



**HAL**  
open science

## Rare sex or out of reach equilibrium? The dynamics of F IS in partially clonal organisms.

Katja Reichel, Jean-Pierre Masson, Florent Malrieu, Sophie Arnaud-Haond,  
Solemn Stoeckel

► **To cite this version:**

Katja Reichel, Jean-Pierre Masson, Florent Malrieu, Sophie Arnaud-Haond, Solemn Stoeckel. Rare sex or out of reach equilibrium? The dynamics of F IS in partially clonal organisms.. BMC Genetics, 2016, 17 (1), pp.76. 10.1186/s12863-016-0388-z . hal-01345820

**HAL Id: hal-01345820**

**<https://hal.science/hal-01345820v1>**

Submitted on 7 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access



# Rare sex or out of reach equilibrium? The dynamics of $F_{IS}$ in partially clonal organisms

Katja Reichel<sup>1</sup>, Jean-Pierre Masson<sup>1</sup>, Florent Malrieu<sup>2</sup>, Sophie Arnaud-Haond<sup>3</sup> and Solenn Stoeckel<sup>1\*</sup>

## Abstract

**Background:** Partially clonal organisms are very common in nature, yet the influence of partial asexuality on the temporal dynamics of genetic diversity remains poorly understood. Mathematical models accounting for clonality predict deviations only for extremely rare sex and only towards mean inbreeding coefficient  $\overline{F_{IS}} < 0$ . Yet in partially clonal species, both  $F_{IS} < 0$  and  $F_{IS} > 0$  are frequently observed also in populations where there is evidence for a significant amount of sexual reproduction. Here, we studied the joint effects of partial clonality, mutation and genetic drift with a state-and-time discrete Markov chain model to describe the dynamics of  $F_{IS}$  over time under increasing rates of clonality.

**Results:** Results of the mathematical model and simulations show that partial clonality slows down the asymptotic convergence to  $F_{IS} = 0$ . Thus, although clonality alone does not lead to departures from Hardy-Weinberg expectations once reached the final equilibrium state, both negative and positive  $F_{IS}$  values can arise transiently even at intermediate rates of clonality. More importantly, such “transient” departures from Hardy Weinberg proportions may last long as clonality tunes up the temporal variation of  $F_{IS}$  and reduces its rate of change over time, leading to a hyperbolic increase of the maximal time needed to reach the final mean  $\overline{F_{IS, \infty}}$  value expected at equilibrium.

**Conclusion:** Our results argue for a dynamical interpretation of  $F_{IS}$  in clonal populations. Negative values cannot be interpreted as unequivocal evidence for extremely scarce sex but also as intermediate rates of clonality in finite populations. Complementary observations (e.g. frequency distribution of multiloci genotypes, population history) or time series data may help to discriminate between different possible conclusions on the extent of clonality when mean  $\overline{F_{IS}}$  values deviating from zero and/or a large variation of  $F_{IS}$  over loci are observed.

**Keywords:** Partial asexuality, Parthenogenesis, Mating system, Inbreeding coefficient, Heterozygote excess, Genetic diversity

## Background

Reproductive systems impact the evolution of genetic diversity at the population level [1, 2], making them an important factor for considerations on the evolvability of species. Partially clonal species, i.e. species that are able to reproduce both sexually and clonally, are common across many phyla and ecosystems [3] and represent an important part of the global biodiversity. They include many species of which evolutions are directly impacted by, or impacting humans, such as cultivated species [4], pathogens [5], invasive

species [6], and species threatened by extinction (e.g. [7–11]). Partially clonal species are therefore frequently the subject of molecular analyses describing their genetic diversity [12], and the conclusions drawn depend on a correct understanding of the effects of their reproductive mode on the genetic composition of their populations.

The interpretation of standard population genetic indices from partially clonal populations can be challenging, as expectations are likely to depend on the rate of clonality, which is usually unknown in natural populations. The estimate of this rate on the basis of

\* Correspondence: solenn.stoeckel@rennes.inra.fr

<sup>1</sup>IGEPP, Agrocampus Ouest, INRA, Université de Rennes 1, 35650 Le Rheu, France

Full list of author information is available at the end of the article



indirect approaches such as population genetics analysis remains elusive. One example of an index that has been suggested to change with the rate of clonality is  $F_{IS}$  [13, 14]. Within diploid individuals, it represents a correlation coefficient among alleles at a particular locus, and depends on their tendency to be randomly associated ( $F_{IS} = 0$ ) or more likely identical ( $F_{IS} > 0$ ) or not identical ( $F_{IS} < 0$ ) at the population level.  $F_{IS}$  is defined either based on population heterozygosity ( $H_e$  – expected heterozygosity,  $H_o$  – observed heterozygosity) or allele identities/homozygosity ( $F$  – allele identity within individuals,  $\Theta$  – allele identity within the population [13]):

$$F_{IS} = \frac{H_e - H_o}{H_e} \cong \frac{F - \Theta}{1 - \Theta}, \quad F_{IS} \in [-1, 1]$$

Results from both definitions differ only for loci with just a single allele remaining (fixation), where  $F_{IS}$  cannot be defined.

To date, only few mathematical models studying  $F_{IS}$  at selectively neutral loci in partially clonal populations have been published. For partially clonal populations otherwise complying with the Hardy-Weinberg conditions,  $F_{IS}$  and the underlying genotype frequencies are expected to be identical to those expected for random mating, except for the approach to the Hardy-Weinberg equilibrium (HWE) being slowed down as the rate of clonality increases [15]. If mutation and genetic drift are taken into account [13], very high rates of sexual reproduction are expected to lead eventually to strongly negative mean  $F_{IS}$  values up to  $\overline{F_{IS, \infty}} = -1$  for completely clonal populations. In addition to this effect on the mean, the shape of the full steady state (i.e. “equilibrium”: the convergence of values toward a given probabilistic and dynamically-stable distribution) distribution of  $F_{IS}$ , measured by its variance, skewness and kurtosis, also changes with the rate of clonality [16]. Based on the results of [13],  $\overline{F_{IS}}$  was suggested as an informative parameter to estimate the rate of clonality [14, 17] in connection with other indices such as linkage disequilibrium or the frequency of repeated multiloci genotypes [18]. However, using the mean of the steady state distribution provided by [13] as a reference for the mean  $\overline{F_{IS}}$  values from field studies often pointed to rates of clonality that were at odds with other indices or even direct observation (e.g. [19, 20]).

While some previous works may be interpreted as demonstrating negative  $F_{IS}$  as a signature of nearly exclusive clonality [13, 14], others underline the influence of clonality not only on the steady state distributions of  $F_{IS}$  but also on the temporal dynamics of this parameter in natural population [15, 16]. Here

we aimed at complementing the results of the previous studies by describing the temporal changes of genotype frequencies over time under the distinct and joint influence of partial clonality, mutation and genetic drift. Understanding how quickly the steady state distribution of  $F_{IS}$  is reached again after a disturbance (e.g. change of reproductive system, stochastic or selective events) due to clonality, mutation and drift, may indeed help to explain the unexpected departures observed in nature, in populations otherwise considered as undergoing rather frequent sexual reproduction.

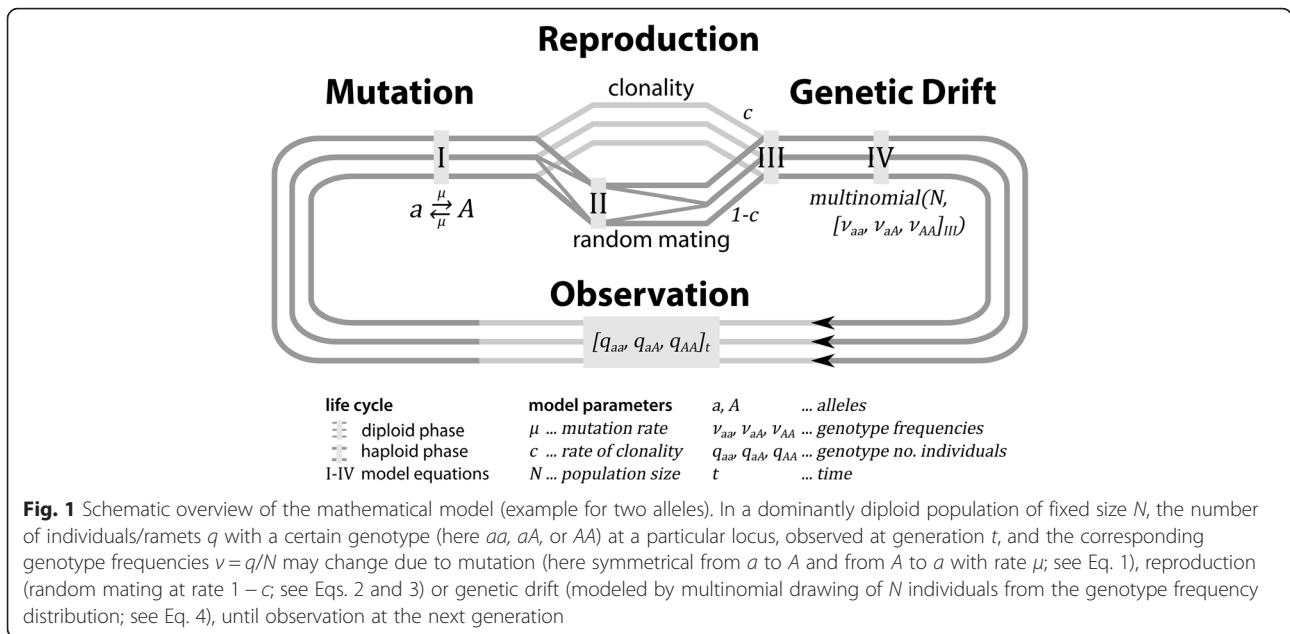
We used a stochastic model to follow the neutral dynamics of genotype frequencies in the basic case of a single locus in a diploid, isolated and panmictic population that combines random mating and clonality. We derived the dynamical effects of population size, mutation rate and rate of clonality separately on genotype frequencies over generations and how long it takes until a steady state distribution is reached again after any disturbance. We subsequently connected these partial results to analyze the “complete” system with the joint effects of reproductive mode, mutation and genetic drift. Finally, we discuss how our results may provide new hypotheses in the interpretation of field data, based on examples from a literature review, and we provide methodological recommendations for future analyses of genotype frequencies in partially clonal populations.

## Methods

### Mathematical model

The biological template for our model is a single population with a finite number of individuals. These individuals correspond to ramets, i.e. factually or potentially physiologically distinct units that may or may not be genetically identical or descended from the same parent. All individuals follow the same life cycle, which consists of a dominant diploid phase during which they can acquire heritable mutations (Fig. 1). All individuals are monoecious and can produce offspring both by clonal and sexual (here defined as random mating including selfing) reproduction. A fixed number of these offspring individuals corresponding to a constant population size survive randomly to replace their parents in the next generation.

