GenAPoPop 1.0 : A user-friendly software to analyse genetic diversity and structure from partially clonal and selfed autopolyploid organisms

Solenn Stoeckel, Ronan Becheler, Ekaterina Bocharova, Dominique Barloy

Title: GENAPOPOP 1.0: a user-friendly software to analyze genetic diversity and structure from partially clonal and selfed autopolyploid organisms

Short running title: Population genetic analyses of autopolyploids

Authors: Solenn Stoeckel[1,2], Ronan Becheler[1,2], Ekaterina Bocharova[3], Dominique Barloy[2]

Affiliations

[1] IGEPP, INRAE, Institut Agro, Université de Rennes, Le Rheu, France.

[2] DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, IFREMER, INRAE, Rennes, France.

[3] Koltzov Institute of Developmental Biology of Russian Academy of Sciences (IDB RAS), Moscow, Russia

Corresponding author: solenn.stoeckel@inrae.fr

Published in ***Molecular Ecology Resources,*** DOI 10.1111/1755-0998.13886

Abstract

Autopolyploidy is quite common in most clades of eukaryotes. The emergence of sequence-based genotyping methods with individual and marker tags now enables confident allele dosage, overcoming the main obstacle to the democratization of the population genetic approaches when studying ecology and evolution of autopolyploid populations and species. Reproductive modes, including clonality, selfing and allogamy, have deep consequences on the ecology and evolution of population and species. Analyzing genetic diversity and its dynamics over generations is one efficient way to infer the relative importance of clonality, selfing and allogamy in populations.

GENAPOPOP is a user-friendly solution to compute the specific corpus of population genetic indices, including indices about genotypic diversity, needed to analyze partially clonal, selfed and allogamous polysomic populations genotyped with confident allele dosage. It also easily provides the posterior probabilities of quantitative reproductive modes in autopolyploid populations genotyped at two-time steps and a graphical representation of the minimum spanning trees of the genetic distances between polyploid individuals, facilitating the interpretation of the genetic coancestry between individuals in hierarchically structured populations.

GENAPOPOP complements the previously existing solutions, including SPAGEDI and POLYGENE, to use genotypings to study the ecology and evolution of autopolyploid populations. It was specially developed with a simple graphical interface and workflow, and comes with a simulator to facilitate practical courses and teaching of population genetics for autopolyploid populations.

Keywords

Polyploidy, Genotypic diversity, Genetic differentiation, Unrooted tree of genetic distances, Bayesian inference of reproductive modes

Introduction

Population genetics is a robust, cost- and time-efficient framework to predict, understand and infer the ecology and evolution of species (Ewens 2004, Ellegren & Galtier 2016). This paradigm at the center of biological evolution theory has stood the test of time to predict and track the ancestral relatedness between individuals at the scale of studied populations (Wakeley 2005). Using changes of genetic variations over time and space, population genetic models allow quantifying evolutionary forces in populations and interpreting them as hypothesized biological and environmental influences on lineages (Ellegren & Galtier 2016). Among all the possible biological features driving evolution, reproductive mode is one of the most significant evolutionary forces impacting the dynamics of genetic diversity and its structure among populations as it determines the transmission of the hereditary DNA signal over generations (Duminil *et al.* 2007). In return, analyzing the genetic diversity within populations allows inferring their reproductive modes, providing a precious knowledge to predict and understand their ecological and biological evolution. It also helps better targeting ecological scenarios and more robust inferences of other evolutionary forces (Fehrer 2010, Yu et al. 2016, Stoeckel et al. 2021). However, to date and despite nearly one century of research, population genetic models and tools were mostly developed for sexual, diploid species (Orive & Krueger-Hadfield 2021, Dufresne et al. 2014).

Eukaryotes with more than two sets of homologous chromosomes (autopolyploids) or duplicated genomic segments are very common in ferns, flowering plant and fungi species (Barker et al. 2015, Albertin & Marullo 2012, Wood et al. 2009). Polyploidy seems less frequent in animals albeit significant in a handful of clades such as in fishes, cnidarians, amphibians and reptiles (Gregory & Mable 2005, Mable *et al.* 2011, Boots et al. 2023). It also occurs in some species only for some chromosomes (aneuploidy), like commonly observed in partially clonal parasitic protozoa (Tibayrenc & Ayala 2013, Rougeron et al. 2015).

Polyploidization influences genetic and phenotypic diversity including potential ecological adaptations and radiation, with a long-term dynamic from whole genome duplication to re-diploidization (Haldane

1930, Baduel et al. 2018, Wu et al. 2019). Interestingly, polyploidy strongly co-occurs with reproductive modes involving partial clonality, both in natural and experimental populations (Herben et al. 2017; Van Drunen & Husband 2019). It also seems to be an influential complementary factor to the more classical Baker's hypothesis of the advantage of uniparental reproductive mode, including selfing and clonality, when peripatric populations establish in new areas (Pandit et al. 2011, Barrett 2018, Rutland et al. 2021). Studying the reciprocal influences of reproductive modes on the ecology and evolution of populations is now usual in diploid populations using their genetic diversity, favored by a wide range of tools adapted to analyze their genetic diversity like GENCLONE (Arnaud-Haond & Belkhir 2007), RMES (David et al. 2007) and RCLONE (Bailleul et al. 2015). However, it is less common in polyploid populations. The lack of adapted and easily accessible analysis solution leads previous studies to consider such datasets as haplotypes or analyze them as diploid.

Indeed, population genetic studies of polyploid organisms were long limited by two main difficulties (Dufresne et al. 2014, Jighly et al. 2018). First, accessing robust genotyping in such populations has long been a true challenge due to the problematic allele dosage in individuals. For example, it was methodologically impractical to distinguish between *AABB*, *ABBB* and *AAAB* individuals at a tetraploid genetic marker with two alleles, A and B, without assuming hypotheses difficult to verify (Dufresne et al. 2014, Bourke et al. 2019). Allele dosage difficulties intensify with increasing ploidy and number of possible alleles at the considered genetic marker, as the number of combinations of alleles determining the number of possible genotypes itself increases. However, recent advances in genotyping methods exploiting deep sequencing with low errors rates combined to individuals and marker tags unlocked the possibility to genotype polyploid individuals with confident allele dosage, even in species with large sets of chromosomes (Delord et al. 2018). These methods benefit both from the advances made on the sequencing process itself that decrease sequencing errors and from the development of upstream molecular processing of genetic samples to tag and target very-specific genomic regions. They increase the sequencing depth of the genotyped marker and allow reproducible replicates. It is now easier to access for a limited cost to more than 20 to hundreds of replicated sequences per SNP (Single

Nucleotide Polymorphism) or microsatellite allele within each individual in a pool of individuals using genotype-by-sequence method. For example, HIPLEX genotyping method allows genotyping ~500 individuals at 100 SNPs using one sequencing run (*e.g.*, MiSeq 2x150 Heflin), with a sequencing depth of ~50 sequences per allele in tetraploids and ~33 sequences per allele in hexaploids, resulting in genotype assignations with a confidence superior to 99% (Delord et al. 2018, Besnard et al. 2023).

Second, we also long lacked adapted models and analysis methods to compute population genetic indices and quantify evolutionary forces in polyploid populations (Dufresne et al. 2014), especially considering that partially clonal and selfed populations can result in repeated genotypes (*i.e.,* the same multi-locus genotype found in different samples, Arnaud-Haond et al. 2007) or patterns of high probabilities of identity between genotypes (David et al. 2007; Jullien et al. 2019). Due to challenges introduced by data formats and difficulties in generalizing the mathematical formula of population genetic indices (Ewens 2004), common population genetics software, such as Genalex (Peakall & Smouse 2012) and GenClone (Arnaud-Haond & Belkhir 2007) are not designed to work with partially clonal populations with more than two allelic copies per gene (Excoffier & Heckel 2006). A handful of libraries and software emerged in the last years, like the command-line SPAGEDI (Hardy & Vekemans 2002), the more user-friendly recent and multiplatform POLYGENE (Huang et al. 2020) or GENODIVE (Meirmans & Tienderen 2004) a software restricted to MACOS X operating system. However, all these programs do not compute all the population genetic indices used to understand and interpret all reproductive modes, including selfing and clonality in populations, such as indices based on genotypic diversity. POLYGENE for example cannot handle repeated genotypes that can be commonly observed in partially clonal populations. POLYSAT (Clarck & Jasieniuk 2011) cannot currently deal with data with confident allele dosage, which becomes a standard with massive sequencing & tagging methods. Some R librairies like POPPR (Kamvar et al. 2014), RCLONE and POLYSAT, and command-line solutions like SPAGEDI may help analyze genotypes of polyploid populations with different modes of reproduction, but they require an exhaustive exploration of their documentation and some training in scripting languages to use them. During practical courses, they involve a preliminary introduction about

scripting or on the reasons for using some options over another, complicating teaching population genetics for polyploid species by dispersing the topic in technical considerations.

GenAPoPop software

Thereby, to provide a user-friendly solution to compute the specific corpus of population genetic indices needed to analyze partially clonal and selfed polysomic populations, we developed and packaged a new portable, multi-operating system, working by itself with no dependency software, named GENAPOPOP (standing for Genetic Analyses of POlyploid POPulations).

GENAPOPOP is written combining PYTHON, FORTRAN and HTML with a graphical user interface coded in Qt. The binary executables for WINDOWS, LINUX and MACOS are provided under the terms of a CC-BY-NC-SA license, version 4, and can be downloaded at https://forgemia.inra.fr/solenn.stoeckel/genapopop1.0/. Each packaged version of GENAPOPOP is tested on X64 CPU systems (including server CPU INTEL XEON E5-2650 V3, AMD THREADRIPPER 3970X and AMD RYZEN 7 5800U) with a LINUX DEBIAN-based distribution and MICROSOFT WINDOWS 10 and 11 uptodate versions; The MACOS version is currently tested on a MACOS BIG SUR, INTEL version.

The idea of this software is to relieve the users of all scripting tasks, and simplify as much as possible the infile formatting. To this aim, GENAPOPOP uses a graphical interface organized in a comprehensive workflow (Fig. 1). This software was also designed to complement the previously cited softwares that can be used easily using GENAPOPOP dataset export feature.

It enables analyzing genotypic datasets with confident allele dosage of autopolyploid species in which we can neglect double-reduction. GENAPOPOP only assumes a random chromosome segregation model (Muller 1914). This model considers that gametes originate from any combination of homologous chromosomes, thus excluding that two sister chromatids segregate in a same gamete. This is the most commonly observed case in polyploids (Wu et al. 2001). GENAPOPOP doesn't consider yet for double-reduction models (see supplementary material). Double-reductions in auto- and allopolypoloids result from multivalent pairing among homologous chromosomes, when two or more sister chromatids

segregate in a same gamete (Wu et al. 2001, Huang et al. 2019, Jiang et al. 2021, Ferreira de Carvalho et al. 2021). The main consequence of double-reduction for population genetics is to increase the probability of identity-by-descent when compared to random chromosome segregation model (Hardy 2016). For example, an autotetraploid individual typed *ABCD* can produce *AA*, *BB*, *CC*, *DD* gametes when double-reduction happens. Without double-reduction, and as currently considered by GENAPOPOP, only *AB*, *AC*, *AD*, *BC*, *BD* and *CD* gametes are produced.

Format of input data and output results

GENAPOPOP was intentionally designed to accept different genotyping text-file format as long as each line codes for one individual genotype, and each allele is reported in one column, with columns separated by tabulation. It also manages files with multiple header lines. The advantage of this GENALEX-like format text file (Peakall & Smouse 2012) is that it is universally handled by spreadsheets and text editors, and it fits the most commonly used output format of many SNP-set callers. GENAPOPOP workflow requires to first upload such data file, and then label the four necessarily present columns in the data file: three columns indicating population name, generation or date of sampling and individual identifier (Table 1). Any character can be used in these columns except tabulation and space. The fourth column indicates the column with the first allele of the first locus, and implies that all the following columns until the last one only contains alleles coding for the individual genotype. Alleles can be SNPs, thus expected to be coded as upper- or lower-case *a*, *c*, *g*, *t* and *n* for missing allele or number *1*, *2*, *3*, *4* and *0* for missing allele. Alleles can also be sequence repeat markers (like micro-, mini- and macro-satellites) or sequence length-based markers, named hereafter SSR-like (for Simple Sequence Repeat) markers in GENAPOPOP software and documentation. In the case of SSR-like markers, each allele is expected to be coded as an integer number of repeats or a sequence size, and, if encountered, missing allele should be coded as zero. For the moment, GENAPOPOP supposes genotypes evolve following a K-allele mutation model (KAM) in which any allele can mutate in any other allele with the same probability, which has the advantage of aptly modelling the mutation of both

microsatellites and SNPs (Weir & Cockerham, 1984), but does not make it possible to exploit the number of repeated DNA segments or the sequence sizes for computing population and individual genetic indices and distances.

GENAPOPOP can work on input file with genotypes of one or multiple populations, with identical ploidy and genotyped with a common marker-set, to analyze them in mass. GENAPOPOP has no limit in the number of populations, of time steps and genotypes it can analyze, out of the classic material and operating system limitations, *i.e.*, the quantity of random-access memory (RAM) to upload the datafiles and the outputs, and the central processing unit clock speed and advancement of its instruction sets.

Implemented methods and workflow

GENAPOPOP is organized by tabs: one homepage, one page to load the dataset and describe its arrangement, three tabs to perform the three different types of analyses and one tab of documentation (Fig. 1).

*** Insert here Figure 1 ***

The software opens on a welcome homepage giving basic information and enabling opening the attached PDF documentation. This can be done either using the built-in browser, which is interesting in situations where the software must be deployed on workstations without administrator rights or with restricted access (such as during practical courses at university), or by using the system's default PDF file reader, which will provide greater reading comfort. Next, users are directed to a tab dedicated to upload and describe at a minimum the composition of the genotype dataset. In this tab, users upload the text file containing the genotypes, inform the header line (after which all lines code for one genotype of one individual), inform the 4 main columns (population, generation, individual identifier, and the column containing the first allele of the first locus), inform the ploidy (from 1 to 50) and the type of markers (SSR-like or SNP-like). Once the dataset is uploaded and the required lines and columns labeled, users are invited to check the data format. If troubles, the verification will report explicit errors

to be corrected, returning the problematic line of the dataset. The verification passed, users are then invited to launch one of the three types of analyses performed by GENAPOPOP by clicking on the corresponding button opening a dedicated new tab.

