Molecular Ecology Resources
November 2017, Volume 17, Issue 6, Pages e251-e267
http://dx.doi.org/10.1111/1755-0998.12698
http://archimer.ifremer.fr/doc/00390/50187/

ClonEstiMate, a Bayesian method for quantifying rates of clonality of populations genotyped at two-time steps

Becheler Ronan [1], Masson Jean-Pierre [2], Arnaud-Haond Sophie [3], Halkett Fabien [4], Mariette Stéphanie [5], Guillemin Marie-Laure [1, 6], Valero Myriam [1], Destombe Christophe [1], Stoeckel Solenn [2, *]

[1] CNRS, Sorbonne Universités, UPMC, Univ. Paris VI, UC, UACH, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff; CS 90074 29688 Roscoff, France
[2] INRA, UMR1349 Institute for Genetics, Environment and Plant Protection; Le Rheu ,France
[3] Laboratory Environnements-Ressources, Ifremer, Bd Jean Monnet BP 171; Sète 34203, France
[4] INRA, UMR 1136 Interactions Arbres-Microorganismes, Nancy Grand-Est, France
[5] UMR1202 BIOGECO, INRA, Univ. Bordeaux, 69 Route d'Arcachon; 33610 Cestas, France
[6] Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Casilla 567; Valdivia, Chile

* Corresponding author : Ronan Becheler, email address : solenn.stoeckel@inra.fr

Abstract :

Partial clonality is commonly used in Eukaryotes and has large consequences for their evolution and ecology. Assessing accurately the relative importance of clonal versus sexual reproduction matters for studying and managing such species.

Here, we proposed a Bayesian approach, *ClonEstiMate*, to infer rates of clonality c from populations sampled twice over a short time interval, ideally one generation time. The method relies on the likelihood of the transitions between genotype frequencies of ancestral and descendent populations, using an extended Wright-Fisher model explicitly integrating reproductive modes. Our model provides posterior probability distribution of inferred c, given the assumed rates of mutation, as well as inbreeding and selfing when occurring.

Tested under various conditions, this model provided accurate inferences of c, especially when the amount of information was modest, i.e. low sample sizes, few loci, low polymorphism and strong linkage disequilibrium. Inferences remained robust when mutation models and rates were misinformed. However, the method was sensitive to moderate frequencies of null alleles and when the time interval between required samplings exceeding two generations. Misinformed rates on mating modes (inbreeding and selfing) also resulted in biased inferences. Our method was tested on eleven datasets covering five partially clonal species, for which the extent of clonality was formerly deciphered. It delivered highly consistent results with previous information on the biology of those species.

ClonEstiMate represents a powerful tool for detecting and inferring clonality in finite populations, genotyped with SNPs or microsatellites. It is freely available at "h t tp://https://w">https://w w w 6.rennes.inra.fr / igepp_eng/ Productions/ Software.

Keywords : Rate of asexuality, instantaneous inference, inbreeding, selfing, population genetics model

INTRODUCTION

Clonality or asexuality is a reproductive mode leading to the production of offspring genetically identical to their single parent, with the exception of somatic mutations (De Meeus *et al.* 2007). Clonal descendants are produced by a variety of mechanisms, from vegetative reproduction, such as rhizome elongation in angiosperms (Sintes *et al.* 2005), thallus fragmentation in seaweeds (Guillemin *et al.* 2008) or fission in Cnidaria (Wiedenmann *et al*. 2000), to different modes of apomixis (*i*.e. production of clonal seeds in plants and parthenogenesis for animals, Schön *et al.* 2009). This life-history trait is ubiquitous among the tree of life and can be commonly found in Fungi (Taylor *et al.* 2015), Algae (Scrosati 1998), flowering plants (Vallejo-Marín *et al.* 2010) and all subkingdom of animals (De Meeus *et al.* 2007). Among Eukaryotes, examples of species reproducing using strict clonality seem rare (Judson & Normark 1996; Neiman *et al.* 2009), supporting the hypothesis that almost all "clonal" species reproduce actually using partial clonality; *i.e.* a mixture of clonal and sexual reproduction. This reproductive mode has multiple shades of grey that can be characterized by the rate of clonality *c* indicating the relative frequency of the descendants resulting from clonal reproduction within populations or species (Marshall & Weir 1979). This rate of clonality deeply influences the dynamics of ecosystems (*e.g.* Cornelissen *et al.* 2014), of communities (Jones *et al.* 2009; Neiman *et al.* 2009), the ecology and evolutionary trajectories of species, populations and individuals (Halkett *et al.* 2005; De Meeus *et al.* 2007; Avise 2008; Schön *et al.* 2009; Tibayrenc & Ayala 2012) and impacts human activities (McKey *et al.* 2010) and health (Tibayrenc *et al.* 1990). Studying and managing the dynamics and ecology of partially clonal organisms requires an accurate quantitative estimate of *c* within populations (Halkett *et al.* 2005; Duminil *et al.* 2007; Fehrer 2010; Villate *et al.* 2010; Allen & Lynch 2012). Partial clonality strongly constraints the range of genetic diversity that can evolve within species by shaping the spatial and temporal distributions of polymorphism within and between individuals (Sunnucks *et al.* 1996; De Meeus *et al.* 2007; Pearson *et al.* 2009; Fehrer 2010; Stoeckel & Masson 2014; Rouger *et al.* 2016). Indeed, clonal reproduction maintains genotypic combinations across larger time scales than sexuality, allowing the other

evolutionary forces to apply on them, and retaining successful combinations that would only be transient under full sexuality. Moreover, when parents unequally contribute to offspring production, partial clonality may create repeated genotypes within populations and some linkage disequilibrium among alleles all over the genome (Halkett *et al.* 2005). These effects of partial clonality on the occurrence of repeated genotypes and on genetic diversity have led scientists to interpret two proxies in population genetics studies to assess the rate of clonality: either by estimating clonal richness ($^{G}/_{N}$ *i.e*. the ratio of different genotypes G on the sample size N, Dorken & Eckert 2001; Becheler *et al.* 2010; Becheler *et al.* 2015) or by using expertise knowledge on some identified population genetics indices (Villate *et al.* 2010; Allen & Lynch 2012). Both approaches present methodological limits. The observed number of discriminated genotypes within samples $^{G}/_{N}$ highly depends on the sampling scheme and effort, and on the fact that parents unequally contribute to clonally producing next generations. The relationship between $^{G}/_{N}$ and *c* is thus multifaceted, and impossible to reliably establish in most circumstances. In consequence, the proportion of identical genotypes within samples based on genetic markers cannot be considered as a reliable estimate of *c* (Arnaud-Haond *et al.* 2007). Indirect appraisal of this rate using other population genetics indices like F_{IS}, F_{ST} or linkage disequilibrium estimates still relies on field expertise but also remains largely inaccurate for low to moderate rates of clonality (Berg & Lascoux 2000; De Meeus & Balloux 2005; Arnaud-Haond *et al.* 2007; De Meeus *et al.* 2007; Navascues *et al.* 2010). Thus, dedicated methods are still missing to indirectly assess the rate of clonality in natural populations.

A recent theoretical study shows that the probabilities of genotypic transitions (*i.e*. change in genotype frequencies over one generation) would be impacted since very low rates of clonality and thus can be useful for indirectly inferring the rates of clonality using temporal genotyping (Stoeckel & Masson 2014). Here, we built on those findings and models to propose an innovative semi-quantitative approach, using the evolution of genotype frequencies sampled over successive generations to infer the instantaneous rate of clonality within populations. Such temporal sampling

is commonly used for directly or indirectly inferring qualitative and quantitative biological properties, like population size (Waples 1989; Wang & Whitlock 2003), dispersal distances (Broquet & Petit 2009), evolutionary forces (Stoeckel *et al.* 2012; Do *et al.* 2014), behaviors (Zamudio & Sinervo 2000; Lander *et al.* 2013) and mating systems (Franco-Trecu *et al.* 2015). Temporal sampling strategy could allow disentangling the dynamic signature of clonality from the effects of other evolutionary forces (Reichel *et al.* 2016, Rouger *et al.* 2016).

In order to exploit those signatures, the present method provides the posterior probabilities of a user-defined range of *c* given the observed transitions of genotype frequencies per locus between two temporal samples. It relies on a Wright-Fisher-like population genetics model, which explicitly takes the clonal rate *c* into account (Stoeckel & Masson 2014), to calculate the likelihoods of tested *c* values. For each tested *c* value, data obtained per locus are combined in a Bayes formula to provide posterior probability distributions of *c*.

We tested the accuracy and robustness of this newly developed method using both pseudo-observed datasets (*pods*) and eleven real genotypic datasets covering a large spectrum of *c* over five species with different life cycles and ecology, which combine multiple life history traits and methodological limitations that were not formalized in our model.

MATERIAL AND METHODS

We developed an explicit inference method to estimate the rate of clonality within populations genotyped over generations that would gain on the last advances in genetic dynamics under partial clonality. We thus develop our mathematical equations for samples collected over two consecutive generations, to rapidly assess current rates of clonality and their changes over time.

