Received: 4 November 2017 | Accepted: 16 November 2017

DOI: 10.1111/2041-210X.12948

RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Sex, size and timing: Sampling design for reliable population genetics analyses using microsatellite data

Quentin Dubois[1] | Christophe Lebigre[1,2] | Nicolas Schtickzelle[1] | Camille Turlure[1]

[1]Biodiversity Research Centre, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

[2]IFREMER Centre Bretagne, Unité Sciences et Techniques Halieutiques, Plouzané, France

Correspondence
Quentin Dubois
Email: quentin.dubois@uclouvain.be

Funding information
F.R.S.-FNRS

Handling Editor: Justin Travis

Abstract

1. Population genetics is used in a wide variety of fields such as ecology and biodiversity conservation. How estimated genetic characteristics of natural populations can be influenced by the sampling design has been a long-standing concern. Multiple simulation and empirical studies illustrated the influence of both sample size and polymorphism of markers. However, our review of studies on butterfly population genetics indicates no consensus on sample size for the estimation of genetic diversity or differentiation. Furthermore, other aspects of sampling design (sex ratio and timing of sampling) were not addressed and their potential impact on genetic parameter estimates rarely explored.
2. Using a large empirical dataset (with spatial and temporal replicates) collected on a butterfly species, *Boloria aquilonaris*, as well as simulated datasets reflecting (1) three scenarios of migration–genetic drift equilibrium and (2) one scenario of parameter stabilization after 100,000 generations, we quantified the impacts of three aspects of genetic sampling design (namely, sample size, sex ratio and timing of sampling) on the estimation of allele frequencies and its potential downstream impact on the estimation of genetic parameters.
3. With empirical data, we found that sample size and timing of sampling strongly affected the accuracy of allele frequencies and the downstream analyses, while sex ratio did not. Our results were consistent across spatial and temporal replicates. Also, with simulated data, we showed that the genetic sampling design had limited effect in systems where dispersal outweighs genetic drift, while it can have major consequences on our understanding of the genetic diversity and population differentiation in systems dominated by genetic drift (such as most study systems with conservation concerns).
4. We advocate for careful consideration of all aspects of the sampling design in population genetics studies, that is a sufficient number of samples, while ensuring similar sex ratio among sampling locations and collecting with timing appropriate to the question under study. This is particularly important when the study aims at species conservation.

KEYWORDS

butterfly, conservation, genetic diversity, metapopulation, number of samples, period of sampling, population structure, sampling scheme optimization, sex ratio

© 2017 The Authors. Methods in Ecology and Evolution © 2017 British Ecological Society

1 | INTRODUCTION

Population genetics has been applied in various research fields such as agronomy, ecology and biodiversity conservation. It is of ultimate importance in conservation, as genetic variability is linked to the evolutionary potential of populations (Frankham, Ballou, & Briscoe, 2002). To estimate this variability, a wide variety of markers have been used, such as amplified fragment length polymorphisms (AFLP), microsatellites, allozymes or single nucleotide polymorphisms (SNP; Allan & Max, 2010). One of the most used markers are microsatellite markers (Selkoe & Toonen, 2006) because they are co-dominant, hypervariable, reproducible and *a priori* selectively neutral (Jarne & Lagoda, 1996). Furthermore, they are based on PCR, which allows the use of small tissue samples (e.g. insect legs or wing clippings), which do not necessarily require the capture of individuals (non-invasive sampling of fur and faeces among others, see Taberlet et al., 1996; Taberlet, Waits, & Luikart, 1999). These markers have been successfully used in research on dispersal and gene flow (e.g. Keyghobadi, Roland, & Strobeck, 2005; Richards, Bernard, Feldheim, & Shivji, 2016), on genetic differentiation and the distribution in space of genetic diversity (e.g. in mammals, Paetkau, Calvert, Stirling, & Strobeck, 1995; in insects, Keyghobadi et al., 2005; in plants, Kikuchi & Isagi, 2002).

Butterflies are excellent study systems in ecology and conservation biology, particularly for studies focusing on global change (Warren et al., 2001) and metapopulation theory (e.g. Mousson, Nève, & Baguette, 1999; Nieminen, Siljander, & Hanski, 2004; Nowicki et al., 2007). They present relatively simple life cycles, often with non-overlapping generations (preventing the blurring or alteration of a genetic signal by mixing generations) and, they are relatively easy to study and their ecology is well known (Boggs, Watt, & Ehrlich, 2003). Furthermore, many species are specialized at the larval stage (Bink, 1992), facilitating the *a priori* delineation of habitat units in the landscape (Dennis, Shreeve, & Van Dyck, 2003; Turlure, Choutt, Van Dyck, Baguette, & Schtickzelle, 2010), a point essential for studies implying a structuring among discrete populations in the landscape such as in metapopulation ecology and population genetics (Hanski & Gaggiotti, 2004). In the context of conservation, many studies have focused on species facing some level of threat, often legally protected (e.g. *Boloria eunomia*, Radchuk, Turlure, & Schtickzelle, 2013; *Maculinea arion*, Thomas, 1995; *Melitaea cinxia*, Ehrlich & Hanski, 2004). Working with such species entails that the genetic data collection must be optimized, even more than for less threatened species. Indeed, in addition to the usual trade-off between the costs (both in terms of time and money) of acquiring genetic samples and the overall reliability of the results, the potential harmful impact of sampling on the species and/or its often fragile habitat must be limited as much as possible.

With this problem in mind, we reviewed the literature published on butterfly species (Scopus database; keywords: butterfl* AND microsatellite*; 50 studies matched a population genetics context on 1 November 2016; Appendix S1). We found no standard practice for genetic sampling design, even on the seemingly simple question of the number of samples collected, which most of the time probably resulted from logistic limitations and/or local population densities. We found substantial variation in the mean number of samples collected across these studies with a median sample size per predefined unit or population of 20 (range: 1–110, *SD*: 15.29). Moreover, most studies neglected other potentially important features of the sampling design: the sex ratio and the timing of sampling. The sex ratio was explicitly considered in only six studies, with conservation concern leading to the sampling of males only. However, this could be problematic as many butterfly species show sex-specific behaviours (e.g. male mate-locating strategy, Rutowski, 1991; harassment of females by males, Baguette, Vansteenwegen, Convi, & Nève, 1998), potentially affecting the distribution in space of individuals by promoting sex-biased movements or dispersal (e.g. Baguette et al., 1998; Ohsaki, 1980; Schtickzelle, Turlure, & Baguette, 2012). Sex-biased dispersal, together with an unbalanced sampling of males or females, may also lead to a substantial discrepancy between true and observed within- and among-population genetic characteristics (e.g. sampling only males after dispersal occurred in systems with male-biased dispersal might lead to an underestimation of the genetic structure compared to a female-only or balanced sampling). If genetic structure at the landscape scale is used to infer dispersal rates, such bias could be detrimental. Finally, only a third of the studies used samples collected on a single generation and the timing of collection within the generation was never mentioned. Yet, collecting newly emerged imagoes that have had no chance to disperse or older individuals with a potential dispersal history can strongly influence the measured genetic characteristics of natural populations, and inferred dispersal rates. Therefore, there are multiple ways through which changes in sampling scheme could influence the measured genetic characteristics of the focal study system and these impacts may be even stronger when genetic drift is dominant as the homogenizing effect of dispersal will be low. Based on a quick review of the ten last studies published on microsatellites using mammals and birds, we observed very similar patterns, with varying sample size (mammals: median sample size = 20.5, *SD* = 51.6; birds: median = 29, *SD* = 26.7), sex ratio mentioned in one or two studies and sampling done over several years in most cases.

Given the potential problems associated with the aforementioned issues, our aim was to quantify the impact of these three components of the sampling design (sample size, sex ratio and timing of sampling) on the estimation of allele frequencies and genetic parameters. We used empirical and simulated data to combine (1) the realism of empirical genetic data, collected across temporal and spatial replicates and (2) the possible generalization of the results across several levels of dispersal-genetic drift equilibrium and parameter stabilization. In a first step, we tested the effect of sampling design using a large genetic dataset (four spatial replicates totalling 1,726 samples, and four temporal replicates totalling 1,890 samples) collected on Belgian populations of the cranberry fritillary, *Boloria aquilonaris* (Stichel 1908), a specialist butterfly species. In a second step, we used simulated datasets to explore the impact of parameter stabilization and three levels of dispersal–genetic drift equilibrium: (1) genetic drift prevailing, (2) dispersal–genetic drift equilibrium and (3) dispersal prevailing.