GenPopPoly tab

This tab allows users to compute a list of population genetic indices suitable to analyze genetic diversity and population structure of polyploid populations with a special focus on reproductive modes. These indices are useful and efficient to estimate rates of clonality, autogamy (selfing) and allogamy on genotypes of populations sampled once (Castric et al. 2002, David et al. 2007, Hardy 2016, Stoeckel et al. 2021). Users select the population(s) to be analyzed, select the analyses to be computed and reported, launch the computation and can directly browse the results for a first sight in the integrated calc viewer. The results are also saved in a text-file (separator tabulation) in the folder containing GENAPOPOP executable. Result files can readily be opened by all spreadsheet applications to be explored and manipulated to do tables and figures. The output file presents first all intra-population indices computed per population, then computed overall populations. It includes genotypic and genetic diversity indices as recommended in Stoeckel et al. (2021), probabilities of identity for diploids and autopolyploids (Jacquard 1970, Evett & Weir 1998, Waits et al. 2001, Huang et al. 2015), the four first moments (*i.e.,* mean, variance, skewness and kurtosis) of inbreeding coefficient F_{IS} in populations (Stoeckel & Masson 2014). It also provides a list of multi-locus genotypes (commonly named MLG in literature or genet) with their shared genotype, and in the last column, the number of repeated genotypes (ramet) found in the considered population. In each and overall populations, it reports genotypic diversity indices including the index of clonal diversity (*R*, Dorken & Eckert, 2001) and the size distribution of lineages (D* of Simpson and *Pareto* β, Arnaud-Haond et al., 2007) computed properly for autopolyploids. We deliberately discarded many other indices to help users robustly interpreting genotypic diversity in their populations. Despite *Pareto* β is far more robust than the *R* to assess genotypic diversity in sampled populations (Stoeckel et al. 2021, Arnaud-Haond et al. 2020), we

still compute *R* for reference, as this one was historically massively reported in past literature. The output also provides the mean correlation coefficient of genetic distances between unordered alleles at all loci, usually named $\bar{r}_d$ as an overall measure of linkage disequilibrium per population and overall populations (Agapow & Burt, 2001). This index, ranging from slightly negative or 0 (no correlation) to 1 (maximum association of alleles over all loci), presents the advantage of limiting the dependency of the correlation coefficient on the number of alleles and loci. GENAPOPOP also provides per population and overall populations a table of classical intra-populational genetic indices per locus: observed heterozygosity, raw and unbiased expected heterozygosity (also name gene diversity), resulting raw and unbiased inbreeding coefficient (*Fis*) accounting for intra-individual genetic variation as a departure from Hardy-Weinberg assumptions of the genotyped populations and the raw and effective number of alleles (*Ae*, Weir 1996). On a side and more experimental part, GENAPOPOP allows computing analysis of molecular variance (AMOVA) computed following Meirmans & Liu (2018) and Weir (1996) equations and recommendations, including the Fis, Fst and Fit per population, over all populations, per marker and over all markers. These results can already be obtained using Polygene and Genodive. GENAPOPOP also provides in this section the overall and pairwise-population *rhost*. *rhost* measures the genetic differentiation between populations as the *Fst* value that would have the same haploid population sizes connected with the same migration rate, and present the advantage to be comparable between species and populations of different ploidy levels (Ronfort et al. 1998, Meirmans & Van Tienderen 2013). These indices of genetic differentiation/structuration are a good complement to the minimum spanning tree of the genetic distances between individuals when colored or tagged by population to get a picture of the genetic structure of genotyped populations (see below). As these indices are also computed in SPAGEDI and POLYGENE, we invite users to also compare their results with these softwares.

GENAPOPOP was thought and designed to complement GENODIVE, POLYGENE that performs hierarchical, Bayesian clustering and parentage analysis, and SPAGEDI that already performs multiple spatial analyses and that can be used to estimate selfing rate. In this tab, GENAPOPOP users can export their datasets

automatically in a SPAGEDI-format file that will be recorded in the same folder under the same imported data name extend with "*_spagedi_ready.txt*". This file that can be easily imported to extend and access complementary analyses in the previously cited software, including SPAGEDI and POLYGENE, and we greatly encourage future GENAPOPOP users to analyze their data with multiple software to get the most complete view of their dataset.

ClonEstiMatePoly tab

This tab allows users to compute the posterior probabilities of joint rates of clonality and selfing in polyploid populations genotyped at, at least, two-time steps. This method was demonstrated to be the most accurate way to quantitatively assess reproductive modes in diploid populations over multiple Eukaryotes species, especially for detecting low rates of clonality (Becheler et al. 2017). It should facilitate the detection of clonal reproduction, the estimation of the rates of clonality in polyploid populations, and promote the study of reproductive modes and their genetic consequences in such species. It should be a nice addition to the method of estimation of selfing rates using multilocus standardized identity disequilibrium coefficient found in SPAGEDI (Hardy 2016).

Here, we extended to autopolyploids the Bayesian formula and method CLONESTIMATE from Becheler et al. (2017). It exploits the likelihood of transitions of genotype frequencies from one generation to another to accurately estimate rates of mutation, clonality and selfing, and thus works well even in the absence of equilibrium between evolutionary forces (genetic drift, mutation and rates of clonality) which is quite common in partially clonal populations (Reichel et al. 2016). This method remains accurate using from about ten polymorphic markers, even physically linked and mutating with other mutation model, and from 30 sampled individuals. It is however sensitive to erroneous assumed or restricted prior values of clonal and selfing rates, null alleles and sampling time interval greater than two generations. Extended equations for autopolyploids can be found in the documentation in supplementary material. This discretized Bayesian method needs an analysis plan listing discretized priors on rates of mutation, clonality and selfing for each population (Fig. S1). Restricted ranges of prior

on each of these parameters allow better inferences on other targeted parameters. Analysis plan can be uploaded or prepared (and saved for future use) using the graphical interface. Analysis plan can be browsed and checked using the integrated browser before launching the computations. To speed up the calculations, computations per locus and population of the analysis plan were parallelized using the maximum number of threads available by the operating system. Results are stored in the folder containing GENAPOPOP in a text-file separator tabulation file that can be readily handled using any spreadsheet application. Results are presented per population between two time steps as a list of discrete joined values of mutation rates, rates of clonality and selfing with the corresponding posterior probabilities of such joined combination of priors. This presentation of the results makes it easy to combine the posterior probability mass functions per population and generations into table and/or into plots of their distributions. If found in the dataset, it also returns the list of monomorphic loci at, at least, one sampling time. Monomorphic loci decrease the inference power of the dataset to assess rates of mutation, clonality and selfing between the two sampled generations.

Minimum spanning tree of genetic distances between individuals tab

This tab allows users to compute the genetic distance between individuals using their identity-in-state (number of shared alleles) and provides the corresponding minimum spanning unrooted tree using the classical equal-angle algorithm (Christopher Meacham in Felsenstein 2004). This network representation is useful to detect multilocus lineages (named MLL in literature) due to clonality that shape typical rosettes or small rosaries, *i.e.,* a group of ramets differing by a limited number of mutations radiating around a main genet (Fig. 2).

*** Insert here Figure 2 ***

Users can get the computed genetic distances between pairwise-individuals in an exported text-file to use them with other software, and they can customize the plot of the minimum spanning unrooted tree using individual colors and tags. The resulting graph can be exported at different resolutions into

research-standard portable document format (PDF) file format, raster (portable network graphics, PNG) or vector (scalable vector graphics, SVG) image formats that can be processed afterward in dedicated software. The resulting graph can be previewed and explored using the integrated browser, using mouse controls (zoom in and out using mouse wheel, move the graph with mouse grab) before exportation.

Consistency and accuracy tests

To test the consistency and accuracy of our software, we used simulated data and empirical datasets as control data. To obtain simulated data, we used the embedded simulator that can also be used for testing and teaching purposes (Tab Simulation, see Documentation).

First, we perform consistency test of GENAPOPOP with the output of SPAGEDI reference software (Hardy & Vekemans 2001) on four basic population genetic indices (Ae, He, Ho, Rhost). We simulated four test datasets simple enough to be checked by hand calculations for unit testing. For further and future unit testing, the raw datasets were deposited on the European general-purpose open repository ZENODO (Barloy et al. 2022). These four scenarios correspond respectively to panmictic (A), highly selfed (B), highly clonal (C) and half-clonal-half-selfed (D) reproductive modes. Quantitative values are explicitly indicated in the two first lines of the datasets. In each scenario, we simulated two connected populations of 100 individuals each with a migration rate of 0.01 and mutating at a rate of 0.01, genotyped at 10 SNPs, 1000 generations after an initial randomly drawing population. For each scenario, we recorded the populations' genotyping states over two consecutive generations (generations 1000 and 1001). In addition, we tested GENAPOPOP on two field datasets genotyped with confident allele dosage, one SNPs set from the autotetraploid genome part of *Ludwigia grandiflora subsp. hexapetala* (hereafter *Lgh*, Genitoni et al. 2020) and one microsatellite set from the autotetraploid arctic sea anemone *Aulactinia stella* (hereafter *As*, Bocharova et al. 2018). These two datasets are genetic samples of larger metapopulations genotyped with confident allele dosage, including missing alleles and genotypes, and including some loci fixed in one of the populations. We

draw attention of users that the different software present different ways to handle fixed, missing alleles and genotypes.

Second, to analyze how population genetic indices of a snapshot of genotyped populations behave in autopolyploids and how they compare to diploids as reported in Stoeckel et al. 2021, we simulated 6300 different datasets following 21 reproductive mode scenarios and three different ploidies (2,4 and 6). Each reproductive scenario at one ploidy level was independently simulated a hundred times to get a confident picture of the range of the possible genetic trajectories. Each of the 21 different reproductive mode scenarios consists on a triplet of values including one rate of clonality, one rate of selfing and one complementary rate of allogamy, the three rates necessarily summing to one. Rates of clonality, selfing and allogamy took complementary values of all the possible combinations within the set [0., 0.2, 0.4, 0.6, 0.8, 1.], constrained to sum to one. For example, one scenario was (rate of clonality=0.2, rate of selfing=0.4, rate of allogamy=0.4). Hereafter, for easier representation, we reported couple of rates of clonality and of selfing in figures and text, implicitly considering that rate of allogamy was one minus the rates of clonality and selfing. In each scenario, we simulated two connected populations of 100 individuals each with a migration rate of 0.01 and mutating at a rate of 0.01, genotyped at 30 markers with 4 possible alleles randomly introduced within individuals in the first generation with the same frequency. Analyzed datasets were recorded 1000 generations after the initial generation, corresponding to 5 times the overall instantaneous population size (N=200). Distributions of population genetic indices obtained with the 21 reproductive modes and 3 ploidy levels were reported as violin plots, each made from one hundred independent simulations.

Third, to test the accuracy and precision of CLONESTIMATEPOLY method to jointly infer rates of clonality, selfing and allogamy in autopolyploid populations genotyped at two-time steps, we simulated again 6300 different datasets following the same 21 reproductive mode scenarios, for diploid, tetraploid and hexaploid populations. For each quantitative reproductive mode (*i.e.*, a precise couple of values of one rate of clonality and one rate of selfing), we simulated 100 couples of populations, each of size N=100, mutating at a rate 1/N and exchanging migrants at a rate of 1/N, over 1000 generations. We submitted

the genotypes found in parents (generation 999) and in their descendants (generation 1000) to CLONESTIMATEPOLY with flat priors to get the inferred posterior distribution of the joint rates of clonality and selfing. The 100 posterior distributions per couple of rates of clonality and selfing were summed and reported as a confusion matrix for ploidy 2, 4 and 6.

All the results were aggregated and deposited in Barloy et al. (2022).

Results

Consistency test

SPAGEDI and GENAPOPOP reported similar values of Ae, He, Ho and Rhost (Table S1). Ae and He were corrected by sample size in SPAGEDI but not in GENAPOPOP, explaining the little differences observed. GENAPOPOP uses double-precision floating-point format (64 bits) while, to our knowledge, SPAGEDI uses a lower precision that can also add up along the calculations. GENAPOPOP intentionally computes estimators with limited 'correction' to avoid giving more weight to some loci rather than other which may bias the global picture of a dataset (see formulas in Documentation).

Variations of snapshot population genetic indices with rates of clonality and selfing, and ploidy

The principal component analysis on the values of genetic diversity using a snapshot of genetic diversity of a population showed two main non-collinear clusters of genetic indices associated with clonality and selfing (Fig. S2, S3 and S4). Rates of clonality were collinear with clonal heterogeneity and evenness indices (including R, Pareto Beta, the complement of the Simpson index and Shannon-Wienner's index), with variance of Fis and Fit, and linkage disequilibrium. Increasing rates of clonality increased linkage disequilibrium and variance of Fis among genotyped loci while it decreased clonal evenness. Rates of selfing were collinear with indices based on allele diversity (including gene diversity, probabilities of identity and genetic structure indices) and heterozygosity. Increasing rates of selfing increased mean Fis and Fit, and variances of panmictic probabilities of identity, of the number of alleles

per loci, of gene diversity and of pairwise Fst. Conversely, it decreased the mean number of alleles per loci, the mean gene diversity and observed heterozygosity.

The distributions of population genetics indices varied with joint rates of clonality and selfing and with increasing ploidy, except the distribution of rhost values between the two simulated populations (Fig. S5). For a fixed ploidy, changes in the range of expected values with reproductive modes remained quite similar to those observed for diploids.

Accuracy of jointly inferred rates of clonality and selfing with increasing ploidy

CLONESTIMATEPOLY, the Bayesian method we propose here to jointly infer rates of clonality, selfing and allogamy in autopolyploid populations genotyped at two-time steps showed high accuracy and limited confusion to jointly infer the true rates of clonality and selfing. The method inferred in the worst cases the true joint rates of clonality and selfing with a precision of ±0.2 (Fig. S6). It occurred when populations reproduced using both intermediate rates of clonality and rates of selfing. Increasing ploidy showed a slight tendency to overestimate selfing rates when populations also reproduced using intermediate rates of clonality.

Recommendations and warning

Most population genetic analyses rely on accurate estimates of real populational genotype frequencies, including here CLONESTIMATEPOLY method. The number of different possible genotypes at one locus increases with the ploidy and the number of alleles (Reichel et al. 2015). We thus draw users' attention on the fact that sample sizes should naturally be larger in polyploid organisms to accurately estimate their genotype frequencies, despite the fact that genotyping more alleles per individual may help assessing allele frequencies.

Missing values and null alleles compromise comparisons between individuals, lineages and populations, and are susceptible to create biases and misinterpretations. Suspected null allele can be coded as unknown allele with their own specific letters or positive integers, and should be clearly

reported before interpretations. Indeed, no “correction” or “assumption” can enhance blurred and incomplete genotyping signals without deep consequences on the computed indices and then their interpretations, whatever the ‘correction’. We thus recommend users to rather remove genetic markers and individuals with missing values and uncertain genotypes.

Conclusion

GENAPOPOP provides a user-friendly, multi-operating systems, efficient mass processing way to analyze autopolyploid (including diploid) genotypings with a special focus on interpreting the genetic diversity and its structure within and between populations in regards with their reproductive modes. It especially allows computing genotypic indices to analyze clonal heterogeneity and clonal evenness for polyploids due to repeated multilocus genotypes (MLGs) in samples. It includes an extension of the robust and efficient CLONESTIMATE Bayesian method to quantitatively infer joint rates of clonality, selfing and allogamy using populations genotyped at two-time steps. It facilitates the interpretation of genetic diversity in partially clonal, partially selfing autopolyploid populations with no or very-limited double reduction. It has no vocation to include or encompass all methods and population genetic indices that can be computed when analyzing autopolyploid genotypings. This is why it allows exporting datasets in format that can be uploaded in other software like SPAGEDI (Hardy & Vekemans 2002), GENODIVE (Meirmans 2020) and POLYGENE (Huang *et al.* 2020). We thus warmly recommend users to use GENAPOPOP in complement to other dedicated analyses that can be found in these other softwares, depending on the tackled questions. GENAPOPOP also answers the need of a population genetic analyzing software for autopolyploid datasets with confident allele dosage that will come growing with the new genotyping-by-sequencing methods with individually tagged sample and locus. It finally answers the need of a user-friendly software for practical course that doesn’t need teaching command-lines or scripting languages as a prior to introduce students to population genetics for polyploid species and to the genetic consequences of reproductive modes on the genetic diversity and structure of populations.

Acknowledgements

We warmly thank Luis Portillo-Lemus for initial discussions during his PhD thesis that encouraged the development of this software. We thank four anonymous reviewers for finding a bug in the AMOVA output and for helping enhance our manuscript and associated results and files. We globally thank students of the International Master in Biodiversity Ecology and Evolution and Modelling in Ecology at The University of Rennes1 and l'Institut Agro for identifying over the year the need of an adapted and user-friendly software to correctly teach polyploid population genetics in practical courses. Finally, we thank participants and organizers of the POLYPLOIDY AND BIODIVERSITY conference (Rennes 11-12 October 2021) for their useful returns on the first version of GENAPOPOP.