1. Models and notations

We developed an explicit inference method based on the theoretical transition probabilities of diploid genotype frequencies between ancestral and descendent generations with finite population sizes. This method required sampling and genotyping individuals in delimited populations sampled at a two-generation interval. We defined, a "delimited population" as a group of individuals that experience a common event of (mixed, *i.e.* clonal and sexual outcrossing or selfing) reproduction across the studied generation(s) and in which most, if not all, the pool of the studied descendants arises from the sampled pool of ancestors (Figure 1). The core idea of this method is that the evolution of genotype frequencies at each locus over one generation (hereafter named a genotypic transition) will differ depending on the rate of clonality *c* (Stoeckel & Masson 2014). This method uses the exact mathematical prediction of genotypic transitions over one generation given *c* to infer $\tilde{c}$ observed from genotyped individuals collected in the field.

We chose to concentrate our first efforts on the development of an inference method for diploid individuals because polyploid populations are characterized by complex inheritance patterns and problems of genotype identification due to allele copy numbers that can vary, for example, from 1 to 3 per locus in tetraploid heterozygotes (Dufresne *et al*. 2014). Nevertheless, most eukaryotes have a life cycle which includes a diploid part (Lott *et al.* 2008) with the vast majority of animals and nearly 70% of the flowering plant species being diploid (Otto & Whitton 2000).

Likelihood of genotypic transition over one generation in partially clonal finite population with mutation

Let consider the evolution of genotype frequencies at one locus with *n* alleles, in a finite and diploid population of size N, under a mutation rate *u,* between two successive, non-overlapping and diploid generations, the ancestral one *t* and the descendent one *t+1*. Let consider that individuals transmit their genetic material to offspring using clonality at a rate *c*, sexuality with consanguinity at an

inbreeding rate $(1-c).\varphi$, sexuality with selfing at a rate $(1-c).s$, and thus sexuality under panmixy at a corresponding rate or $(1-c).(1-\varphi).(1-s)$. The present equations on genotypic transition predict the composition of genotype pools rather than the individual genotypes forming them.

Let $\tilde{X}_t = \{\tilde{p}_{A_1A_1}, \dots, \tilde{p}_{A_nA_n}\}$ denotes the set of genotype frequencies observed at one locus within the ancestral population at generation *t* from a sample of N_t individuals and $\tilde{X}_{t+1} = \{\tilde{q}_{A_1A_1}, \dots, \tilde{q}_{A_nA_n}\}$ the set of genotype frequencies observed at the same locus one generation after in a descendent sample of N_{t+1} individuals. Note that sample size at *t* and *t+1* can differ, and that N_{t+1} formalizes the observed genetic drift that shaped genotypic transitions, provided that $\tilde{X}_t$ and $\tilde{X}_{t+1}$ are good estimates of the ancestral and descendent genotype frequencies, or, at least that samples at *t* are unbiased pictures of the ancestral genotype frequencies that sired the descendent sample at *t+1*. Whatever the generation, genotypes can be grouped as homozygote genotypes $A_iA_i = \{(A_1A_1), \dots, (A_nA_n)\}$ and heterozygotes $A_iA_j = \{(A_1A_2), \dots, (A_{n-1}A_n)\}$. Over one generation, ancestral genotype frequencies at one locus evolve under the influence of genetic drift N_{t+1}, mutation rate *u* and rate of clonality *c* into a pool of genotype frequencies expected under panmictic reproduction $X_t^* = \{p^*_{A_1A_1}, \dots, p^*_{A_nA_n}\}$ (*i.e.* pool of genotype frequencies in the possible zygotes). The likelihood to observe a genotypic transition over one generation between $\tilde{X}_t$ to $\tilde{X}_{t+1}$, knowing the quantitative strength of evolutionary forces (N_σ, u, s, c), can thus be written as:

$$L(\tilde{X}_t, \tilde{X}_{t+1}) = \frac{N_{t+1}!}{\prod_{i=1}^{i=n}(q_{A_iA_i})! \prod_{i=1,j\neq i}^{i=n}(q_{A_iA_j})!} . \prod_{i=1}^{i=n}(p^*_{A_iA_i})^{q_{A_iA_i}} . \prod_{i=1,j\neq i}^{i=n}\left(p^*_{A_iA_j}\right)^{q_{A_iA_j}} \quad \text{eq.1}$$

with $N_{t+1} = \sum_i q_{A_iA_i} + \sum_{i,j} q_{A_iA_j}$ and $\forall\, i,j \in \{1, \dots, n\}$.

In populations, if sexuality occurs with selfing and inbreeding, genotype frequencies after reproduction are:

$$\begin{cases} p^*_{A_iA_i} = c.p'_{A_iA_i} + (1-c).\left[(1-\varphi).(1-s).\left(f'_{A_i}\right)^2 + (1-\varphi).s.\left(p'_{A_iA_i} + \frac{1}{2}\sum_{j\neq i} p'_{A_iA_j}\right) + \varphi.(1-s).f'_{A_i}\right] \\ p^*_{A_iA_j} = c.p'_{A_iA_j} + (1-c).\left[2.(1-\varphi).(1-s).f'_{A_i}.f'_{A_j} + (1-\varphi).\frac{s}{2}.p'_{A_iA_j}\right] \end{cases}$$

eq. 2

in which $p'_{A_iA_i}$ and $p'_{A_iA_j}$ are the homozygote and heterozygote sets of ancestral genotype frequencies impacted by mutation, f'_{A_i} and f'_{A_j} are the sets of ancestral allele frequencies impacted by mutation, s the populational rate of homogamy and φ the populational inbreeding rate that would be equivalent to an estimate of the coefficient of inbreeding in populations.

Raw entities in previous equations are genotype frequencies in isolated finite population. What we name *mutation* hereafter acts actually as a *disturbing factor of gene frequencies* (in the sense of Wright 1931) integrating all processes that modifies inherited genotype frequencies with no specific direction, out of reproductive modes and genetic drift, *i.e*. mutation in isolated populations, low migration coming from indistinct and unknown populations outside the studied populations and all technical issues that create random genotyping errors between temporal samples. Though different mutation models varying with marker types can be applied in our equations, we choose to integrate by default the K-alleles mutation model (KAM) as it better proxies changes that mix such *disturbing factor* accounting for mutation and migration, and is commonly used for modeling microsatellite and SNP evolution (Cockerham 1984; Putman & Carbone 2014). KAM considers that the *n* alleles observed in the data are the only possible alleles. Mutation changes any allele to any other allele with equal probability. Four different alleles *i*, *j*, *k* and *l* have to be considered to generalize this model for our equations. Nonetheless the model still works with one, two or three alleles, which simplify equations below. The rate of mutation from one allele to another among the *n* possible alleles is $\mu = \frac{u}{n-1}$ and the non-mutation rate is $\nu = 1 - u$. Therefore, the ancestral genotype frequencies only impacted by mutation X'_t calculate as

$$\begin{cases} p'_{A_iA_i} = \nu^2.p_{A_iA_i} + \mu^2\left(\sum_{i\neq j} p_{A_jA_j} + \sum_{i\neq j,k} p_{A_jA_k}\right) + \ \mu.\nu.\sum_{i\neq j} p_{A_iA_j} \\ p'_{A_iA_j} = (\mu^2+\nu^2).p_{A_iA_j} + 2.\mu^2.\left(\sum_{k\neq i,j} p_{A_kA_k} + \sum_{k,l\neq i,j} p_{A_kA_l}\right) + (\mu.\nu+\mu^2).\left(\sum_{k\neq i,j} p_{A_iA_k} + \sum_{k\neq i,j} p_{A_jA_k}\right) + 2.\mu.\nu.\left(p_{A_iA_i} + p_{A_jA_j}\right) \end{cases}$$

eq.3

From the pool of genotype frequencies expected under panmictic population of finite size, whatever the mutation model, we can now derive the ancestral pool of frequencies after mutation as

$$f'_{A_i} = p'_{A_iA_i} + \frac{p'_{A_iA_j}}{2}$$ eq.4

that can be used together with the ancestral genotype frequencies to predict the filial genotype frequencies in equation1.

Semi-quantitative inference of rates of clonality considering the likelihoods of genotypic transitions

To infer *c* from genetic data sampled over two generations, we use the likelihood equation defined above as classifier in a naïve Bayes approach to update the likelihoods obtained at all loci. This supervised learning method is one of the most efficient and successful, but mathematically simple and extensively studied technique for constructing classifiers (Zhang 2004; Webb *et al.* 2005). It presupposes, as a sufficient but not necessary condition, that random variables (here the genotypic transitions) are independent and identically distributed. This conditional independence assumption (also called "naïve") is rarely true in population genetic models as in most real-world applications. Yet, it has been shown to be an empirically reasonable approach in many identical situations (Hand & Yu 2001) and can be theoretically applied if: *i)* dependences are evenly distributed, *ii)* they cancel each other out over inferred classes, *iii)* distributions of random variables sufficiently segregate over their means per class (Hand & Yu 2001; Zhang 2004; Webb *et al.* 2005). On a mathematical point of view, the naïve Bayes approach is more convenient since it avoids the complex mathematical formula and intensive calculations required to ascertain the likelihoods over all loci as coming from a joint distribution that would take into account the possible dependences between transitions. For a naïve approach, the product of the marginal likelihoods formalized per locus is sufficient. To

strengthen this parsimonious choice as a first step, rather than formalizing the whole complex dependence among loci, we thus cross-validated our inference method using pseudo-observed simulations (see below). We can gain information on our estimate of the posterior probability of rates of clonality conditionally to transitions by stacking likelihoods computed on each locus into a Bayes's theorem equation. In this supervised learning method, estimated rates of clonality and mutation rates are discretized and assessed as values among κ rates of clonality $\tilde{c} \in \{c_1, \dots, c_\kappa\}$ and ρ values of mutation rates $\tilde{u} \in \{u_1, \dots, u_\rho\}$. This discretization limits the computing cost of integrating continuous functions and fits with the inference method.