Most genetic parameters used in a diversity or differentiation context are based on the estimation of allele frequencies (Hartl &

Clark, 2007), and some authors argued that it is sufficient to focus exclusively on obtaining reliable estimates of allele frequencies (Hale, Burg, & Steeves, 2012). Indeed, if the local populations' genetic composition is characterized accurately, the downstream analyses relying on it should produce robust and consistent results. Yet the degree to which changes in allele frequencies can influence the estimation of parameters commonly used in population genetics will also depend on the number of loci and their level of polymorphism. Moreover, fully understanding the effect of different sampling designs on genetic parameters requires quantifying their effect not only on the allele frequencies but also on the actual downstream genetic parameter estimations. For this reason, we also focused on genetic parameters describing (1) local-scale genetic diversity: allelic richness (AR; El Mousadik & Petit, 1996), observed heterozygosity (H_o, Nei, 1987), expected heterozygosity (H_s, Nei, 1987), and within-population heterozygote deficit (F_{IS}; Nei, 1987), as well as (2) inter-group features: population differentiation (F_{ST} and D_{est}; Jost, 2008; Nei, 1987) and dispersal (number of first-generation migrants; Piry et al., 2004).

Our expectations concerning the impact of sampling features were as follows. First, we expected a strong impact of the sample size on the precision of the allele frequencies estimates, with a nonlinear increase in precision (faster increase in precision at lower sample size than at larger sample size). This has been reported in multiple studies (e.g. Hale et al., 2012) and this effect should affect similarly the estimation of the derived genetic parameters (H_o, H_s, F_{IS}, F_{ST} and D_{est}). We also expected a nonlinear gain of accuracy in the estimation of AR with increasing sample size, given that the probability of sampling rare alleles increases with the number of individuals. Second, we expected the sex ratio to have an effect on allele frequencies and subsequent genetic parameter estimates in case there is a strong genetic differentiation among sexes. This could be due to sex-linked markers, sex-biased dispersal, strong selection pressure acting on one sex and/or a biased sex ratio of the offspring produced by dispersing individuals. Third, we expected the estimates of allele frequencies and of genetic parameters to be affected by the timing of sampling within the flight season. Indeed, an early sampling would give a representation of demographic processes and dispersal that took place over the previous generation(s), while the genetic characteristics measured using a late sampling would result from past and current dispersal processes (with the hypothesis that the probability of sampling dispersing individuals increases during the flight period). In particular, we expected this to lead to an increase in AR and H_s due to the introduction of "foreign" genotypes, and a decrease in $F_{IS,}$ F_{ST} and D_{est}, due to a homogenization effect of potential dispersal. All these expectations regarding the effect of sample size, sex ratio and timing of sampling concern cases where populations are at dispersal–genetic drift equilibrium. We expect these effects to be stronger when genetic drift prevails and weaker when dispersal prevails or when the parameter value is stabilized. Finally, we conclude on recommendations for future sampling designs of population genetics in general, and for further study of *B. aquilonaris* population health and metapopulation structure in particular.

2 | MATERIALS AND METHODS

2.1 | Study species and system

The cranberry fritillary *B. aquilonaris* is a specialist butterfly inhabiting peat bogs where its larval food plant, the cranberry *Vaccinium oxycoccos* (L. 1753), grows. In Belgium, the species is restricted to naturally fragmented habitats in uplands where climatic conditions are cold and wet, allowing the development of its larval host plant. Adults fly in one generation a year in June–July in this area, showing a mild protandry as many other butterflies do (Wiklund & Fagerström, 1977). The species occurs in thirty-nine sites in Belgium (Figure 1) that are aggregated in space, with thirteen populations in the "Hautes Fagnes" area (hereafter referred to as Group 1: G1),

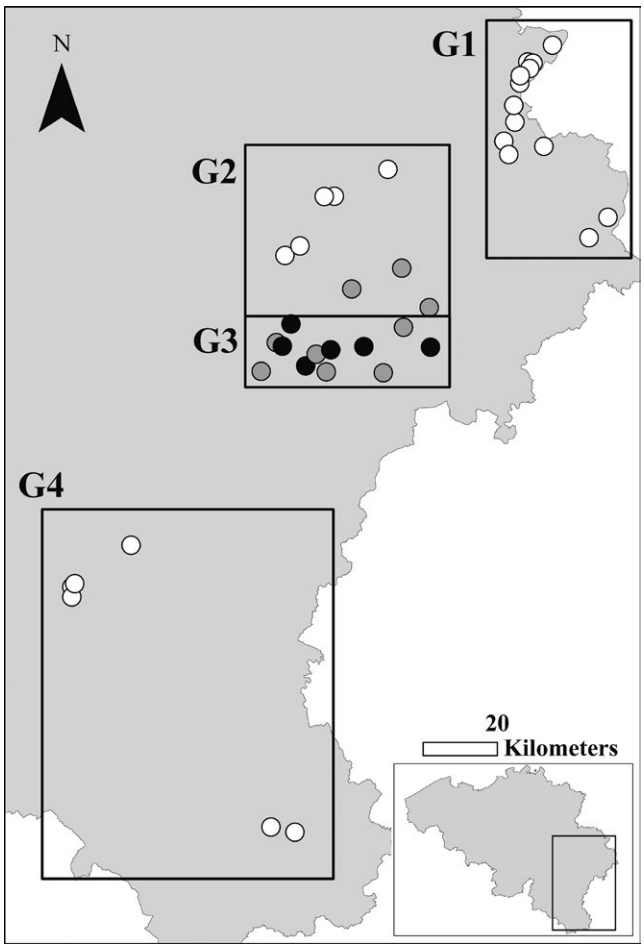

FIGURE 1 Sites sampled between 2010 and 2015 across south-eastern Belgium, as represented in the bottom right frame. We grouped the populations into four geographical clusters, represented by black rectangles, from the Hautes Fagnes (G1), Ardennes liégeoises (G2), Plateau des Tailles (G3) and Plateau de Libin and Lorraine (G4). These clusters were used as spatial replicates. Individuals from the populations represented with black and grey dots were grouped to define four temporal replicates in 2010, 2013, 2014 and 2015. These were used to study the impact of sample size and sex ratio. The populations represented by black dots were used in the study of the impact of the timing of sampling. The grey surface represents the Belgian territory

eight populations in the "Ardennes liégeoises" area (G2), twelve populations in the "Plateau des Tailles" area (G3) and six populations spread throughout the "Plateau de Libin" and "Lorraine" areas (G4). The groups were defined on a geographical basis, taking into account the potential natural barriers between them (the sites are mostly present on uplands separated by valleys). These groups present four levels of habitat network fragmentation, with sites highly aggregated (G3) to loosely distributed (G4). The species is able to disperse over long distances, as revealed by capture–mark–recapture (up to 13 km recorded in Baguette 2003, and up to 37 km recorded by ourselves in 2014).

2.2 | Genetic data collection

DNA collection was based on the sampling of one leg per individual on at least the 15 first males and 15 first females seen in each site (i.e. we kept a balanced sex ratio). Individuals were sampled at the beginning of the flight period (1) in fifteen populations present in G2 and G3 in 2010, 2013, 2014 and 2015 (temporal replicates), and (2) in all populations of the four groups defined above (G1–G4) in 2013 and 2014 (spatial replicates; see Figure 1 for details on sampling). These data were used to test the effect of sample size and sex ratio. Additionally, in six of the G3 populations, we sampled approximately 20 females late in the 2010 flight season (i.e. sampling was performed after the peak in population size), as mostly females were available at that period due to protandry. These data were used to test the effect of timing of sampling. Each leg was immediately placed in a 1.5-ml labelled tube filled with pure ethanol; tubes were kept at −20°C from the end of the field day.