This work was supported by CLONIX2D ANR-18-CE32-0001; the INVAMAT PROJECT (Plant Health and Environment Division of the French National Institute of Agricultural Research) and the French Embassy in the Russian Federation, for funding this project by a METCHNIKOV 2019 grant. The SNP data on *Ludwigia grandiflora subsp. hexapetala* populations were acquired using FEDER funds from Région Centre-Val de Loire and by Agence de l'eau Loire-Bretagne, grant Nature 2045, programme 9025 (AP 2015 9025). The SSR data on *Aulactinia stella* was obtained by EB under the IDB RAS Government basic research program № 0088-2021-0019.

Authors' contributions

DB, EB and SS laid the foundation of this work, identifying its need and were responsible for funding applications. SS formalized the mathematical equations, formalized the methods, coded the software, packaged the software and coded the simulator for unit testing and for testing the consistency with other software. RB and SS contributed the code testing and interface enhancement, the output exploration, the test of consistency with other software and performed the literature researches. SS wrote the core manuscript. All authors read, edited and approved the final manuscript.

Data Accessibility and Benefit-Sharing Section

The latest packaged binaries of GenAPoPop1.0 can be downloaded on the long-term academic Gitlab server of INRAE (French National Research Institute for Agriculture, Food and Environment): https://forgemia.inra.fr/solenn.stoeckel/genapopop1.0

Pseudo-observed and field dataset (*Ludwigia grandiflora subs hexapetala* and *Aulactinia stella*) used for consistency tests are available on Zenodo (DOI: 10.5281/zenodo.8164531).

Conflict of interest

The authors of this preprint declare that they have no financial conflicts of interest based on the content of this article.

ORCID

Solenn Stoeckel, 0000-0001-6064-5941

Ronan Becheler, 0000-0001-9322-0771

Ekaterina Bocharova, 0000-0001-9978-3006

Dominique Barloy, 0000-0001-5810-4871

References

Agapow, P.-M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, *1*(1–2), 101–102. https://doi.org/10.1046/j.1471-8278.2000.00014.x

Albertin, W., & Marullo, P. (2012). Polyploidy in fungi: Evolution after whole-genome duplication. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2497–2509. https://doi.org/10.1098/rspb.2012.0434

Arnaud-Haond, S., & Belkhir, K. (2007). genclone: A computer program to analyze genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, *7*(1), 15–17. https://doi.org/10.1111/j.1471-8286.2006.01522.x

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Arnaud-Haond, S., Stoeckel, S., & Bailleul, D. (2020). New insights into the population genetics of partially clonal organisms: When seagrass data meet theoretical expectations. *Molecular Ecology*, *29*(17), 3248–3260. https://doi.org/10.1111/mec.15532

Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., & Yant, L. (2018). The "Polyploid Hop": Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications. *Frontiers in Ecology and Evolution*, *6*. https://www.frontiersin.org/articles/10.3389/fevo.2018.00117

Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: A package to identify MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and Evolution*, *7*(8), 966–970. https://doi.org/10.1111/2041-210X.12550

Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, *210*(2), 391–398. https://doi.org/10.1111/nph.13698

Barloy, D., Bocharova, E., Harang, M., Portillo, L., & Stoeckel, S. (2022). Reference datasets for consistency tests of GENAPOPOP 1.0 software: a user-friendly software to analyze genetic diversity and structure in partially clonal and selfed polyploid organisms. (1.0) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.8164531

Barrett SCH (2018) Why reproductive systems matter for the invasion biology of plants. In: Fifty Years of Invasion Ecology: The Legacy of Charles Elton (ed. DM Richardson). Oxford University Press, Oxford, UK.

Becheler, R., Masson, J.-P., Arnaud-Haond, S., Halkett, F., Mariette, S., Guillemin, M.-L.,

Valero, M., Destombe, C., & Stoeckel, S. (2017). ClonEstiMate, a Bayesian method for quantifying rates of clonality of populations genotyped at two-time steps. *Molecular Ecology Resources*, *17*(6), e251–e267. https://doi.org/10.1111/1755-0998.12698

Besnard A.-L., Park D. J., Pope B. J., Hammet F., Michon-Coudouel S., Biget M., Krueger-Hadfield S.A., Mauger S., Petit E.J. 2023. Workflow for SNP genotyping using the Hi-Plex method. Protocols.io, dx.doi.org/10.17504/protocols.io.8epv5jnnnl1b/v1

Bocharova, E. S., Sergeev, A. A., & Volkov, A. A. (2018). Identification of microsatellite loci in sea anemones Aulactinia stella and Cribrinopsis albopunctata (family Actiniidae). *F1000Research*, *7*, 232. https://doi.org/10.12688/f1000research.13724.1

Booth, W., Levine, B. A., Corush J. B., Davis, M. A., Dwyer, Q., De Plecker, R., & Schuett, G. W. (2023). Discovery of facultative parthenogenesis in a new world crocodile. *Biology Letters*, *19*(6), 20230129. https://doi.org/10.1098/rsbl.2023.0129

Bourke, P. M., Hackett, C. A., Voorrips, R. E., Visser, R. G. F., & Maliepaard, C. (2019). Quantifying the Power and Precision of QTL Analysis in Autopolyploids Under Bivalent and Multivalent Genetic Models. *G3: Genes|Genomes|Genetics*, *9*(7), 2107–2122. https://doi.org/10.1534/g3.119.400269

Castric, V., Bernatchez, L., Belkhir, K., & Bonhomme, F. (2002). Heterozygote deficiencies in small lacustrine populations of brook charr Salvelinus Fontinalis Mitchill (Pisces, Salmonidae): a test of alternative hypotheses. *Heredity*, 89, 27–35. https://doi.org/10.1038/sj.hdy.6800089

Clark, L. V., & Jasieniuk, M. (2011). polysat: An R package for polyploid microsatellite analysis. *Molecular Ecology Resources*, *11*(3), 562–566. https://doi.org/10.1111/j.1755-0998.2011.02985.x

David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, *16*(12), 2474–2487.

https://doi.org/10.1111/j.1365-294X.2007.03330.x

Delord, C., Lassalle, G., Oger, A., Barloy, D., Coutellec, M.-A., Delcamp, A., Evanno, G., Genthon, C., Guichoux, E., Le Bail, P.-Y., Le Quilliec, P., Longin, G., Lorvelec, O., Massot, M., Reveillac, E., Rinaldo, R., Roussel, J.-M., Vigouroux, R., Launey, S., & Petit, E. J. (2018). A cost-and-time effective procedure to develop SNP markers for multiple species: A support for community genetics. *Methods in Ecology and Evolution*, *9*(9), 1959–1974. https://doi.org/10.1111/2041-210X.13034

De Meeûs, T. & Balloux, F. (2005), F-statistics of clonal diploids structured in numerous demes. *Molecular Ecology*, 14, 2695-2702. https://doi.org/10.1111/j.1365-294X.2005.02643.x

Dorken, M. E., & Eckert, C. G. (2001). Severely reduced sexual reproduction in northern populations of a clonal plant, Decodonverticillatus (Lythraceae). *Journal of Ecology*, *89*(3), 339–350. https://doi.org/10.1046/j.1365-2745.2001.00558.x

Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology*, *23*(1), 40–69. https://doi.org/10.1111/mec.12581

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J. (2007). Can Population Genetic Structure Be Predicted from Life-History Traits? *The American Naturalist*, *169*(5), 662–672. https://doi.org/10.1086/513490

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), Article 7. https://doi.org/10.1038/nrg.2016.58

Ewens, W. J. (2004). *Mathematical Population Genetics* (Vol. 27). Springer. https://doi.org/10.1007/978-0-387-21822-9

Excoffier, L., & Heckel, G. (2006). Computer programs for population genetics data analysis:

a survival guide. *Nature Reviews Genetics*, 7, 745–758. https://doi.org/10.1038/nrg1904

Evett, I. W., & Weir, B. S. (1998). Interpreting DNA evidence: statistical genetics for forensic scientists. Sinaur Associates Inc, Sunderland, Massachusetts. ISBN 0878931554.

Fehrer, J. (2010). Unraveling the mysteries of reproduction. *Heredity*, *104*(5), Article 5. https://doi.org/10.1038/hdy.2010.12

Felsenstein, J. (2004) Inferring Phylogenies. Sinauer Associates Inc., Sunderland. ISBN: 9780878931774

Ferreira de Carvalho, J., Stoeckel, S., Eber, F., Lodé-Taburel, M., Gilet, M.-M., Trotoux, G., Morice, J., Falentin, C., Chèvre, A.-M., & Rousseau-Gueutin, M. (2021). Untangling structural factors driving genome stabilization in nascent Brassica napus allopolyploids. *New Phytologist*, *230*(5), 2072–2084. https://doi.org/10.1111/nph.17308

Genitoni, J., Vassaux, D., Delaunay, A., Citerne, S., Portillo Lemus, L., Etienne, M.-P., Renault, D., Stoeckel, S., Barloy, D., & Maury, S. (2020). Hypomethylation of the aquatic invasive plant, Ludwigia grandiflora subsp. Hexapetala mimics the adaptive transition into the terrestrial morphotype. *Physiologia Plantarum*, *170*(2), 280–298. https://doi.org/10.1111/ppl.13162

Gregory, T. R., & Mable, B. K. (2005). CHAPTER 8—Polyploidy in Animals. In T. R. Gregory (Ed.), *The Evolution of the Genome* (pp. 427–517). Academic Press. https://doi.org/10.1016/B978-012301463-4/50010-3

Haldane, J. B. S. (1930). Theoretical genetics of autopolyploids. *Journal of Genetics*, *22*(3), 359–372. https://doi.org/10.1007/BF02984197

Hardy, O. J. (2016). Population genetics of autopolyploids under a mixed mating model and the estimation of selfing rate. *Molecular Ecology Resources*, *16*(1), 103–117. https://doi.org/10.1111/1755-0998.12431

Hardy, O. J., & Vekemans, X. (2002). spagedi: A versatile computer program to analyze spatial

genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*(4), 618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Herben, T., Suda, J., & Klimešová, J. (2017). Polyploid species rely on vegetative reproduction more than diploids: A re-examination of the old hypothesis. *Annals of Botany*, *120*(2), 341–349. https://doi.org/10.1093/aob/mcx009

Huang, K., Guo, S., Shattuck, M., Chen, S. T., Qi, X. G., Zhang, P., & Li, B. G. (2015). A maximum-likelihood estimation of pairwise relatedness for autopolyploids. *Heredity, 114*, 133–142. https://doi.org/10.1038/hdy.2014.88

Huang, K., Dunn, D. W., Ritland, K., & Li, B. (2020). polygene: Population genetics analyses for autopolyploids based on allelic phenotypes. *Methods in Ecology and Evolution*, *11*(3), 448–456. https://doi.org/10.1111/2041-210X.13338

Huang, K., Wang, T., Dunn, D. W., Zhang, P., Cao, X., Liu, R., & Li, B. (2019). Genotypic Frequencies at Equilibrium for Polysomic Inheritance Under Double-Reduction. *G3 Genes|Genomes|Genetics*, *9*(5), 1693–1706. https://doi.org/10.1534/g3.119.400132

Jacquard, A. (1970). Structures génétiques des populations. Masson et cie, Paris. ISBN 9782733220238

Jiang, L., Ren, X., & Wu, R. (2021). Computational characterization of double reduction in autotetraploid natural populations. *The Plant Journal*, *105*(6), 1703–1709. https://doi.org/10.1111/tpj.15126

Jighly, A., Lin, Z., Forster, J. W., Spangenberg, G. C., Hayes, B. J., & Daetwyler, H. D. (2018). Insights into population genetics and evolution of polyploids and their ancestors. *Molecular Ecology Resources*, *18*(5), 1157–1172. https://doi.org/10.1111/1755-0998.12896

Jullien, M., Navascués, M., Ronfort, J., Loridon, K., & Gay, L. (2019). Structure of multilocus genetic diversity in predominantly selfing populations. *Heredity*, *123*(2), 176-191.

https://doi.org/10.1038/s41437-019-0182-6Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Mable, B. K., Alexandrou, M. A., & Taylor, M. I. (2011). Genome duplication in amphibians and fish: An extended synthesis. *Journal of Zoology*, *284*(3), 151–182. https://doi.org/10.1111/j.1469-7998.2011.00829.x

Meirmans, P.G. (2020). GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids, *Molecular Ecology Resources*, 20, 1126-1131. https://doi.org/10.1111/1755-0998.13145

Meirmans, P. G., & Liu, S. (2018). Analysis of Molecular Variance (AMOVA) for Autopolyploids. Frontiers in Ecology and Evolution, 6. https://www.frontiersin.org/articles/10.3389/fevo.2018.00066

Meirmans, P. G., & Van Tienderen, P. H. (2004). genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*(4), 792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Meirmans, P. G., & Van Tienderen, P. H. (2013). The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity*, *110*(2), 131–137. https://doi.org/10.1038/hdy.2012.80

Muller, H. J. (1914). A New Mode of Segregation in Gregory's Tetraploid Primulas. *The American Naturalist*, *48*(572), 508–512. https://doi.org/10.1086/279426

Orive, M. E., & Krueger-Hadfield, S. A. (2021). Sex and Asex: A Clonal Lexicon. *Journal of Heredity*, *112*(1), 1–8. https://doi.org/10.1093/jhered/esaa058

Pandit, M. K., Pocock, M. J. O., & Kunin, W. E. (2011). Ploidy influences rarity and invasiveness in plants. *Journal of Ecology*, *99*(5), 1108–1115.

https://doi.org/10.1111/j.1365-2745.2011.01838.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Reichel, K., Bahier, V., Midoux, C., Parisey, N., Masson, J.-P., & Stoeckel, S. (2015). Interpretation and approximation tools for big, dense Markov chain transition matrices in population genetics. *Algorithms for Molecular Biology*, *10*(1), 31. https://doi.org/10.1186/s13015-015-0061-5

Reichel, K., Masson, J.-P., Malrieu, F., Arnaud-Haond, S., & Stoeckel, S. (2016). Rare sex or out of reach equilibrium? The dynamics of FISin partially clonal organisms. *BMC Genetics*, *17*(1), 76. https://doi.org/10.1186/s12863-016-0388-z

Ronfort, J., Jenczewski, E., Bataillon, T., & Rousset, F. (1998). Analysis of population structure in autotetraploid species. *Genetics*, *150*(2), 921–930.

Rougeron, V, De Meeûs, T, & Bañuls, A. L. (2015). A primer for Leishmania population genetic studies. Trends in Parasitology, *31*(2), 52-9. https://doi.org/10.1016/j.pt.2014.12.001

Rutland, C. A., Hall, N. D., & McElroy, J. S. (2021). The Impact of Polyploidization on the Evolution of Weed Species: Historical Understanding and Current Limitations. *Frontiers in Agronomy*, *3*. https://www.frontiersin.org/articles/10.3389/fagro.2021.626454

Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species—Prunus avium L. *Molecular Ecology*, *15*(8), 2109–2118. https://doi.org/10.1111/j.1365-294X.2006.02926.x

Stoeckel, S., & Masson J.P. (2014) The Exact Distributions of F_{IS} under Partial Asexuality in

Small Finite Populations with Mutation. PLOS ONE 9(1): e85228. https://doi.org/10.1371/journal.pone.0085228

Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: Improving our estimation of clonal rates. *Molecular Ecology Resources*, *21*(4), 1068–1084. https://doi.org/10.1111/1755-0998.13316

Tibayrenc, M., & Ayala, F. J. (2013). How clonal are Trypanosoma and Leishmania?. *Trends in Parasitology*, 29(6), 264–269. https://doi.org/10.1016/j.pt.2013.03.007

Van Drunen, W. E., & Husband, B. C. (2019). Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. *New Phytologist*, *224*(3), 1266–1277. https://doi.org/10.1111/nph.15999

Villate, L., Esmenjaud, D., Van Helden, M., Stoeckel, S., & Plantard, O. (2010). Genetic signature of amphimixis allows for the detection and fine scale localization of sexual reproduction events in a mainly parthenogenetic nematode. *Molecular Ecology*, *19*(5), 856–873. https://doi.org/10.1111/j.1365-294X.2009.04511.x

Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Molecular Ecology*, *10*(1), 249–256. https://doi.org/10.1046/j.1365-294x.2001.01185.x

Wakeley, J. (2005). The Limits of Theoretical Population Genetics. *Genetics*, *169*(1), 1–7.