Considering that $\tilde{Y}_t = \left\{\widetilde{X_t^1}, \dots, \widetilde{X_t^m}\right\}$ and $\tilde{Y}_{t+1} = \left\{\widetilde{X_{t+1}^1}, \dots, \widetilde{X_{t+1}^m}\right\}$ represent the respective set of genotypic frequencies at generation *t* and *t+1* genotyped from N_t and N_{t+1} sampled individuals at *m* markers/loci, the log-posterior of assessed rate of clonality conditionally to the transition from $\tilde{Y}_t$ to $\tilde{Y}_{t+1}$ and the assumed rates of mutation, inbreeding and selfing is:

$$log\ \_P\left(c \mid \tilde{Y}_t, \tilde{Y}_{t+1}, N_t, N_{t+1}, u, \varphi, s\right) =$$

$$\sum_{\alpha=1}^{\alpha=m} log\ \_L_\alpha\left(\widetilde{X_t^\alpha}, \widetilde{X_{t+1}^\alpha}, N_t, N_{t+1} \mid c, u, \varphi, s\right) - log\left(\sum_{i=1}^{i=\kappa}\left(\prod_{\alpha=1}^{\alpha=m} L_\alpha\left(\widetilde{X_t^\alpha}, \widetilde{X_{t+1}^\alpha}, N_t, N_{t+1} \mid c_i, u, \varphi, s\right)\right)\right) \text{ (eq.5)}$$

We provide equation 5 as a joint distribution over genotypic transitions sampled from *m* loci with a uniform prior distribution over a bounded vector of rates of clonality. We deliberately restricted this equation to only one parameter, the rate of clonality, given assumed rates of mutation, inbreeding and selfing. Indeed the identifiability of the effects of mutation, inbreeding and selfing has not been properly studied yet and our ability to decipher their effect on genotypic transitions from the one of partial clonality seems to be limited, which thus impede a joint inference of all those parameters.

As a complementary approach, the method also provides the Maximum *a posteriori* estimations of $\tilde{c}$ for each locus as $arg\ max_{\tilde{c}}\ P\left(c \mid \tilde{Y}_t, \tilde{Y}_{t+1}, N_t, N_{t+1}, u, \varphi, s\right)$.

2. Application to simulated data

To assess the efficiency and the functionality of the method in practice and how it can be generalized to field and experimental datasets, we tested it on *pods* in which the true values of *c* and all other studied evolutionary forces are known parameters that were used for simulations. Figure 1 illustrates the approach we used to test the method and summarizes the demo-genetic parameters of *pods*.

To test the reliability of our method, we chose basic population features (panmictic sexuality, 1000 individuals, mutation rate 10^{-3} under KAM) and basic parameters sets (two samples of 40 individuals at one generation interval, 10 loci, 4 alleles per locus, independence between loci) which are conventionally used to study empirical populations. We assessed the accuracy of the method as the proportion of times our method inferred the true *c* as the best posterior probability among 100 replicated *pods* and by providing the 90% precision intervals (5% excluded on both sides of the mode), *i.e*. intervals containing 90% of the maximum a posteriori inferences over 100 replicates.

For each *pods*, genotypes at each locus were drawn by randomly associating alleles (frequency of each allele: $1/n_a$, genotypes at Hardy-Weinberg proportions, *i.e.* maximal genetic diversity possible under panmixia) to constitute the initial population. Ancestral and descendent datasets in *pods* were saved at generations 50 and 51, respectively. Those conditions were chosen so that our tests are conservative. Indeed, in those conditions, the number of genotype frequencies to be estimated per sampled population is maximal which increases the average and total sampling error on genotype frequencies, and thus on estimated genotypic transitions. Moreover, we cannot expect to have reached the equilibrium which would compromise the use of genetic parameters to estimate *c*. To assess the robustness of the method (*i.e.* its ability to perform well even if its mechanistic and mathematical assumptions are somewhat violated) we then tested our method with *pods* in which we altered, one by one: *i)* the population properties: increasing the genetic drift ($N = 100$), changing mutation rates ($\mu = 10^{-3}, 10^{-9}$), introducing inbreeding ($\varphi = 0.1, 0.5, 0.9$) and selfing ($s = 0.1, 0.5, 0.9$) during sexual events; *ii)* the dataset used to infer rates of clonality: changing the

sampling effort in both temporal samples ($N_t = N_{t+1} = 20, 30, 40, 50, 100, 500$), increasing the number of generations between ancestral and descendent samples ($t+2, t+3, t+4, t+5$); *iii)* the marker properties: changing the mutation model (IAM and SMM), increasing the number of loci ($m = 50, 100, 500$), changing the number of alleles per loci ($n = 3, 4, 5, 10$), introducing physical linkage between loci ($Pr = 0.45, 0.25, 0.05$), random genotyping errors ($\mu = 1\%, 10\%$) and null allele frequency ($f(na) = 1\%, 2\%, 5\%, 10\%, 25\%$). In our study, by "null allele" we mean the missing data due to non-amplified loci (short allele dominance, large allele dropout and null allele *sensu stricto*). All *pods* were simulated using 15 rates of clonality ($c = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.98, 0.99, 0.999, 1$).

Among those factors, some can lead to difficulties in accurately identifying parameters, and thus should be dealt with increased caution at the time of providing assumed values and interpreting data. First, mutation and sexuality with inbreeding or selfing must be provided as given rates to the method. We thus analyzed the consequences of providing erroneous quantitative values, *i.e.* which would seriously depart from the real parameters in the population being analyzed. To this end, assumed mutation rates were provided as widely under or overestimated (by a factor $\pm 10^6$), assumed inbreeding rates as under or overestimated by ±20% and assumed selfing rates as under or overestimated by ±10%. Second, depending on the type of markers, different mutation models may apply and empirically determining the most relevant mutation model is challenging (Putman & Carbone 2014). Thus, we tested the robustness of inferences of $\tilde{c}$ when *pods* evolved under two extreme alternative mutation models, Infinite-Alleles (IAM) and Stepwise-Mutation (SMM) models. In SMM, the rate of mutation from one allele size to its two possible other sizes becomes $\mu = 0.5u$. Under SMM, the ancestral genotype frequencies are only impacted by mutation, as follows:

$$\begin{cases} r'_{A_iA_i} = \nu^2 . r_{A_iA_i} + \mu^2 \sum_{j=i\pm1} r_{A_jA_j} + \mu . \nu . \sum_{j=i\pm1} r_{A_iA_j} \\ r'_{A_iA_j} = \nu^2 . r_{A_iA_j} + 2.\mu^2 . \sum_{k=i\pm1, j\pm1} r_{A_kA_k} + \mu^2 . \sum_{k=i\pm1, l=j\pm1, k=l\pm2} r_{A_kA_l} + \mu . \nu \left(\sum_{k=j\pm1, i\neq j\pm1} r_{A_iA_k} + \sum_{k=i\pm1, j\neq i\pm1} r_{A_jA_k} \right) + \mu^2 \left(\sum_{k=j\pm1, i=j\pm1} r_{A_iA_k} + \sum_{k=i\pm1, j=i\pm1} r_{A_jA_k} \right) \end{cases}$$

eq.6

Where $\omega = \begin{cases} 2.\mu.\nu\left(r_{A_iA_i} + r_{A_jA_j}\right) & \text{if i} = \text{j} \pm 1 \\ 0 & \text{if i} \neq \text{j} \pm 1 \end{cases}$

In IAM, every time one mutation occurred, one new allele was created.

Third, many current datasets, especially those obtained using high-throughput sequencing, may involve the existence of physically linked markers or a standard background linkage disequilibrium due to certain technical or biological features other than clonality. We thus tested the reliability of our method when increasing the degree of dependence among loci used for genotyping individuals. We computed *pods* with reduced probability of recombination between markers, Pr. In previous *pods*, loci were simulated as fully independent markers, implying $Pr = {}^{1}/_{2}$. The interest of Pr is that it can be approximately linked to d (Haldane 1919), the genetic distance between loci in Morgan or map units using Haldane's map function: $Pr = 0.5.\left(1 - e^{-2d}\right)$ (Haldane 1919) and thus easily interpreted for applied datasets.

Finally, as for all methods using genetic markers, improper and biased observations like the ones caused by null alleles could result in biased inferences. We thus tested the influence of null alleles by generating pods where loci contained 1, 2, 5, 10 and 25% of null alleles that match the classical distinction between negligible, moderate and high frequency of null alleles (Chapuis & Estoup 2007).