Genomic DNA was extracted from each leg separately using a protocol which is a slightly modified version of the one used by Vandewoestijne, Turlure, and Baguette (2012). The leg was first homogenized in 100 μl of extraction buffer (50 mM Tris–HCl pH 8.0, 20 mM EDTA pH 8.0, 2% SDS) and one unit of proteinase K. After 2 hr of incubation at 55°C, 40 μl of a 6 M NaCl and 150 μl of chloroform were added and mixed vigorously. The sample was then spun at 20,800 g at 4°C for 5 min. The collected supernatant was mixed with an equal volume of ice-cold isopropanol. After a 15-min spin at 20,800 g, all the liquid was poured off and 150 μl of room temperature ethanol added. The sample was spun again for 5 min and the ethanol poured off and dried. RNAse treatment was applied after resuspension in 20 μl of ultrapure water. Amplification of 16 microsatellite markers was conducted with QIAGEN® Multiplex PCR kits as described in Vandewoestijne et al. (2012). The genotypes were scored using Genemapper® 3.7 (Applied Biosystems®), and manually verified to correct for automatic scoring errors when necessary. Four microsatellite markers (Baq-30, Baq-44, Baq-74 and Baq-77) were discarded due to linkage disequilibrium.

2.3 | Simulated datasets

We simulated four datasets using EASYPOP 2.0.1 (Balloux, 2001) characterized by (1) three levels of dispersal–genetic drift equilibrium, created using three constant dispersal rates and (2) a relative stabilization of the genetic parameters on the long run (hereafter labelled as "parameter stabilization"). The dispersal rates used to create the datasets were 0 (genetic drift prevailing), 0.01 (dispersal–genetic drift equilibrium and parameter stabilization), 0.05 (dispersal prevailing). First, to test the effect of sample size, sex ratio (only in the case of long-term parameter stabilization) and timing of sampling, we simulated five populations of 500 individuals linearly distributed at coordinates 1, 2, 4, 8, 16. The genotypes were generated for 10 independent microsatellite markers (single-step mutation) with 10 alleles, presenting a maximal original diversity at the start of the simulations. Second, to test the effect of sex ratio in the three dispersal–genetic drift equilibrium scenarios, we simulated 10 populations of 250 individuals each, placed by pairs at the previous coordinates to force the emergence of some random levels of differentiation within pairs. All other simulation specifications were identical to the previous case. After the simulations, the populations with even labels were considered as females, and the ones with odd labels were considered as males. In each case, males and females were allowed to mate only once per generation, the mean distance travelled by individuals was four units, and simulations were run for 100 generations (dispersal–genetic drift equilibrium) or 100,000 generations (parameter stabilization). The state of equilibrium was assessed graphically, using the shape of the relation between pairwise F_{ST} and distance (Appendix S2; Beebee & Rowe, 2008): (1) no relation between F_{ST} values and distance with largely scattered values indicating genetic drift prevailed, (2) positive correlation with cone-shaped scatter indicating dispersal–genetic drift equilibrium and (3) no relation between F_{ST} and distance with only low values indicating dispersal prevailed. For each state of dispersal–genetic drift equilibrium, we produced 10 simulated datasets and selected the one showing the best concordance with the criteria presented in Beebee and Rowe (2008). The changes in genetic parameters along the generations were followed using the information provided by EASYPOP regarding allelic richness, observed and expected heterozygosity, within-population heterozygote deficit and genetic differentiation to confirm their stabilization (Appendix S3).

2.4 | Subsampling the datasets

We used a general subsampling approach to study the impact of sampling quality on allele frequencies estimation and the subsequent calculation of genetic parameters, except for timing of sampling on empirical data (details below). The subsamples were created by drawing samples out of the original datasets without replacement. The original datasets are (1) the four geographical replicates (i.e. G1, G2, G3 and G4, each pooling samples collected in 2013 and 2014), (2) the four temporal replicates in the empirical data (i.e. datasets collected in G2 and G3: 2010, 2013, 2014 and 2015) and (3) the populations present in the four scenarios in the simulated data. First, to test for sample size effect in both the empirical and simulated datasets, we subsampled the original datasets

from four to 76 individuals, with an increment of four individuals, keeping a balanced sex ratio and using only individuals collected early in the flight season. Second, to test for sex ratio effect in the empirical datasets, we subsampled a fixed sample size of 28 individuals while varying the sex ratio from 0 (no females) to 1 (females only), by steps of 0.25 (seven individuals) again using samples collected early in the season. In the simulated datasets, the individuals were drawn from the two created subpopulations at each location that we arbitrarily defined as male or female. Third, to test for the timing of sampling effect, in empirical data, we compared two datasets comprising 20 females each collected either early or late (i.e. before or after the demographic peak) in six populations within the 2010 flight period. In the simulation analyses, we subsampled a fixed number of 28 individuals, taking most of them from a focal population (the population located at position 1) and up to seven samples from the other populations (the populations located at positions 2–5), one population at a time. We did this to simulate the increasing probability of sampling dispersing individuals along the flight period, as the chances of successful dispersal should increase with time. This allowed us to test the effect of an increasing original differentiation between the populations as closer populations were less genetically differentiated. Subsampling of the data was performed 100 times independently each time it was involved. We assessed the potential problems of overlap of individuals between subsamples and found that this was not a major concern with our empirical dataset (Appendix S4).

2.5 | Genetic and statistical analyses

We quantified the impact of sampling design at three levels: (1) the estimation of allele frequencies, (2) the description of local genetic variation and (3) the estimation of intergroup features.

First, we calculated allele frequencies observed in each temporal and geographical replicates and each simulated populations, for each dataset with different levels of sample size, sex ratio and timing of sampling. Calculations were made using the R package HIERFSTAT (Goudet, 2005; R version 3.2.0, R Core Team, 2014).

Second, to highlight the consequences of potentially erroneous estimation of allele frequencies, we estimated four parameters commonly used to describe the local genetic characteristics of a population. Namely, we calculated the observed heterozygosity (H_o; Nei, 1987), the expected heterozygosity (H_s; Nei, 1987), the within-population heterozygote deficit (F_{IS}; Nei, 1987) and the allelic richness (AR; El Mousadik & Petit, 1996) for each dataset with different levels of sample size, sex ratio and timing of sampling. The estimation of each parameter was done at the locus level and was summarized using the mean over the population in each dataset.

Third, we calculated two measures of overall genetic differentiation using HIERFSTAT (Goudet, 2005): the common diversity partitioning statistic F_{ST} and the real differentiation D_{est}. This was done for the groups of temporal and spatial replicates and the group of populations used to study the impact of the timing of sampling using empirical data. This was done for each simulation scenario by calculating the genetic differentiation between the total focus population and each subsample, that is composed mainly of individuals of the focus population and a variable number of dispersers. F_{ST} is calculated as the amount of gene diversity among samples divided by the overall gene diversity present in the dataset (Nei, 1987). D_{est} has been introduced by Jost (2008) as an unbiased measure of population differentiation based on the effective number of alleles (which corresponds to the number of alleles of equal frequencies it would take to achieve a given level of gene diversity; Kimura & Crow, 1964). Furthermore, and only for the dataset containing females collected early and late in the 2010 season (in a subset of G3 populations), we estimated the probability of each individual being a first-generation migrant using the software GENECLASS2 (Piry et al., 2004). We used the Bayesian criterion derived from Rannala and Mountain (1997) to estimate this likelihood. The probability estimation has been done using Monte–Carlo resampling on 10,000 simulated individuals and a type I error rate of 0.01, following Paetkau, Slade, Burdens, and Estoup (2004). We used all the individuals sampled (both early and late in the flight season) to discriminate the individuals identified as first-generation migrants on a common basis.

A "reference value" was defined for each replicate (sample size and sex ratio) and population (timing of sampling) in empirical data, and in each population in simulated data, as the value of a given genetic parameter calculated for the full dataset (i.e. presenting the maximum sample size). The reference value for genetic differentiation in simulated data was calculated as the differentiation between the focus population and the population of origin of the "dispersers." We report the results as the absolute deviation from the reference value (calculated as |computed − reference|), with the exception of the detection of first-generation migrants for which we give the raw results, because we were mainly interested in quantifying the magnitude of the error in the estimation, rather than its direction. Whenever we considered it useful, we added graphical information about the direction of the effect in appendices. The maximum sample sizes of our empirical genetic datasets were as follows: (1) in the temporal replicates: 726 individuals in 2010, 408 in 2013, 372 in 2014, 384 in 2015; (2) in the spatial replicates: 528 for G1, 324 for G2, 658 for G3 and 216 for G4; and (3) from 29 to 33 samples taken early in the flight season in the six populations of G3 considered to study the timing effect.

