**Title:** GENAPOPOP 1.0: a user-friendly software to analyse genetic diversity and structure in partially clonal and selfed polyploid organisms

**Short running title:** Population genetic analyses of autopolyploids

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

**Abstract**

Autopolyploidy is quite common in most clades of eukaryotes. The emergence of sequence-based genotyping methods with individual and marker tags enables now 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. Partial clonality, allogamy and selfing are reproductive modes commonly that have deep consequences on the ecology and evolution of population and species. Analysing 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 needed to analyse partially clonal, allogamous and selfed 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 to facilitate practical course and teaching of population genetics for autopolyploid populations.

**Keywords**

Polyploidy, AMOVA, 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 a set of biological entities (i.e., a population) being studied (Wakeley 2005). Using changes of genetic variations over time and space, population genetic models allow quantifying evolutionary forces in genotyped 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 of a population is one of the most significant evolutionary force impacting the dynamics of genetic diversity and its structure among populations as it determines the transmission of the hereditary DNA signal over time (Duminil *et al.* 2007). In return, analysing the genetic diversity within populations allows inferring the reproductive modes of populations, providing a precious knowledge to predict and understand the ecological and biological evolution of studied populations and 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 or amphibians (Gregory & Mable 2005, Mable *et al.* 2011). Polyploidization influences genetic and phenotypic diversity including potential ecological adaptations and radiations, with a long-term dynamic from whole genome duplication to re-diploidization (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). If studying the reciprocal influences of reproductive modes on the ecology and evolution of populations is now usual in diploid populations using their genetic diversity, favoured by a wide range of tools adapted to analyse their genetic diversity like GENCLONE (Arnaud-Haond & Belkhir 2007), RMES (David et al. 2007) and RCLONE (Bailleul et al. 2015), it is less common in polyploid populations; And when rarely achieved, the lack of adapted , easily accessible analysis solution leads studies to consider such datasets as haplotypes or analyse 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 genotypings 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 and 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 genotyping 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 to 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 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).

Second, we also long lacked of adapted models and analysis methods to compute population genetic indices and quantify evolutionary forces in polyploid populations (Dufresne et al. 2014). Due to challenges introduced by data formats and difficulties in generalizing the mathematical formula of population genetic indices (Ewens 2004), the most commonly-used population genetics software solutions were not designed to work with more than two allelic copies per gene, leaving aside researchers and teachers willing to study polyploid species. A handful of library and software emerged in the last years, like the command-line SPAGEDI (Hardy & Vekemans 2002), the more user-friendly recent and multiplateforme 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 interprete reproductive modes, including selfing and clonality in populations, such as indices based on genotypic diversity and individual probabilities of identities. POLYGENE for example cannot handle replicated genotypes like commonly observed in partially clonal populations. POLYGENE and 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 analysing genotypes of polyploid populations with different modes of reproduction, but they require an exhaustive exploration of their documentation and some trainings in scripting language 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 analyse 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). It can be downloaded 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/. 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, by computing indices and methods that were not yet proposed.

It enables analysing genotyped dataset with confident allele dosage of autopolyploid species in which we can neglect double-reduction, i.e., neglecting that a gamete can inherit of a single allele more than once. GENAPOPOP assumes a random chromosome segregation model (Muller 1914), that considers 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). In consequence, users should thus be warned that GENAPOPOP doesn't consider yet for double-reduction. It 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) 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). These possible 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, while without it, as supposed in GenAPoPop in its current version, only *AB*, *AC*, *AD*, *BC*, *BD*, *CD* gametes are considered.

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.