Weir, B.S. (1996) Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Inc., Sunderland.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the*

National Academy of Sciences, *106*(33), 13875–13879. https://doi.org/10.1073/pnas.0811575106

Wu, R., Gallo-Meagher, M., Littell, R. C., & Zeng, Z. B. (2001). A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. *Genetics*, *159*(2), 869–882.

Wu, S., Cheng, J., Xu, X., Zhang, Y., Zhao, Y., Li, H., & Qiang, S. (2019). Polyploidy in invasive Solidago canadensis increased plant nitrogen uptake, and abundance and activity of microbes and nematodes in soil. *Soil Biology and Biochemistry*, 138, 107594. https://doi.org/10.1016/j.soilbio.2019.107594.

Yu, F.-H., Roiloa, S. R., & Alpert, P. (2016). Editorial: Global Change, Clonal Growth, and Biological Invasions by Plants. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/articles/10.3389/fpls.2016.01467

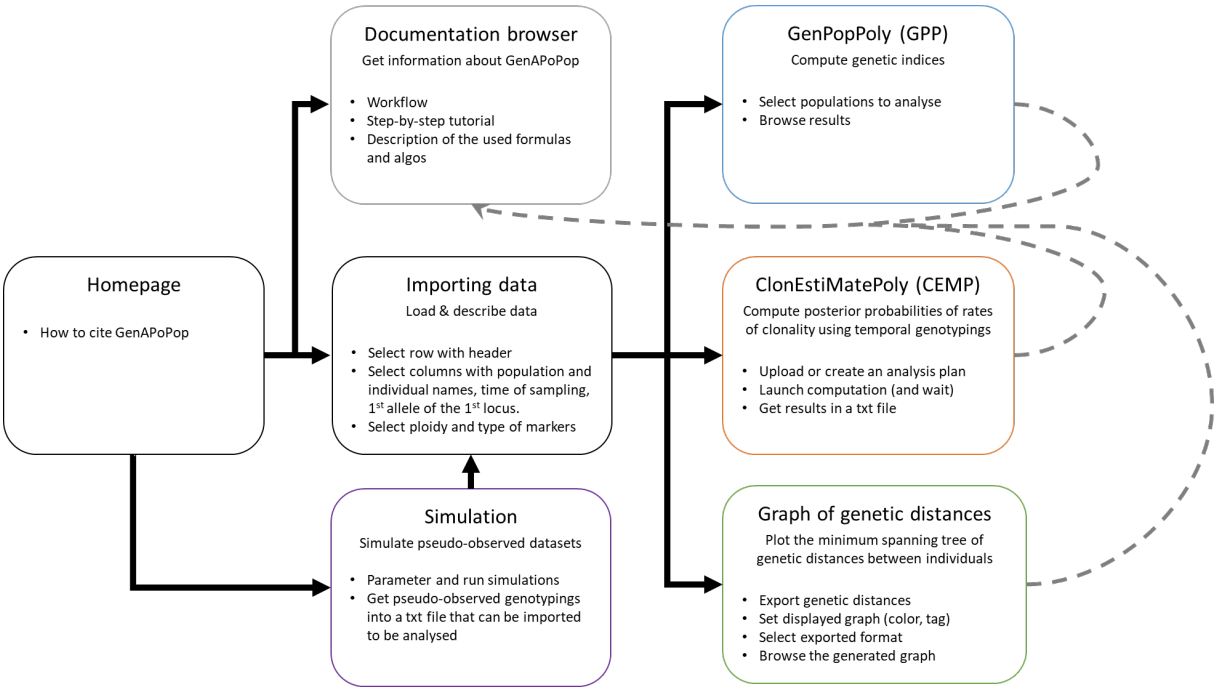


Figure 1: Workflow in GENAPOPOP. Users first import dataset, either from the embedded simulator or from external sources; Second, describe the data structure, and then launch at least one of the three types of analyses. Full connectors indicate the possible workflows, dashed connectors indicate optional possibility to consult documentation using the embedded light PDF reader. Results can be browsed within the software and by opening the exported files using common spreadsheets and text editors.

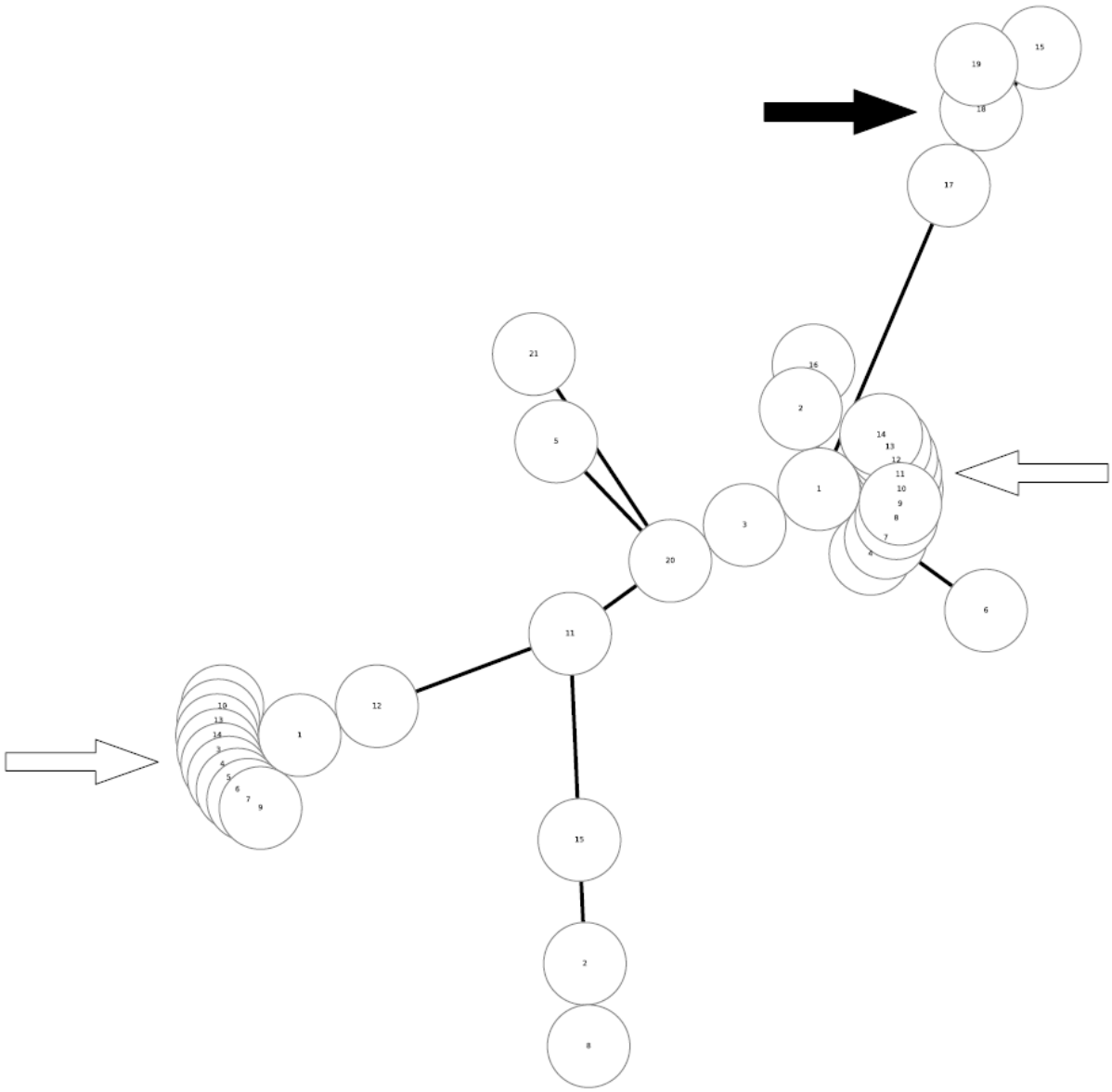


Figure 2: Minimum spanning tree of the genetic distances in the AS dataset. White arrows indicate rosettes of multilocus genotypes differing from one allele from a central multilocus genotype, suspected to be recent mutants of a same multilocus lineage. Black arrow indicates a rosary pattern of multilocus genotypes differing from few alleles, suspected to be clones of a same multilocus lineage which would have accumulated a small number of mutations over the clonal generations.

Table1: An example of formatted triploid dataset ready to be analyzed by GENAPOPOP. Bold headers indicate the minimum required columns per individual; italic headers optional columns expected to format the minimum spanning tree of the genetic distances between individuals. One or multiple additional columns with custom information like ecological, physiological, traits, latitude and longitude, etc. can figure anywhere before the column containing the first allele of the first locus.

Pop	**Gen**	**ID**	Info	*Col*	*Tag*	**1A_1L**	**2A_1L**	**3A_1L**	**1A_L2**
pop1	**1**	**Ind1**	…	*Blue*	*p1_i1*	**A**	**A**	**G**	**T**
pop1	**2**	**Ind2**	…	*Red*	*p1_i2*	**A**	**G**	**G**	**T**
…	…	…	…	…	…	…	…	…	…
popn	**2**	**Ind30**	…	*orange*	*pn_i30*	**A**	**A**	**A**	**C**

Supplemental Information for:

GENAPOPOP 1.0: a user-friendly software to analyze genetic diversity and structure from partially clonal and selfed autopolyploid organisms

Solenn Stoeckel, Ronan Becheler, Ekaterina Bocharova, Dominique Barloy

Table of Contents:

Genetic segregation model in autopolyploids

GENAPOPOP doesn't consider yet for double-reduction as it assumes a *random chromosome segregation model* (Muller 1914). GENAPOPOP thus ignores *pure random chromatid segregation model* where chromatids randomly segregate into gamete resulting in a rate of double-reduction of 1/7 for tetrasomic inheritance (Haldane 1930, 1935) and *complete and partial equational segregation model* where whole arms of sister chromatids are exchanged by recombination into different chromosomes, resulting in a rate of double-reduction of 1/6 when complete equational segregation occurs (Mather 1935, Huang et al. 2019). Even if less commonly observed (Wu et al. 2001), these segregation mechanisms may have deep implications for population genetics analyses (Huang et al. 2019, Jiang et al. 2021).

References

Haldane, J. B. S. (1930). Theoretical genetics of autopolyploids. Journal of Genetics, 22(3), 359–372. https://doi.org/10.1007/BF02984197

Haldane, K. (1935). Reductional and equational separation of the chromosomes in bivalents and multivalents. Journal of Genetics, 30(1), 53–78. https://doi.org/10.1007/BF02982205

Huang, K., Wang, T., Dunn, D. W., Zhang, P., Cao, X., Liu, R., & Li, B. (2019). Genotypic Frequencies at Equilibrium for Polysomic Inheritance Under Double-Reduction. G3 Genes|Genomes|Genetics, 9(5), 1693–1706. https://doi.org/10.1534/g3.119.400132

Jiang, L., Ren, X., & Wu, R. (2021). Computational characterization of double reduction in autotetraploid natural populations. The Plant Journal, 105(6), 1703–1709. https://doi.org/10.1111/tpj.15126

Mather, K. (1935). Reductional and equational separation of the chromosomes in bivalents and multivalents. *Journal of Genetics*, *30*(1), 53–78. https://doi.org/10.1007/BF02982205

Muller, H. J. (1914). A New Mode of Segregation in Gregory's Tetraploid Primulas. The American Naturalist, 48(572), 508–512. https://doi.org/10.1086/279426

Wu, R., Gallo-Meagher, M., Littell, R. C., & Zeng, Z. B. (2001). A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics, 159(2), 869–882.

Table S1: Comparison of four classic population genetic indices computed to compare the consistency of GENAPOPOP with the output of SPAGEDI reference software (Hardy & Vekemans 2001) on four autotetraploid simulated datasets, each obtained simulating two populations of 100 individuals connected with a migration rate of 0.01 and mutating at a rate of 0.01, 1000 generations after an initial randomly drawing population. A, B, C and D scenarios respectively stand for panmixia; high selfing; high clonality; half-clonal half-selfed reproductive modes. *Lgh* and *As* are two tetraploid field datasets each composed of two populations. *Lgh* includes two populations of 75 genotypes each genotyped with 36 SNPs. As includes one population of 21 individuals and one population of 15 individuals genotyped with 10 microsatellites. Raw data are available on ZENODO (Barloy et al. 2022, DOI: 10.5281/zenodo.8164531). *Ae* stands for the average effective number of alleles on the whole dataset (Weir 1996), *He* for the overall genetic diversity, *Ho* for the observed heterozygosity and *rhost* for the genetic differentiation between populations being independent of double-reduction and ploidy level.