3 | RESULTS

3.1 | Sampling design effects on allele frequencies

In both empirical and simulated data, increasing the sample size led to a clear nonlinear reduction in the absolute deviation from reference value (Figure 2 first column). Sample sizes of about 40 (empirical data) and 30 (simulated data) individuals were generally required to estimate allele frequencies with an error <0.05 although the gain in precision was already limited above 20 individuals. The magnitude of the deviation was also clearly affected by the original allele frequency

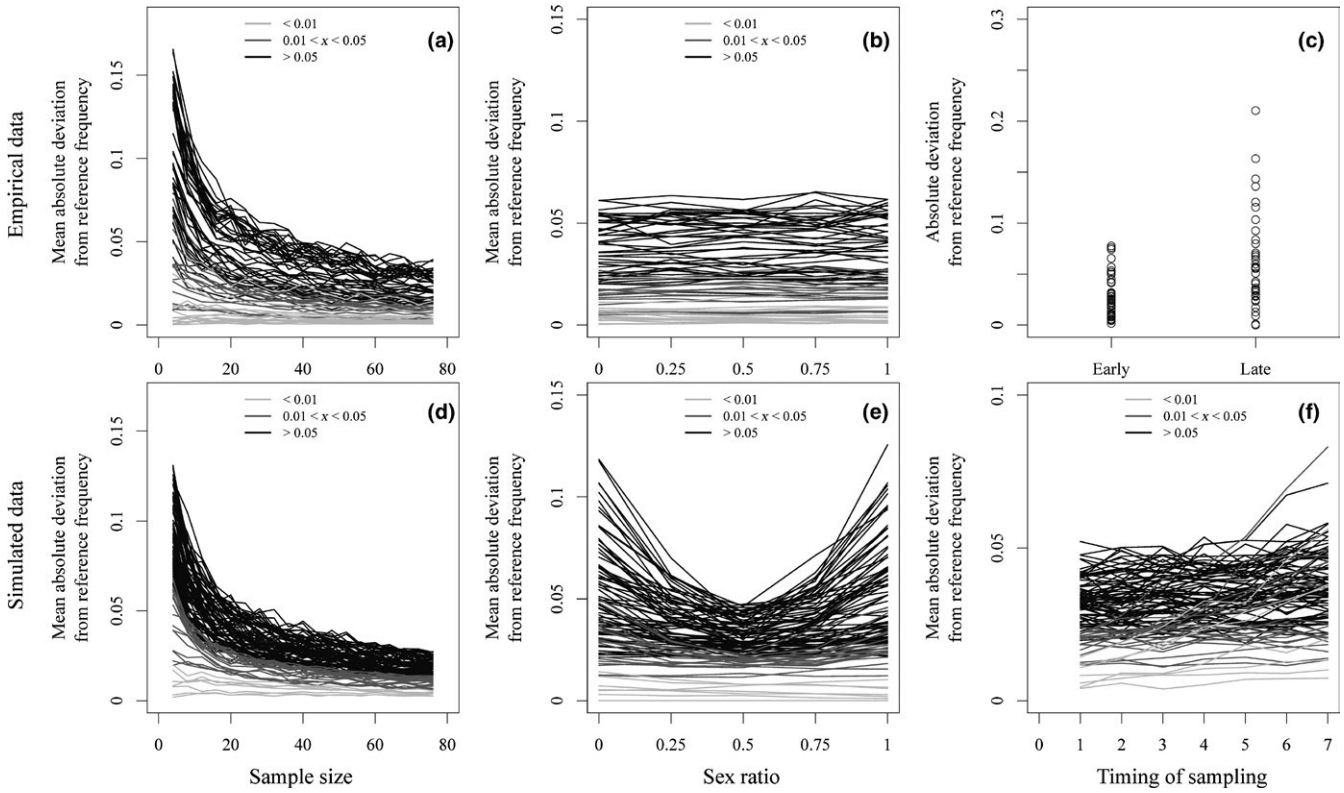

FIGURE 2 Effects of the sampling design on the estimation of allele frequencies (empirical data: first line; simulated data: second line): sample size (first column), sex ratio (second column) and timing of sampling (third column). In all panels, each line represents the mean (over the 100 subsampled datasets) absolute (|computed − reference|) deviation from reference frequency (full dataset) calculated for one allele. In panel c, each circle represents the absolute deviation from reference frequency calculated for one allele. The effects were consistent across the temporal (i.e. in all years of G3) and spatial (i.e. in all spatial groups) replicates (empirical data), and populations (simulated data), so only representative examples are displayed here (2010 dataset for sample size and sex ratio; Mirenne population for timing of sampling). We show only the figure for the dispersal–genetic drift equilibrium (simulated data). Results for all datasets are available in Appendices S5, S6. The ticks on the *Y* axes are separated by 0.05 units to ease comparison

in the group: the more frequent the allele, the bigger the deviation (Appendices S5–S6).

Variation in the sex ratio had contrasting effects on the estimation of allele frequencies. It did not influence the absolute deviation from reference frequency in empirical data (flat pattern) while it did in simulated ones (U-shaped pattern; Figure 2 second column). The effect detected in simulated data was especially strong when genetic drift was stronger than dispersal, a situation creating strong differences between sexes (Appendix S6). In simulations where no genetic differences were intentionally forced, we also detected an effect of sex ratio in several populations (Appendix S6).

There was a clear effect of the timing of sampling on the estimated allele frequencies. In empirical data, the absolute deviation associated with an early sampling was smaller (mostly < 0.10) compared with a late sampling (mostly over 0.15 and sometimes up to 0.25; Figure 2c). In simulated data, the more individuals arrive, the more divergent the allele frequencies of the reference dataset and the ones estimated for the subsamples will be (Figure 2f), which was exacerbated when genetic drift prevailed (Appendix S6). The complete results are provided in Appendices S5–S6.

3.2 | Sampling design effects on local genetic parameters

In both empirical and simulated data, the deviation from reference value for AR decreased with sample size, while the range of observed values was practically not affected, except when the maximum number of alleles was reached (Figure 3a for empirical data, Figure 3m for simulated data). The speed of saturation was positively associated with dispersal predominance (Appendix S8). The absolute deviation from reference values decreased substantially along the sample sizes considered for H_o, H_s and F_{IS}, with little gain in precision above 30 individuals in empirical data and 20–24 individuals with simulated data (Figure 3b–d for empirical data, Figure 3n–p for simulated data). The deviation from reference value for H_o, H_s and F_{IS} was greater when genetic drift prevailed (Appendices S10, S12 and S14).

The sex ratio did not affect the absolute deviation from reference value for any parameter in empirical data (Figure 3e–h), while it clearly did for AR, H_s and F_{IS} (but not on H_o) in simulated data (Figure 3q–t). In the latter case, the most accurate estimates were obtained with a balanced sex ratio. These effects were greater in scenarios where genetic drift was stronger than dispersal, and were detectable even