**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. Each time uploading, GENAPOPOP workflow requires to 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 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 repeats 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 analyse them in mass. GENAPOPOP has not limit in the number of populations, of time-steps and genotypes it can analyse 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 either using the integrated GENAPOPOP's browser, interesting in situations where the software must be deployed on workstations without administrator's rights or with restricted access like during practical courses at university, or using the default system PDF file reader, that often provides more comfort and accessibility options than found the basic integrated browser provided within GENAPOPOP. 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 genotype dataset, 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). When the dataset is uploaded and the required line and columns labelled, 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 analyse genetic diversity and population structure of polyploid populations with a special focus on reproductive modes. Users select the population(s) to be analysed, 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 (Waits et al. 2001), the four first moment of inbreeding coefficient $F_{IS}$ in populations, i.e. mean, variance, skewness and kurtosis (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 (*Pareto* $\beta$, Arnaud-Haond et al., 2007). We deliberately discarded many other indices to help users robustly interpreting genotypic diversity in their populations. Despite *Pareto* $\beta$ 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). This table also provides the unbiased probabilities of identity under panmixia and between sibs ($P_{ID}$ and $P_{ID\text{-}SIB}$, Waits et al. 2001) that enable assessing the power of a marker set considering the number of sampled individuals in a population to distinguish between genotypes. These indices give a populational measure of unique genotype probability when *pgen* and *psex* indices provide this measure at the scale of an individual (Arnaud-Haond et al. 2007; Stoeckel et al. 2006; Villate et al. 2010). Finally, it provides over all loci the four first moments of the distributions of the population genetic indices computed per loci.

Then, when selected, the output file provides the results of the analysis of molecular variance (AMOVA) computed following Meirmans & Liu (2018) and Weir (1996) equations and recommendations, including the Fis, Fst, rhost and Fit per population, over all population, per marker and over all marker. It then provides the pairwise-population table of 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 coloured or tagged by population to get a picture of the genetic structure of genotyped populations (see below).

To our knowledge, GENAPOPOP (tab GENPOPPOLY) is the first software to allow computing the *Pareto* $\beta$, unbiased probabilities of identities hypothesizing panmixia and reproduction between sibs, and the four moments of Fis values over loci per population. These indices are useful and efficient to estimate rates of clonality, autogamy (selfing) and allogamy on genotypes of sampled populations sampled at one time (Castric et al. 2002, David et al. 2007, Hardy 2016, Stoeckel et al. 2021).

For easily extend the exploration of the analysed dataset with other existing software, the dataset can also be exported in a SPAGEDI format in the same folder under the same imported data name extend with "*_spagedi_ready.txt*". This file that can be easily imported in other software, including SPAGEDI and POLYGENE, and we greatly encourage future GENAPOPOP users to analyse their data with multiple other analysing 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 step. 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. This method remains accurate 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), using from about ten 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 can be found in the supplementary material. This discretized Bayesian method needs an analysis plan listing discretized priors on rates of

mutation, clonality and selfing for each population. Restricted ranges of prior on each of these parameters allows 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 sampling time as a list of discrete joined values of mutation rate, 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.

To our knowledge, GENAPOPOP (tab CLONESTIMATEPOLY) is the first user-friendly software allowing computing the posterior probabilities of rates of clonality and selfing in polyploid (including diploid) populations genotyped at, at least, two-time step. This method was demonstrated to be the most accurate way to quantitatively assess reproductive modes in populations over multiple Eukaryotes species, especially for detecting low rates of clonality (Becheler et al. 2017). It should facilitate the identification of clonal reproduction and 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).

**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, and customize the plot of the minimum spanning unrooted tree using individual color and tags. The resulting graph can be exported at different resolution into research-standard portable document format (PDF) file format, raster (portable network graphics, PNG) or vector (scalable vector graphics, SVG) image formats. 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.

To our knowledge, GENAPOPOP (tab CLONESTIMATEPOLY) is the simplest and fastest way to get a custom unrooted minimum spanning tree of the identity-in-state between individuals colorized by custom (physiological, ecological, spatial, etc.) features.

## Results

To test the consistency and accuracy of our software, we used simulated data and empirical datasets as control data to determine how GENAPOPOP compares with the referent existing software SPAGEDI on computed population genetic indices. We simulated four test datasets available for further and future unit testing on the European general-purpose open repository ZENODO (Barloy et al. 2022). 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. For each scenario, we recorded the populations' genotyping states over two consecutive generations. These four scenarios correspond

respectively to panmictic (A), highly clonal (B), highly selfed (C) and half-clonal-half-selfed (D) reproductive modes. In addition, we tested GENAPOPOP on two field datasets 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 artic 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 incomplete and missing genotypes, and including some loci fixed in one of the populations.