		Dataset						
Index	**Program**	**A**	**B**	**C**	**D**	***As***	***Lgh***	**Mean difference (index)**
Ae	Spagedi	2.97	1.94	2.88	2.82	1.21	1.61	0.008
	GenAPoPop	2.97	1.93	2.88	2.82	1.21	1.57	
He	Spagedi	0.6608	0.4202	0.6291	0.6236	0.1269	0.2689	0.0017
	GenAPoPop	0.6604	0.4200	0.6287	0.6232	0.1267	0.2601	
Ho	Spagedi	0.659	0.126	0.613	0.517	0.152	0.246	0.0
	GenAPoPop	0.659	0.126	0.613	0.517	0.152	0.246	
Rhost	Spagedi	0.0157	0.089	0.0454	0.0801	0.037	0.5182	0.0
	GenAPoPop	0.0157	0.089	0.0454	0.0801	0.037	0.5182	

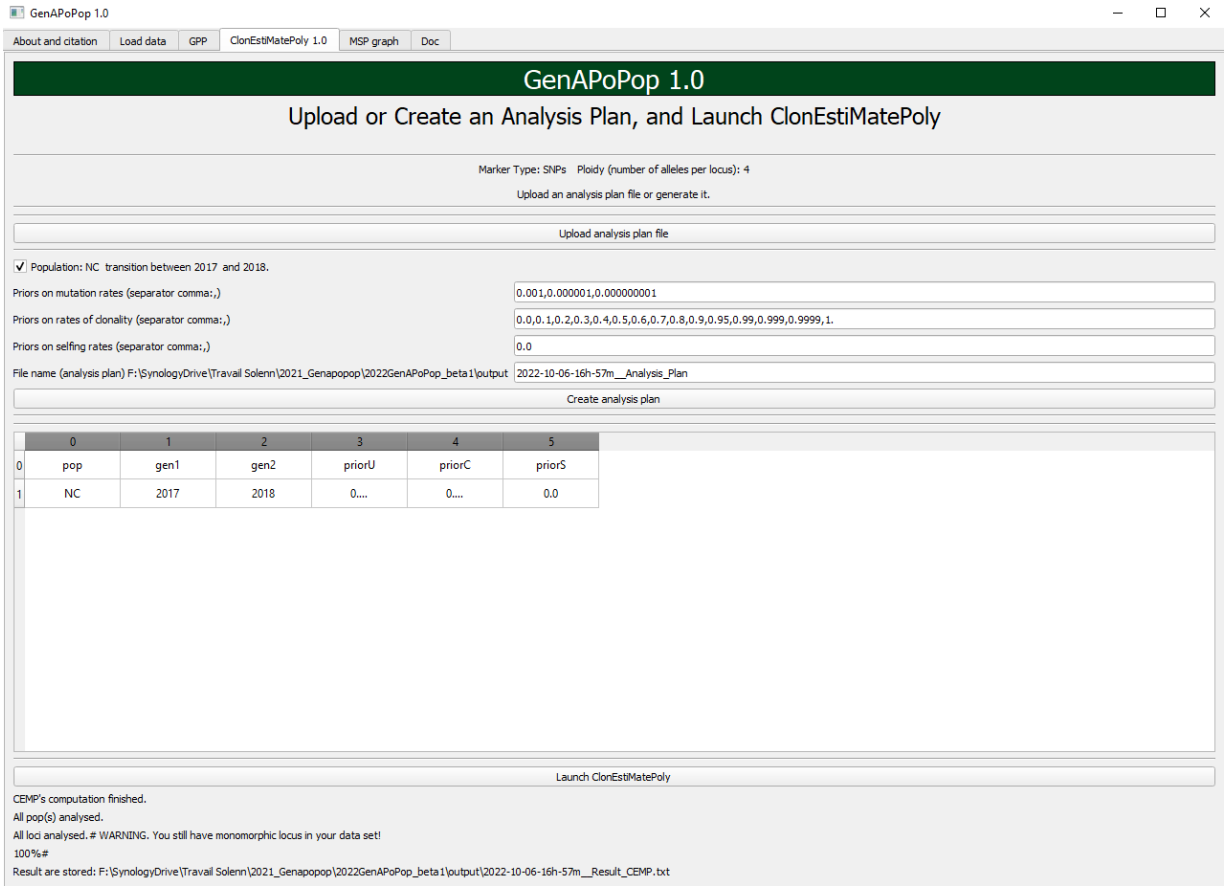
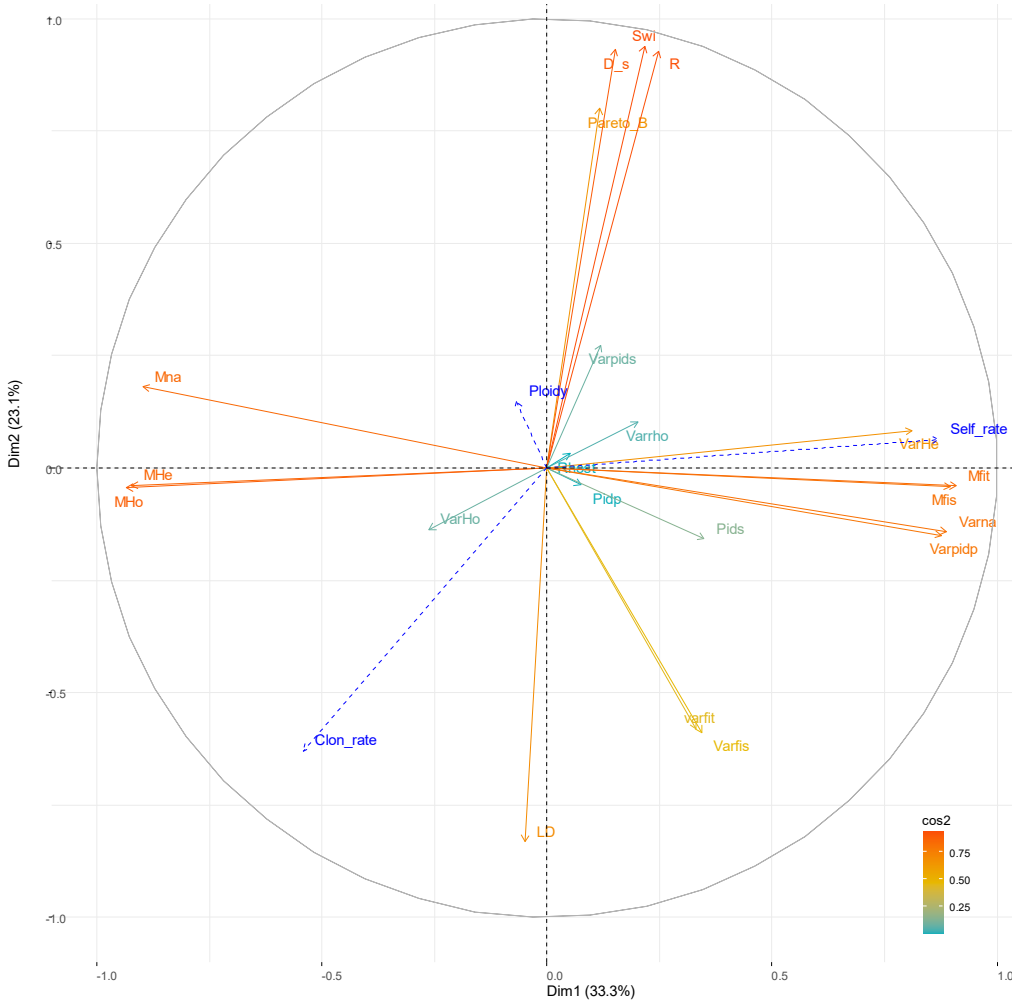


Figure S1: screenshot of the ClonEstiMatePoly tab and its menus to select populations genotyped at two time-step to be analysed and the list of discretized priors of mutation rate, rates of clonality and selfing that will be evaluated. The analysis plan can be previewed in the browser part of the windows before launching the Bayesian computation. This is an extension of the Bayesian method proposed in Becheler et al. (2017) to autopolyploids.



R: genotypic diversity (Arnaud-Haond et al. 2007)

BetaPareto : clonal evenness fitted as a Pareto distribution (Arnaud-Haond et al. 2007)

Dstar: Complement of the Simpson index then describes the probability of sampling distinct MLGs when randomly drawing two units in the sample (Arnaud-Haond et al. 2007).

SW_index: Shannon Wienner's index of clonal diversity (Arnaud-Haond et al. 2007)

rbarD: Mean Linkage disequilbrium over all loci (Agapow & Burt 2000).

Mna and **varna**: Mean number of alleles per locus and its variance over loci.

MHe and **VarHe**: Mean gene diversity over all loci and its variance among loci.

MHo and **VarHo**: Mean observed heterozygosity over all loci and its variance among loci.

pid_p and **Varpidp**: probability of identity of a pair of individuals expected under panmixia and its variance over loci.

pid_sib and **Varpids**: probability of identity of a pair of siblings expected under panmixia and its variance over loci.

Mfis and **Varfis**: Mean Fis value and variance of Fis values among loci.

Mfst and **Varfst**: Mean Fst value and variance of Fst values among loci.

Mfit and **Varfit**: Mean Fit value and variance of Fit values among loci.

Rhost and **Varrho**: Mean rhost value and variance of rhost values among loci (Ronfort et al. 1998).

Figure S2: Correlation circles of the principal component analysis on population genetic indices from simulated populations. The first dimension accounting for 33.3% of the total variance varies with rates of selfing while the second, accounting for 23.1% of the total variance is rather colinear to rates of clonality. Contributions of each population indices are reported with the color code 'cos2' ranging from red (high contribution) to green blue (low contribution).

Prediction of the ploidy (Ploidy), rates of clonality (Clon_rate) and rates of selfing (Self_rate) that produced the simulated values are reported as independent variables.

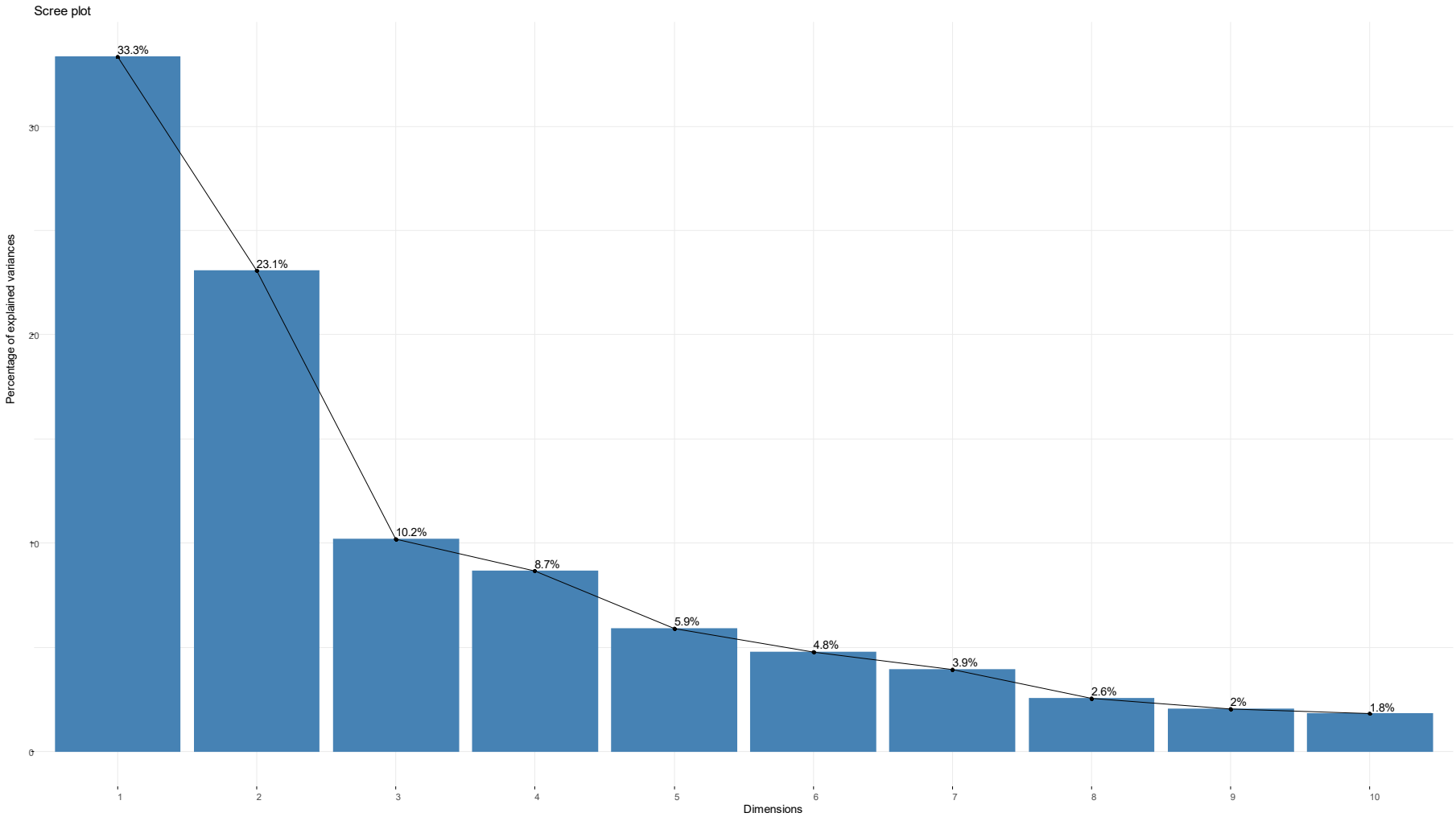
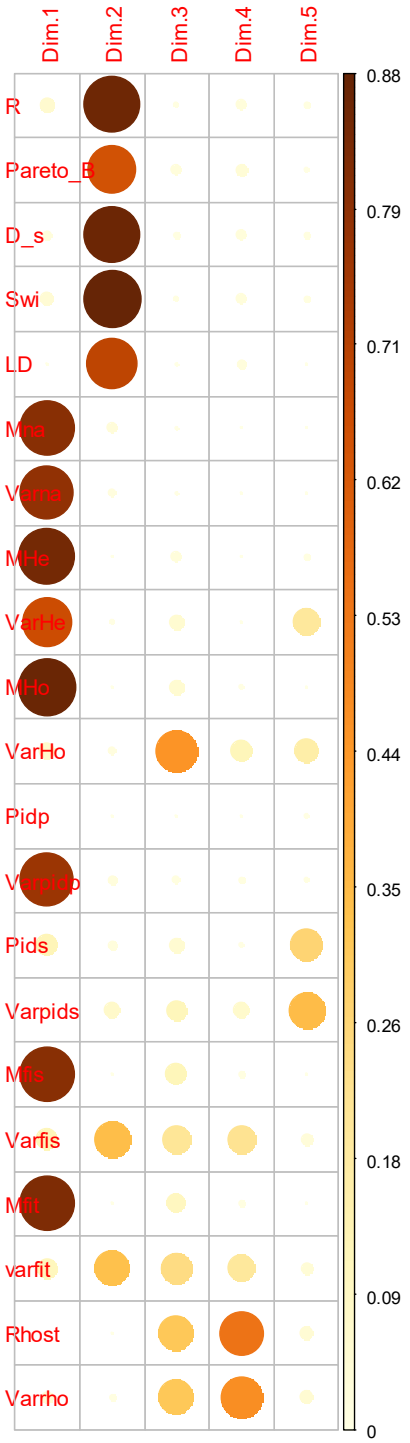


Figure S3: Scree plot of the percentage of the explained variances of each dimension. We see that the two first dimensions, colinear to rates of selfing and rates of clonality respectively, account for most of the explained variance.



R: genotypic diversity (Arnaud-Haond et al. 2007)

Pareto_B : clonal evenness fitted as a Pareto distribution (Arnaud-Haond et al. 2007)

D_s: Complement of the Simpson index then describes the probability of sampling distinct MLGs when randomly drawing two units in the sample (Arnaud-Haond et al. 2007).

Swi_index: Shannon Wienner's index of clonal diversity (Arnaud-Haond et al. 2007)

LD: Mean Linkage disequilbrium over all loci (Agapow & Burt 2000).

Mna and Varna: Mean number of alleles per locus and it variance over loci.

MHe and VarHe: Mean gene diversity over all loci and its variance among loci.

MHo and VarHo: Mean observed heterozygosity over all loci and its variance among loci.

Pidp and Varpidp: probability of identity of a pair of individuals expected under panmixia and its variance over loci.

