population - number of individuals sampled to calculate the statistic - number of markers sampled to calculate the statistic	- pair wise linkage disequilibrium values for all markers - pair wise R^2 values for all markers	Calculates all pairwise LD and R^2 values for a set of markers
`calculate_tajima_D`	- input population - number of individuals sampled to calculate the statistic - vector with locations of markers used	- Tajima's D - π : the average pairwise difference between markers - S (the number of segregating sites) - $\hat{\theta}$: the expectation of π	Calculates Tajima's D
`calculate_marker_frequency`	- input population - locations of markers	- a table containing the frequency of each ancestor at the provided marker locations.	Estimates the frequency of ancestors at a marker location within the population
`calculate_allele_frequencies`			Estimates the frequency of ancestors in the population for

			sliding windows across the genome.
Visualization			
`plot.individual`	- genome of an individual (result of `simulate_admixture`)	Visualization of ancestry along both chromosomes, with different ancestry represented by different colors.	Plots both chromosomes of an individual, where colors represent ancestors
`plot_chromosome`	- chromosome of an inividual	Visualization of ancestry along one chromosome, with different ancestry represented by different colors.	Visualize ancestry long a chromosome using different colors for each ancestor.
`plot_joyplot_frequencies`	- tracked frequencies over time (output object of simulate_admixture) - vector of timepoints to include in the plot - selection of ancestors to plot	A so called 'joyplot' or 'ridge plot', which visualizes density distributions over time.	Plots the distribution of genomic frequencies within a region as a joyplot, where the change over time in the distribution is visualized.
`plot_start_end`	- output of `simulate_admixture` - selection of ancestors to plot	Plots both the initial and final frequency of each ancestor along the genome.	Plot the distribution of genomic frequencies at the start and at the end of a simulation together in one plot
`plot_difference_frequencies`			Plots the difference in frequency, compared between the start and the end.
`plot_frequencies`	- output of `simulate_admixture` - a vector with marker locations	Plots the frequency of each ancestor at the end of the simulation	Plot the frequency of each ancestor along the genome
`plot_over_time`	- output of `simulate_admixture` - focal marker to focus on	Plot ancestor frequencies over time.	Plot the frequency of each ancestor at a specific marker, over time.
Utilities			
`save_population`	- population object - file name		Save a population to file
`load_population`	- file name	Population object	Load a population from file

521