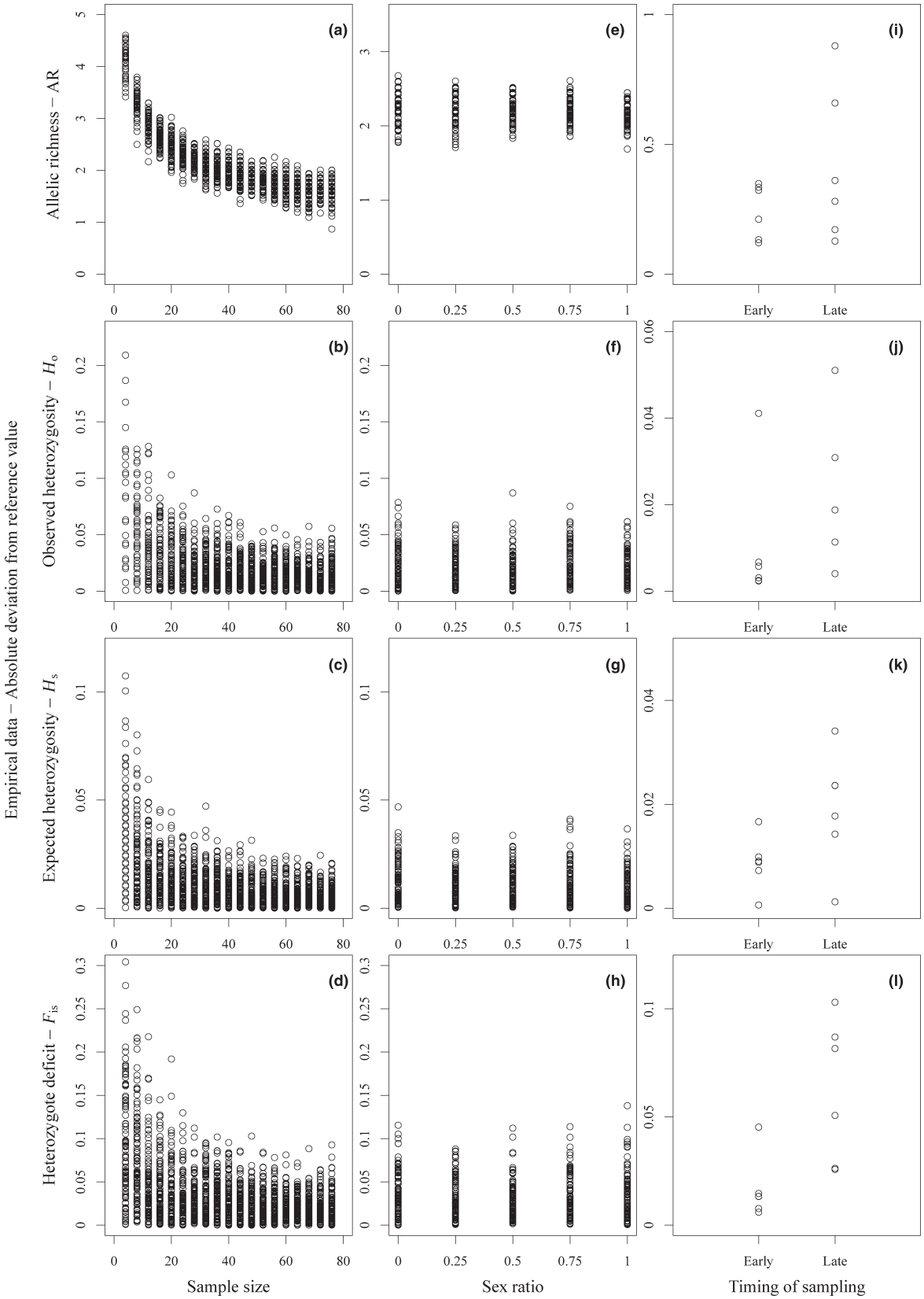

FIGURE 3 Effects of the sampling design on local genetic parameters (empirical data: a–l; simulated data: m–x): sample size (first column), sex ratio (second column) and timing of sampling (third column). Each point represents the absolute (|computed − reference|) deviation from reference value (full dataset) for each individual subsampled dataset. The effects were consistent across temporal and spatial replicates (empirical data), and populations (simulated data) so only representative examples are displayed here (2010 dataset for sample size and sex ratio; all populations for timing of sampling). We show only the data when populations are at dispersal–genetic drift equilibrium (simulated data). Results for all datasets and scenarios are available in Appendices S7–S14