We translated this system into a time and state discrete Markov chain model, conceptually similar to [16] to follow the population dynamics of genotype frequencies at one locus with multiple alleles, as in the classic models of Wright [21] and others [22, 23]. Each time step of the model corresponds to one generation, i.e. the time between two consecutive observations of the population (Fig. 1). The model states



represent all possible distributions of the  $N$  individuals on  $g$  genotypes (a genotypic state): For a single locus  $n$  alleles, the number of genotypes corresponds to  $g = n(n + 1)/2$  and the number of states to  $(N + g - 1)! / (N! \cdot (g - 1)!)$ . For example, a locus of 3 alleles ( $A1$ ,  $A2$  and  $A3$ ) results in 6 possible genotypes ( $A1A1$ ,  $A1A2$ ,  $A1A3$ ,  $A2A2$ ,  $A2A3$ ,  $A3A3$ ) and a population composed of 100 individuals could evolve among the 96 560 646 single possible repartitions of 100 individuals within the 6 genotypes (hereafter named a population state) [16, 24].

At each time step, the population makes a transition from its current state to a next state (where current and next state can be the same), based on a vector of transition probabilities. These probabilities depend on the genotype frequencies  $v_{ii}, v_{ij}$  (with the indices  $i \neq j$  denoting different alleles) derived from the current population state, and on the three constant model parameters: population size  $N$ , mutation rate  $\mu$  and rate of clonality  $c$ , according to the following equations (Fig. 1; simplified for the special case of two alleles in Additional file 1, 1.1):

**I Mutation.** The theoretical frequencies  $v_{ii,t}, v_{ij,t}$  of each genotype after mutation are derived as: where  $\alpha = 1 - \mu$ , the probability that an allele does not mutate, and  $\beta = \frac{\mu}{n-1}$ , the probability that an allele

mutates into one of the  $n - 1$  others during one generation. This corresponds to a classic  $k$ -alleles or Jukes-Cantor substitution model [25].

**II Gamete formation (allele segregation then fusion).** The gamete frequencies in the gamete pool after sexual reproduction  $v_{i,I}$  are calculated as:

$$v_{i,I} = v_{ii,I} + \frac{1}{2} \sum_{j \neq i} v_{ij,I} \tag{2}$$

There is no difference in the allele frequencies between sexes, mating types etc., and all individuals contribute equally to the gamete pool (pangamy).

**III Reproduction.** The genotype frequencies  $v_{ii,III}, v_{ij,III}$  after reproduction are calculated as:

$$\begin{cases} v_{ii,III} = cv_{ii,I} + (1-c)v_{i,I}^2 \\ v_{ij,III} = cv_{ij,I} + 2(1-c)v_{i,I}v_{j,I} \end{cases} \tag{3}$$

based on the results from Eqs. 1 and 2. The rate of clonality  $c$  thus corresponds to the proportion of offspring per generation that is the result of clonal reproduction. The remainder of the offspring (“rate of sexuality”,  $1 - c$ ) is derived from random mating including autogamy, assuming panmixis.

$$\begin{cases} v_{ii,t} = \alpha^2 v_{ii,t} + \alpha\beta \sum_{j \neq i} v_{ij,t} + \beta^2 \left( \sum_{j \neq i} v_{jj,t} + \sum_{k, j \neq i} v_{jk,t} \right) \\ v_{ij,t} = (\alpha^2 + \beta^2) v_{ij,t} + 2\alpha\beta (v_{ii,t} + v_{jj,t}) + (\alpha\beta + \beta^2) \left( \sum_{k \neq i,j} v_{ik,t} + \sum_{l \neq i,j} v_{jl,t} \right) + 2\beta^2 \left( \sum_{k \neq i,j} v_{kk,t} + \sum_{k,l \neq i,j} v_{kl,t} \right) \end{cases} \tag{1}$$

**IV Genetic Drift.** The vector of genotype frequencies  $\vec{v}_{t+1}$  at the next generation depending on the population size is derived from:

$$\vec{v}_{t+1} = \frac{X_{t+1}}{N} \text{ where } X_{t+1} \sim \mathcal{M}(N, \vec{v}_{III}) \quad (4)$$

where  $X_{t+1}$  is the state of the model at the next generation, drawn from a multinomial distribution  $M$  that is based on  $N$ , the population size counting all potentially reproducing individuals (mathematically the number of samples), and  $\vec{v}_{III}$ , the vector of genotype frequencies derived from Eq. 3 (mathematically the probabilities of the genotype “categories”). Transition probabilities  $P$  between any two model states  $X_t, X_{t+1}$  can then be calculated based on:

$$P(X_{t+1}|X_t) = \frac{N!}{\prod_i q_{ii, t+1}! \prod_{i,j} q_{ij, t+1}!} \prod_i v_{ii,III}^{q_{ii, t+1}} \prod_{i,j} v_{ij,III}^{q_{ij, t+1}} \quad (5)$$

where  $q_{ii, t+1}, q_{ij, t+1} \in \mathbb{N}_0$  are the counts (natural numbers) of individuals per genotype in the presumed next state  $X_{t+1}$  which sum to  $N$ . Note that our description of genetic drift is based on genotype frequencies rather than allele frequencies. As explained in [22], describing population genetic processes based on allele frequencies is a mathematical convenience justified by HWE (i.e. assuming exclusively sexual reproduction), which assures that allele frequencies can always be directly translated into genotype frequencies. For partially clonal populations, we cannot automatically assume HWE and thus modeled all population genetic processes, including genetic drift, at the genotypic level.

**Model analysis and description of biological consequences**

First, we studied the effect of each of the three model parameters ( $c, \mu, N$ ) on the genotype frequencies by itself. Setting the other two parameters to have no influence on the model result, i.e.  $c = 1, \mu = 0$  and/or  $N = \infty$  (or no random drawing in Eq. 4), and substituting Eq. 2 into Eq. 3, the model reduces to one equation per process, i.e. Eq. 1 for  $\mu$ , Eq. 3 (with 2) for  $c$ , and Eq. 4 for  $N$ . For each equation/process, we then determined the steady state distributions of genotype frequencies at one locus, i.e. the probabilistic combination of population states for which  $\vec{q}_{t+1} = \vec{q}_t$ . We also derived the respective maximal convergence times  $t_c, t_\mu$  and  $t_N$  of each separate evolutionary process to reach steady state distributions of genotype frequencies. While  $t_N$  could only be approximated from numerical results (Markov chain first passage time approach and

simulations), for  $c$  and  $\mu$  convergence to the steady states is asymptotical as it can be described by geometric progressions (details of derivation in Additional file 1, 1.2). We therefore defined a universal “acceptable error”  $\varepsilon = 1/(2N)$ , corresponding to one half the minimal change in genotype frequency that would be measurable by exhaustive sampling in a population of finite size  $N$ , below which the distance from the steady states has to pass (convergence criterion). Using the maximal convergence times  $t_c, t_\mu$  and  $t_N$  as measures for the “strength” with which each process acts upon the genotype frequencies, we could then use this analytical basis to partition the parameter space of the full model into regions where either process dominates the genotype frequency dynamics.

Secondly, we approached the full model by following the dynamics of  $F_{IS}$  over time from three different start states for combinations of  $c, \mu$  and  $N$  representative of the different regions of the parameter space. Aggregating the transition probabilities between all model states in a transition matrix  $M$  (same current state per column, i.e. columns summing to one), the probability distribution of the model states (and consequently the probability distribution of  $F_{IS}$ ) at time  $t$ , given by the vector  $\vec{x}_t$ , is derived by matrix multiplication:

$$\vec{x}_t = M^t \vec{x}_0 \quad (6)$$

where  $\vec{x}_0$  describes the start state (vector of zeros except for a single one at  $X_0$ ). The steady state distribution of population states corresponds to the dominant eigenvector of the transition matrix  $M$ . Based on the transition matrix  $M$ , we also calculated the time to the steady state distribution (Markov chain mixing time, Additional file 1, 1.5).

We illustrated the numerical result of our model using three start states:  $F_{IS,0} \in \{-1; 0; 1\}$  under isoplethic allele frequencies (equally frequent,  $v_a = v_A = \frac{1}{n}$ ), standing for HWE ( $F_{IS,0} = 0$ ) and the most extreme deviations from it (complete homozygosity,  $F_{IS,0} = 1$ ; complete heterozygosity,  $F_{IS,0} = -1$ ). These start states were chosen to represent the range of  $F_{IS}$  values, and not because of their biological meaning or how they may be reached in nature. Deviations from steady state distributions may derive from a recent change in the rate of clonality (e.g. from full sexuality with  $F_{IS,0} = 0$ ), or full adaptation to past selection for ( $F_{IS,0} = -1$ ) or against ( $F_{IS,0} = 1$ ) heterozygotes, or changes in population size (demographic bottleneck, founder event) [26, 27], or secondary contact between two populations in which different alleles become fixed ( $F_{IS,0} = 1$ ) or



hybridization with subsequent reproductive isolation from the parents ( $F_{IS,0} = -1$ ) (e.g. [28–30]). Based on the transition matrix  $M$ , we also calculated the time to the final distribution of states (= steady states), which is also the time until the exact final distribution of  $\overline{F_{IS,\infty}}$  (Markov chain mixing time, Additional file 1, 1.5).

To link our results with those obtained by previous authors, we calculated the final mean  $\overline{F_{IS,\infty}}$  from Eq. 10 in [13] formalized for metapopulation:

$$F_{IS} = \frac{\gamma(q_s - c(\gamma(q_s - q_d) - 1) - 1)}{2N(1 - \gamma c)(\gamma(q_s - q_d) - 1) - \gamma(q_s - c(\gamma(q_s - q_d) - 1) - 1)}$$

where  $\gamma = (1 - \mu)^2$ ,  $q_s$  is the probabilities that two individuals taken at random in the same reproductive subpopulation after migration were sired in the same reproductive subpopulation one generation before, and  $q_d$  the probability that two individuals taken at random in different reproductive subpopulations after migration originated from the same subpopulation one generation before.

By setting  $q_s = 1$ ,  $q_d = 0$  (a single finite population, no migration) and  $s = 1/N$  (random mating):

$$\overline{F_{IS,\infty}} = \frac{1}{(2N - 1) - 2N/c(1 - \mu)^2} \tag{7}$$

We derived the corresponding expected convergence time of  $\overline{F_{IS}}$  iteratively obtained from Eq. 5 in [13] (detailed in Additional file 1, 1.6)

$$\begin{aligned} \begin{bmatrix} 1 - H_{o,t+1} \\ 1 - H_{e,t+1} \end{bmatrix} &= \begin{bmatrix} F_{t+1} \\ \Theta_{t+1} \end{bmatrix} \\ &= (1 - \mu)^2 \left( \begin{bmatrix} c + \frac{1-c}{2N} & (1-c)\left(1 - \frac{1}{N}\right) \\ \frac{1}{2N} & 1 - \frac{1}{N} \end{bmatrix} \begin{bmatrix} F_t \\ \Theta_t \end{bmatrix} + \begin{bmatrix} \frac{1-c}{2N} \\ \frac{1}{2N} \end{bmatrix} \right) \end{aligned} \tag{8}$$

Where  $F$  and  $\theta$  are the allele identities within individuals and between individuals respectively. In contrast to our own model, these equations do not contain the number of alleles. This is because they are based on an infinite alleles model, and treat the expected and observed hetero-/homo-zygosity as continuous variables whatever the population size considered.