***insert here Table 2***

Globally, SPAGEDI and GENAPOPOP reported similar values, especially considering that GENAPOPOP uses double-precision floating-point format (64 bits) while SPAGEDI uses a lower precision that can add up along the calculations. Looking at the marginal mean absolute difference between SPAGEDI and GENAPOPOP reports (Table 2), the worst inconsistencies occurred in *As* populations genotyped with a higher number of alleles per marker (microsatellites). GENAPOPOP, intentionally computes estimators with limited 'correction' to avoid to give more weight to some loci rather than other which may biais the global picture of a dataset (see formulas in Supplementary Material). Marginal mean absolute differences between SPAGEDI and GENAPOPOP on genetic indices showed that Ae, Fis and Fit showed more differences. These indices imply divisions of genotype and allele frequencies, which are the more susceptible to be impacted by float precision. GENAPOPOP was designed to complement Genodive, Polygene that performs hierarchical and Bayesian clustering and PARENTAGE analysis, and SPAGEDI that already performs multiple spatial analyses and that can be used to estimate selfing rate. All these three softwares input SPAGEDI -format files. GENAPOPOP allows users to export automatically their data in a SPAGEDI -format file and thus easily extend and access these complementary analyses using these softwares.

## Recommendations and warning

Most, if all population genetic analyses rely on accurate estimates of real populational genotype frequencies, including here CLONESTIMATEPOLY method. The number of 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.

Missing value and null allele compromise comparisons between individuals, lineages and populations, and are susceptible to create biases and misinterpretations. Suspected null allele can be coded as unknow allele with their own specific letters or positive integers, and should be clearly reported before interpretations. Indeed, no “correction” or “assumptions” can enhance a 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 provide a user-friendly, multi-operating systems, efficient mass processing way to analyse polyploid (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 interpreting the genetic diversity in partially clonal, partially selfing autopolyploid populations with no or very-limited double reduction. It includes an extension of the robust and efficient CLONESTIMATE Bayesian method to quantitatively infer rates of clonality (and selfing rate with adapted prior on rates of clonality) using populations genotyped at two-time steps. It has no vocation to include or encompass all methods and population genetic indices that can be computed when analysing autopolyploid genotypings, as its main purpose is to ease analyses helping interpreting reproductive modes in autopolyploids. That’s why we warmly recommend users to use GENAPOPOP in complement to other dedicated analysing software like SPAGEDI (Hardy & Vekemans

2002), GENODIVE (Meirmans 2020) and POLYGENE (Huang *et al.* 2020), depending on the tackled questions.

GENAPOPOP also answers the need of a population genetic analysing software for polyploid dataset 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 globally thank students of the International Master in Biodiversity Ecology and Evolution 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 formalised the mathematical equations, formalized the methods, coded the software and coded the simulator to test 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**

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.7299914

**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*

421 *Ecology Notes*, *1*(1–2), 101–102. https://doi.org/10.1046/j.1471-8278.2000.00014.x
422 Albertin, W., & Marullo, P. (2012). Polyploidy in fungi: Evolution after whole-genome
423 duplication. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738),
424 2497–2509. https://doi.org/10.1098/rspb.2012.0434
425 Arnaud-Haond, S., & Belkhir, K. (2007). genclone: A computer program to analyse genotypic
426 data, test for clonality and describe spatial clonal organization. *Molecular Ecology*
427 *Notes*, *7*(1), 15–17. https://doi.org/10.1111/j.1471-8286.2006.01522.x
428 Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods
429 to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139.
430 https://doi.org/10.1111/j.1365-294X.2007.03535.x
431 Arnaud-Haond, S., Stoeckel, S., & Bailleul, D. (2020). New insights into the population
432 genetics of partially clonal organisms: When seagrass data meet theoretical
433 expectations. *Molecular Ecology*, *29*(17), 3248–3260.
434 https://doi.org/10.1111/mec.15532
435 Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., & Yant, L. (2018). The "Polyploid Hop":
436 Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome
437 Duplications. *Frontiers in Ecology and Evolution*, *6*.
438 https://www.frontiersin.org/articles/10.3389/fevo.2018.00117
439 Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: A package to identify
440 MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and*
441 *Evolution*, *7*(8), 966–970. https://doi.org/10.1111/2041-210X.12550
442 Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative
443 abundance of autopolyploids and allopolyploids. *New Phytologist*, *210*(2), 391–398.
444 https://doi.org/10.1111/nph.13698
445 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 analyse genetic diversity and structure in partially clonal and selfed polyploid organisms. (1.0) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.7299914