3. Application to empirical datasets combining multiple deviations from model constraints

To assess both the reliability and feasibility of our method in complex natural and realistic situations, we used eleven genotypic transitions derived from genetic data on five partially clonal species, for which the prevalence of clonality was previously estimated using both genetic interpretations and direct field measures (Table 1). These tests are also the opportunity to present real cases of the analysis output and the methodology to interpret them. When analyzing one population, the method provides a distribution of posterior probabilities of possible $\tilde{c}$ values. Classically in Bayesian inference, users must identify the mode of the posterior distribution as the best inferred $\tilde{c}$ with the

maximum a posteriori probability, and delimit a credible interval, *i.e.* typically the range of *c* values that show higher posterior probabilities than flat/uninformative prior.

The tested species are the cultivated seaweed Rhodophyta *Gracilaria chilensis*, the engineer marine monocot plant *Zostera marina*, the scattered long-lived dicot tree *Prunus avium*, the pathogenic fungus *Melampsora larici-populina* and the cyclical parthenogenetic insect *Rhopalosiphum padi*. This applied set of species was selected because they are distributed all over the tree of life, present very distinct modes of clonal reproduction (Table 1) and ecological interests, and have the advantage to be otherwise intensively studied. Together, they also encompass a broad range of *c*, from very low to very high values (Table 1). Those datasets were based on very different numbers of microsatellite markers and levels of polymorphism. In addition, such data from the "real world" do not perfectly respect all model's assumptions, and presents additional biological features that were not included in our population genetics model (*e.g.* long overlapping generations as in *P. avium* and *Z. marina*). Applying our method on such diverse sets of data and species for which rates of clonality were estimated by other means, aimed at providing a fair appraisal of its versatility.

4. Comparison with CloNcaSe (Ali et al. 2016)

The CloNcaSe method was very recently published, and is, to our knowledge, the only other method aiming at quantifying the amount of sexual reproduction versus the amount of clonal reproduction within partially clonal populations using genetic information. It was initially extended from the equations of Burt *et al.* (1996) and developed for cyclical parthenogenetic populations (N1 and N2 numbers of events of pure clonal reproduction, separated by one event of mixed sexual/asexual reproduction), and relies on the probability of resampling repeated genotypes within and across generations to infer the proportion of sexual vs asexual generations. Ali *et al.* (2016) also recommend its use to quantitatively infer the proportion of sexual versus clonal reproduction in other modes of clonality. To assess the accuracy of our method and compare it to the only available

method dedicated to quantitatively infer the rates of clonality, we thus applied the CloNcaSe R package to 40 replicated pods representing fourteen rates of clonality ($c = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.98, 0.99, 1$). Those pods were obtained in population sizes of 100 and 1000 individuals (not sub-sampled to provide the best possible conditions ClonCaSe requires). Each individual was simulated with 100 independent loci of 10 alleles (KAM) that avoid obtaining low probabilities of identity, as expected in classical population genetics studies with about ten microsatellites loci with few alleles. Low probabilities of identity may create confusing repeated genotypes despite the fact that they would not share the same ancestral clonal lineage. Neither null allele, nor physical linkage between markers, nor selfing nor inbreeding were integrated in the simulations, thus none of those interfering parameters should affect the genetic and genotypic composition of the *pods*. *Pods* were saved at generations 50 and 51 as when testing ClonEstiMate. We provided the method with the known number of fully clonal cycles before and after the mixed (*i.e.* clonal-sexual) cycle between samples, setting both parameters respectively to N1=0 and N2=0. As equations in CloNcaSe do not take mutations into account, we only tested CloNcaSe for simulation with our lowest mutation rate, *i.e.* 10^{-9}, which over one generation in a population of N=100 matches CloNcaSe constraint. We also applied CloNcaSe R package to the 15 real data sets we used to test ClonEstiMate, providing the right parameters N1 and N2, *i.e.* 0 for all species apart for the aphid *R. padi* that truly reproduced using cyclical parthenogenesis. Data is available from the Dryad Digital Repository http://dx.doi.org/10.5061/dryad.32qh8 (Becheler *et al.* 2017).

RESULTS & DISCUSSION

1. Inferences from simulated data

Model results in general conditions

Raw temporal distributions of genotype frequencies within populations allowed inferring the true rates of clonality *c* within populations, even at low rates of clonality, while using standard minimal population genetics datasets. Indeed, only 40 sampled individuals per temporal sample genotyped with 10 low polymorphic loci (max. 4 alleles per loci) were sufficient to accurately infer the true *c* values with the maximum *a posteriori* probability estimates in 69 percent of cases on average, and 100% of the maximum *a posteriori* probabilities inferred *c* values in a precision interval of ±0.1 around their true values (Figure 2). In the rare cases where the maximum *a posteriori* over- or under-estimated *c* values, the posterior probability distributions always included the true *c* values *i)* with similar, even if slightly lower, posterior probabilities, *ii)* as the second most probable point estimates of the posterior distributions, whatever the true rates of clonality, even if the posterior probability distributions were more spread than when the true *c* values were inferred (Figure 3).

Improved performances when increasing datasets (individuals and loci) and polymorphism

Increasing the sample sizes from 20 to 30, 40, 50, 100, 250 and 500 sampled individuals increased the accuracy and the precision of the method to infer the true *c* values (Figure SI1). It also increased the peakedness of the posterior probability distributions (Figure SI2). However, sampling only 20 individuals tended to overestimate *c* (Figure SI1).

Increasing the quantity of genotyped loci, *i.e.* the number of genotypic transition, logically increased the accuracy of our method and the peakedness of the posterior probability distributions (Figure SI3 and Figure 4).

Interestingly, increasing the number of alleles per loci increased the accuracy of inferences up to a certain threshold (Figure SI4 and SI5), above which biased inferences occurred for the lowest rates of clonality. This threshold is defined by the balance between sample size and the number of alleles per locus that gives the number of possible genotypes whose frequency has to be estimated. With only 40 individuals sampled, increasing the number of alleles per locus from 2 to 4 alleles increased the accuracy of inferences while using loci with 5 to 10 alleles (and more) biased inferences toward an overestimate of *c* when the true rates were low. Indeed, in situations where loci can have more possible genotypes than the number of individuals sampled, insufficient sampling would tend to give more weight to the more frequent genotypes and miss the less frequent ones in both temporal samples. New, therefore less frequent, genotypes are expected to mainly appear through sexuality, thus insufficient sampling minimizing changes in genotypic transition would lead to an overestimate of *c*.

Limited impacts of drift, mutation and linkage among loci on inferences

Performance of our model was maintained for populations evolving under higher genetic drift ($N = 100$, Figure 2) and mutating at lower rate ($\mu = 10^{-9}$, Figure 2). As expected, increasing the genetic drift increased the stochastic changes of genotype frequencies over one generation but, ultimately, only marginally increased the width of the interval of inferences (and thus the precision of estimates).

The method seemed robust to infer true *c* values when populations mutated with different mutation models (Figure SI6). It also appeared to be robust to erroneous assumed rates of mutation (Figure SI7). Given values exceeding or underestimating the true mutation rate by six orders of magnitude ($\mu = 10^{-3} \leftrightarrow 10^{-9}$) only shifted the worst inferences to ±0.2 from the true *c* values, while the majority of inferences remained accurate resulting in the similar precision intervals as the ones found for the reference set based on correct values.

Despite the fact that our equations consider loci as independent, the method still inferred the true *c* values when the 10 loci were in strong physical linkage (Figure SI8). This robustness to linkage among markers is consistent with the fact that our inference method is based not on genetic states but on genotypic transitions, which properties definitely meet the naïve Bayes constraints (Hand & Yu 2001).

Impacts of sexuality occurring with inbreeding and selfing

The method still showed good performance to infer the true *c* values when populations reproduced sexually under inbreeding ($\varphi = 0.1, 0.5, 0.9$) or low to moderate rates of selfing ($s = 0.1, 0.5, 0.9$) when those rates were properly informed (Figures SI9 and SI10). It only showed difficulties to distinguish low rates of clonality from full sexuality at very high rates of selfing ($s \geq 0.9 \cap 0 \leq c \leq 0.4$) which resulted to underestimate *c* in those populations. However, the method seemed sensitive to erroneous information given on inbreeding and selfing rates (Figure SI11). Providing undervalued rates of inbreeding tended to result in the underestimation of *c* values in populations reproducing at low rates of clonality, and the other way around. It can be noted, however, that overvaluing the rate of inbreeding seemed to less bias inferences than undervaluing it. Interestingly, in selfing populations, providing either over- and under-valued selfing rates of ±0.1 all tended to lead to an overestimated *c* in populations experiencing low and moderate clonality ($0 \leq c \leq 0.8$).

Technical issues: impacts of random genotyping errors or migrant gametes, null alleles and erroneous timing in sampling

The performances of the method maintained when introducing disturbing factors fixed at 1% and 10%, such as random genotyping errors or migrant gametes between temporal samples (Figure SI12). Providing the method the appropriate estimates of disturbing factor rates (μ) was key for the method to remain as accurate and precise as using datasets without disturbance. The method seemed robust to infer true *c* values when providing underestimated rates of disturbing factors.

Underestimating the disturbing factor by two orders of magnitude (μ=10^{-3} instead of 10^{-1}) only marginally increased the width of the intervals of inferences.