Finally, to get a better idea how our theoretical results are comparable to those published for field data, we looked at the sampling effect of using different numbers of polymorphic loci  $L$  to estimate the mean  $\overline{F_{IS}}$  of the population at time  $t$ ,  $\overline{F_{IS,t,L}}$ . Assuming that each locus represents an independent estimate of this mean (no confounding effect of linkage), and that the genotype frequencies are known exactly (exhaustive sampling of all individuals/ramets), it is derived as:

$$\overline{F_{IS,t,L}} = \frac{1}{L} \sum_{z=0}^L F_{IS,t,z} \tag{9}$$

Both assumptions are usually violated [14, 18], so that our results represent a conservative estimate of the true error of this method. We randomly sampled both the steady state distribution  $\overline{F_{IS,\infty}}$  and the instantaneous distribution  $\overline{F_{IS,50}}$  of a population that was 50 generations ago at HWE at all loci with equal allele frequencies (isoplethy for two alleles per locus), for the same parameter combinations that we previously used to illustrate the dynamics of the full model. Based on  $10^5$  random samples of size  $L$ , we then calculated the mean signed deviation of the sample means  $F_{IS,t}$  from the true mean  $\overline{F_{IS,t}}$ :

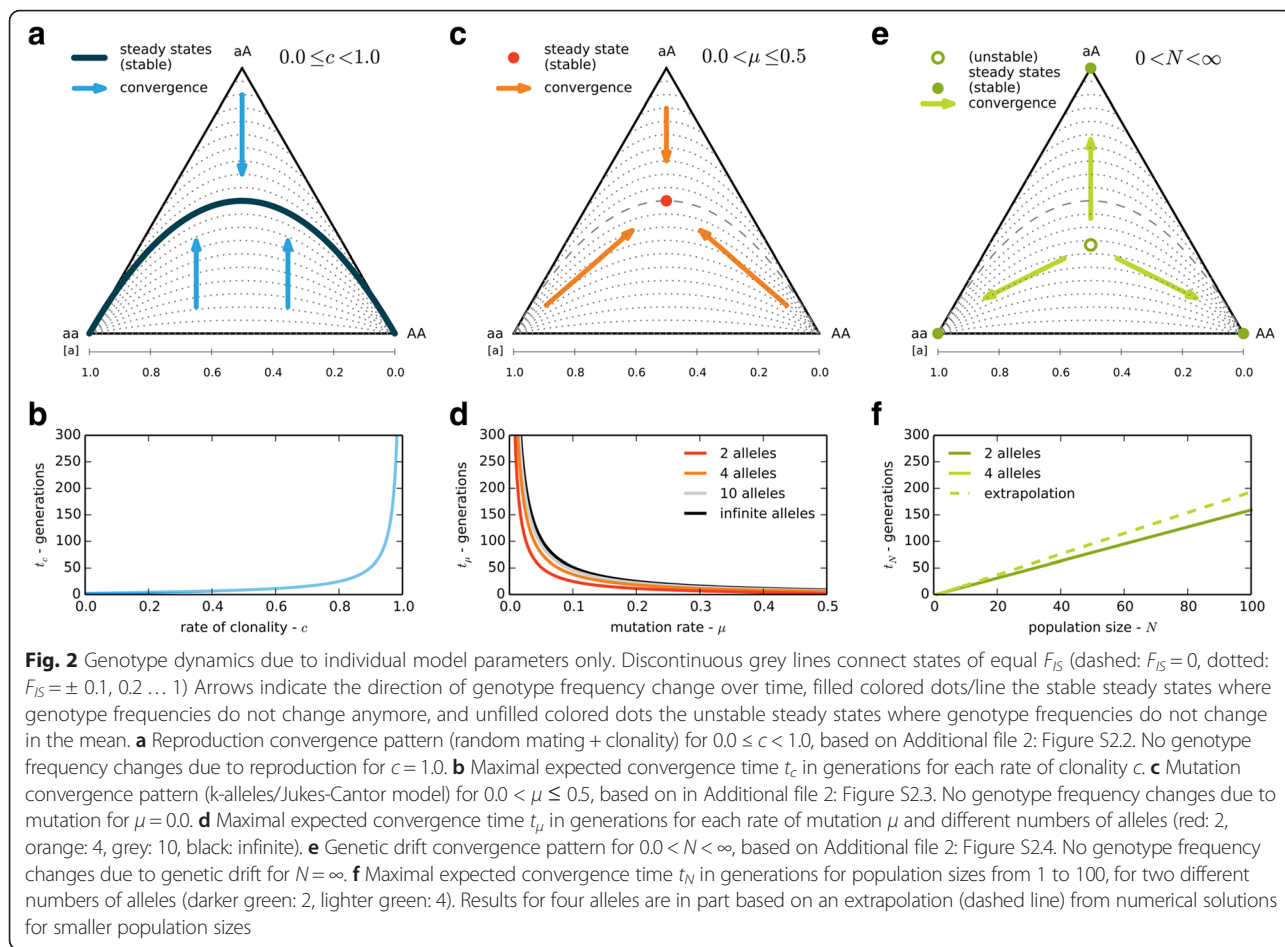
$$\Delta \overline{F_{IS,t}} = \begin{cases} \frac{1}{z_1} \sum_{F_{IS,t} \geq \overline{F_{IS,t}}} F_{IS,t} - \overline{F_{IS,t}} \\ \frac{1}{z_2} \sum_{F_{IS,t} < \overline{F_{IS,t}}} F_{IS,t} - \overline{F_{IS,t}} \end{cases}, \quad z_1 + z_2 = 10^5 \tag{10}$$

where  $z_1$  and  $z_2$  represent the number of positive and negative deviations, respectively. Loci at or near fixation are typically not used in population genetic studies, since they are especially affected by genotyping errors. We therefore excluded all loci where the frequency of one allele exceeds  $1 - \sqrt{1/(2N)}$  (near fixation; Additional file 2: Figure S2.1 for the derivation of this value, and compare similar considerations in [31]) from the calculation of values for this analysis.

All computations were performed in Python 2.7 with 64 bit precision, using the modules *numpy*, *scipy* [32], *networkx* [33] and *matplotlib* [34] (basic code in Additional file 3). We illustrate some of our results with *de Finetti* diagrams [24, 35] (Fig. 2a, c, e), which are ternary plots of the genotype frequencies [ $v_{aa}$ ,  $v_{aA}$ ,  $v_{AA}$ ] at one locus with two alleles within a population (see Additional file 2: Figures S2.1 to S2.4 for more information). Details for the literature review in the discussion are given in Additional file 4.

### Examples for empirical $F_{IS}$ values of partially clonal populations from published field studies

To get a summary view of the kind of  $F_{IS}$  values that can be encountered in real field populations suspected to reproduce using partial clonality, we compiled the data for 51 populations belonging to 13 species (seven angiosperms, four protists, a sponge and a nematode), based on 13 previously published studies (see references in Additional file 3) selected for their near fit with the assumptions of our model from a Web of Science search for [(microsatellite OR “SSR” OR “simple sequence repeat” OR “SNP” OR



“single nucleotide polymorphism”) AND (clonal OR asexual OR vegetative OR apomictic OR apomixis OR agamospermy OR parthenogenesis)]. All studies are based on SSR data. When  $F_{IS}$  values were not directly provided by publications, we calculated  $F_{IS}$  values per locus from the reported  $H_o$  and  $H_e$ . We grouped populations into three classes according to the information given by the authors about their putative rate of clonality: *i*) rarely clonal, *ii*) frequent clonality and sexuality (including unknown) and *iii*) rarely sexual. We included populations for which preferential outbreeding (self-incompatibility system) was expected.

**Results**

The maximal times until the genotypic diversity at one locus has converged to its steady state distribution changed greatly with the rates of clonality. During this convergence, the steady state distributions, the dynamics of  $F_{IS}$  and the underlying genotype frequencies through time were also affected by the rate of clonality. However, both maximal convergence times and the dynamics of genotypic diversity at one locus strongly depend on the interactions of the rate

of clonality with mutation rate and population size. Indeed, each of these evolutionary processes has its own maximal convergence time toward the steady state distribution of genotype frequencies and  $F_{IS}$  ( $t_c$ ,  $t_\mu$ ,  $t_N$ ). The overall convergence time of a population to its steady state distribution, which is shaped by mutation, genetic drift and the reproduction mode, depends on the relative “strength” of each of those three forces.

The maximal convergence time due to reproduction only,  $t_c$  (derived in Additional file 1, 1.2) can be approximated as

$$t_c = 1 + \log_c \varepsilon = 1 + \frac{\log \varepsilon}{\log c} \tag{11}$$

with  $\varepsilon = \frac{1}{2N}$  corresponding to a small error term as convergence to the steady state distribution (HWE: Fig. 2a, Additional file 2: Figure S2.2) is asymptotic. Between  $t_c = 1$  for exclusively sexual and  $t_c = \infty$  for exclusively clonal populations, the dependence of  $t_c$  on  $c$  is not linear, but shapes like an hyperbola of type  $y = 1 + \frac{\text{constant}}{x}$  (Fig. 2b). Consequently, the time

required to reach a steady state distribution if only reproduction is considered increases hyperbolically as  $c$  increases.

The maximal convergence time due to mutation only  $t_\mu$  (derived in Additional file 1, 1.2) can be approximated as

$$t_\mu = 1 + \log_{(1-\mu\frac{n}{n-1})}\epsilon = 1 + \frac{\log\epsilon}{\log(1-\mu\frac{n}{n-1})}, \tag{12}$$

which simplifies for two alleles to

$$t_\mu = 1 + \log_{(1-2\mu)}\epsilon = 1 + \frac{\log\epsilon}{\log(1-2\mu)} \tag{13}$$

The steady state distribution for mutation corresponds to  $F_{IS} = 0$  and, due to the symmetry of mutation between alleles in our models, equal allele frequencies (Additional file 1, 1.3; Additional file 2: Figure S2.3). For  $\mu = \frac{1}{n}$  in theory the highest mutation rate (each allele has the same chance to mutate or not to mutate into other alleles),  $t_\mu$  is only one generation. For realistic mutation rates ranging from  $10^{-3}$  to  $10^{-18}$  [36, 37], convergence due to mutation can take much longer: setting  $\epsilon = 0.005$  and assuming a mutation rate of  $\mu = 10^{-6}$ ,  $t_\mu$  would be around  $2.6 \cdot \mu^{-1}$  or 2.6 million generations.