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

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

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

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

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.

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 analyse 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., 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

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

Kamvar, 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

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

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

546 Meirmans, P. G., & Liu, S. (2018). Analysis of Molecular Variance (AMOVA) for
547 Autopolyploids. Frontiers in Ecology and Evolution, 6.
548 https://www.frontiersin.org/articles/10.3389/fevo.2018.00066
549 Meirmans, P. G., & Van Tienderen, P. H. (2004). genotype and genodive: Two programs for
550 the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*(4),
551 792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x
552 Meirmans, P. G., & Van Tienderen, P. H. (2013). The effects of inheritance in tetraploids on
553 genetic diversity and population divergence. *Heredity*, *110*(2), 131–137.
554 https://doi.org/10.1038/hdy.2012.80
555 Muller, H. J. (1914). A New Mode of Segregation in Gregory's Tetraploid Primulas. *The*
556 *American Naturalist*, *48*(572), 508–512. https://doi.org/10.1086/279426
557 Orive, M. E., & Krueger-Hadfield, S. A. (2021). Sex and Asex: A Clonal Lexicon. *Journal of*
558 *Heredity*, *112*(1), 1–8. https://doi.org/10.1093/jhered/esaa058
559 Pandit, M. K., Pocock, M. J. O., & Kunin, W. E. (2011). Ploidy influences rarity and
560 invasiveness in plants. *Journal of Ecology*, *99*(5), 1108–1115.
561 https://doi.org/10.1111/j.1365-2745.2011.01838.x
562 Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population
563 genetic software for teaching and research—an update. *Bioinformatics*, *28*(19), 2537–
564 2539. https://doi.org/10.1093/bioinformatics/bts460
565 Reichel, K., Bahier, V., Midoux, C., Parisey, N., Masson, J.-P., & Stoeckel, S. (2015).
566 Interpretation and approximation tools for big, dense Markov chain transition matrices
567 in population genetics. *Algorithms for Molecular Biology*, *10*(1), 31.
568 https://doi.org/10.1186/s13015-015-0061-5
569 Reichel, K., Masson, J.-P., Malrieu, F., Arnaud-Haond, S., & Stoeckel, S. (2016). Rare sex or
570 out of reach equilibrium? The dynamics of FISin partially clonal organisms. *BMC*