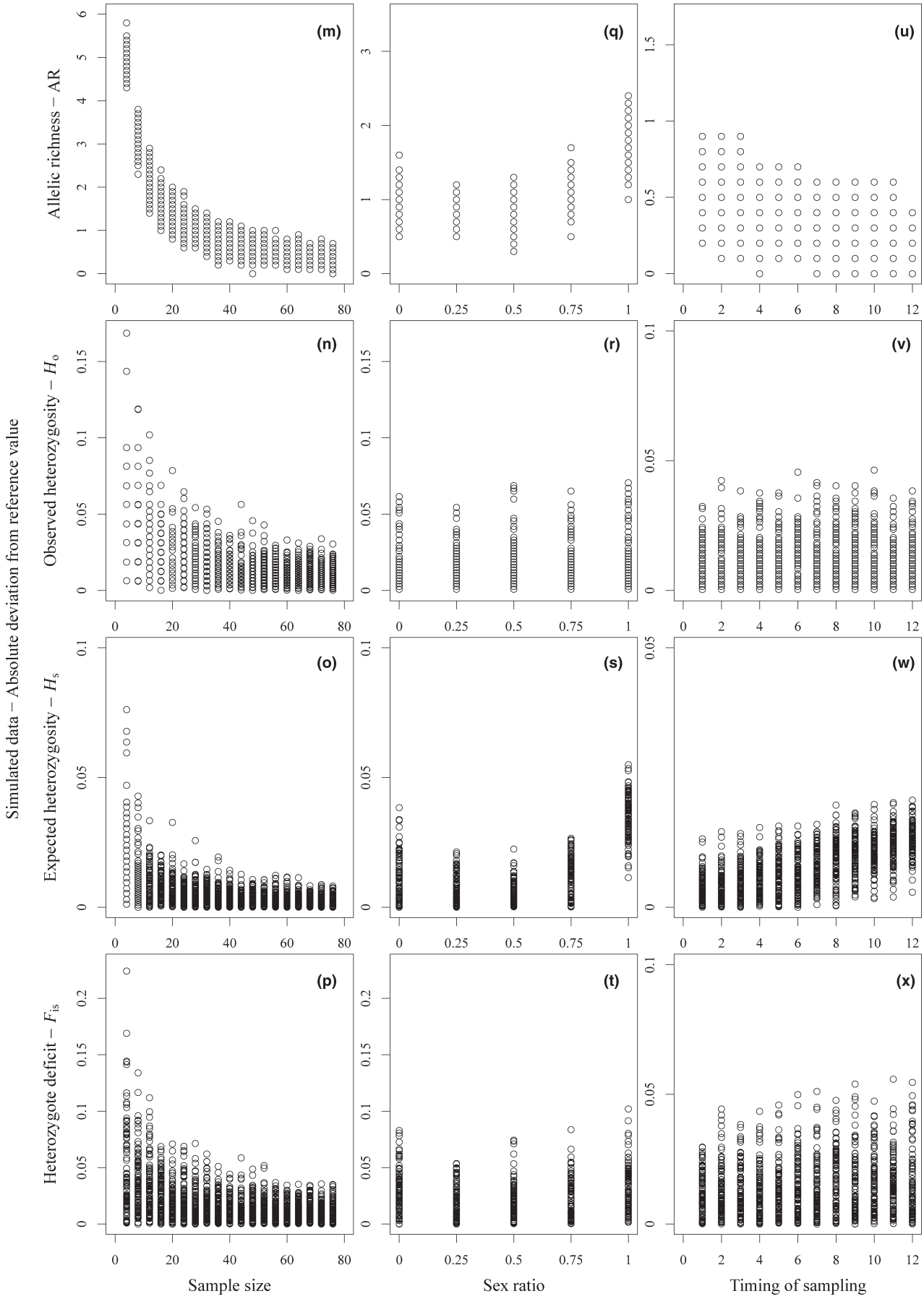

FIGURE 3 Continued

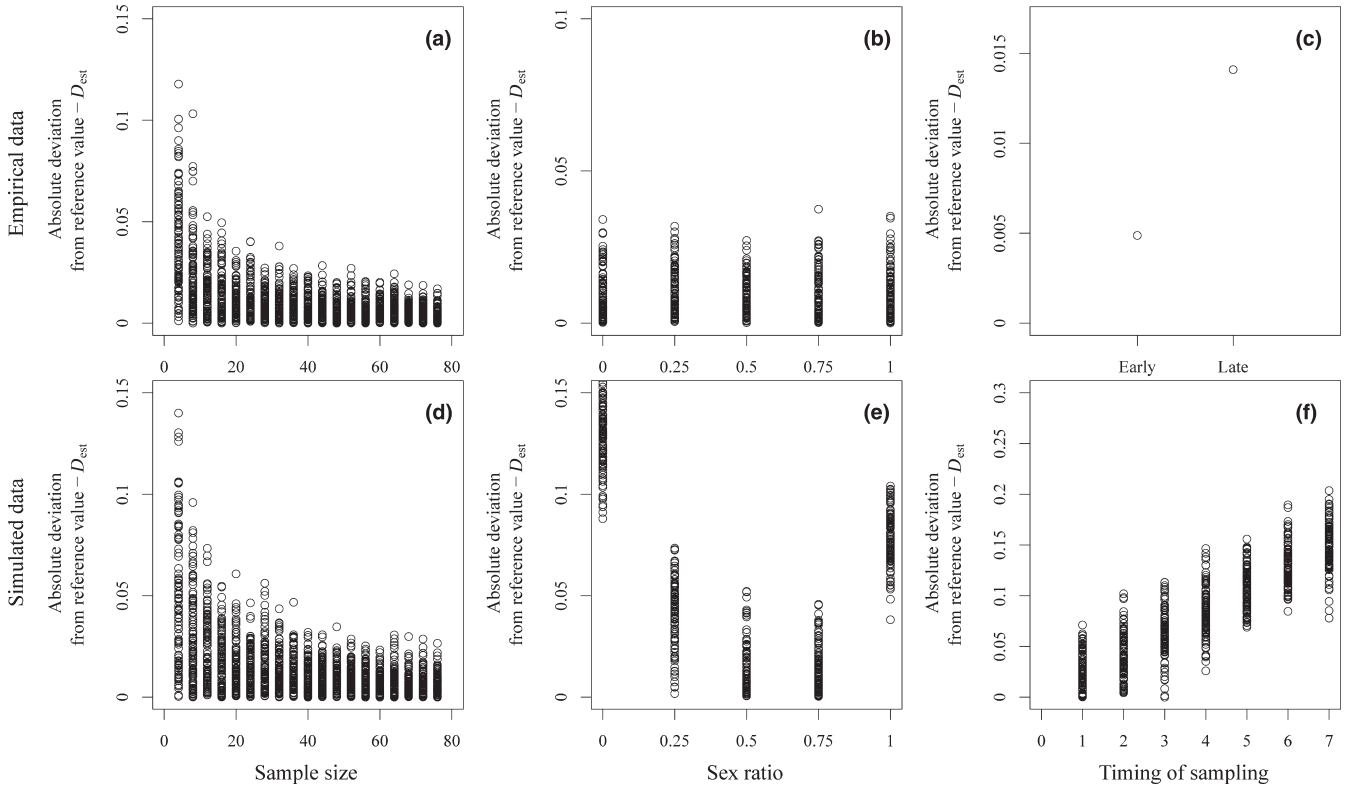

FIGURE 4 Effects of the sampling design on the genetic differentiation measured as D_{est} (empirical and simulated data): sample size (first column), sex ratio (second column) and timing of sampling (third column). In all panels except for c, each circle represents the absolute (|computed − reference|) deviation from reference value (full dataset) for each group of populations within a given dataset (four temporal replicates in empirical data and the pair composed of the focus and fourth populations in simulated data). In panel c, the circles represent the same values, but obtained for the group of six populations. The results for F_{ST} were similar and are available in Appendices S15, S16

when no genetic differences were intentionally forced by the simulation protocol (Appendices S8, S12 and S14).

In empirical data, the timing of sampling affected the absolute deviation from reference value for AR, H_o, H_s and F_{IS} estimates, with smaller deviations associated with an early sampling (Figure 3i–l). Nevertheless, these deviations always lay in the range of values observed across the spatial and temporal replicates when the sample size was 20 individuals, with the exception of F_{IS} (see Appendix S13). In simulated data, the timing of sampling affected the absolute deviation of AR, H_s, and F_{IS} estimates (but no effect on H_o), with higher deviations associated with more immigrants in the population. The magnitude of the effect was positively associated with genetic drift dominance, and totally disappeared when dispersal prevailed (Figure 3u–x; Appendices S8, S12, and S14). The complete results are provided in Appendices S7–S14.

3.3 | Sampling design effects on genetic differentiation and assessment of first-generation migrants

The estimators of intergroup genetic differentiation (i.e. F_{ST} and D_{est}) showed similar behaviours concerning the sampling design in both empirical and simulated data. The decrease in absolute deviation from reference value was very slow above 30 (empirical data) and 20 individuals (simulated data; Figure 4a,d). These deviations were greater when the initial differentiation was high (i.e. in empirical data when considering the spatial replicates; in simulated data when drift prevailed; Appendices S15 and S16). Changes in sex ratio did not affect the absolute deviations from reference value in empirical data (Figure 4b) while they did in simulated data (Figure 4e). This effect was visible at all levels of dispersal–genetic drift equilibrium and was detectable even when no genetic differences were intentionally forced by the simulation protocol (Appendix S16). The timing of sampling affected estimated genetic differentiation, with higher deviations from reference value associated with a late sampling in empirical data and with more immigrants in simulated data (Figure 4c,f), always with a clear decrease in genetic differentiation. Finally, a late sampling allowed the detection of first-generation migrants in two populations that did not receive individuals early in the flight season (Table 1).

4 | DISCUSSION

In any research field, the sampling design quality directly influences the measure quality, and consequently the reliability of the statistics and the resulting conclusions. In population genetics, the aspect of sampling design that has received much attention in the past is sample size, in both simulation (e.g. Kalinowski, 2005) and empirical

TABLE 1 Detection of first-generation migrants in six populations when sampling early and late in the flight season, using GENECLASS2 (Piry et al., 2004). The acceptance threshold to identify the migrants was set to 0.01

	Number of first-generation migrants	
Population of arrival	**Early**	**Late**
Commanster	1	1
Crépale	1	1
Grand Passage	0	0
Mirenne	1	1
Nazieufa	0	1
Sacrawé	0	1

studies (e.g. Hale et al., 2012; Miyamoto, Fernández-Manjarrés, Morand-prieur, Bertolino, & Frascaria-Lacoste, 2008; Pruett & Winker, 2008). Our study, based on empirical and simulated genetic datasets, contributes to the existing evidence of sample size influence on estimated genetic parameters. With the empirical data, we found that the timing of sampling also strongly affected the accuracy of allele frequencies and the downstream analyses, while the sex ratio did not. The effects of sampling design described hereafter were consistent across temporal and spatial replicates for sample size and sex ratio, and across populations for the timing of sampling. With simulated data, we showed that the sampling design had limited effect on genetic measures in systems where dispersal outweighs genetic drift, while it can have major consequences on our understanding of the genetic diversity and differentiation of local populations in systems dominated by genetic drift (such as most study systems with conservation concerns). We also showed the importance of the sampling design, even when the genetic parameter values were stabilized. We discuss those results in details hereafter and conclude with integrative guidelines for future sampling design in population genetic studies.

4.1 | Sample size effect

Our results from empirical and simulated data are in line with those of previous studies regarding the estimation of allele frequencies. The larger the sample size, the better the estimates of allele frequencies. Nevertheless, as stated by Hale et al. (2012), the gain of precision may, at some point, become negligible compared to the increased cost of processing more samples (i.e. financial and labour cost, as well as the impact of invasive sampling on species with conservation concerns inhabiting fragile habitats). Here, the greatest gains in precision occurred below 20 samples, and most of the allele frequencies were estimated with an average deviation less than 0.05 at approximately 40 samples in the empirical data and 30 in simulated data. This difference probably arises from uncontrolled difference between the two sets of markers: the simulation data are homogenous and highly polymorphic while the empirical dataset is more heterogeneous and less polymorphic. The original allele frequency greatly influenced the magnitude of the deviation, the absolute deviation being more important for more common alleles, that is the ones that have the most important impact on the genetic parameter estimation. These larger deviations at small sample sizes are primarily due to a lack of accuracy (Hale et al., 2012; Pruett & Winker, 2008).

These deviations directly led to an important uncertainty in the estimation of local genetic parameters, although some parameters seem less sensitive than others. Indeed, we did not observe any important gain of precision above 30 samples in empirical data and 20 in simulated data for H_o, and above 20 samples in both empirical and simulated data for H_s. This represents additional evidence that H_s is a more robust measure of diversity compared to H_o, especially with empirical data (as in, e.g., Beebee & Rowe, 2008). Alternatively, 30–40 individuals are needed to observe similar patterns in F_{IS} with empirical data, and 24 individuals with simulated data, while it depends directly on the two previous parameters. The deviation from the reference value may still reach 0.1 units at 76 samples in empirical data, and 0.05 units in simulated data. In the empirical case, it must be kept in mind that this deviation may be due to the grouping of populations presenting some levels of genetic differentiation. AR was always estimated with an increasing accuracy and a similar precision along the explored range of sample sizes, except when the maximum richness was reached (only in simulated data). The rates at which accuracy increased depended largely upon the diversity present in the group in both empirical and simulated data. It was also reached faster as the importance of dispersal increased in simulation data. Indeed, a plateau is already reached at approximately 25 samples in a smaller and *a priori* less diverse group as G4 (Appendix S7.A.h), but not reached at sample sizes of 76 samples in large and diverse populations (in G3 mainly; Appendix S7.A.a–d;g). Hence, if the objective is to detect most of the alleles, the required sample sizes would largely depend on the inherent diversity of the studied populations.