The maximal convergence time due to reach a steady state distribution (Fig. 2e, Additional file 2: Figure S2.4) due to genetic drift only  $t_N$  depends on the population

size and on the number of initial genotypes (Fig. 2f). It grows approximately linearly with  $N$  and the slope of this dependence  $\sigma$  increases with the number of genotypes/alleles (Additional file 1, 1.2):

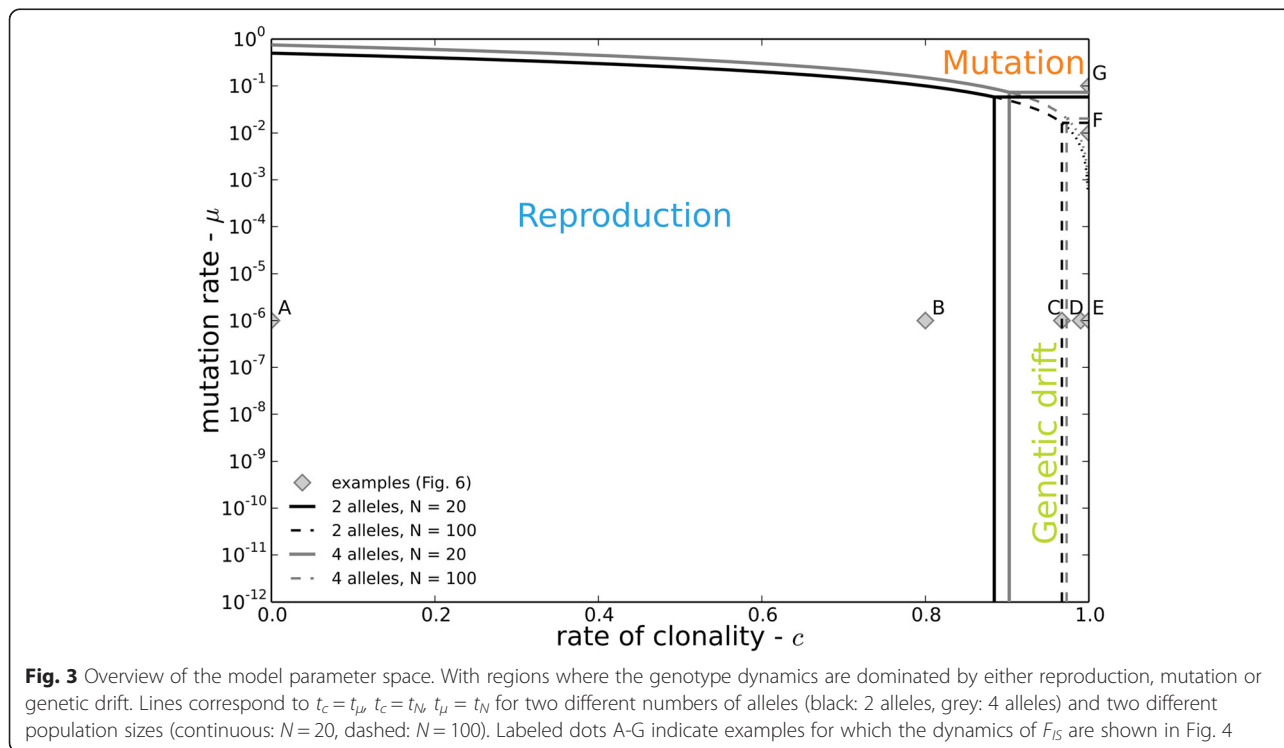
$$t_N = \sigma N - \tau \tag{14}$$

with  $\tau \geq 1$  and  $1 < \sigma < 2$  determined from numerical results (Additional file 2: Table S2.2) and simulations (Additional file 2: Table S2.3). For example, in a population of 100 individuals, up to about 160 generations can be required until only one genotype remains at a biallelic locus, and loci with more alleles take even a few generations more.

The lowest maximal convergence time amongst  $t_c$ ,  $t_\mu$ ,  $t_N$ , can serve as approximation to determine the evolutionary process that dominates the dynamics of genotype frequencies and its associated steady state. Pairwise equality between the convergence times can thus be used to partition the parameter space of our model distribution (eqs. 1 to 4,  $c \in [0, 1]$ ,  $\mu \in ]0, 0.5]$ ,  $N \in [1, \infty[$ ) into three domains where either process dominates (Fig. 3). Based on eqs. 12 to 14, these pairwise equalities resolve to:

$$c = \mu \frac{n}{n-1} \tag{15}$$

$$c = e^{1/(\sigma N - \tau - 1)} \tag{16}$$





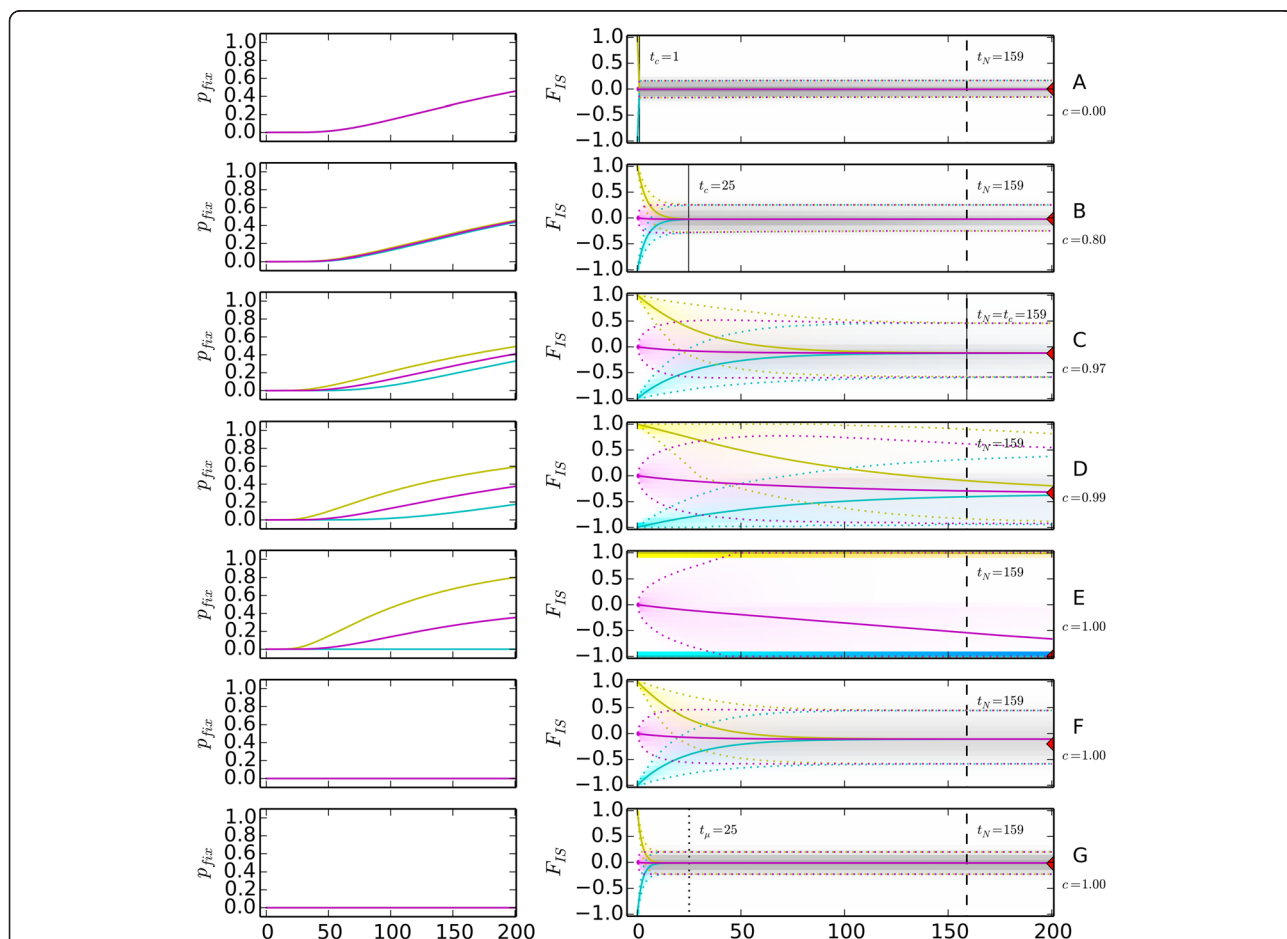
$$\mu = \frac{n-1}{n} \left( 1 - e^{1/(\sigma N - \tau - 1)} \right) \tag{17}$$

If  $t_c \ll t_\mu, t_N$ , as is usually the case for strictly sexually reproducing populations, genotype frequencies quickly converge to HWE, i.e.  $F_{IS,\infty} \cong 0$ . In very large and highly clonal populations where  $t_c \gg t_\mu, t_N$ , genotype frequencies also converge, although more slowly, to HWE so that  $t_\mu \ll t_c, t_N$  is equally awaited in this case. This contrasts with small and highly clonal populations where  $F_{IS,\infty} \cong 0$ , as dominant genetic drift does not lead to convergence toward  $t_N \ll t_\mu, t_c$ , but rather to a successive loss of genotypes and eventually to genotypic uniformity (fixation of one homozygous or heterozygous genotype  $F_{IS,\infty} = 0$  or  $-1$ ). In conclusion, both random mating and mutation lead to a random association of alleles (Additional file 2: Table S2.1, [16]), so that the dynamics of  $F_{IS}$  in (partially) clonal

populations is mostly driven by the relative “strength” of genetic drift.

A closer look at the transitions between the predominance of either process shows that changes are actually gradual. This is because different processes don’t globally compensate each other, as each convergence pattern is different (Fig. 2a, c, e, Additional file 2: Table S2.1). We therefore examined the joint action of all three processes in more details.