571 *Genetics*, *17*(1), 76. https://doi.org/10.1186/s12863-016-0388-z
572 Ronfort, J., Jenczewski, E., Bataillon, T., & Rousset, F. (1998). Analysis of population
573 structure in autotetraploid species. *Genetics*, *150*(2), 921–930.
574 Rutland, C. A., Hall, N. D., & McElroy, J. S. (2021). The Impact of Polyploidization on the
575 Evolution of Weed Species: Historical Understanding and Current Limitations.
576 *Frontiers in Agronomy*, *3*.
577 https://www.frontiersin.org/articles/10.3389/fagro.2021.626454
578 Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., &
579 Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal
580 forest tree species—Prunus avium L. *Molecular Ecology*, *15*(8), 2109–2118.
581 https://doi.org/10.1111/j.1365-294X.2006.02926.x
582 Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of
583 clonality on the genotypic and genetic states of populations: Improving our estimation
584 of clonal rates. *Molecular Ecology Resources*, *21*(4), 1068–1084.
585 https://doi.org/10.1111/1755-0998.13316
586 Van Drunen, W. E., & Husband, B. C. (2019). Evolutionary associations between polyploidy,
587 clonal reproduction, and perenniality in the angiosperms. *New Phytologist*, *224*(3),
588 1266–1277. https://doi.org/10.1111/nph.15999
589 Villate, L., Esmenjaud, D., Van Helden, M., Stoeckel, S., & Plantard, O. (2010). Genetic
590 signature of amphimixis allows for the detection and fine scale localization of sexual
591 reproduction events in a mainly parthenogenetic nematode. *Molecular Ecology*, *19*(5),
592 856–873. https://doi.org/10.1111/j.1365-294X.2009.04511.x
593 Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among
594 genotypes in natural populations: Cautions and guidelines. *Molecular Ecology*, *10*(1),
595 249–256. https://doi.org/10.1046/j.1365-294x.2001.01185.x

596 Wakeley, J. (2005). The Limits of Theoretical Population Genetics. *Genetics*, *169*(1), 1–7.
597 Weir, B.S. (1996) Genetic Data Analysis II: Methods for Discrete Population Genetic Data.
598 Sinauer Associates, Inc., Sunderland.
599 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of
600 Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641
601 Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L.
602 H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of*
603 *the National Academy of Sciences*, *106*(33), 13875–13879.
604 https://doi.org/10.1073/pnas.0811575106
605 Wu, R., Gallo-Meagher, M., Littell, R. C., & Zeng, Z. B. (2001). A general polyploid model
606 for analyzing gene segregation in outcrossing tetraploid species. *Genetics*, *159*(2),
607 869–882.
608 Wu, S., Cheng, J., Xu, X., Zhang, Y., Zhao, Y., Li, H., & Qiang, S. (2019). Polyploidy in
609 invasive Solidago canadensis increased plant nitrogen uptake, and abundance and
610 activity of microbes and nematodes in soil. *Soil Biology and Biochemistry*, 138,
611 107594. https://doi.org/10.1016/j.soilbio.2019.107594.
612 Yu, F.-H., Roiloa, S. R., & Alpert, P. (2016). Editorial: Global Change, Clonal Growth, and
613 Biological Invasions by Plants. *Frontiers in Plant Science*, *7*.
614 https://www.frontiersin.org/articles/10.3389/fpls.2016.01467
615