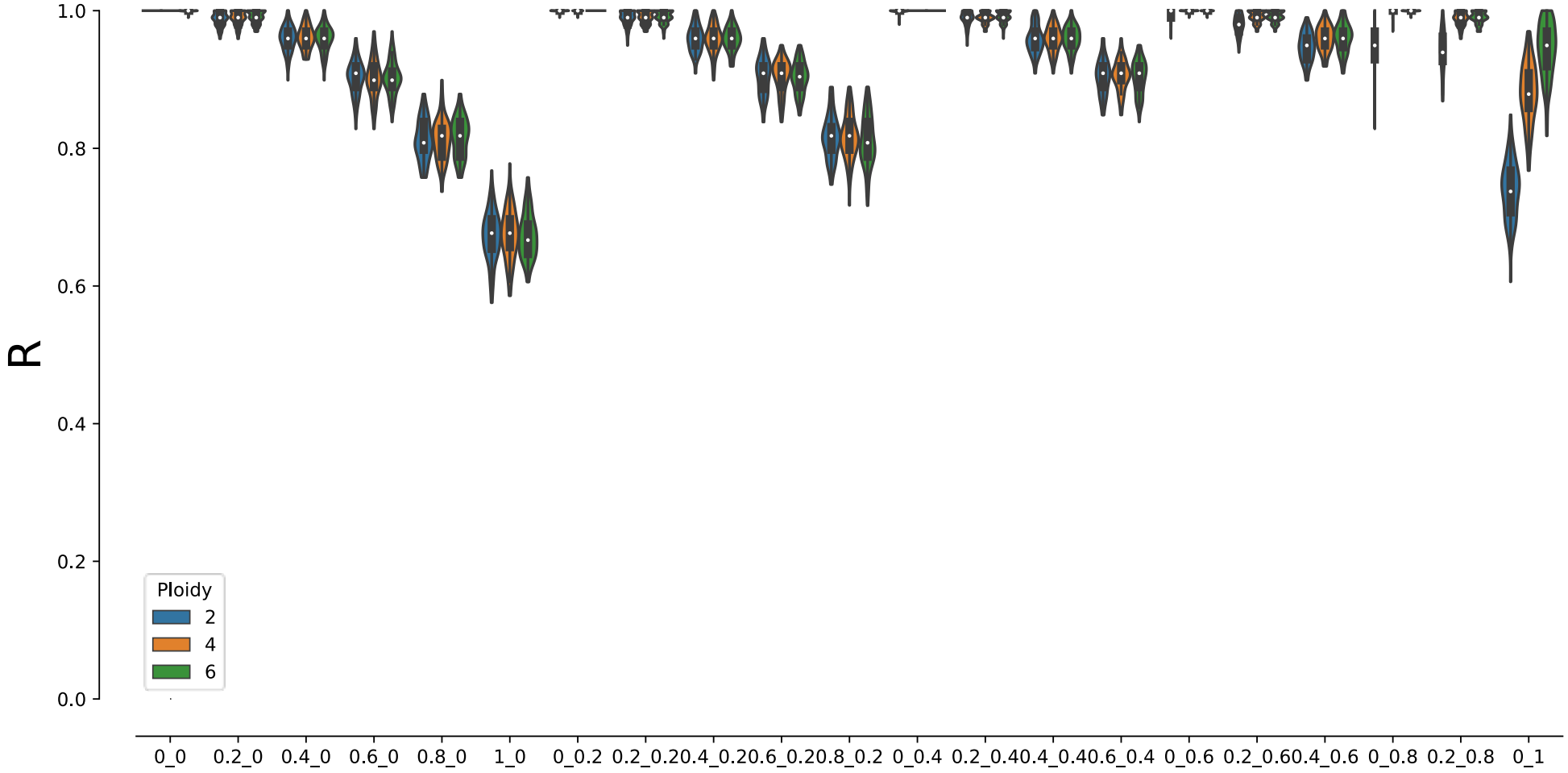
Pids and Varpids: probability of identity of a pair of siblings expected under panmixia and its variance over loci.

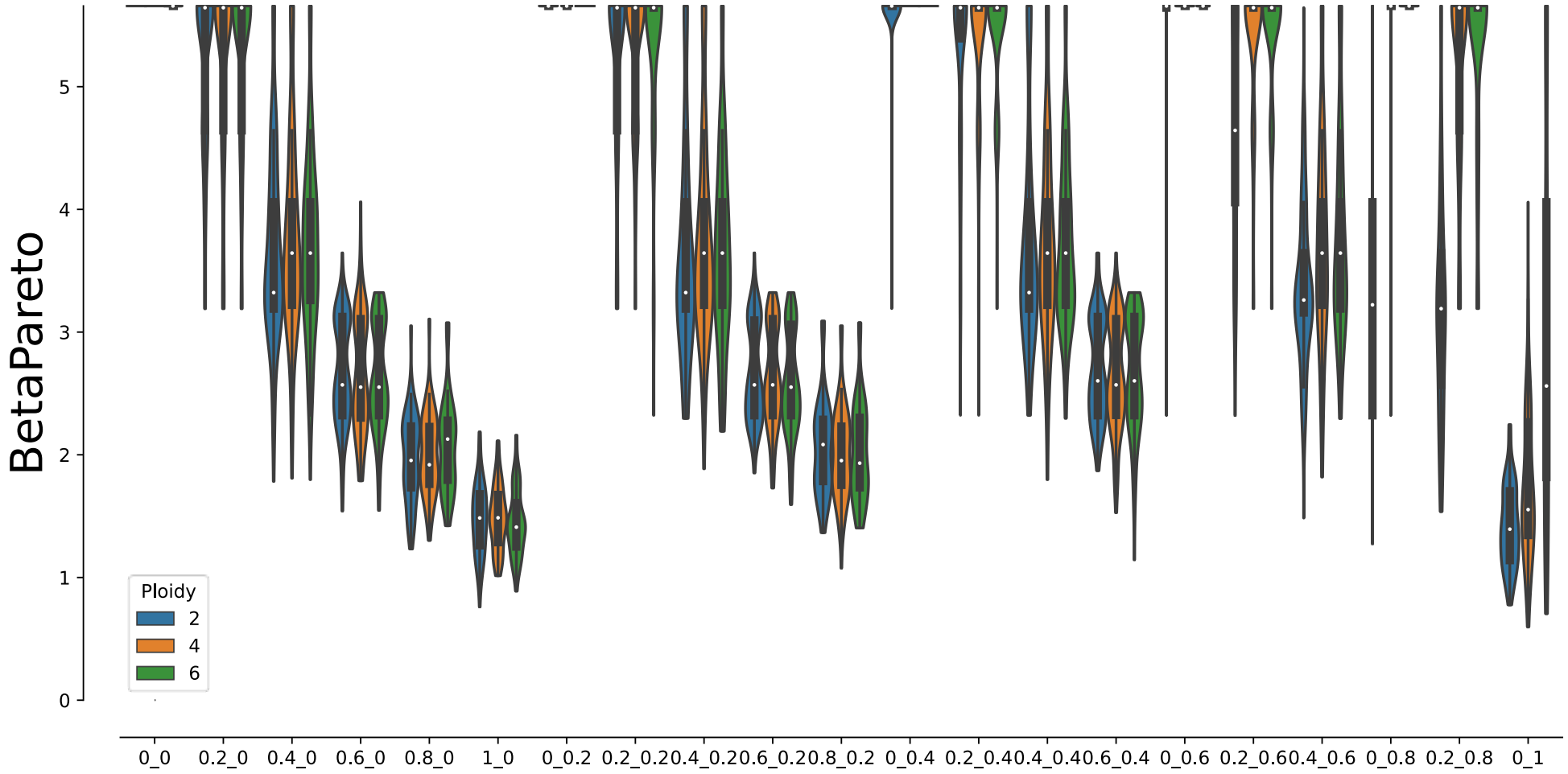
Mfis and Varfis: Mean Fis value and variance of Fis values among loci.

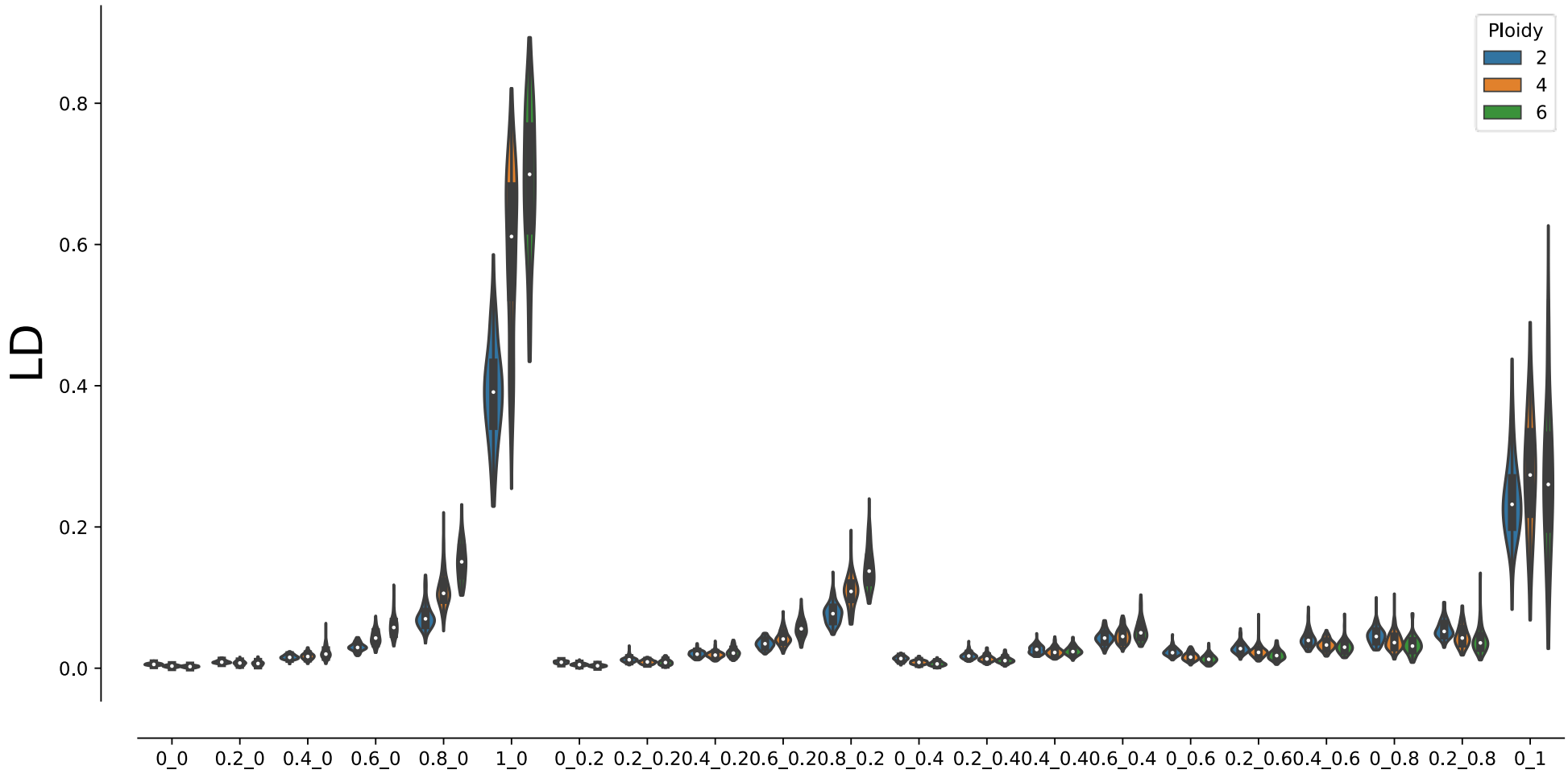
Mfit and Varfit: Mean Fit value and variance of Fit values among loci.

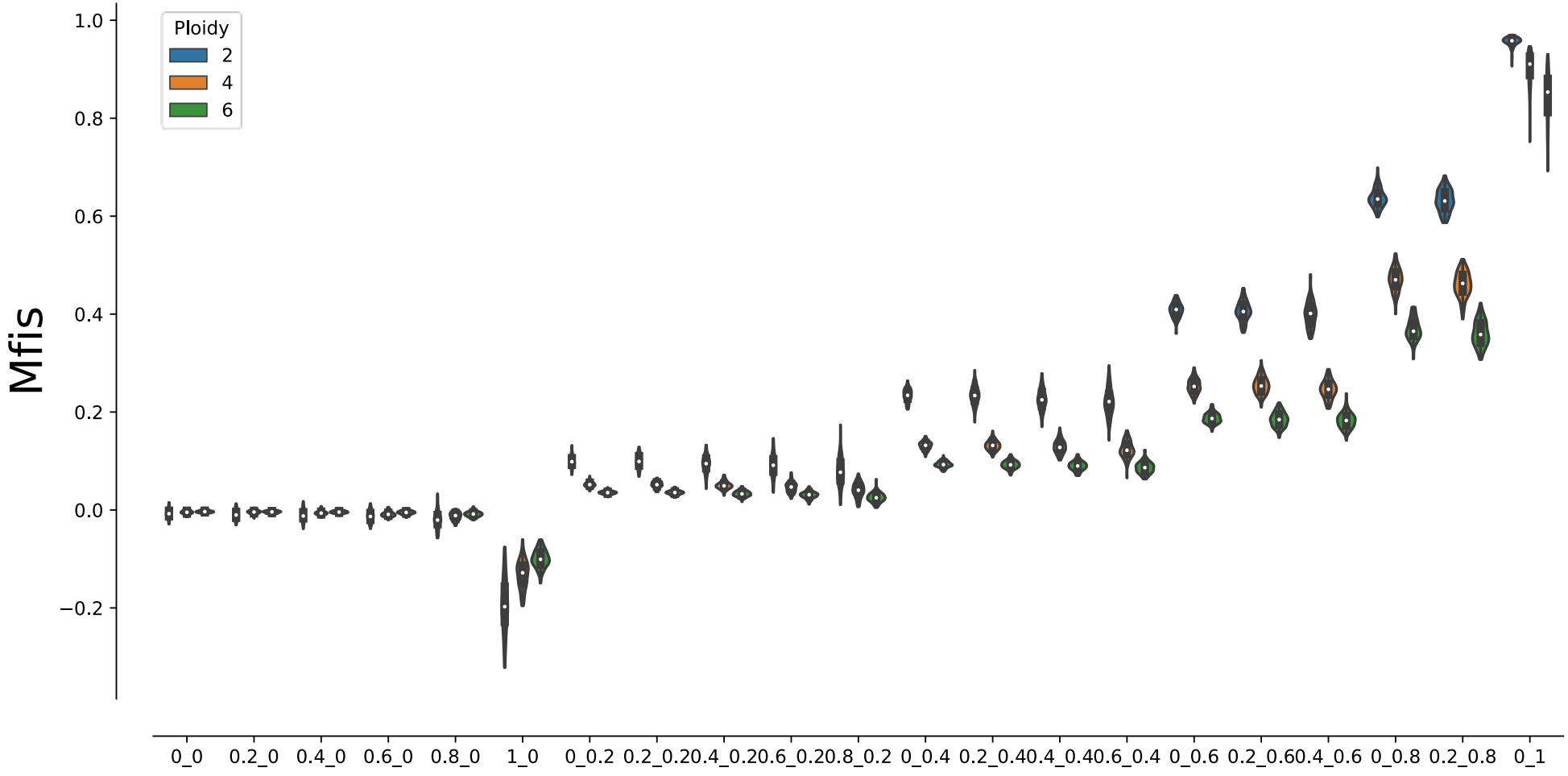
Rhost and Varrho: Mean rhost value and variance of rhost values among loci (Ronfort et al. 1998).