Keeping population size and mutation rate constant ( $N = 100, \mu = 10^{-6}$ ) while successively increasing the rate of clonality illustrates the changes in the dynamics of  $F_{IS}$  as genetic drift takes over (Fig. 4a to e, Additional file 2: Figures S2.6 and S2.7, Table S2.1). At low to intermediate rates of clonality (here  $0 < c \leq 0.8$ ), the dynamics of  $F_{IS}$  appear almost identical to those expected for a purely sexual population (Fig. 4a and b). However, variation around the final mean



**Fig. 4** Dynamics of probability of fixation  $p_{fix}$  and  $F_{IS}$  through time for seven representative example parameter sets. Single loci with two alleles. Colors represent different start states (yellow:  $F_{IS,0} = 1$  for  $v_a = v_A$ , magenta:  $F_{IS,0} = 0$  for  $v_a = v_A$ , cyan:  $F_{IS,0} = -1$ ), with their respective  $F_{IS,t}$  distributions (shading), mean (continuous line) and 95 % confidence interval (dotted lines). Vertical lines represent  $t_c$  (continuous),  $t_N$  (dashed) and  $t_\mu$  (dotted). Red triangles at  $t = 200$  indicate the mean  $\bar{F}_{IS,\infty}$  according to [13]. Model parameters – **a**  $c = 0, \mu = 10^{-6}, N = 100$ ; **b**  $c = 0.8, \mu = 10^{-6}, N = 100$ ; **c**  $c \approx 0.97$  ( $t_c = t_N$ ),  $\mu = 10^{-6}, N = 100$ ; **d**  $c = 0.99, \mu = 10^{-6}, N = 100$ ; **e**  $c = 1.0, \mu = 10^{-6}, N = 100$ ; **f**  $c = 1.0, \mu = 10^{-2}, N = 100$ ; **g**  $c = 1.0, \mu = 10^{-1}, N = 100$

$\overline{F_{IS,\infty}}$  is severely increased, and extreme initial  $F_{IS,0}$  values which would be lost in one generation under exclusive sexuality, can now be traced over a significant number of generations. These tendencies (increasingly negative  $\overline{F_{IS,\infty}}$ , increased variation of  $F_{IS}$  values, and increased time/start value dependence) continue until  $t_c$  crosses  $t_N$ , leading to negative mean  $F_{IS}$  values ( $c \approx 0.97$ :  $\overline{F_{IS,\infty}} \approx -0.14$ , Fig. 4c) and then gain even further momentum as sexual reproduction becomes very rare, a situation in which  $F_{IS}$  reaches more negative mean  $F_{IS}$  values ( $c = 0.99$ :  $\overline{F_{IS,\infty}} \approx -0.33$ , Fig. 4d). Indeed, transition probabilities that decrease  $F_{IS}$  raise while those that increase  $F_{IS}$  reduce (compare transition probabilities from (25,50,25) to (25,49,26) and to (25,51,24) in Additional file 2: Table S2.1). Moreover, probabilities to stay on the same genotypic state increase with clonality, except for states near fixation when mutation is high.

For loci starting with a high genotypic diversity, a marked difference in the positive and negative ranges of the  $F_{IS}$  distribution, namely the progressive disappearance of positive values, appears after some generations ( $t \sim 50$ , Fig. 4c-e). In parallel, the probability of allele fixation becomes increasingly dependent on the initial  $F_{IS,0}$  value (Fig. 4a-e, Additional file 2: Figures S2.6 and S2.7, left columns). The first reason for this effect, which also explains the convergence to negative  $\overline{F_{IS,\infty}}$  in highly clonal populations, is the trend towards randomly increasing the frequency of the heterozygous genotype(s) at HWE and approximately evenly distributed allele frequencies (Additional file 1, 1.4). Indeed, HWE is a concave function in the  $F_{IS}$  space (for one locus, 2 alleles, see Additional file 2: Figure S2.1) which in probability facilitates slightly negative  $F_{IS}$  as a result of stochastic changes even for random mating populations (see Additional file 2: Figure S2.1 and Table S2.1, e.g. transition probabilities from (25,50,25) to (25,49,26) and (25,51,24)). Partial and full clonality only prevents populations, partly or completely, to get back to HWE, thus to express this feature over more generations. The second reason is the dependence of the rate of fixation (of one allele) on the initial  $F_{IS,0}$  at extreme clonal rate (Fig. 4a-e). All loci reaching positive  $F_{IS,t}$  values then have a higher chance to move quickly toward fixation, helped by each random mating event (as the HWE parabola in *de Finetti* diagrams connects fixation states) and thus to be usually excluded from  $F_{IS}$  calculation. Over time, this only leaves analyzable loci with negative values. It can be noted that contrastingly, loci shifting to negative  $F_{IS,t}$  (higher probabilities than shifting to positive, due to the concave shape of HWE) are unlikely to reach  $\overline{F_{IS,\infty}} = -1.0$  by pure stochastic genetic drift, as

each sexual and homoplasmy events bring back populations toward the HWE concave parabola (Fig. 2a, c, e). Interestingly in our example, the mean  $\overline{F_{IS,t}}$  (starting from  $F_{IS,0} = 0$  in sexual populations) has not reached the final  $\overline{F_{IS,\infty}}$  even after 200 generations while  $t_N = 159$  because convergence to drift steady states is slowed down by “weak” mutation.

Highly clonal populations dominated by mutation rather than genetic drift present a different picture (Fig. 4f, g:  $c = 1.0$ ,  $\mu = \{10^{-2}, 10^{-1}\}$ ,  $N = 100$ , or simulation for  $c = 1.0$ ,  $\mu = 10^{-3}$ ,  $N = 10^3$  in Additional file 2: Figure S2.5). As in predominantly sexually reproducing populations,  $F_{IS}$  values converge to only slightly negative final  $\overline{F_{IS,\infty}}$  (Fig. 4g:  $\overline{F_{IS,\infty}} \approx -0.02$ ) but at a speed that mainly depends on  $t_\mu$  and with a limited variation of  $F_{IS}$  values across generations. In contrast, the instantaneous  $\overline{F_{IS,t}}$  distributions appears more symmetrical as the fixation of single alleles is very rare (Fig. 4f, g). As the maximal allelic diversity (number of possible alleles) increases beyond two,  $t_\mu$  increases accordingly and mutation is “weakened” even further in comparison to the other processes (Fig. 2d).

The time until the exact final  $\overline{F_{IS,\infty}}$  distribution is reached depends on the lowest force acting in the population. In realistic biological conditions, it typically depends on  $t_\mu$  (Table 1, Additional file 1, 1.5). However, the mean  $\overline{F_{IS,\infty}}$  may be reached much earlier than the time needed to reach the exact final  $\overline{F_{IS,\infty}}$  distribution (Fig. 4, Table 1). Indeed, reaching the final mean  $\overline{F_{IS,\infty}}$  only requires that populations have reached the final heterozygosity which depends on mutation and sexual reproduction. This final heterozygosity is quickly reached with few sexual events. Reaching the exact final  $\overline{F_{IS,\infty}}$  distribution also implies having reached the steady state allele frequencies which only depends on mutation.

**Table 1** Convergence times under partial clonality for seven representative example parameter sets

$N = 100$	$c$	$\mu$	$t_N$	$t_c$	$t_\mu$	$t_I$	$t_{II}$	$t_{III}$
<b>A</b>	0.0	$10^{-6}$	159	<b>1</b>	$2.6 \times 10^6$	1	1	$2.6 \times 10^6$
<b>B</b>	0.8	$10^{-6}$	159	<b>25</b>	$2.6 \times 10^6$	27	27	$2.6 \times 10^6$
<b>C</b>	0.97	$10^{-6}$	159	<b>159</b>	$2.6 \times 10^6$	174	177	$2.6 \times 10^6$
<b>D</b>	0.99	$10^{-6}$	159	<b>529</b>	$2.6 \times 10^6$	464	498	$2.6 \times 10^6$
<b>E</b>	1.0	$10^{-6}$	159	$\infty$	<b><math>2.6 \times 10^6</math></b>	38,366	$\gg 40,000$	$2.6 \times 10^6$
<b>F</b>	1.0	$10^{-2}$	159	$\infty$	<b>263</b>	234	138	264
<b>G</b>	1.0	$10^{-1}$	159	$\infty$	<b>25</b>	25	14	25

Population size  $N = 100$  throughout. Columns:  $c$  – rate of clonality,  $F_{IS,\infty} = 0$  – mutation rate,  $t_N$  – genetic drift maximal expected convergence time,  $t_c$  – reproduction maximal convergence time,  $t_\mu$  – mutation maximal convergence time,  $t_I$  – convergence time to the mean  $\overline{F_{IS,\infty}}$  based on the model in [13],  $t_{II}$  – convergence time to the mean  $\overline{F_{IS,\infty}}$  based on our model,  $t_{III}$  – convergence time to full final distribution of  $F_{IS,\infty}$ . Rows: example parameter sets (compare Fig. 4). Bold: min ( $t_c, t_\mu$ )

Thus, when comparing different rates of clonality in populations with the same size and realistic mutation rate, the increase of the convergence time to  $\overline{F_{IS,\infty}}$  is directly related to  $t_c$  until  $t_c > t_\mu$ .

Finally, an important observation is also that when approaching the steady state in partially clonal populations, the variations of  $F_{IS}$  increase again as compared to variations expected at the steady states, as the “evolutionary memory” of clonality (sensu [38]) ease the influence of genetic drift on genotype frequencies (Fig. 4). Partial clonal populations that have not yet reached their steady state distribution of  $\overline{F_{IS,\infty}}$  may thus show greater deviations from their current exact mean  $\overline{F_{IS,t}}$  value than expected for the final  $\overline{F_{IS,\infty}}$  values, especially in small population sizes (Additional file 2: Figures S2.8 and S2.9). For example in highly clonal populations (Fig. 4d, e at  $t = 50$ ), the average deviations from the current exact mean  $\overline{F_{IS,50}}$  based on ten loci even exceeded  $\pm 0.1$  the variation yet expected at the steady states.

## Discussion

Our results on the dynamics of  $F_{IS}$  in partially clonal populations add a new dimension – time – to the description of the steady state distribution  $\overline{F_{IS,\infty}}$  and its mean  $\overline{F_{IS,\infty}}$  derived from previous models [13, 16]. They allow estimating the time needed to reach steady state distributions of genotype frequencies, be there are different or not from the classical Hardy-Weinberg proportions defined for sexual populations. Those estimates not only take into account the hyperbolic relationship between the rate of clonality and the convergence times to HWE [15], but also the effects of two other major evolutionary forces, genetic drift and mutation (Fig. 3).

Thus far, the frequent occurrence of departure from HWE observed in partially clonal species was interpreted as fitting a narrow range of scenario where clonality (heterozygote excesses) and/or selfing (homozygote excesses) would dominate. This study offer a new perspective on these observations, by associating significant  $F_{IS}$  to a much broader range of stable and transient scenario including modest rates of clonality, even in the absence of selfing. These results thus call for increased caution when interpreting these field data alone to make inferences on the importance of clonality or selfing, in the absence of other biological knowledge of the mating system.

Indeed, our work confirms the previous findings of departure from the usual average  $\overline{F_{IS,\infty}} = 0$  only at very high rates of clonality [13, 14, 17]. However, it also recalls and supports the pioneering work from Marshall and Weir [15] on the speed of convergence towards this  $\overline{F_{IS,\infty}}$  in partially clonal populations. We

extend the previous results by elucidating how the interplay between rate of clonality, population size (i.e. genetic drift) and mutation affects the nature of the steady state, the time required to reach it and the transient variation of the distribution of  $F_{IS}$  before reaching it. The results reported here show that owing to an increased “evolutionary memory” (sensu [38]) of past genotypic diversity in partially clonal populations, population history such as changes in reproductive mode can produce a transient but possibly long-lasting overrepresentation of  $F_{IS}$  values departing from 0, even under modest rates of clonality. Additionally the variation of  $F_{IS}$ , which changes dynamically until the steady state is reached, increases the risk to misestimate the mean  $\overline{F_{IS,t,L}}$  unless a high number of loci is used. These findings thus contribute to explaining the frequent reports of negative mean  $\overline{F_{IS}}$  in the literature even for species expected to undergo clonal reproduction at intermediate rates. Their otherwise restricted interpretation as a signature of extreme clonality, conceived under the narrow light of the steady state distributions of  $\overline{F_{IS,\infty}}$ , is thus alleviated. In the following, we further discuss the way this increased “evolutionary memory” and the interaction between rate of clonality, genetic drift and mutation affect the interpretation of  $\overline{F_{IS}}$  in natural populations of partially clonal organisms.