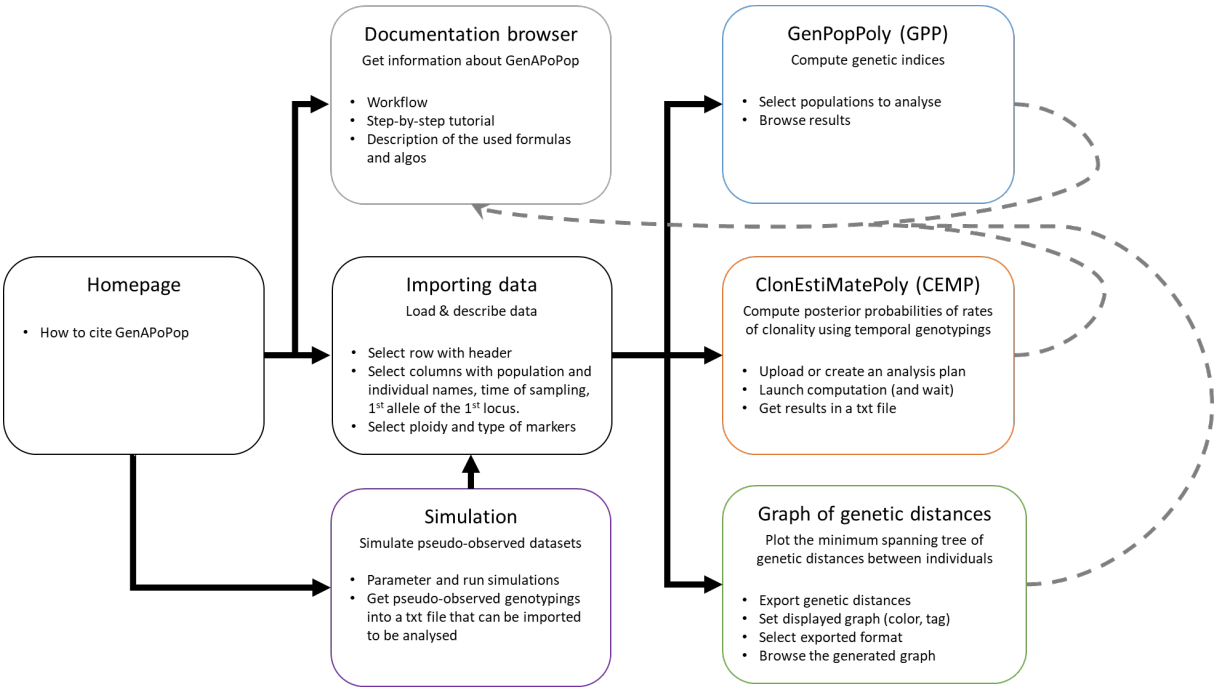


**Figure 1:** Workflow in GENAPOPOP. Users first import dataset using the embedded simulator or external datasets, describe the data structure, and then can launch at least one of the three types of analyses mediated by the graphical interface. Full connectors indicate the possible workflows, dashed connectors indicate optional possibility to consult documentation using the embedded light PDF reader. Results can be browsed within the software and retrieved in exported files directly importable into common spreadsheets and text editors.