As for the local genetic parameters, the inaccuracies in allele frequency estimation evidently affected the estimation of genetic differentiation, and the deviation from the reference value seemed to be affected by the expected existing differentiation. Indeed, we did not observe much increase in precision above 30 (low differentiation; temporal replicates) to 40 (high differentiation; spatial replicates) in empirical data and 20 individuals in simulated data. The behaviour of F_{ST} and D_{est} were very similar, with regard to precision: the only difference is that the values obtained for F_{ST} are smaller than D_{est}, which is due to the differences in the upper bounds of the metrics (D_{est} has an upper bound of 1 while F_{ST} will nearly always have an upper bound lower than 1; Jost, 2008).

4.2 | Sex ratio effect

Sex ratio has received little, if any, attention in the sampling design of genetic studies probably because it is *a priori* difficult to know the real sex ratio of a given population. In addition, researchers may consider sex ratio unimportant because microsatellite markers

found on sex chromosomes would be discarded as they hold high proportions of null alleles for one sex. Still, sex-biased processes, such as dispersal, are detectable through the use of "usual" microsatellite markers (Prugnolle & De Meeûs, 2002). Sex ratio can therefore affect the genetic estimates not because of an inherent genetic difference among sexes, but because sexes behave differently, which affects genetic parameters *in fine*, as if males and females were two subpopulations diverging along the flight season. We did not observe such a phenomenon with our empirical data. This is what was expected as the samples were collected on individuals as soon as possible after they emerged, leaving very little time for sex-specific behaviours, if any, to create variation in genetic parameters among sexes. However, when some levels of differentiation are present, as in our simulated data, we showed that an uncontrolled sex ratio can affect the estimation of allele frequencies and local and global genetic parameters. As expected, this effect was exacerbated when genetic drift prevailed (i.e. when genetic differentiation between sexes within a population was stronger), but was still detectable in some populations although no genetic differences were intentionally forced by the simulation protocol (i.e. in the simulated dataset in which the parameters value was stabilized). We are aware that our empirical dataset was, by design, limited to the study of the impact of the sex ratio early in the flight season.

4.3 | Timing of sampling effect

The timing of sampling has received little attention in the past, with the exception of Basset, Balloux, and Perrin (2001) and Balloux and Lugon-Moulin (2002), to the best of our knowledge. Using empirical data, we showed an effect of the timing of sampling (early vs. late sampling in the season), with an increase in the deviation from the reference allele frequency for later sampling. This translated into a general and consistent increase in the deviation for all genetic parameters. The most likely explanation for this effect is the progressive arrival of immigrants from other populations along the season. The arrival of a few individuals may induce profound changes in the genetic variability of inbred populations over only a few generations (Madsen, Shine, Olsson, & Wittzell, 1999). Our results suggest that even in well-connected natural populations, the progressive arrival of immigrants from other populations along the season can induce substantial changes in allelic frequencies and genetic parameter estimates. With our simulated data, at the local scale, the progressive adjunction of immigrants increased the deviation from reference frequencies, especially for alleles rare in the focus population. This led to changes in the estimation of local genetic parameters (particularly H_s and F_{IS}) when drift prevailed and when the populations were at dispersal–genetic drift equilibrium. This effect was much weaker when dispersal prevailed, but was still observable in relatively homogenous systems such as the one simulated over 100,000 generations.

At the intergroup scale, the deviation of genetic differentiation increased with later sampling in empirical data and with the adjunction of immigrants in simulated data, with a reduction in genetic differentiation. Basset et al. (2001) showed similar results. In empirical data, this is again likely linked to the arrival of dispersing individuals, as showed by the results of first-generation migrant detection. The two populations receiving a migrant only when the sampling was done late are well-connected populations in an area where the individuals emerge later in the flight season, allowing the arrival of individuals from populations with earlier emergences. This result also illustrates the risk of not considering the timing of sampling, as those populations would be considered isolated in the case of an early sampling. Hence, even though assignment methods allow the study of dispersal on a very short time-scale, which is of great interest in ecology, one clear drawback is that the results will be directly affected by the timing of sampling. Another drawback, seemingly more important (in populations confined to discrete and relatively well-separated sites), is that females could mate in their natal population, disperse and lay their eggs in a new population, thereby introducing non-admixed genotypes that did not actually disperse. This potential behaviour has been shown to be evolutionary competitive compared to dispersal right after emergence, so that its consequences should not be neglected (Lakovic, Poethke, & Hovestadt, 2015).

4.4 | An optimal sampling design?

If it is obvious that the number of samples will affect the precision of the genetic parameter estimates, it is impossible to derive a single minimum number of samples that will guarantee their reliability. With our set of markers, a minimum of 30 samples (20 in simulated data) are sufficient as not much information will be gained above. Determining the optimal sample size is then ultimately an optimization problem depending on the marker polymorphism and the costs of acquiring and processing genetic samples. Sex ratio also had an effect, in cases where a relatively important level of differentiation existed between the sexes, irrespective of the process creating this differentiation. Males and females should be used interchangeably only in cases where the phenomenon of interest is not reasonably affected by sex, and one should not assume this or elude the problem. Therefore, we advise future research to at least aim for consistent sex ratio (i.e. sampling only one sex or considering a balanced design). Regarding the timing of sampling, our results clearly show that individuals should be sampled within a single generation before dispersal, as it may alter the genetic composition of the unit(s) under study. However, conducting two periods of sampling within the same generation is appropriate to investigate dispersal based on genetic assignment methods. Overall, the various aspects of sampling design are to be considered as a whole as their effects may be interactive.

We strongly believe those recommendations to be generalizable at least for insects and sexual organisms producing non-overlapping generations as we accounted for the four following points. First, we explicitly considered the problem of overlap between subsampled datasets from empirical datasets of limited size. Second, we considered in a single study the impact of sampling design on several genetic parameters, both at the local and intergroup scales. Third, we explored the temporal and spatial repeatability of our empirical results. Fourth, we explicitly considered the state of dispersal–genetic drift

equilibrium and the long-term parameter stabilization in our simulations. Nevertheless, specific aims and/or estimation methods might require different, more elaborate or specific sampling designs. We therefore encourage the explicit description and justifications of the sampling design, and if needed, an acknowledgement of the sampling limitations in further genetic studies.

ACKNOWLEDGEMENTS

We thank three anonymous reviewers and the Editors for their help in improving the quality of this manuscript, L. Dhondt, C. Noël and O. Devroye for their help in laboratory work, M.-C. Flamant and M.-C. Eloy (UCL, Institute of Life Sciences) for their work in genotyping, V. Radchuk and G. Bataille for advice on the use of R and B. Colling (UCL, Support en Méthodologie et Calcul Statistique) for statistical advice in the estimation of overlap between samples. Q.D. (PhD), C.L. (postdoctoral fellow), C.T. (postdoctoral fellow) and N.S. (research associate) acknowledge financial support from the F.R.S.-FNRS. Site access and a permission to study the species in the field and in the laboratory were granted by the Ministère de la Région Wallonne. This study is contribution BRC413 of the Biodiversity Research Centre at UCL. We have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

All authors conceived the ideas and designed methodology; Q.D. and C.T. collected the data; Q.D. analysed the data with the help of coauthors; Q.D. led the writing of the manuscript. All authors contributed critically to the manuscript and gave final approval for publication.

DATA ACCESSIBILITY

The allele frequencies and genetic parameters data (both from the empirical and simulated datasets) are available on the Dryad repository (DOI: https://doi.org/10.5061/dryad.5j13n) (Dubois, Lebigre, Schtickzelle, & Turlure, 2017).

ORCID

Quentin Dubois http://orcid.org/0000-0003-3792-1279
Camille Turlure http://orcid.org/0000-0002-8313-8528

REFERENCES

Allan, G. J., & Max, T. L. (2010). Molecular genetic techniques and markers for ecological research. *Nature Education Knowledge*, *3*, 2.