## The respective effects of reproduction, mutation and genetic drift in partially clonal populations

Results on the dynamics of genotype frequencies due to each evolutionary process (model parameters) formally demonstrate that:

- Compared to the standard expectation for exclusively sexual populations, clonality only slows down the approach to HWE (Fig. 2b), at a rate depending on other processes such as mutation and genetic drift to which it grants an increased influence.
- mutation, if acting independently at each allelic copy, leads towards  $F_{IS,\infty} = 0$  (Fig. 2c and Additional file 2: Figure S2.3) even for extreme clonal rates, when genetic drift is comparatively negligible ( $\mu > 1/N$ )
- genetic drift, if not negligible, tends towards genotypic uniformity, i.e. either.  $F_{IS,\infty} = -1$  (one heterozygous genotype remains) or fixation (one homozygous genotype remains), depending on the initial genotype frequencies (Fig. 2e).

Thus the negative  $\overline{F_{IS,\infty}}$  predicted by previous models, and observed in many empirical studies, strongly depend not only on the rate of clonality but

also on its interplay with the population size and, to a lesser extent, the mutation rate (Fig. 3).  $\overline{F_{IS,\infty}} = 0$  is expected at steady state even under high rates of clonality and pure clonality in large populations (Additional file 2, 2.6), as well as in smaller populations when mutation rates are high, as e.g. for microsatellites markers (Fig. 3, “mutation” part of the diagram; Fig. 4f, g). Contrastingly, negative  $\overline{F_{IS,\infty}}$  as suggested by [13] are also expected, yet only in smaller clonal populations or when back mutations are neglected (as in [13]), thus only where genetic drift dominates the dynamics of genotype frequencies (Fig. 4d, e; Fig. 3, “genetic drift” part of the diagram).

The maximal convergence time for mutation,  $t_\mu$ , to its steady state situated on the HWE (i.e.  $\overline{F_{IS,\infty}} = 0$ ) increases with the number of possible alleles (Fig. 2d). This steady state due to mutation exists regardless of the number of alleles or of the mutation scheme as long as all alleles mutate at constant rates (e.g. [39, 40], rates need not be equal among alleles, compare Additional file 1, 1.3). This result is only violated if mutation depends on the other alleles at the same locus within the same individual, as for gene conversion [41, 42]. In this particular case, which we did not assess here but which promotes homozygote excess, populations should converge with a similar dynamics to  $F_{IS} = 1$  or to the fixation of one allele.

According to our results, clonal populations dominated by mutation and with no gene conversion (Fig. 3, “mutation” part of the diagram) may distinguish themselves from their counterparts dominated by random mating (Fig. 3, “reproduction” part of the diagram) by the rarity of loci that are fixed for one allele throughout the whole population, rather than by a different mean  $\overline{F_{IS}}$ . In order to assess the relative importance of clonality, the examination of the distribution of  $F_{IS}$  values among multiple loci and of the proportion of fixed loci would thus be more informative than the mean  $F_{IS}$  value itself.

### The dynamics of $F_{IS}$ in partially clonal populations: the implications of the long lasting temporal dynamics and of the large interloci variance

Our equations and numerical results demonstrated that the dynamics of genotype frequencies and  $F_{IS}$  are slowed down in partially clonal populations, which therefore retain traces of their past for much longer than their exclusively sexual counterparts. This questions the generic value of the final mean  $\overline{F_{IS,\infty}}$  as derived in previous studies [13, 16] as a basis for the interpretation of field data. For at least intermediate rates of clonality, the genetic and genotype composition of the population may indeed reflect population

history rather than the present day reproductive system and thus mostly depend on the time since the last disturbance of the population (Fig. 4, Additional file 2, 2.6).

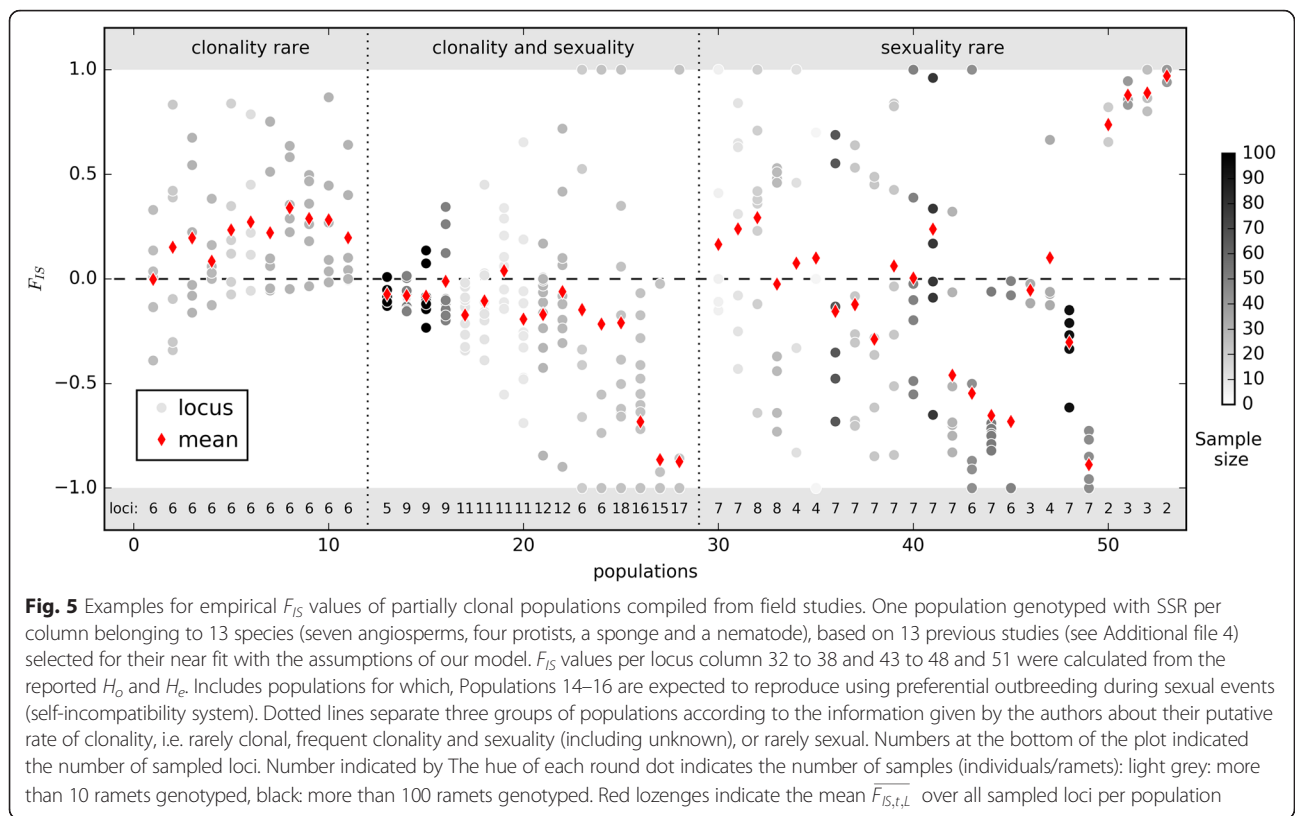
The deceleration of the dynamics of  $F_{IS}$  during its approach to the steady state is connected to the hyperbolic increase of  $t_c$  (Fig. 2b) and thus much stronger under high rates of clonality. We demonstrated that a comparison of the maximal expected convergence times  $t_c$ ,  $t_\mu$ ,  $t_N$  can be an efficient means to predict the overall pattern of  $F_{IS}$  dynamics (Figs. 3, 4 and 5, Additional file 2: Figure S2.5). The times  $t_c$  and  $t_\mu$  can even be used to estimate convergence times of the complete model (Table 1): While the time until the steady state distribution of  $\overline{F_{IS,\infty}}$  is reached nearly always depends on  $t_\mu$  in realistic biological conditions, the convergence time to the final mean  $\overline{F_{IS,\infty}}$  can be estimated by the minimum of  $t_c$  and  $t_\mu$  (i.e. usually  $t_c$  in small populations). If  $t_N \ll \min(t_c, t_\mu)$ , loci with different initial genotype frequencies may not converge to the same final  $F_{IS,\infty}$  value (convergence to genotypic uniformity), so that the expected final  $\overline{F_{IS,\infty}} = -1$  cannot be reached within biologically realistic time spans even under nearly pure clonality (Fig. 4e, Additional file 2: Figure S2.5). Though not yet included in our model, perennality leading to overlapping generations (partial survival of the individuals across generations) is expected to slow down  $F_{IS}$  dynamics even further. If disturbances are sufficiently frequent, e.g. in very instable environments or in populations cyclically changing between exclusive sexual and clonal reproduction [43, 44], the final  $\overline{F_{IS,\infty}}$  and  $\overline{F_{IS,\infty}}$  based on the currently observed rate of clonality may even never be reached.

Finally, during the convergence toward the final  $\overline{F_{IS,\infty}}$ , loci may go through intermediate distributions that would be definitely unusual in exclusively sexual populations both in terms of genotype frequencies and  $F_{IS}$  (Fig. 4, Additional file 2: Figure S2.6). Even at modest rates of clonality, the variation of  $F_{IS}$  is increased compared to exclusively sexual populations. Consequently, information from more loci is required to reach an accurate estimate of the mean  $\overline{F_{IS,t}}$  in partially clonal populations compared to strictly sexual ones. We found that the variation of  $F_{IS}$  observed during the approach to the steady state distribution may be even greater than predicted based on the final  $\overline{F_{IS,\infty}}$ , especially when  $t_N \gg t_\mu$  (Additional file 2: Figures S2.8 and S2.9).

### Transient $F_{IS}$ values: a new hypothesis to account for values observed in field data?

We performed a literature review to illustrate the frequent observation of a very wide variety of  $F_{IS}$





values, positive as well as negative, in partially clonal populations (Fig. 5; details in Additional file 4). Field data may be influenced by technical biases, including sampling bias due to an unknown spatial structure of clones, missing rare genotypes due to non-exhaustive sampling, genotyping errors (e.g. undetected null alleles for SSRs) or preferential sampling of loci with near-isoplethy (thus increasing the probability to find negative  $F_{IS}$ ). Moreover, biological processes other than those in our model may have acted on the data we collected. We therefore applied strict criteria to standardize the dataset we used and retain only those populations that fitted our model best. We present only studies that did include repeated multiloci genotypes in their calculations, published  $F_{IS}$  values (or  $H_o$  and  $H_e$ ) per locus and clearly isolated population, and reported data on organisms whose life cycle fits with our model (i.e. dominantly diploid, no cyclic clonality as e.g. in aphids). Only few studies matched these strict criteria [20, 45–56]. We kept studies on species with self-incompatibility systems, as this system of preferential outbreeding has been shown to have very little effect on  $F_{IS}$  at loci physically (nearly) unlinked to the SI genes [57, 58].