Introducing null alleles in datasets resulted in an overestimation of *c* values in moderate clonal populations beyond null allele frequencies as low as 2% (Figure SI13). Indeed, null alleles lead to the spurious increase of homozygote genotypes and maintain them over genotypic transition against heterozygotes, which is one of the signatures of clonality (rough prediction: stable homozygote frequencies over time) against sexuality (rough prediction: homozygote frequencies converging toward Hardy-Weinberg expectations over time, Stoeckel & Masson 2014, Reichel *et al.* 2016).

Sampling the descendent populations after more than one generation resulted in an underestimation of the true *c* values (Figure SI14). Those underestimations increased both in frequency and in value as the number of generations between the ancestral and descendent samples increased. Indeed, an excess of changes in genotype frequencies due to the reshuffling of genotypes through more than one generation will be interpreted by the model as larger amounts of recombination. Yet, the bias was not evenly distributed along the spectrum of *c*: while negligible for extreme rates (low or high) of clonality because both respectively maintain stable Hardy-Weinberg proportions and stable ancestral proportions over time (Reichel *et al.* 2016), it is maximal for intermediate *c* values. This result confirms the fact that partial clonality has its own specific and nonlinear genetic dynamics. After only 2 generations, inferred *c* values were underestimated, but the bias was lower than ±0.3 from their true values for the worst inferences and of ~0.1 on average for $c \in [0.1 - 0.99]$, which still allowed a good appraisal of the rates of clonality. However, after 3, 4 or 5 generations, biases were prohibitive for $c \in [0.1 - 0.99]$ with strong risks to drastically underestimate the true *c* values.

2. Inferences from real data

Inferred *c* from real-world datasets were in line with previously proposed appraisals of the importance of clonality in those populations based on different knowledge of their biology (Table 2; Figure 4).

The cultivated seaweed G. chilensis. This haploid-diploid red algae is cultivated along the Chilean coasts (Buschmann *et al.* 2008). Its production in farm is essentially based on replanting cuttings of crops selected for diploidy and vegetative propagation (Guillemin *et al.* 2008; Guillemin *et al.* 2013). This human-induced clonality depressed significantly the level of genotypic diversity of farmed populations as compared to wild populations (Guillemin *et al.* 2008), suggesting a *c* close to 1. We used a dataset from a farmed population in Chile (Guillemin *et al. unpublished*), sampled in 2009 and 2010, in which the input of sexual recruits was estimated as more than enough negligible (absence of reproductive individuals in the farm and the haploid generation of the sexual life cycle is missing, Guillemin *et al. unpublished*). The method correctly inferred *c*=1, despite the low number of genetic markers used (5 microsatellites).

The engineer marine plant Z. marina. This dataset comes from a seagrass meadow in Brittany (France), sampled in 2009 and 2012. This species reproduces sexually only once a year through flowers that blossoms in spring. In 2012, a large clone was sampled, with at least 11m between two ramets, suggesting an active rhizomatous growth. Along these 3 years, the clonal richness decreased from 0.93 to 0.65 (Becheler *et al.* 2014). We thus expected intermediary rates of clonality for this species. Our method inferred *c*=0.5. Yet, as the number of generations between 2009 and 2012 is likely higher than one, considering the effects we identified on *pods* (Figure SI13), the true rate of clonality should thus be probably above 0.5.

The pathogenic fungus M. larici-populina. This fungal pathogen is responsible for foliar rust disease on poplar trees and is a typical example of obligate cyclical parthenogenesis (Xhaard *et al.* 2012), in which around ten rounds of clonal reproduction alternate once a year with a unique and synchronized event of sexual reproduction. Here we chose a population living in quasi-isolation upstream the Durance River (Xhaard *et al.* 2012), sampled twice at a year interval. In accordance with the obligate cyclical parthenogenetic life cycle, a single event of sex separates the two successive samplings. However, about 10 generations of clonal reproduction have occurred between parental and descended samplings. Accordingly, the model inferred c=0.2, highlighting the slight but significant occurrence of clonal reproduction, even though no genotype was shared between sampling periods.

The scattered long-lived tree P. avium. Wild cherry trees combine clonality through sprouting and sexual reproduction controlled by a gametophytic self-incompatibility system. Clonality was extensively estimated in the species using isozymes (Frascaria *et al.* 1993; Ducci & Santi 1997; Gömöry & Paule 2001) and microsatellites (Schueler *et al.* 2006; Stoeckel *et al.* 2006; Vaughan *et al.* 2007a; Vaughan *et al.* 2007b) leading to contrasting values of the G/N index. The parental population was composed by adult trees of different ages and fitnesses (Stoeckel *et al.* 2012). In 2002 and 2003, seeds were collected as the production of one parental perennial population. Groups of these seeds originated from a single mother (substructure of inheritance) and long-distance migration was detected due to insect pollination (Stoeckel *et al.* 2008; Lander *et al.* 2013; Stoeckel *et al.* 2012). In 2002 and 2003, the expected c was 0 but our method inferred a slightly higher value (c=0.1). The weak overestimation may be linked to the genetic substructure within the seed pools. In 2004, a population of wild cherry seedlings was collected on the ground as young shoot leaves. These seedlings mainly originated from seeds (visible at roots) and more marginally from vegetative propagation (phalanx strategy from seeds and sprouting). Our method inferred c=0.2, a slightly higher value than found in seeds which is thus highly consistent with field information.

The parasitic insect R. padi. Aphids have complex mechanisms of coexistence and inheritance of sexuality and clonality in populations (Simon *et al.* 2010; Jaquiery *et al.* 2014). In *R. padi*, two kinds of lineages coexist. On the one hand, sexual lineages reproduce through obligate cyclical parthenogenesis and have a full commitment to the production of sexual forms during the studied period. On the other hand, facultative asexual lineages have a mixed investment in the production of both sexual and parthenogenetic forms, and hence survive clonally from year to year (Halkett *et al.* 2005). In this system, the sexual forms migrate to a secondary host where only sexual reproduction takes place, which enables to collect them specifically. A previous study, Halkett *et al.* (2006) drew the first assessment of the rate of clonality of a given population. The abundance of both kinds of lineages and their contribution to the reproductive effort were monitored for 11 weeks. The results show that the abundance of sexual lineages collapsed after week 5 (Halkett *et al.* 2006). We thus expect an increase of the inferred rates of clonality from low to high levels as time elapsed. The instantaneous variations of the rate of clonality perfectly match the field observations from Halkett *et al.* (2006). During the first weeks of survey, clonality was medium (between 0.3 and 0.5). We then observe a clear shift toward higher values of *c*. At the end of the survey, the model indeed inferred strong rate of clonality between 0.8 and 0.95. Those last estimations are consistent with the fact that clonal lineages are in fact facultative sexual and produce under stress some sexual descendants among a majority of clonal progeny.

3. Comparison with CloNcaSe inferences

CloNcaSe (Ali *et al.* 2016) was the first published method allowing quantitative estimates of the rate of sex versus clonality within populations subjected to cyclical parthenogenesis. To summarize briefly, this method used the probability to observe repeated genotypes as *a proxy* for rates of clonality. Additionally, it provides estimate for population size. Like our method, ClonNcaSe requires temporal samples and users have to provide information on the number of generations that separates samples.

We tested CloNCase estimates on the whole simulated populations containing 100 and 1000 individual genotyped with 100 independent loci. Results are provided in Table 2 (all details are available on Dryad repository DOI:10.5061/dryad.32qh8). Based on our *pods*, when $c \leq 0.9$, the CloNCase method assessed *c* with difficulty. It provided highly variable inferred effective sizes, *e.g.* ranging for example from 3 to $3*10^{26}$ in populations of 100 individuals (Table 2). The discrepancies between true and inferred values of both *c* and *Ne* may come from the fact that CloNcaSe estimates clonality from repeated genotypes within and/or among generations, while ClonEstiMate infers clonality using the likelihoods of changes in genotype frequencies over generations. CloNcaSe was formalized from the observed distribution and life cycle of *Puccinia striiformis*, a cyclical parthenogenetic species which mostly reproduce using full clonality in expanding populations and sporadically using few synchronized sexual events (Ali *et al.* 2016), thus using concepts that may appear to be true for high clonal species but are not generally applicable to other partially clonal species. Even if trivial, it is important to remember that clonality does not equal the rate of occurrence of repeated genotypes G/N (Arnaud-Haond *et al.* 2007), and depending on sampling density many genotypes would appear unique while in reality they are not. For example, if one ancestor provides one clonal and one sexual descendant per generation on average, it is unlikely to find many repeated genotypes due to clonality even though the population rate of clonality is indeed 50%. Only the accumulation of stochastic variations along generations in clonal production can lead to the detection, after many generations, of a power-law (Pareto) distribution of the number of ramets within genets that would indeed indicate the significant occurrence of clonal reproduction. Yet, even in such case, inference methods based on repeated genotypes would likely miss the background *c*=50% of the mass of genets with no detected repeated ramets.

Moreover, repeated genotypes in species with life cycle similar to *Puccinia striiformis* will be still hard to observe with samples of realistic size when taken from very large populations (*i.e*. blooming planktonic populations of marine unicellular dinoflagellates, Dia *et al.* 2014; extended sea grass meadows, Arnaud-Haond *et al.* 2007). This limitation may explain why ClonCaSe mis-inferred more

sex than expected in our aphid dataset on the last sample (i.e. sample "week 5", Table 1). In fact, repeated genotypes can be caused by clonality and unequal production of clonal descents between genets but also in purely sexual populations by, for example, twinning, inbreeding or selfing. However, ClonCaSe, if applicable, should perform better than ClonEstiMate on datasets with null alleles because repeated genotypes can be expected to be evenly impacted by such technical artefact.