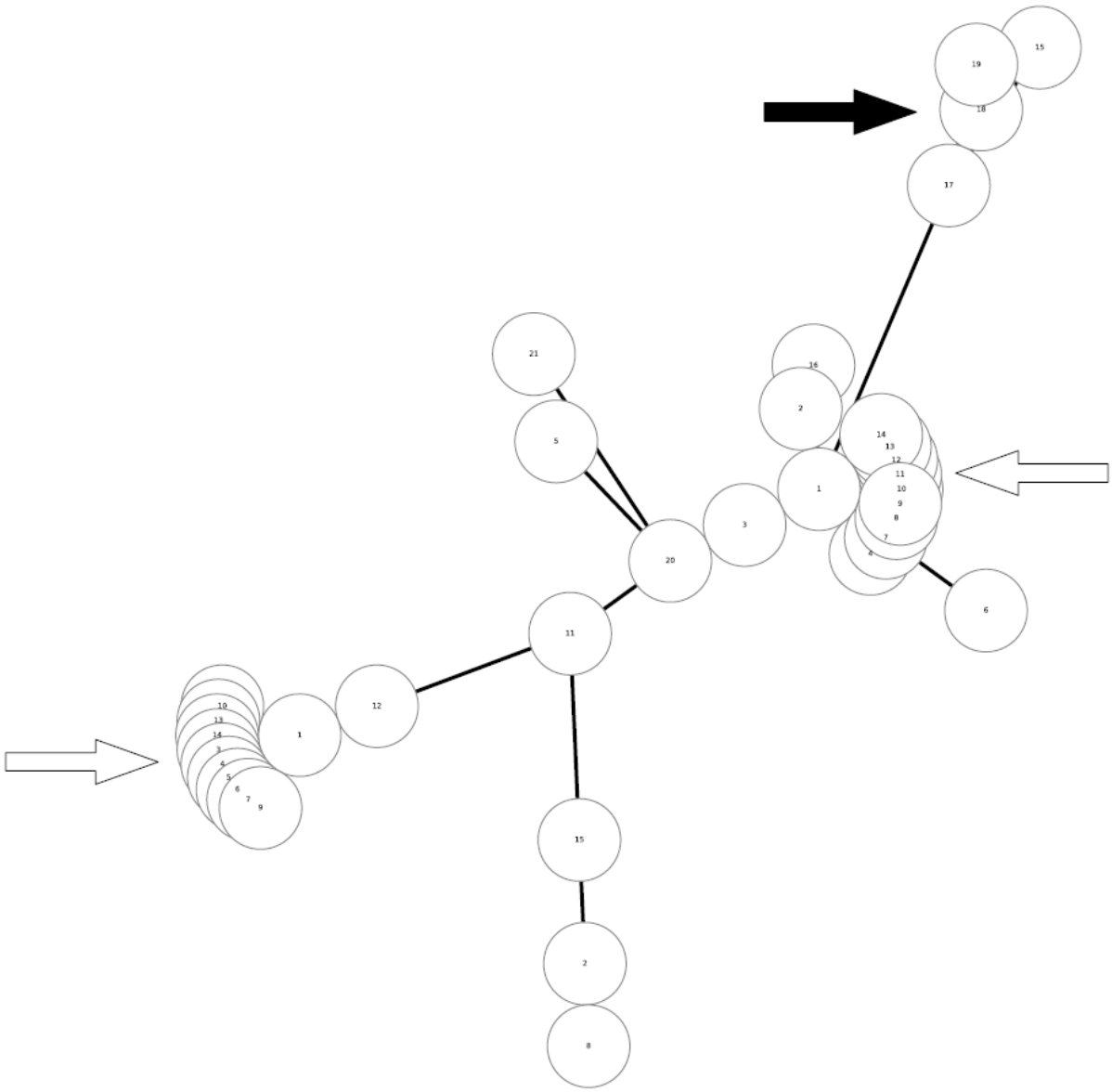


**Figure 2:** Minimum spanning tree of the genetic distance in the AS dataset. White arrows indicate one rosette of multiple multilocus genotypes differing from one allele of a central multilocus genotype, suspected to be recent mutants of a same multilocus lineage. Black arrow indicates a rosary pattern of multilocus genotypes differing from few alleles, suspected to be clones of a same multilocus lineage in which other sampled clones may have accumulated few mutations over generations.

| **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** |

631

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

637

| | | **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 | 0.066 |
| | 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 | 0.016 |
| | 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 | 0.013 |
| | GenAPoPop | 0.659 | 0.126 | 0.613 | 0.517 | 0.152 | 0.246 | |
| **Fis** | Spagedi | -0.0003 | 0.6839 | 0.016 | 0.1522 | -0.1976 | -0.0221 | 0.059 |
| | GenAPoPop | -0.0669 | 0.6094 | -0.0530 | 0.0673 | -0.2289 | -0.0818 | |
| **Fst** | Spagedi | 0.004 | 0.0693 | 0.0123 | 0.0307 | 0.0039 | 0.2063 | 0.004 |
| | GenAPoPop | 0.0032 | 0.0646 | 0.0099 | 0.0255 | 0.0030 | 0.1687 | |
| **Fit** | Spagedi | 0.0037 | 0.7058 | 0.0281 | 0.1782 | -0.193 | 0.1887 | 0.061 |
| | GenAPoPop | -0.0635 | 0.6346 | -0.0426 | 0.0910 | -0.2252 | 0.1007 | |
| **Rhost** | Spagedi | 0.0157 | 0.089 | 0.0454 | 0.0801 | 0.037 | 0.5182 | 0.000 |
| | GenAPoPop | 0.0157 | 0.089 | 0.0454 | 0.0801 | 0.037 | 0.5182 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Mean difference (pop)** | 0.035 | 0.020 | 0.019 | 0.023 | 0.013 | 0.110 |

**Table 2:** comparison of seven classic population genetic indices computed to compare the consistency of GENAPOPOP with 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 clonality; high selfing; 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 (Stoeckel et al. 2022, DOI). *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, then four components of the F-statistics *Fis* for the mean inbreeding coefficient value over all the populations and loci, *Fst* for the mean classical genetic differentiation and *rhost*, its analogue independent of double-reduction and ploidy level. F-statistics were computed using AMOVA framework.