Owing to the increased evolutionary memory of past demographic fluctuations, the results of our study open up

new possible explanations for the presence, but also the absence, of positive and negative  $F_{IS}$  values (both at individual loci or the mean) in partially clonal populations. On their way back to the steady state distribution, even intermediately clonal populations may transiently exhibit  $F_{IS,t}$  values that should be rare in the final steady state distribution  $\overline{F_{IS,\infty}}$ . Such values can be due to the increased variation of  $F_{IS}$  or echo the past departure from equilibrium due to population history (e.g. demographic bottleneck, change in the rate of clonal reproduction). As an example of how to apply our results, slightly negative mean  $\overline{F_{IS,t,L}}$  over loci in some wild cherry populations (populations 14–16 in Fig. 5, [20]), would have suggested almost exclusive clonality when taking the expected mean  $\overline{F_{IS,\infty}}$  under equilibrium as the reference. For example,  $\overline{F_{IS,t,L}} = -0.083$  the exhaustively sampled population 16 (Fig. 5) with  $N = 247$ ,  $\mu \in [10^{-3}, 10^{-12}]$  suggest  $c \approx 0.98$ . However, the proportion of repeated multiloci genotypes and inferences from parentage analysis suggested an intermediate rate of clonality instead ( $c \sim 0.5$ ) [20, 59]; based on this value, genotype dynamics would be dominated by random mating ( $t_c \approx 10$ ). Using our theoretical results, the observed  $F_{IS}$  distribution and its negative mean could be explained



either by extensive logging in the past (population history) or by the fact that only nine loci were analyzed in small populations of about a hundred individuals (increased variation; compare Additional file 2, 2.8 and 2.9).

Our results also suggest ways to further improve the population genetic inferences in natural populations of partial clonals, based on  $F_{IS}$  in connection with other parameters as proposed in [14]. First, other parameters than  $F_{IS}$  may be much more informative when attempting to assess the relative importance of sexual versus clonal reproduction in the dynamics and evolution of partially clonal population. Maximizing the number of loci studied by moving from population genetics to population genomics may help to improve the statistical basis of inferences of population parameters. Second, if pursuing the investigation of the influence of rates of clonality on  $F_{IS}$ , rather than focusing exclusively on the mean  $\overline{F_{IS,t,L}}$  over loci, the full distribution of  $\overline{F_{IS,t,L}}$  values per population should be reported and interpreted. Collecting time series of samples may also provide valuable information, as field data normally represent only a “snapshot” of genotype frequencies at a particular point in time, that may or may not be representative of the steady state distribution of the parameters being studied. Using the Markov chain model implemented here, it is not only possible to statistically analyze example trajectories, but also to analytically derive the transition probabilities between two consecutive sets of genotype frequencies for a range of clonal rates, based on population size, mutation rate and number of generations between time series samples. The results of this study open up perspectives for the development of a unified statistical method to infer rates of clonal reproduction, or other population genetic parameters of our model if  $c$  would be known, based on the analysis of temporal samples. Taking into account the temporal dynamics of genetic descriptors of the populations, including  $F_{IS}$  would therefore help to improve the biological interpretation of values from field data, or to refine methods for estimating the rate of clonality based on a collection of population genetic indices.

## Conclusions

Our results allow reconciling predictions for  $F_{IS}$  under partial clonality from theoretical models, which suggest departures from  $F_{IS} = 0$  only at nearly pure clonality, with some empirically observed values, which show such departures also where sex is known or suspected to be frequent. Examining the dynamic effects of clonality under the varying influence of mutation

and drift showed three main implications for interpreting  $F_{IS}$  in partially clonal populations:

- non-negative  $F_{IS}$ , including null values, are not a reliable indicator of the absence of clonal reproduction, as they may occur i) even at steady state under highly frequent but not exclusive clonality, provided the influence of mutation compared to drift in populations is large ( $\mu \gg 1/N$ ). Or they may occur ii) transiently under all rates of clonality, a likely situation for many wild populations considering the hyperbolic relationship between the rate of clonality and the time toward convergence reported here.
- negative  $F_{IS}$  values are not a reliable indicator of nearly exclusive clonal reproduction, as significant deviations from  $F_{IS} = 0$  for multiple loci are also expected after departures from the steady state distribution in partially clonal populations, which generally last longer even if the rate of clonality is only intermediate.
- An increased number of loci are required to maintain the accuracy of DNA-based estimates of population genetic parameters in partially clonal compared to exclusively sexual populations in general, and the study of time series rather than single snapshots of genetic data may lead to more accurate estimates of the rate of clonal reproduction in particular.

## Nomenclature

$F_{IS}$ , inbreeding coefficient represents a correlation coefficient among alleles at a particular locus within polyploid (diploid here) individuals

$t$ , current generation (discrete time)

$\tilde{F}_{IS,t}$ , exact distribution of  $F_{IS}$  at a time  $t$

$\bar{F}_{IS,t}$ , mean  $F_{IS}$  at a time  $t$

$H_o$ , observed heterozygosity

$H_e$ , expected heterozygosity

$E$ , allele identity within individuals

$\Theta$ , allele identity within the population

HWE, Hardy-Weinberg equilibrium

$N$ , population size

$n$ , number of alleles at one locus. Alleles can be named as  $A$ ,  $a$  for a biallelic locus or  $A_1, A_2, \dots, A_i$  for locus with more than two alleles

$i \neq j \neq k \neq l$ , indices referring to alleles

$\nu_b$ , allele frequency of  $A_i$

$g$ , number of different genotypes at one locus

$q_{ij}$ , number of individual of genotype  $A_i A_j$

$\nu_{ij}$ , genotype frequency of genotype  $A_i A_j$

$c$ , rate of clonal reproduction

$s$ , rate of selfing, set here at  $1/N$  as expected under random mating

$\mu$ , mutation rate

$\alpha$ , the probability that an allele does not mutate

$\beta$ , the probability that an allele mutates into one of the  $n - 1$  others during one generation

$X$ , random variable

$P(X)$ , probability of a random variable  $X$

$M$ , transition matrix

$\mathcal{M}$ , multinomial distribution

$L$ , numbers of studied polymorphic loci

$t_c$ , maximal number of generations to convergence due to rate of clonality

$t_\mu$ , maximal number of generations to convergence due to mutation rate

$t_N$ , maximal number of generations to convergence due to genetic drift

$\min(t_c, t_\mu)$ , minimum convergence time between two evolutionary forces

$\epsilon$ , universal “acceptable error” corresponding to one half the minimal change in genotype frequency that would be measurable by exhaustive sampling in a population of finite size  $N$

$q_s$ , probabilities that two individuals taken at random in the same reproductive subpopulation after migration were sired in the same reproductive subpopulation one generation before

$q_d$ , probability that two individuals taken at random in different reproductive subpopulations after migration originated from the same subpopulation one generation before

## Additional files

**Additional file 1:** Mathematical Background. 1.1 Model equations, 1.2 Convergence times – individual parameters, 1.3 Mutation and heterozygosity – multiple alleles, asymmetric mutation rate, 1.4 Genetic drift and heterozygosity – multiple alleles, 1.5 Convergence times – full model, 1.6 General solution and mixing time of Balloux et al. [13] recurrence equations. (DOCX 73 kb)

**Additional file 2:** Figures and Tables. 2.1. Interpretation of *de Finetti* diagrams, 2.2 *De Finetti* landscapes for reproduction, 2.3 *De Finetti* landscapes for mutation, 2.4 *De Finetti* landscapes for genetic drift, 2.5: Dynamics of probability of fixation and distributions of  $F_{IS}$  through time for large population size, low mutation rate  $10^{-6}$  at locus with 10 alleles 2.6: Dynamics of probability of fixation and distributions of  $F_{IS}$  through time for large population size, high mutation rate  $10^{-3}$  at locus with 10 alleles 2.7 Example trajectories over time. 2.8 Sampling error of the mean  $\overline{F_{IS,t,L}}$  according to number of loci (Markov chain). 2.9 Sampling error of the mean  $\overline{F_{IS,t,L}}$  according to number of loci (simulations) Additional tables. 2.1 transition probabilities conditionally to rates of clonality, mutation rates and previous genotypic state. 2.2 Convergence time of genetic drift  $t_N$  based on Markov chain absorption time. 2.3: Convergence time of genetic drift  $t_N$  based on simulations. 2.4 Effects of different rates of clonality on the dynamics of  $F_{IS}$ . (DOCX 10902 kb)

**Additional file 3:** Basic model code in Python. Input  $c, \mu, N$ , start state and  $t$  to get the expected probability distribution of model states. (DOCX 30 kb)

**Additional file 4:** Literature review. References and supplementary data for the literature review. “PN” refers to the number of the dataset(s) in Fig. 5, \*:  $F_{IS}$  was calculated from  $H_e$  and  $H_o$ . (DOCX 20 kb)

## Acknowledgements

We thank Jurgen Angst, Valentin Bahier, Fabien Halkett, Cédric Midoux, Stéphane De Mita, Martin Lascoux and two anonymous reviewers for helpful discussions and comments on the manuscript. We greatly profited from the exchange of ideas within the CLONIX group.

Authors thank the *French National Research Agency*, CLONIX project (ANR-11-BSV7-0007), the *French National Institute for Agricultural Research, Plant Health and the Environment Department* (INRA-SPE) and the *Région Bretagne, France* for supporting this research.

## Funding

This work was supported by the *French National Research Agency*, CLONIX project (ANR-11-BSV7-0007). Katja Reichel worked in this project during her PhD granted by the *French National Institute for Agricultural Research, Plant Health and the Environment Department* (INRA-SPE) and the *Région Bretagne, France*.

## Availability of data and materials

The data supporting the results of this article are included within the article and its Additional files. A basic version of the model code is available as Additional file 3 and from [http://www6.rennes.inra.fr/igepp\\_eng/Productions/Software](http://www6.rennes.inra.fr/igepp_eng/Productions/Software).

## Authors' contributions

JPM and SS laid the foundation for this work. JPM, FM, KR and SS conceived the mathematical model. KR and SS performed the scientific programming and generated the numerical data. KR did the literature review, most of the data analyses and all illustrations. KR, SAH and SS wrote the manuscript. All authors contributed to editing, managed and performed by SS. SAH and SS were responsible for funding and grant applications. All authors have read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Author details

<sup>1</sup>IGEPP, Agrocampus Ouest, INRA, Université de Rennes 1, 35650 Le Rheu, France. <sup>2</sup>Université de Tours, CNRS-UMR7350 LMPT, F-37200 Tours, France. <sup>3</sup>IFREMER, UMR5240 MARBEC, F-34203 Sète, France.