4. Methodological recommendations on sampling and marker properties for precise estimates

Identifying the way genes are transmitted in space and time in natural populations and obtaining accurate estimates of the relative importance of alternative reproductive modes such as sex and clonality is crucial for understanding the ecology and evolution of species (Duminil *et al.* 2007, Fehrer 2010). The method presented here is the first comprehensive attempt to pave the way for a reliable estimate of the clonality rate, despite some limitations and recommendations to which users should be aware.

In terms of sampling size, we recommend all users sample at least 30 individuals per population and sample period, and to target samples of 50 individuals to ensure confident estimates of genotype frequencies. Those sample sizes are congruent with classical standard recommendations for population genetic studies (Kalinowski 2005, Hale *et al.* 2012, Fung & Keenan 2014). Inference of *c* values appears acceptable when up to two generations separate putative ancestral and descendent populations but were seriously biased for wider time lag. We therefore recommend users to respect as much as possible the modelling assumption that only one generation separates the ancestral and descendant populations. This objective can be addressed using basic naturalist approaches, either *i)* by classifying individuals by age using adapted biometrics or physiological methods, or *ii)* to sample the descendent populations after one average generation-time or after a time equivalent to one average breeding-season. If no naturalistic or only vague information is available to disentangle

generations, users must be aware that by using samples separated by more than one generation, they risk to seriously overestimate sex when rates of clonality are intermediate. In case this is suspected, the inferred $\tilde{c}$ has to be interpreted as a minimal rate of clonality. As detailed in equations, our method does not necessarily need a direct parentage between the genotyped individuals of the two temporal samples. Both samples rather have to give a representative picture of the genotype frequencies within both temporal populations at all genotyped loci.

About ten loci with 4 alleles seem sufficient to get accurate inferences, even if more markers enhance the peakedness of the posterior probability distribution on its best inference. Loci do not need to be physically independent on genomes. However, users should notice that, in case of high linkage, like in sequences or for physically close SNPs, the naïve Bayes approach would yield to infer the correct *c* but with misleading shapes of the posterior distributions. Expectedly when the same information is spuriously multiplicated, posterior probability distribution will appear peakier than they should be if conditional likelihoods over all loci had been explicitly taken into account because of physical linkage. Our analyses on mutation models showed that our method can be used with different types of markers as long as they are codominant (*e.g*. microsatellites and SNP). Providing the method erroneous rates of disturbing factor (which embed mutation, migrant gametes and random genotyping errors) did not seem to affect much the robustness of estimates. However, loci with more than 1% of null allele should be discarded as they tend to induce an overestimation of the rates of clonality.

Locus polymorphism must be adjusted to the sample sizes to optimize the precision of estimates. Indeed, the number of possible genotypes ($N_{genotypes}$) rapidly increases with the number of alleles (n_A) $N_{genotypes} = \frac{n_A.(n_A+1)}{2}$ (*e.g.* Reichel *et al.* 2015). From our current results, we empirically recommend ensuring that the sample size of both temporal samples exceeds at least twice the possible number of different genotypes at each locus. Markers with few alleles, like SNP and many microsatellites with fewer than 5 alleles, should be preferentially selected, especially when using low sample size ($N_{sample} \leq 50$), as they would ensure more precise estimates of genotype frequencies

than highly polymorphic ones. Since the consequences of inbreeding, selfing and partial clonality cannot be disentangled using genotypic transitions to the date, we recommend independently assessing inbreeding and selfing rates, when occurring, before using our method, to provide it with the best assumed values.

Our results analyzed the behavior of the method in various conditions using distributions of maximum *a posteriori* inferences and precision intervals. Future studies will mostly have to interpret one *posterior* probability distribution *per* population and only one maximum *a posteriori*. We propose future studies to report inferences as typical Bayesian credible intervals, *i.e.* ranges of *c* values that would have higher posterior probabilities than a non-informative prior. We showed that, in the rare but existing cases of mis-inference, posterior probability distributions were flatter than usual but always included the true *c* value (*i.e*. true *c* value generally had a posterior probability similar to its maximum, or was ranked as the second-highest probabilities). Finally, we also recommend discussing and interpreting the credible interval of inferences provided by ClonEstiMate in the light of *i)* biological knowledge available, *ii)* inferences obtained from other methods, currently CloNcaSe (Ali *et al.* 2016) and qualitative classes considering that studied populations are at equilibrium proposed by De Meeus *et al.* (2007).

CONCLUSION

We proposed a Bayesian method based on the likelihoods of transitions of genotype frequencies between two samples, one ancestral and one descendent, which provide accurate quantitative inference of clonality in populations. This method works even in the absence of equilibrium between drift and mutation and rates of clonality, which is of primary importance for some combinations of those evolutionary forces (Reichel *et al.* 2016). This method proved to be efficient and functional even when using few makers (*i.e.* ten microsatellites) or physically linked markers, and when using reduced sampling effort. It seems robust when providing misleading assumed mutation rate or when applied on genetic markers that do not mutate under the specified KAM but remains sensitive to

erroneous assumed values of inbreeding and selfing rates, null alleles and sampling time interval greater than two generations. Genomes, SNP or microsatellites data can thus be used and even mixed when analyzed with this method. We believe that, living in a world where clonality is ubiquitous, our method will be useful for population geneticists dealing with partially clonal species, to make rational quantitative interpretations of genetics data. Its performance to estimate low rates of clonality enables for the first time the possibility for future studies to screen natural populations, at low cost, for the natural occurrence of suspected but infrequent clonal reproduction events (*e.g*. vertebrates: Avise 2015, Dudgeon *et al.* 2017).

Acknowledgements

We wish to thank Stéphane De Mita and Nicolas Parisey for their valuable comments on earlier versions of this manuscript. We are also grateful to the three referees whose highly relevant comments deeply improved the manuscript. This work was supported by the French National Research Agency, CLONIX project (ANR-11-BSV7-0007). *Gracilaria chilensis* dataset gathering was supported by CONICYT FONDECYT/REGULAR N° 1130868 and 1170541 to MLG.

References

Ali S, Soubeyrand S, Gladieux P*, et al.* (2016) CloNcaSe: Estimation of sex frequency and effective population size by clonemate resampling in partially clonal organisms. *Molecular Ecology Resources* **16**, 845-861.

Allen DE, Lynch M (2012) The effect of variable frequency of sexual reproduction on the genetic structure of natural populations of a cyclical parthenogen. *Evolution* **66**, 919-926.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology* **16**, 5115-5139.

Avise J (2008) *Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals* Oxford University Press.

Avise JC (2015) Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8867-8873.

Becheler R, Benkara E, Moalic Y, Hily C, Arnaud-Haond S (2014) Scaling of processes shaping the clonal dynamics and genetic mosaic of seagrasses through temporal genetic monitoring. *Heredity* **112**, 114-121.

Becheler R, Cassone A-L, Noël P*, et al.* Low incidence of clonality in cold water corals revealed through the novel use of a standardized protocol adapted to deep sea sampling. *Deep Sea Research Part II: Topical Studies in Oceanography*.

Becheler R, Diekmann O, Hily C, Moalic Y, Arnaud-Haond S (2010) The concept of population in clonal organisms: mosaics of temporally colonized patches are forming highly diverse meadows of Zostera marina in Brittany. *Molecular Ecology* **19**, 2394-2407.

Becheler R, Masson J, Arnaud-Haond S, Halkett F, Mariette S, Guillemin M, Valero M, Destombe C, Stoeckel S (2017). Data from: ClonEstiMate, a Bayesian method for quantifying rates of

clonality of populations genotyped at two-time steps. *Dryad Digital Repository*. doi:10.5061/dryad.32qh8

Berg LM, Lascoux M (2000) Neutral genetic differentiation in an island model with cyclical parthenogenesis. *Journal of Evolutionary Biology* **13**, 488-494.

Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. In: *Annual Review of Ecology Evolution and Systematics*, pp. 193-216.

Buschmann AH, Hernandez-Gonzalez MD, Varela D (2008) Seaweed future cultivation in Chile: perspectives and challenges. *International Journal of Environment and Pollution* **33**, 432-456.

Cockerham CC (1984) Drift and mutation with a finite number of allelic states. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**, 530-534.

Cornelissen JHC, Song YB, Yu FH, Dong M (2014) Plant traits and ecosystem effects of clonality: a new research agenda. *Annals of Botany* **114**, 369-376.

De Meeus T, Balloux F (2005) F-statistics of clonal diploids structured in numerous demes. *Molecular Ecology* **14**, 2695-2702.

De Meeus T, Prugnolle F, Agnew P (2007) Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences* **64**, 1355-1372.

Dia A, Guillou L, Mauger S, Bigeard E, Marie D, Valero M & Destombe C (2014). Spatiotemporal changes in the genetic diversity of harmful algal blooms caused by the toxic dinoflagellate *Alexandrium minutum*. *Molecular ecology*, **23**(3), 549-560.