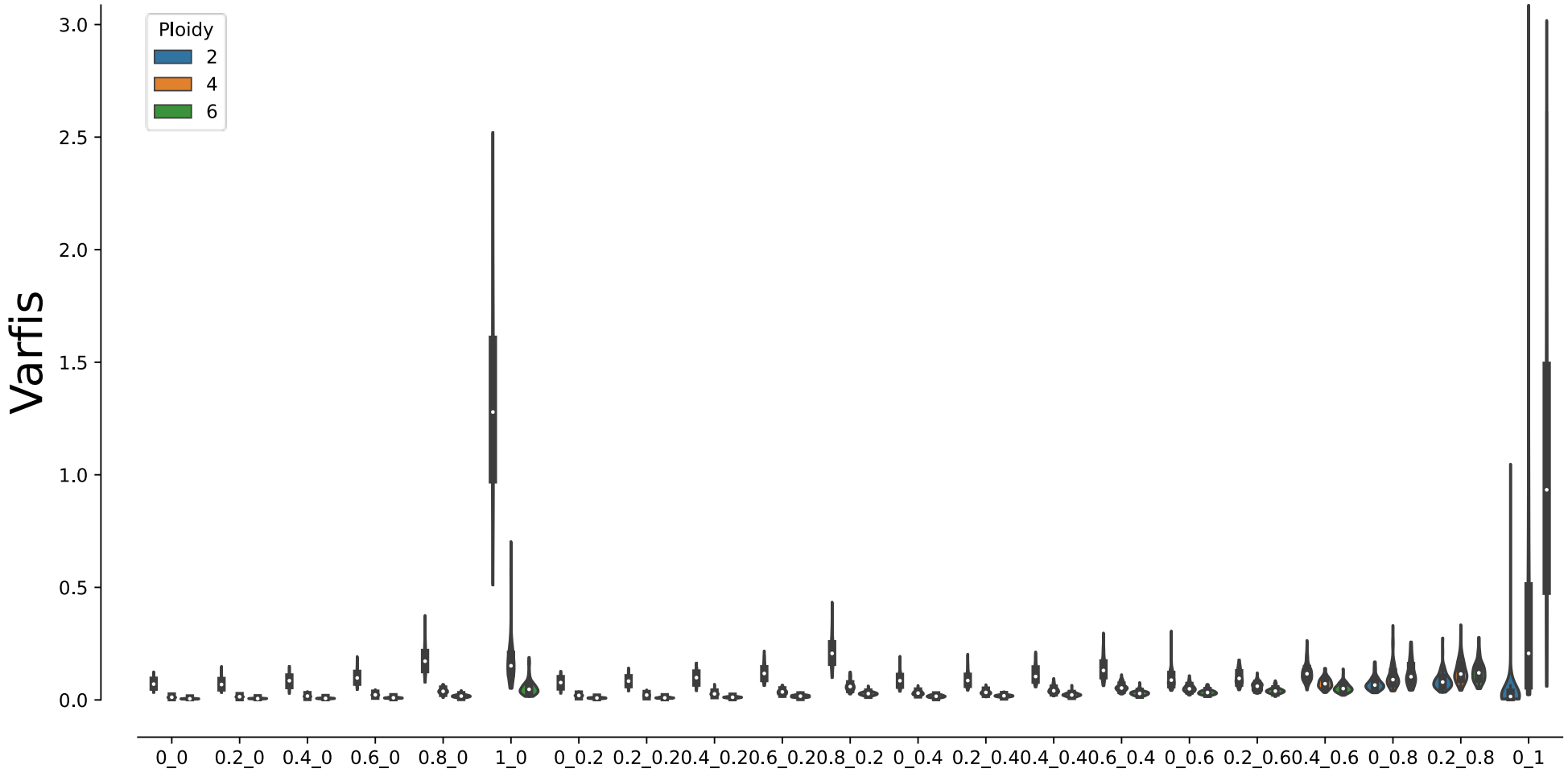
Figure S4: Contributions of population genetic indices to the total variance of genetic diversity in simulated datasets. Contributions of indices to dimensions are reported following the color scale on the right of the matrix.

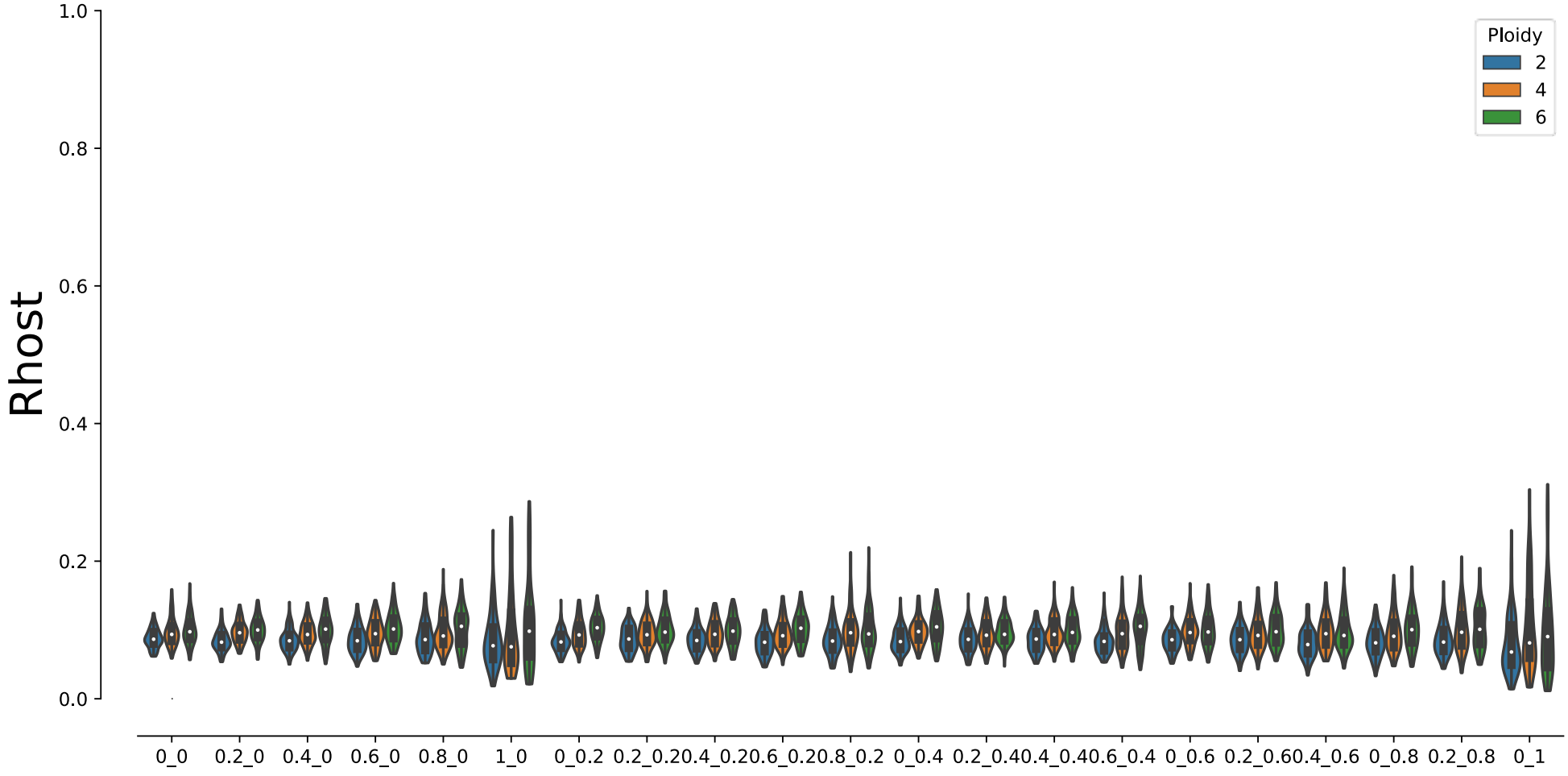

R
1.0
0.8
0.6
0.4
0.2
0.0
Ploidy
2
4
6
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1


BetaPareto
0
1
2
3
4
5
Ploidy
2
4
6
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1


Ploidy
2
4
6
LD
0.0
0.2
0.4
0.6
0.8
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1


Ploidy
2
4
6
Mfis
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1


Ploidy
2
4
6
Varfis
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1


Rhost
1.0
0.8
0.6
0.4
0.2
0.0
Ploidy
2
4
6
0_0
0.2_0
0.4_0
0.6_0
0.8_0
1_0
0_0.2
0.2_0.2
0.4_0.2
0.6_0.2
0.8_0.2
0_0.4
0.2_0.4
0.4_0.4
0.6_0.4
0_0.6
0.2_0.6
0.4_0.6
0_0.8
0.2_0.8
0_1

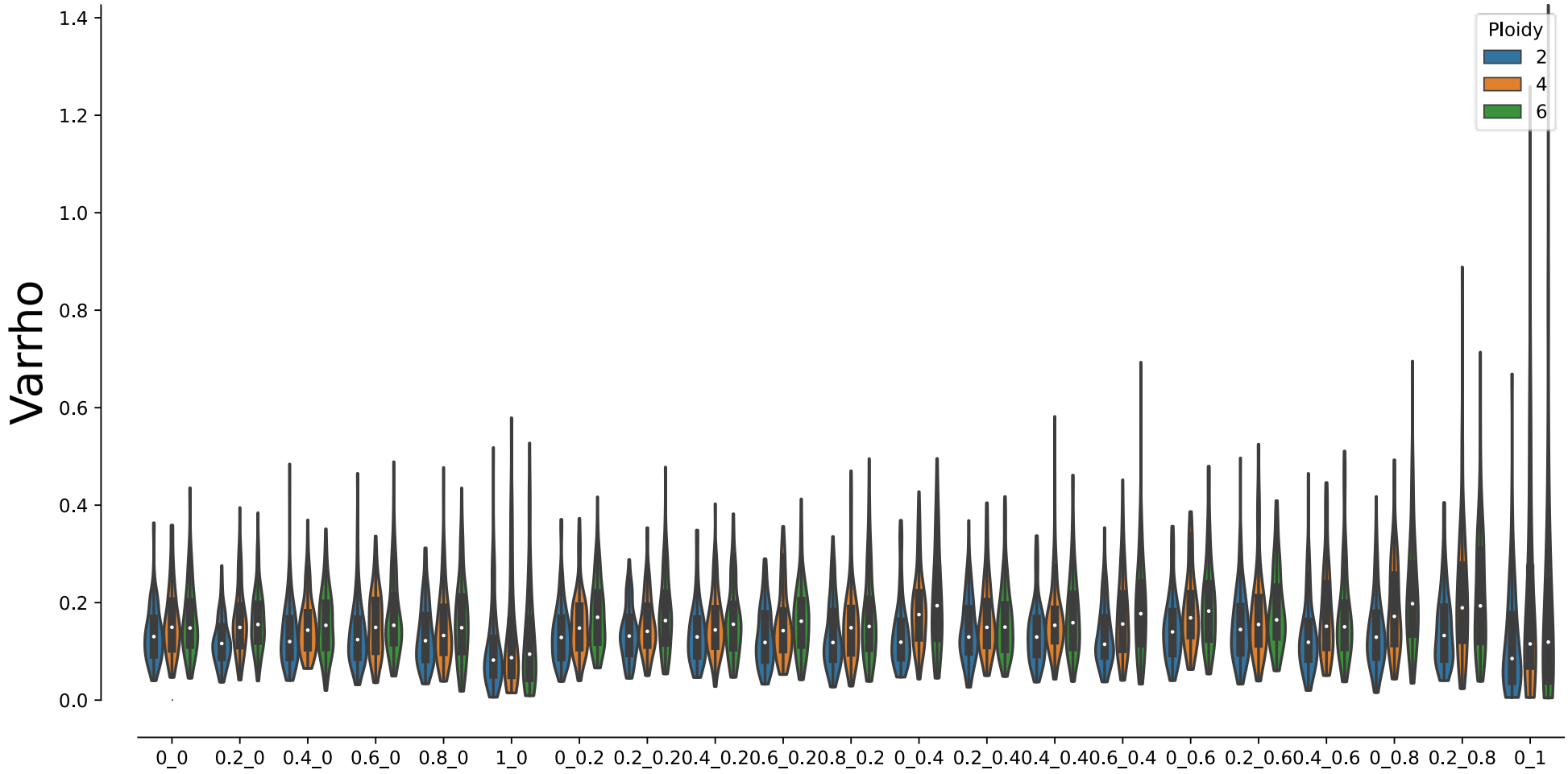
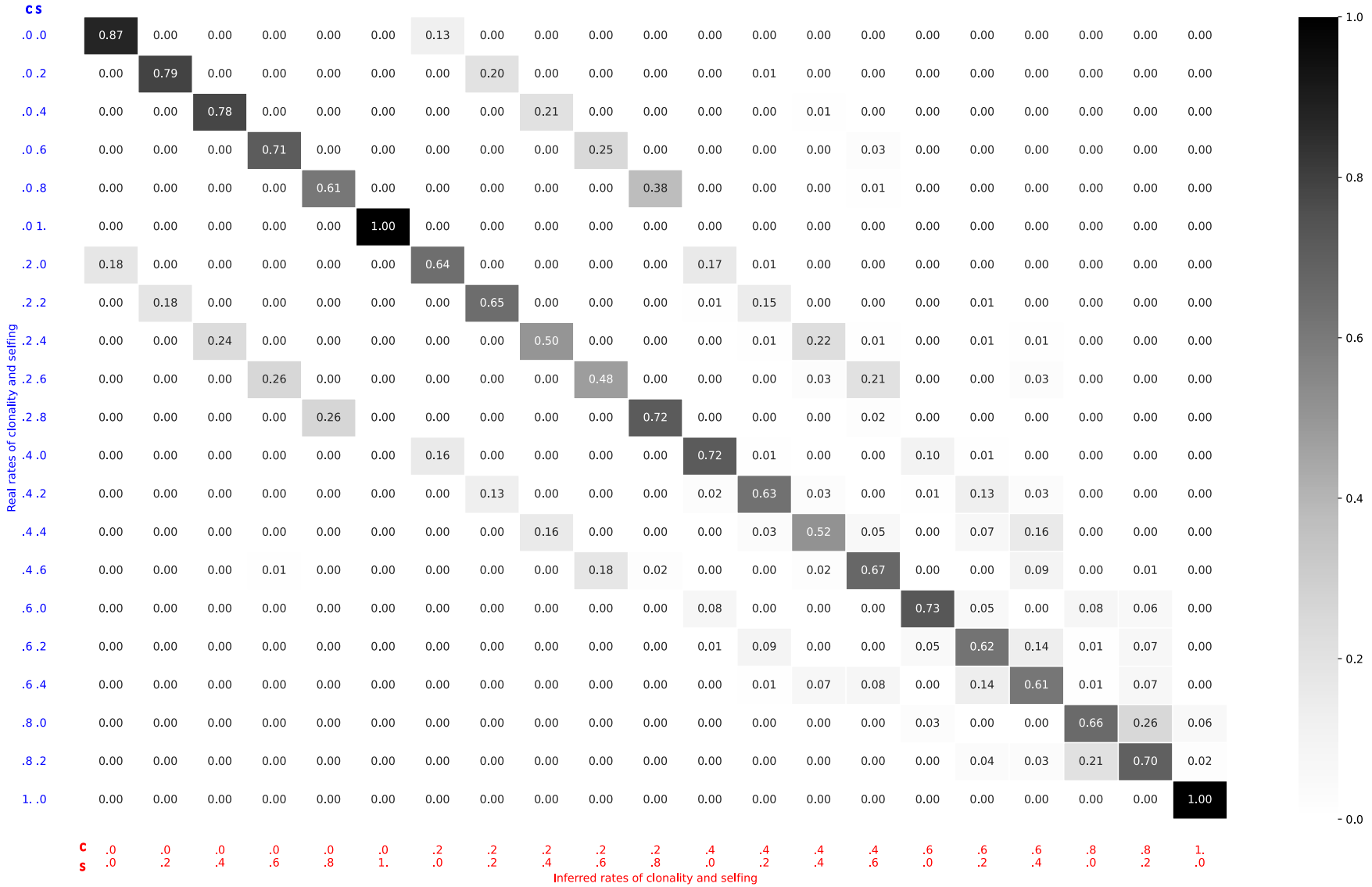