Received: 29 January 2016 Accepted: 1 June 2016

Published online: 10 June 2016

## References

- Duminil J, Fineschi S, Hampe A, Jordano P, Salvini D, Vendramin GG, Petit RJ. Can population genetic structure be predicted from life-history traits? *Am Nat*. 2007;169:662–72.
- Duminil J, Hardy OJ, Petit RJ. Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC Evol Biol*. 2009;9:177.
- de Meeüs T, Prugnolle F, Agnew P. Asexual reproduction: Genetics and evolutionary aspects. *Cell Mol Life Sci*. 2007;64:1355–72.
- McKey D, Elias M, Pujol B, Duputié A. The evolutionary ecology of clonally propagated domesticated plants: Tansley review. *New Phytol*. 2010;186:318–32.
- Tibayrenc M, Ayala FJ. Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proc Natl Acad Sci*. 2012;109:E3305–13.
- Liu J, Dong M, Miao SL, Li ZY, Song MH, Wang RQ. Invasive alien plants in China: role of clonality and geographical origin. *Biol Invasions*. 2006;8:1461–70.

7. Luijten SH, Oostermeijer JGB, van Leeuwen NC, den Nijs HC. Reproductive success and clonal genetic structure of the rare *Arnica montana* (Compositae) in the Netherlands. *Plant Syst Evol.* 1996;201:15–30.
8. Sydes MA, Peakall R. Extensive clonality in the endangered shrub *Haloragodendron lucasii* (Haloragaceae) revealed by allozymes and RAPDs. *Mol Ecol.* 1998;7:87–93.
9. Brzosko E, Wróblewska A, Ratkiewicz M. Spatial genetic structure and clonal diversity of island populations of lady's slipper (*Cypripedium calceolus*) from the Biebrza National Park (northeast Poland). *Mol Ecol.* 2002;11:2499–509.
10. Setsuko S, Ishida K, Tomaru N. Size distribution and genetic structure in relation to clonal growth within a population of *Magnolia tomentosa* Thunb. (Magnoliaceae). *Mol Ecol.* 2004;13:2645–53.
11. Brzyski JR, Culley TM. Genetic variation and clonal structure of the rare, riparian shrub *Spiraea virginiana* (Rosaceae). *Conserv Genet.* 2011;12:1323–32.
12. Schön I, Martens K, Dijk P, editors. *Lost Sex*. Dordrecht: Springer Netherlands; 2009.
13. Balloux F, Lehmann L, de Meeüs T. The population genetics of clonal and partially clonal diploids. *Genetics.* 2003;164:1635–44.
14. Halkett F, Simon J, Balloux F. Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol.* 2005;20:194–201.
15. Marshall DR, Weir BS. Maintenance of genetic variation in apomictic plant populations. *Heredity.* 1979;42:159–72.
16. Stoeckel S, Masson J-P. The exact distributions of  $F_{IS}$  under partial asexuality in small finite populations with mutation. *PLoS One.* 2014;9:e85228.
17. de Meeüs T, Lehmann L, Balloux F. Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. *Infect Genet Evol.* 2006;6:163–70.
18. Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA. Standardizing methods to address clonality in population studies. *Mol Ecol.* 2007;16:5115–39.
19. Motoie G, Ferreira GEM, Cupolillo E, Canavez F, Pereira-Chioccola VL. Spatial distribution and population genetics of *Leishmania infantum* genotypes in São Paulo State, Brazil, employing multilocus microsatellite typing directly on dog infected tissues. *Infect Genet Evol.* 2013;18:48–59.
20. Stoeckel S, Grange J, Fernández-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S. Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. *Mol Ecol.* 2006;15:2109–18.
21. Wright S. Evolution in Mendelian populations. *Genetics.* 1931;16:97–159.
22. Ewens WJ. *Mathematical population genetics: I. Theoretical Introduction*. 2nd ed. New York: Springer; 2004.
23. Gale JS. *Theoretical population genetics*. London: Springer: Unwin Hyman; 1990.
24. Reichel K, Bahier V, Midoux C, Masson J-P, Stoeckel S. Interpretation and approximation tools for big, dense Markov chain transition matrices in ecology and evolution. *Algorithms Mol Biol.* 2016;10:1–14.
25. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, editor. *Mammalian protein metabolism*. Volume III. New York: Academic; 1969. p. 21–132.
26. Nei M. *Molecular population genetics and evolution*. Amsterdam: North Holland Publishing Company; 1975.
27. Brakefield PM. The variance in genetic diversity among subpopulations is more sensitive to founder effects and bottlenecks than is the mean: a case study. In: Fontdevila A, editor. *Evolutionary biology of transient unstable populations*. Berlin: Springer Berlin Heidelberg; 1989. p. 145–61.
28. Edmands S, Feaman HV, Harrison JS, Timmerman CC. Genetic consequences of many generations of hybridization between divergent copepod populations. *J Hered.* 2005;96(2):114–23.
29. Zalapa JE, Brunet J, Guries RP. The extent of hybridization and its impact on the genetic diversity and population structure of an invasive tree, *Ulmus pumila* (Ulmaceae). *Evol Appl.* 2010;3(2):157–68.
30. Lindtke D, Buerkle CA, Barbará T, Heinze B, Castiglione S, Bartha D, Lexer C. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Mol Ecol.* 2012;21:5042–58.
31. Graffelman J, Camarena JM. Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Hum Hered.* 2008;65:77–84.
32. Oliphant TE. Python for scientific computing. *Comput Sci Eng.* 2007;9:10–20.
33. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, SciPy Edited by: Varoquaux G, Vaught T, Millman J. CA USA: Pasadena; 2008:11–15.
34. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9:90–5.
35. de Finetti B. Considerazioni matematiche sull'ereditarietà mendeliana. *Metron.* 1926;6:3–41.
36. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics.* 1998;148:1667–86.
37. Hile SE, Yan G, Eckert KA. Somatic mutation rates and specificities at TC/AG and GT/CA microsatellite sequences in nontumorigenic human lymphoblastoid cells. *Cancer Res.* 2000;60:1698–703.
38. Desai MM. Reverse evolution and evolutionary memory. *Nat Genet.* 2009;41:142–4.
39. Estoup A, Jarne P, Cornuet J-M. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol.* 2002;11:1591–604.
40. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet.* 2004;5:435–45.
41. McMahill MS, Sham CW, Bishop DK. Synthesis-dependent strand annealing in meiosis. *PLoS Biol.* 2007;5:e299.
42. Flot J-F, Hespels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B, Koszul R, Aury J-M, Barbe V, Barthélémy R-M, Bast J, Bazykin GA, Chabrol O, Couloux A, Da Rocha M, Da Silva C, Gladyshev E, Gouret P, Hallatschek O, Hecox-Lea B, Labadie K, Lejeune B, Piskurek O, Poulain J, Rodriguez F, Ryan JF, Vakhrusheva OA, Wajnberg E, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature.* 2013;500:453–7.
43. Allen DE, Lynch M. The effect of variable frequency of sexual reproduction on the genetic structure of natural populations of cyclical parthenogen. *Evolution.* 2012;66:919–26.
44. Berg LM, Lascoux M. Neutral genetic differentiation in an island model with cyclical parthenogenesis. *J Evol Biol.* 2000;13:488–94.
45. Duran S, Pascual M, Estoup A, Turon X. Strong population structure in the marine sponge *Crambe crambe* (Poecilosclerida) as revealed by microsatellite markers. *Mol Ecol.* 2004;13:511–22.
46. Nagamitsu T, Ogawa M, Ishida K, Tanouchi H. Clonal diversity, genetic structure, and mode of recruitment in a *Prunus ssp.* population established after volcanic eruptions. *Plant Ecol.* 2004;174:1–10.
47. Corral JM, Molins MP, Aliyu OM, Sharbel TF. Isolation and characterization of microsatellite loci from apomictic *Hypericum perforatum* (Hypericaceae). *Am J Bot.* 2011;98:e167–9.
48. Jiang K, Gao H, Xu N-N, Tsang EPK, Chen X-Y. A set of microsatellite primers for *Zostera japonica* (Zosteraceae). *Am J Bot.* 2011;98:e236–8.
49. Tew JM, Lance SL, Jones KL, Fehlberg SD. Microsatellite development for an endangered riparian inhabitant, *Lilaeopsis schaffneriana* subsp. *recurva* (Apiaceae). *Am J Bot.* 2012;99:e164–6.
50. Liu W, Zhou Y, Liao H, Zhao Y, Song Z. Microsatellite primers in *Carex moorcroftii* (Cyperaceae), a dominant species of the steppe on the Qinghai-Tibetan Plateau. *Am J Bot.* 2011;98:e382–4.
51. Barnabe C, Buitrago R, Bremond P, Aliaga C, Salas R, Vidaurre P, Herrera C, Cerqueira F, Bosseno M-F, Waleckx E, Breniere SF. Putative panmixia in restricted populations of *Trypanosoma cruzi* isolated from wild *Triatoma infestans* in Bolivia. *PLoS One.* 2013;8:e82269.
52. Tesson SVM, Borra M, Kooistra WHCF, Procaccini G. Microsatellite primers in the planktonic diatom *Pseudo-nitzschia multistriata* (Bacillariophyceae). *Am J Bot.* 2011;98:e33–5.
53. Villate L, Esmenjaud D, Van Helden M, Stoeckel S, Plantard O. Genetic signature of amphimixis allows for the detection and fine scale localization of sexual reproduction events in a mainly parthenogenetic nematode. *Mol Ecol.* 2010;19:856–73.
54. Gao H, Jiang K, Geng Y, Chen X-Y. Development of microsatellite primers of the largest seagrass, *Enhalus acoroides* (Hydrocharitaceae). *Am J Bot.* 2012; 99:e99–101.
55. McInnes LM, Dargantes AP, Ryan UM, Reid SA. Microsatellite typing and population structuring of *Trypanosoma evansi* in Mindanao, Philippines. *Vet Parasitol.* 2012;187:129–39.
56. Vilas R, Cao A, Pardo BG, Fernández S, Villalba A, Martínez P. Very low microsatellite polymorphism and large heterozygote deficits suggest founder effects and cryptic structure in the parasite *Perkinsus olseni*. *Infect Genet Evol.* 2011;11:904–11.
57. Navascués M, Stoeckel S, Mariette S. Genetic diversity and fitness in small populations of partially asexual, self-incompatible plants. *Heredity.* 2009;104:482–92.
58. Stoeckel S, Klein EK, Oddou-Muratorio S, Musch B, Mariette S. Microevolution of S-allele frequencies in wild cherry populations: respective

impacts of negative frequency dependent selection and genetic drift. *Evolution*. 2012;66:486–504.

59. Stoeckel S. Impact de la propagation asexuée et du système d'auto-incompatibilité gamétophytique sur la structuration et l'évolution de la diversité génétique d'une essence forestière entomophile et disséminée, *Prunus avium* L. France: Cemagref; 2006.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