Do C, Waples RS, Peel D*, et al.* (2014) NeESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (N-e) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

Dorken ME, Eckert CG (2001) Severely reduced sexual reproduction in northern populations of a clonal plant, Decodon verticillatus (Lythraceae). *Journal of Ecology*, **89**, 339–350.

Dudgeon CL, Coulton L, Bone R, Ovenden JR, Thomas S (2017). Switch from sexual to parthenogenetic reproduction in a zebra shark. *Scientific Reports* **7**, 40537.

Dufresne F, Stift M, Vergilino R, Mable BK (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**, 40-69.

Duminil J, Fineschi S, Hampe A*, et al.* (2007) Can population genetic structure be predicted from life-history traits? *American Naturalist* **169**, 662-672.

Eckert CG, Dorken ME, Barrett SCH (2016) Ecological and evolutionary consequences of sexual and clonal reproduction in aquatic plants. *Aquatic Botany* **135**, 46-61.

Fehrer J (2010) Unraveling the mysteries of reproduction. *Heredity* **104**, 421-422.

Franco-Trecu V, Costa-Urrutia P, Schramm Y, Tassino B, Inchausti P (2015) Tide line versus internal pools: mating system and breeding success of South American sea lion males. *Behavioral Ecology and Sociobiology* **69**, 1985-1996.

Frascaria N, Santi F, Gouyon PH (1993) Genetic differentiation within and among populations of chestnut (Castanea-sativa Mill) and wild cherry (Prunus-avium L). *Heredity* **70**, 634-641.

Fung T, Keenan K (2014) Confidence intervals for population allele frequencies: the general case of sampling from a finite diploid population of any size. *Plos One* **9**.

Gomory D, Paule L (2001) Spatial structure and mating system in wild cherry (Prunus avium) population. *Biologia* **56**, 117-123.

Guillemin ML, Faugeron S, Destombe C*, et al.* (2008) Genetic variation in wild and cultivated populations of the haploid diploid red alga Gracilaria chilensis: How farming practices favor asexual reproduction and heterozygosity. *Evolution* **62**, 1500-1519.

Guillemin ML, Sepulveda RD, Correa JA, Destombe C (2013) Differential ecological responses to environmental stress in the life history phases of the isomorphic red alga Gracilaria chilensis (Rhodophyta). *Journal of Applied Phycology* **25**, 215-224.

Haldane JBS (1919) The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299-309.

Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *Plos One* **7**.

Halkett F, Kindlmann P, Plantegenest M, Sunnucks P, Simon JC (2006) Temporal differentiation and spatial coexistence of sexual and facultative asexual lineages of an aphid species at mating sites. *Journal of Evolutionary Biology* **19**, 809-815.

Halkett F, Plantegenest M, Prunier-Leterme N*, et al.* (2005a) Admixed sexual and facultatively asexual aphid lineages at mating sites. *Molecular Ecology* **14**, 325-336.

Halkett F, Simon JC, Balloux F (2005b) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution* **20**, 194-201.

Hand DJ, Yu KM (2001) Idiot's Bayes - Not so stupid after all? *International Statistical Review* **69**, 385-398.

Jaquiery J, Stoeckel S, Larose C*, et al.* (2014) Genetic control of contagious asexuality in the pea aphid. *Plos Genetics* **10**.

Jones LE, Becks L, Ellner SP*, et al.* (2009) Rapid contemporary evolution and clonal food web dynamics. *Philosophical Transactions of the Royal Society B-Biological Sciences* **364**, 1579-1591.

Judson OP, Normark BB (1996) Ancient asexual scandals. *Trends in Ecology & Evolution* **11**, A41-A46.

Kalinowski ST (2005) Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* **94**, 33-36.

Lander TA, Klein EK, Stoeckel S*, et al.* (2013) Interpreting realized pollen flow in terms of pollinator travel paths and land-use resistance in heterogeneous landscapes. *Landscape Ecology* **28**, 1769-1783.

Marshall DR, Weir BS (1979) Maintenance of genetic variation in apomictic plant populations. *Heredity*, **42,** 159–72.

McKey D, Elias M, Pujol B, Duputie A (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* **186**, 318-332.

Navascues M, Stoeckel S, Mariette S (2010) Genetic diversity and fitness in small populations of partially asexual, self-incompatible plants. *Heredity* **104**, 482-492.

Neiman M, Meirmans S, Meirmans PG (2009) What can asexual lineage age tell us about the maintenance of sex? In: *Year in Evolutionary Biology 2009* (eds. Schlichting CD, Mousseau TA), pp. 185-200.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics* **34**, 401-437.

Pearson T, Okinaka RT, Foster JT, Keim P (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection Genetics and Evolution* **9**, 1010-1019.

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* **4**, 4399-4428.

Reichel K, Bahier V, Midoux C*, et al.* (2015) Interpretation and approximation tools for big, dense Markov chain transition matrices in population genetics. *Algorithms for Molecular Biology* **10**.

Reichel K, Masson JP, Malrieu F, Arnaud-Haond S, Stoeckel S (2016) Rare sex or out of reach equilibrium? The dynamics of F-IS in partially clonal organisms. *Bmc Genetics* **17**.

Rouger R, Reichel K, Malrieu F, Masson JP, Stoeckel S (2016) Effects of complex life cycles on genetic diversity: cyclical parthenogenesis. *Heredity* **117**, 336-347.

Schon I, Martens K, VanDijk P (2009) *Lost Sex: The Evolutionary Biology of Parthenogenesis* Springer, The Netherlands.

Schueler S, Tusch A, Scholz F (2006) Comparative analysis of the within-population genetic structure in wild cherry (Prunus avium L.) at the self-incompatibility locus and nuclear microsatellites. *Molecular Ecology* **15**, 3231-3243.

Scrosati R (1998) Population structure and dynamics of the clonal alga Mazzaella cornucopiae (Rhodophyta, Gigartinaceae) from Barkley Sound, Pacific coast of Canada. *Botanica Marina* **41**, 483-493.

Simon JC, Stoeckel S, Tagu D (2010) Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, **333**, 488–496.

Sinervo A, Zamudio KR (2001) The evolution of alternative reproductive strategies: Fitness differential, heritability, and genetic correlation between the sexes. *Journal of Heredity* **92**, 198-205.

Sintes T, Marba N, Duarte CM, Kendrick GA (2005). Nonlinear processes in seagrass colonisation explained by simple clonal growth rules. *Oikos*, **108**(1), 165-175.

Stoeckel S, Castric V, Mariette S, Vekemans X (2008) Unequal allelic frequencies at the self-incompatibility locus within local populations of *Prunus avium* L.: an effect of population structure? *Journal of Evolutionary Biology* **21**, 889-899.

Stoeckel S, Grange J, Fernandez-Manjarres JF*, et al.* (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species - Prunus avium L. *Molecular Ecology* **15**, 2109-2118.

Stoeckel S, Klein EK, Oddou-Muratorio S, Musch B, Mariette S (2012) Microevolution of S-allele frequencies in wild cherry populations: respective impacts of negative frequency dependent selection and genetic drift. *Evolution* **66**, 486-504.

Stoeckel S, Masson JP (2014) The exact distributions of F-IS under partial asexuality in small finite populations with mutation. *Plos One* **9**.

Sunnucks P, England PR, Taylor AC, Hales DF (1996) Microsatellite and chromosome evolution of parthenogenetic Sitobion aphids in Australia. *Genetics* **144**, 747-756.

Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE (2015) Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8901-8908.

Tibayrenc M, Ayala FJ (2012) Reproductive clonality of pathogens: A perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3305-E3313.

Tibayrenc M, Kjellberg F, Ayala FJ (1990) A clonal theory of parasitic protozoa - the population structures of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas, and Trypanosoma and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2414-2418.

Vallejo-Marin M, Dorken ME, Barrett SCH (2010) The ecological and evolutionary consequences of clonality for plant mating. In: *Annual Review of Ecology, Evolution, and Systematics, Vol 41* (eds. Futuyma DJ, Shafer HB, Simberloff D), pp. 193-213.

Vaughan SP, Cottrell JE, Moodley DJ, Connolly T, Russell K (2007) Clonal structure and recruitment in British wild cherry (Prunus avium L.). *Forest Ecology and Management* **242**, 419-430.

Villate L, Esmenjaud D, Van Helden M, Stoeckel S, Plantard O (2010) Genetic signature of amphimixis allows for the detection and fine scale localization of sexual reproduction events in a mainly parthenogenetic nematode. *Molecular Ecology* **19**, 856-873.

Wang JL, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**, 429-446.

Waples RS (1989) A generalized-approach for estimating effective population-size from temporal changes in allele frequency. *Genetics* **121**, 379-391.

Webb GI, Boughton JR, Wang ZH (2005) Not so naive Bayes: Aggregating one-dependence estimators. *Machine Learning* **58**, 5-24.

Wiedenmann J, Kraus P, Funke W, Vogel W (2000) The relationship between different morphs of Anemonia aff. sulcata evaluated by DNA fingerprinting (Anthozoa, Actinaria). *Ophelia* **52**, 57-64.

Wright S (1931) Evolution in mendelian populations. *Genetics* **16**, 97–159.