Baguette, M., Vansteenwegen, C., Convi, I., & Nève, G. (1998). Sex-biased density-dependent migration in a metapopulation of the butterfly *Proclossiana eunomia*. *Acta Oecologica*, *19*, 17–24. https://doi.org/10.1016/S1146-609X(98)80004-0

Balloux, F. (2001). EASYPOP (version 1.7): A computer program for the simulation of population genetics. *Journal of Heredity*, *92*, 301–302. https://doi.org/10.1093/jhered/92.3.301

Balloux, F., & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, *11*, 155–165. https://doi.org/10.1046/j.0962-1083.2001.01436.x

Basset, P., Balloux, F., & Perrin, N. (2001). Testing demographic models of effective population size. *Proceedings of the Royal Society B: Biological Sciences*, *268*, 311–317. https://doi.org/10.1098/rspb.2000.1359

Beebee, T. J. C., & Rowe, G. (2008). *An introduction to molecular ecology* (2nd ed.). Oxford, UK; New York, NY: Oxford University Press.

Bink, F. A. (1992). *Ecologische atlas van de dagvlinders van Noordwest-Europa*. Haarlem, the Netherlands: Schuyt.

Boggs, C. L., Watt, W. B., & Ehrlich, P. R. (Eds.) (2003). *Butterflies: ecology and evolution taking flight*. Chicago, IL: University of Chicago Press.

Dennis, R. L. H., Shreeve, T. G., & Van Dyck, H. (2003). Towards a functional resource-based concept for habitat: A butterfly biology viewpoint. *Oikos*, *102*, 417–426.

Dubois, Q., Lebigre, C., Schtickzelle, N., & Turlure, C. (2017). Data from: Sex, size and timing: Sampling design for reliable population genetics analyses using microsatellite data. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.5j13n

Ehrlich, P. R., & Hanski, I. (Eds.) (2004). *On the wings of checkerspots: A model system for population biology*. Oxford, UK; New York, NY: Oxford University Press.

El Mousadik, A., & Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, *92*, 832–839. https://doi.org/10.1007/BF00221895

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). *Introduction to conservation genetics*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511808999

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE*, *7*, e45170. https://doi.org/10.1371/journal.pone.0045170

Hanski, I., & Gaggiotti, O. E. (Eds.) (2004). *Ecology, genetics, and evolution of metapopulations*. Burlington, MA: Elsevier.

Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (4th ed.). Sunderland, MA: Sinauer Associates.

Jarne, P., & Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, *11*, 424–429. https://doi.org/10.1016/0169-5347(96)10049-5

Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, *17*, 4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Kalinowski, S. T. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*, *94*, 33–36. https://doi.org/10.1038/sj.hdy.6800548

Keyghobadi, N., Roland, J., & Strobeck, C. (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology*, *14*, 1897–1909. https://doi.org/10.1111/j.1365-294X.2005.02563.x

Kikuchi, S., & Isagi, Y. (2002). Microsatellite genetic variation in small and isolated populations of *Magnolia sieboldii* ssp. *japonica*. *Heredity*, *88*, 313–321. https://doi.org/10.1038/sj.hdy.6800047

Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, *49*, 725–738.

Lakovic, M., Poethke, H.-J., & Hovestadt, T. (2015). Dispersal timing: Emigration of insects living in patchy environments. *PLoS ONE*, *10*, e0128672. https://doi.org/10.1371/journal.pone.0128672

Madsen, T., Shine, R., Olsson, M., & Wittzell, H. (1999). Restoration of an inbred adder population. *Nature*, *402*, 34–35. https://doi.org/10.1038/46941

Miyamoto, N., Fernández-Manjarrés, J. F., Morand-prieur, M.-E., Bertolino, P., & Frascaria-Lacoste, N. (2008). What sampling is needed for reliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)? *Annals of Forest Science*, *65*, 403. https://doi.org/10.1051/forest:2008014

Mousson, L., Nève, G., & Baguette, M. (1999). Metapopulation structure and conservation of the cranberry fritillary *Boloria aquilonaris* (Lepidoptera, Nymphalidae) in Belgium. *Biological Conservation*, *87*, 285–293. https://doi.org/10.1016/S0006-3207(98)00081-0

Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.

Nieminen, M., Siljander, M., & Hanski, I. (2004). Structure and dynamics of *Melitaea cinxia* metapopulations. In P. R. Ehrlich & I. Hanski (Eds.), *On the wings of checkerspots: A model system for population biology* (pp. 63–91). New York, NY: Oxford University Press.

Nowicki, P., Pepkowska, A., Kudlek, J., Skórka, P., Witek, M., Settele, J., & Woyciechowski, M. (2007). From metapopulation theory to conservation recommendations: Lessons from spatial occurrence and abundance patterns of *Maculinea* butterflies. *Biological Conservation*, *140*, 119–129. https://doi.org/10.1016/j.biocon.2007.08.001

Ohsaki, N. (1980). Comparative population studies of three *Pieris* butterflies, *P. rapae*, *P. melete*, and *P. napi*, living in the same area. II. Utilization of patchy habitats by adults through migratory and non-migratory movements. *Researches on Population Ecology*, *22*, 163–183. https://doi.org/10.1007/BF02513543

Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, *4*, 347–354. https://doi.org/10.1111/j.1365-294X.1995.tb00227.x

Paetkau, D., Slade, R., Burdens, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. *Molecular Ecology*, *13*, 55–65. https://doi.org/10.1046/j.1365-294X.2004.02008.x

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*, 536–539. https://doi.org/10.1093/jhered/esh074

Pruett, C., & Winker, K. (2008). The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. *Journal of Avian Biology*, *39*, 252–256. https://doi.org/10.1111/j.0908-8857.2008.04094.x

Prugnolle, F., & De Meeûs, T. (2002). Inferring sex-biased dispersal from population genetic tools: A review. *Heredity*, *88*, 161–165. https://doi.org/10.1038/sj.hdy.6800060

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radchuk, V., Turlure, C., & Schtickzelle, N. (2013). Each life stage matters: The importance of assessing the response to climate change over the complete life cycle in butterflies. *Journal of Animal Ecology*, *82*, 275–285. https://doi.org/10.1111/j.1365-2656.2012.02029.x

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, *94*, 9197–9201. https://doi.org/10.1073/pnas.94.17.9197

Richards, V. P., Bernard, A. M., Feldheim, K. A., & Shivji, M. S. (2016). Patterns of population structure and dispersal in the long-lived "redwood" of the coral reef, the giant barrel sponge (*Xestospongia muta*). *Coral Reefs*, *35*, 1097–1107. https://doi.org/10.1007/s00338-016-1435-y

Rutowski, R. L. (1991). The evolution of male mate-locating behavior in butterflies. *The American Naturalist*, *138*, 1121–1139. https://doi.org/10.1086/285273

Schtickzelle, N., Turlure, C., & Baguette, M. (2012). Temporal variation in dispersal kernels of the bog fritillary butterfly. In J. Clobert, M. Baguette, T. G. Benton, & J. M. Bullock (Eds.), *Dispersal ecology and evolution* (pp. 231–239). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199608898.001.0001

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*, 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., … Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*, 3189–3194. https://doi.org/10.1093/nar/24.16.3189

Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution*, *14*, 323–327. https://doi.org/10.1016/S0169-5347(99)01637-7

Thomas, J. A. (1995). The ecology and conservation of *Maculinea arion* and other European species of large blue butterfly. In A. S. Pullin (Ed.), *Ecology and conservation of butterflies* (pp. 180–197). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-011-1282-6

Turlure, C., Choutt, J., Van Dyck, H., Baguette, M., & Schtickzelle, N. (2010). Functional habitat area as a reliable proxy for population size: Case study using two butterfly species of conservation concern. *Journal of Insect Conservation*, *14*, 379–388. https://doi.org/10.1007/s10841-010-9269-3

Vandewoestijne, S., Turlure, C., & Baguette, M. (2012). Development of novel microsatellite markers for a specialist species of Lepidoptera, *Boloria aquilonaris* (Nymphalidae), based on 454 sequences. *European Journal of Entomology*, *109*, 129–134. https://doi.org/10.14411/eje.2012.016

Warren, M. S., Hill, J. K., Thomas, J. A., Asher, J., Fox, R., Huntley, B., … Thomas, C. D. (2001). Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, *414*, 65–69. https://doi.org/10.1038/35102054

Wiklund, C., & Fagerström, T. (1977). Why do males emerge before females? A hypothesis to explain the incidence of protandry in butterflies. *Oecologia*, *31*, 153–158. https://doi.org/10.1007/BF00346917

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Dubois Q, Lebigre C, Schtickzelle N, Turlure C. Sex, size and timing: Sampling design for reliable population genetics analyses using microsatellite data. *Methods Ecol Evol*. 2017;00:1–13. https://doi.org/10.1111/2041-210X.12948