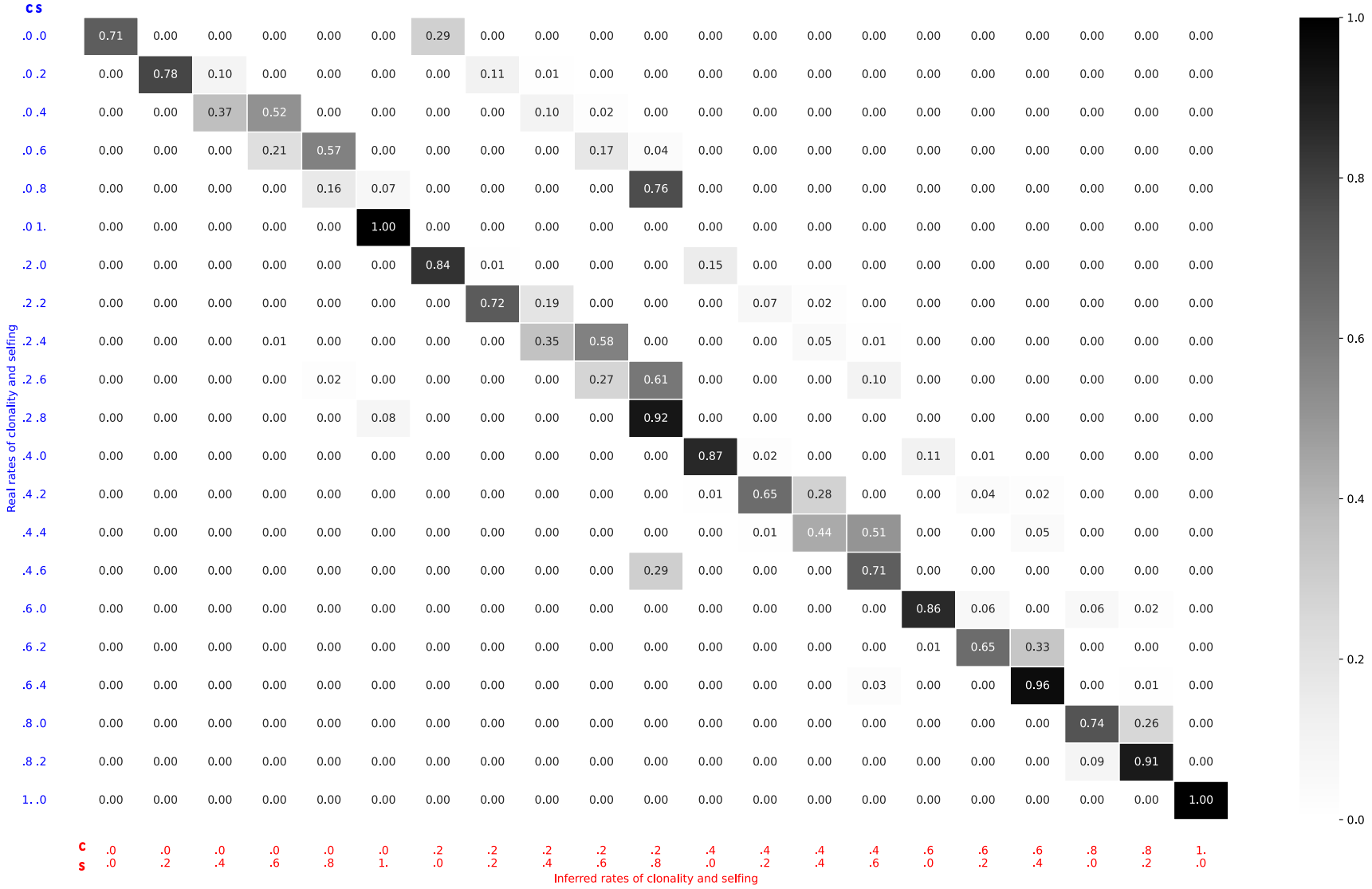
Figure S5: Distributions of population genetic indices with joint rates of clonality (*c*) and selfing (*s*) reported as couple of *c_s* on the x-axis and with ploidy (diploids in blue, tetraploids in orange and hexaploids in green). *R* for genotypic diversity, *BetaPareto* for Pareto β, *LD* for linkage disequilibrium over all loci,

Mfis for mean Fis value, *Varfis* for variance of Fis among genotyped loci, *Rhost* for mean rhost value over all loci between the two simulated populations and *Varrho* for variance of rhost values among loci. Each distribution is reported as a violin obtained from 100 values from 100 independent simulations.

1 Ploidy = 2


c s
.0 .0 0.87 0.00 0.00 0.00 0.00 0.00 0.13 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .2 0.00 0.79 0.00 0.00 0.00 0.00 0.00 0.20 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .4 0.00 0.00 0.78 0.00 0.00 0.00 0.00 0.00 0.21 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .6 0.00 0.00 0.00 0.71 0.00 0.00 0.00 0.00 0.00 0.25 0.00 0.00 0.00 0.00 0.03 0.00 0.00 0.00 0.00 0.00 0.00
.0 .8 0.00 0.00 0.00 0.00 0.61 0.00 0.00 0.00 0.00 0.00 0.38 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00
.0 1. 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .0 0.18 0.00 0.00 0.00 0.00 0.00 0.64 0.00 0.00 0.00 0.00 0.17 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .2 0.00 0.18 0.00 0.00 0.00 0.00 0.00 0.65 0.00 0.00 0.00 0.01 0.15 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00
.2 .4 0.00 0.00 0.24 0.00 0.00 0.00 0.00 0.00 0.50 0.00 0.00 0.00 0.01 0.22 0.01 0.00 0.01 0.01 0.00 0.00 0.00
.2 .6 0.00 0.00 0.00 0.26 0.00 0.00 0.00 0.00 0.00 0.48 0.00 0.00 0.00 0.03 0.21 0.00 0.00 0.03 0.00 0.00 0.00
.2 .8 0.00 0.00 0.00 0.00 0.26 0.00 0.00 0.00 0.00 0.00 0.72 0.00 0.00 0.00 0.02 0.00 0.00 0.00 0.00 0.00 0.00
.4 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.16 0.00 0.00 0.00 0.00 0.72 0.01 0.00 0.00 0.10 0.01 0.00 0.00 0.00 0.00
.4 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.13 0.00 0.00 0.00 0.02 0.63 0.03 0.00 0.01 0.13 0.03 0.00 0.00 0.00
.4 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.16 0.00 0.00 0.00 0.03 0.52 0.05 0.00 0.07 0.16 0.00 0.00 0.00
.4 .6 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.18 0.02 0.00 0.00 0.02 0.67 0.00 0.00 0.09 0.00 0.01 0.00
.6 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.08 0.00 0.00 0.00 0.73 0.05 0.00 0.08 0.06 0.00
.6 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.09 0.00 0.00 0.05 0.62 0.14 0.01 0.07 0.00
.6 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.07 0.08 0.00 0.14 0.61 0.01 0.07 0.00
.8 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.00 0.00 0.66 0.26 0.06
.8 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.04 0.03 0.21 0.70 0.02
1. .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
Real rates of clonality and selfing
c s .0 .0 .0 .0 .0 .0 .0 .2 .2 .2 .2 .2 .4 .4 .4 .4 .6 .6 .6 .8 .8 1.
.0 .2 .4 .6 .8 1. .0 .2 .4 .6 .8 .0 .2 .4 .6 .0 .2 .4 .0 .2 .0
Inferred rates of clonality and selfing
1.0
0.8
0.6
0.4
0.2
0.0

2 Ploidy = 4


c s
.0 .0 0.71 0.00 0.00 0.00 0.00 0.00 0.29 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .2 0.00 0.78 0.10 0.00 0.00 0.00 0.00 0.11 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .4 0.00 0.00 0.37 0.52 0.00 0.00 0.00 0.00 0.10 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .6 0.00 0.00 0.00 0.21 0.57 0.00 0.00 0.00 0.00 0.17 0.04 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .8 0.00 0.00 0.00 0.00 0.16 0.07 0.00 0.00 0.00 0.00 0.76 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 1. 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.84 0.01 0.00 0.00 0.00 0.15 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.72 0.19 0.00 0.00 0.00 0.07 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .4 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.35 0.58 0.00 0.00 0.00 0.05 0.01 0.00 0.00 0.00 0.00 0.00 0.00
.2 .6 0.00 0.00 0.00 0.00 0.02 0.00 0.00 0.00 0.00 0.27 0.61 0.00 0.00 0.00 0.10 0.00 0.00 0.00 0.00 0.00 0.00
.2 .8 0.00 0.00 0.00 0.00 0.00 0.08 0.00 0.00 0.00 0.00 0.92 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.4 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.87 0.02 0.00 0.00 0.11 0.01 0.00 0.00 0.00 0.00
.4 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.65 0.28 0.00 0.00 0.04 0.02 0.00 0.00 0.00
.4 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.44 0.51 0.00 0.00 0.05 0.00 0.00 0.00
.4 .6 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.29 0.00 0.00 0.00 0.71 0.00 0.00 0.00 0.00 0.00 0.00
.6 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.86 0.06 0.00 0.06 0.02 0.00
.6 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.65 0.33 0.00 0.00 0.00
.6 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.00 0.00 0.96 0.00 0.01 0.00
.8 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.74 0.26 0.00
.8 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.09 0.91 0.00
1. .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
Real rates of clonality and selfing
c .0 .0 .0 .0 .0 .0 .2 .2 .2 .2 .2 .4 .4 .4 .4 .6 .6 .6 .8 .8 1.
s .0 .2 .4 .6 .8 1. .0 .2 .4 .6 .8 .0 .2 .4 .6 .0 .2 .4 .0 .2 .0
Inferred rates of clonality and selfing
1.0
0.8
0.6
0.4
0.2
0.0

3

Ploidy = 6

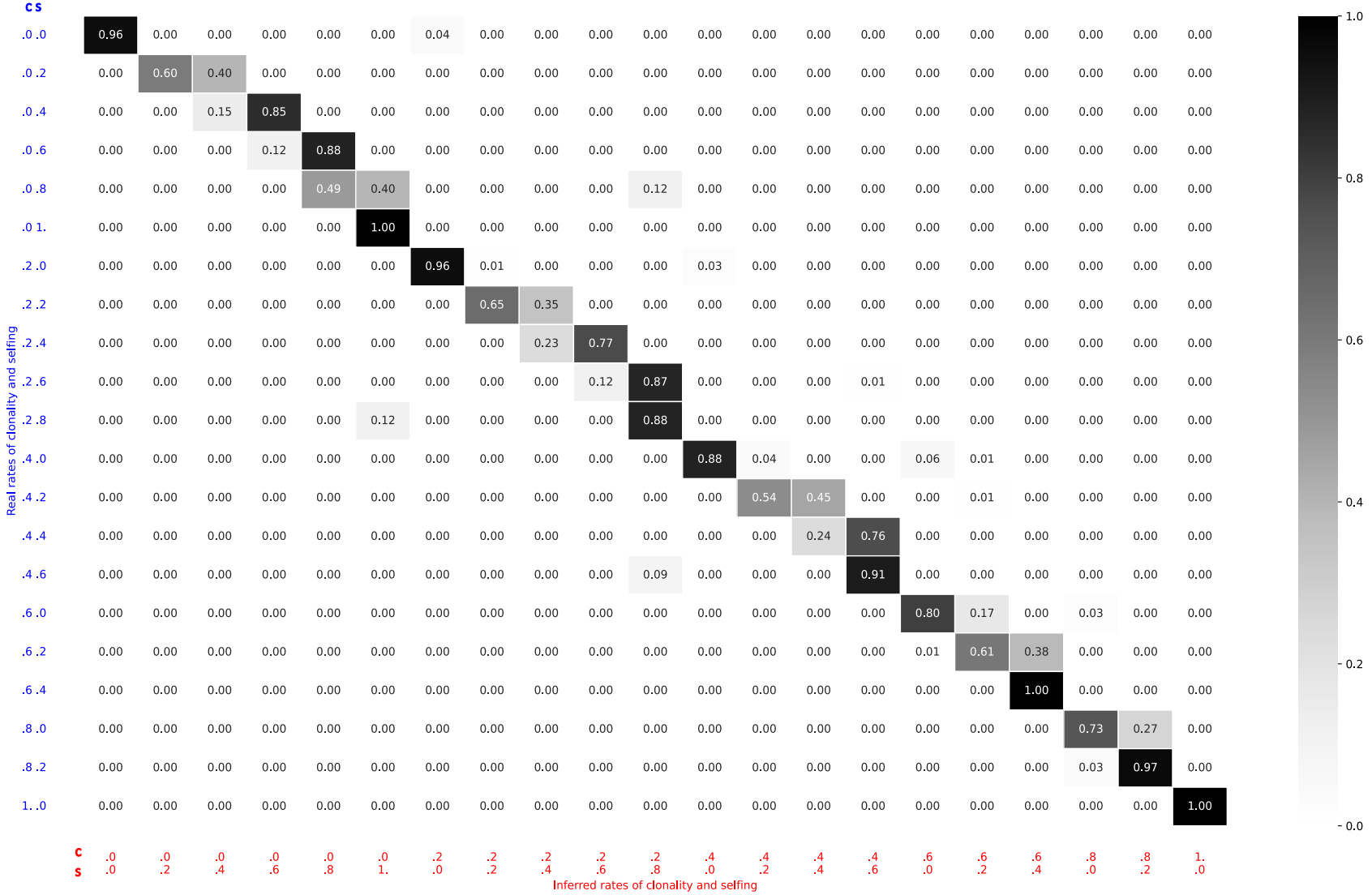

c s
.0 .0 0.96 0.00 0.00 0.00 0.00 0.00 0.04 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .2 0.00 0.60 0.40 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .4 0.00 0.00 0.15 0.85 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .6 0.00 0.00 0.00 0.12 0.88 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 .8 0.00 0.00 0.00 0.00 0.49 0.40 0.00 0.00 0.00 0.00 0.12 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.0 1. 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.96 0.01 0.00 0.00 0.00 0.03 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.65 0.35 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.23 0.77 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.2 .6 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.12 0.87 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00
.2 .8 0.00 0.00 0.00 0.00 0.00 0.12 0.00 0.00 0.00 0.00 0.88 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
.4 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.88 0.04 0.00 0.00 0.06 0.01 0.00 0.00 0.00 0.00
.4 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.54 0.45 0.00 0.00 0.01 0.00 0.00 0.00 0.00
.4 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.24 0.76 0.00 0.00 0.00 0.00 0.00 0.00
.4 .6 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.09 0.00 0.00 0.00 0.91 0.00 0.00 0.00 0.00 0.00 0.00
.6 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.80 0.17 0.00 0.03 0.00 0.00
.6 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.61 0.38 0.00 0.00 0.00
.6 .4 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00
.8 .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.73 0.27 0.00
.8 .2 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.97 0.00
1. .0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
Real rates of clonality and selfing
c .0 .0 .0 .0 .0 .0 .2 .2 .2 .2 .2 .4 .4 .4 .4 .6 .6 .6 .8 .8 1.
s .0 .2 .4 .6 .8 1. .0 .2 .4 .6 .8 .0 .2 .4 .6 .0 .2 .4 .0 .2 .0
Inferred rates of clonality and selfing
1.0
0.8
0.6
0.4
0.2
0.0

Figure S6: Percentage of jointly inferred rates of clonality and selfing for each real couple of rates of clonality and selfing on 100 simulated datasets per real couple of rates of clonality and selfing. The method shows high accuracy to jointly infer true rates of clonality and selfing admitting a precision of ±0.2 (*i.e.*, one step precision of the prior range) of the inferred values.
The method indeed presents a slight tendency to overestimate selfing rates when occurring with intermediate rates of clonality in tetraploid and hexaploidy populations (see for example true c=0.2 and s=0.6 in hexaploids that is inferred with a sum of posteriors of 12% while untrue c=0.2 and s=0.8 is inferred with a sum of posteriors of 87%) and little difficulties to disentangle between no and low selfing when clonality occurs (see for example in tetraploids true c=0.8 and s=0. that is inferred with a sum of posteriors of 16% while untrue c=0.8 and s=0.2 is inferred with a sum of posteriors of 76%).