Xhaard C, Barres B, Andrieux A*, et al.* (2012) Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. *Molecular Ecology* **21**, 2383-2398.

Zhang H (2004) The Optimality of Naive Bayes, 562–567 in Proceedings of the 17th international Florida Artificial Intelligence Research society conference, edited by V. Barr and Z. Markov. American Association for Artificial Intelligence, Miami Beach, FL.

Data accessibility

The method is available at http://www6.rennes.inra.fr/igepp_eng/Productions/Software as binaries for Windows and GNU/Linux operating systems. It has no restrictions on the number of markers that can be analyzed and should work in minutes on all current laptop and desktop computers, without restriction other than computation time that will depend on the number of markers and of populations analyzed. Data files must be formatted as in the example files and outputs can be read as explained in the user guide.

Data of *1)* the cross-validation of ClonEstiMate using *pods*, *2)* the distributions of posterior probabilities of *c* values provided by ClonEstiMate when true *c*=0.4 and 3) the estimations of *c* by CloNcaSe, are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.32qh8

Tables

Table 1 Basic information about the five partially clonal species included in this analysis. R corresponds to the clonal richness, assessed for each time-step (R = [G-1]/[N-1], G and N being the number of distinct clonal lineages and the number of sampling units, respectively). All of these species were genotyped with microsatellite markers. Genotyping errors, inbreeding and selfing were assumed null. Genotypic datasets are available on the Dryad repositories indicated on each reference papers (*Z. marina*: populations of Sainte-Marguerite. *G. chilensis*: population of Chaica. *M. larici-populina*: populations of Prelles. *P. avium*: populations of Saint-Gobain. *R. padi*, population of the tree B).

Species	Type of partial clonality	reference	number of markers	sampling date	sample size *N*	clonal richness *R*	reproductive rhythm	expected rate of clonality	ClonEstiMate inferred *c*	CloNcaSe		
										Ne	*s*	*c=1-s*
Marine phanerogam (*Zostera marina*)	rhizomatic elongation	Bechelet *et al.* 2014	9	2009	30	0.93	annual	medium	0.5	66	1.00	0.00
				2012	30	0.65						
Red alga (*Gracilaria chilensis*)	fragmentation	Guillemin *et al.* Unpublished data	5	2009	26	0.20	annual	fully clonal	1	9.34E+08	0.18	0.82
				2010	27	0.23						
Poplar Rust (*Melampsora larici-populina*)	clonal sporulation	Xhaard *et al.* 2012	21	2008	24	1.00	annual	low	0.2	6.52E+65	1.00	0.00
				2009	42	1.00						
**Wild Cherry tree (*Prunus*	clonal sproutin g	Stoeckel *et al.*	7	2000	235	0.47	annual	null	0.1	5.95E+04	0.99	0.01
				2002	904	0.98						

***avium*)**		2006		2003	654	0.99		null	0.1	6.29E+04	0.46	0.54
				2004	154	0.94		low	0.2	1583	0.25	0.75
Aphid (*Rhopalosiphum padi*)	cyclical parthenogenesis	Halkett *et al.* 2006	8	3rd week	31	1.00	annual	increasing continuously from medium to high	0.5	20	0.25	0.75
				5th week	40	0.46			0.3	591	0.45	0.55
				8th week	47	0.61			0.8	2.68E+02	0.00	1.00
				9th week	17	0.69			0.95	5.00E+09	0.04	0.96
				10th week	19	0.67			0.9	1.78E+09	0.32	0.68
				11th week	10	0.89						

Table 2 Inferences from CloNcaSe, based on our standard simulated populations (for N=100 and an exhaustive sampling, 100 independent loci evolving under KAM, random mating). As the model of CloNcaSe does not take mutations into account, we provided only results based on populations simulated for our lowest mutation rate, *i.e.* 10-9. Additional CloNcaSe's outputs are available on Dryad (Doi:10.5061/dryad.32qh8). Averaged values were assessed through 40 repetitions for each combination of model's parameters.

c	***N***	***μ*/assumed *μ***	**rate of sex *s***			**deduced rate of clonality *c* = *1-s***			**averaged deviation from true c**
			Mean *s*	**SE *s***	**IC(95%)**	**mean *c***	**SE *c***	**IC(95%)**	
0	100	1e-09	0.00	1.28E-29	4.09E-30	1.00	1.12E-16	3.60E-17	1.00
0.1	100	1e-09	0.20	2.23E-01	7.13E-02	0.80	2.23E-01	7.13E-02	0.70
0.2	100	1e-09	0.18	1.92E-01	6.16E-02	0.82	1.92E-01	6.16E-02	0.62
0.3	100	1e-09	0.08	1.17E-01	3.74E-02	0.92	1.17E-01	3.74E-02	0.62
0.4	100	1e-09	0.08	1.05E-01	3.35E-02	0.92	1.05E-01	3.35E-02	0.52
0.5	100	1e-09	0.08	1.15E-01	3.67E-02	0.92	1.15E-01	3.67E-02	0.42
0.6	100	1e-09	0.05	6.64E-02	2.12E-02	0.95	6.64E-02	2.12E-02	0.35
0.7	100	1e-09	0.05	6.39E-02	2.04E-02	0.95	6.39E-02	2.04E-02	0.25
0.8	100	1e-09	0.03	5.28E-02	1.69E-02	0.97	5.28E-02	1.69E-02	0.17
0.9	100	1e-09	0.03	4.35E-02	1.39E-02	0.97	4.35E-02	1.39E-02	0.07
0.95	100	1e-09	0.03	4.01E-02	1.28E-02	0.97	4.01E-02	1.28E-02	0.02
0.98	100	1e-09	0.02	2.74E-02	8.77E-03	0.98	2.74E-02	8.77E-03	0.00
0.99	100	1e-09	0.02	2.92E-	9.34E-	0.98	2.92E-	9.34E-	-0.01

1	100	1e-09	0.01	2.10E-02	6.71E-03	0.99	2.10E-02	6.71E-03	-0.01
0	1000	1e-09	0.00	1.28E-29	1.27E-31	1.00	1.12E-16	1.11E-18	1.00
0.1	1000	1e-09	0.15	1.63E-01	1.61E-03	0.85	1.63E-01	1.61E-03	0.75
0.2	1000	1e-09	0.04	7.01E-02	6.95E-04	0.96	7.01E-02	6.95E-04	0.76
0.3	1000	1e-09	0.05	6.59E-02	6.53E-04	0.95	6.59E-02	6.53E-04	0.65
0.4	1000	1e-09	0.02	2.97E-02	2.94E-04	0.98	2.97E-02	2.94E-04	0.58
0.5	1000	1e-09	0.02	2.50E-02	2.48E-04	0.98	2.50E-02	2.48E-04	0.48
0.6	1000	1e-09	0.03	3.30E-02	3.27E-04	0.97	3.30E-02	3.27E-04	0.37
0.7	1000	1e-09	0.01	2.06E-02	2.04E-04	0.99	2.06E-02	2.04E-04	0.29
0.8	1000	1e-09	0.02	2.49E-02	2.47E-04	0.98	2.49E-02	2.47E-04	0.18
0.9	1000	1e-09	0.01	1.65E-02	1.64E-04	0.99	1.65E-02	1.64E-04	0.09
0.95	1000	1e-09	0.01	1.80E-02	1.79E-04	0.99	1.80E-02	1.79E-04	0.04
0.98	1000	1e-09	0.01	1.87E-02	1.85E-04	0.99	1.87E-02	1.85E-04	0.01
0.99	1000	1e-09	0.01	1.67E-02	1.66E-04	0.99	1.67E-02	1.66E-04	0.00
1	1000	1e-09	0.00	4.17E-03	4.13E-05	1.00	4.17E-03	4.13E-05	0.00

Legends of figures:

Figure 1 Summarized approach and notations of the model.

Figure 2 Relationships between real and inferred *c*, using pseudo-observed data generated in respect of the model's assumptions (*pods* of reference, see Material & Methods for details). Over 100 repetitions, the median values were provided for each tested rate of clonality (grey dots). Intervals of confidence at 90% (black dashed lines) were determined by excluding the 10 extremal inferred values of c (*i.e*. the 5 minimal and 5 maximal).

Figure 3 Posterior distribution of inferred rates of clonality in *pods* reproducing at c=0.4, in a population of 1000 individuals mutating at $\mu=10^{-3}$ in KAM, two samples of 40 individuals at one generation interval, 10 loci, 4 alleles per locus, physical independence between loci. Posterior distribution with black-filled points is the typical posterior distribution obtained in 56% of the replicates that were accurately inferred to be c=0.4; with grey-filled points, the typical posterior

distribution obtained in 23% of the replicates that were overinferred as *c*=0.5; with blank-filled points, the typical posterior distribution obtained in 21% of the replicates that were underinferred as *c*=0.3.

Figure 4 Posterior distributions of rates of clonality in *pods* reproducing at *c*=0.4 with an increasing number of simulated loci (from the darker to lighter shades of grey: 10, 50, 100 and 500 loci respectively).

Figure 5 Distributions of posterior probabilities assessed on the real datasets. On those eleven datasets from five very different species, maximum a posteriori probabilities identified rates of clonality consistent with the ones previously deciphered using both genetic interpretations and direct field measures (for details, see Table 1).