Effects of sample size and full sibs on genetic diversity characterization: a case study of three syntopic Iberian pond-breeding amphibians

Gregorio Sánchez-Montes[1,2,4,6] · Arturo H. Ariño[1] · José L. Vizmanos[3] · Jinliang Wang[4] · Íñigo Martínez-Solano[2,5,6*]

[1] Department of Environmental Biology, School of Sciences, University of Navarra, c/ Irunlarrea, 1, 31008 Pamplona, Spain

[2] Museo Nacional de Ciencias Naturales, CSIC, c/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain

[3] Department of Biochemistry and Genetics, School of Sciences, University of Navarra, c/ Irunlarrea, 1, 31008 Pamplona, Spain

[4] Institute of Zoology, Zoological Society of London, London NW1 4RY, UK

[5] Instituto de Investigación en Recursos Cinegéticos (IREC-CSIC-UCLM-JCCM), Ronda de Toledo, s/n, 13071 Ciudad Real, Spain

[6] Ecology, Evolution, and Development Group, Department of Wetland Ecology, Doñana Biological Station, CSIC, c/ Américo Vespucio, s/n, 41092 Seville, Spain

*corresponding author: Íñigo Martínez-Solano.

Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, CSIC, c/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain

E-mail: inigomsolano@gmail.com, Phone: (+34) 91 411 1328, Fax: (+34) 91 564 5078

Running title: Sample size and full sibs in genetic diversity

Abstract

Accurate characterization of genetic diversity is essential for understanding population demography, predicting future trends and implementing efficient conservation policies. For that purpose, molecular markers are routinely developed for non-model species, but key questions regarding sampling design, like calculation of minimum sample sizes or the effect of relatives in the sample, are often neglected. We used accumulation curves and sibship analyses to explore how these two factors affect marker performance in the characterization of genetic diversity. We illustrate this approach with the analysis of an empirical dataset including newly optimized microsatellite sets for three Iberian amphibian species: *Hyla molleri*, *Bufo calamita* and *Pelophylax perezi*. We studied 17-21 populations per species (total *n* = 547, 652 and 516 individuals, respectively), including a reference locality in which the effect of sample size was explored using larger samples (77-96 individuals). As expected, F_{IS} and tests for Hardy-Weinberg equilibrium and linkage disequilibrium were affected by the presence of full sibs, and most initially inferred disequilibria were no longer statistically significant when full siblings were removed from the sample. We estimated that to obtain reliable estimates, the minimum sample size (potentially including full sibs) was close to 20 for expected heterozygosity (H_E), and between 50 and 80 for allelic richness (AR). Our pilot study based on a reference population provided a rigorous assessment of marker properties and the effects of sample size and presence of full sibs in the sample. These examples illustrate the advantages of this approach to produce robust and reliable results for downstream analyses.

Keywords: Accumulation curves, Allelic richness, Diversity profile, Expected heterozygosity, Minimum sample size, Sibship analysis.

Introduction

Accurate characterization of genetic diversity is a key step towards understanding the ecological and evolutionary histories of populations and, consequently, to predict future trends and implement efficient conservation measures (Hamilton 2009; Habel et al. 2015). The continuous improvement of molecular techniques and computation power, associated with the development of model-based statistical analysis methods, are greatly expanding our ability to estimate demographic parameters and the universe of hypotheses that can be tested about genetic processes (Excoffier and Heckel 2006; Buckley 2009; Guichoux et al. 2011). As a consequence, complex questions regarding the detection of cryptic diversity, quantification of gene flow and population status assessment have become approachable in recent times (Broquet and Petit 2009; Segelbacher et al. 2010; Luikart et al. 2010; Marko and Hart 2011; Arntzen et al. 2013; Fahey et al. 2014). In a scenario of global biodiversity loss, the possibility of early identification of genetically impoverished and/or isolated populations is paramount for informing management policies (Tallmon et al. 2004; Scherer et al. 2012). Thus, accurate evaluation of the amount and spatial distribution of genetic diversity is essential for research and conservation issues. For that purpose, new molecular markers are routinely optimized for non-model species (Guichoux et al. 2011; Gallardo et al. 2012; Habel et al. 2014). However, questions of sampling design with potential consequences on the reliability of inferences, like calculation of the minimum sample size or the effect of excessive relatives in the sample, are often neglected.

Different indexes are commonly used to summarize genetic diversity. Most of these indexes rely either on allele counts, like allelic richness (AR), or on allelic frequencies, like observed and expected heterozygosity (H_O and H_E). Indeed, AR and H_E represent two particular cases of a potentially continuous diversity measurement profile, in which rare alleles are more or less accounted for (Chao and Jost 2015). While AR can be more useful to evaluate the evolutionary potential of populations (Petit et al. 1998; Leberg 2002; Pruett and Winker 2008), accurate estimation of allelic and genotypic frequencies is more important for many other downstream analyses (Allendorf and Phelps 1981; Cornuet and Luikart 1996; Jones and Wang 2010a). It has been documented that AR is heavily dependent on sample size (Banks et al. 2000; Foulley and Ollivier 2006; Miyamoto et al. 2008; Pruett and Winker 2008). Comparing AR across populations with different sample sizes is possible by means of rarefaction methods (El Mousadik and Petit 1996; Kalinowski 2004; Pruett and Winker 2008), but

the accuracy of estimates is still limited by the smallest sample in the dataset. In contrast, 20-30 genetic samples have proven sufficient for estimating H_E in some empirical studies (Miyamoto et al. 2008; Pruett and Winker 2008; Hale et al. 2012). However, these studies assessed the 'sufficiency' of sample either visually for separate markers (Hale et al. 2012) or by exploring the approximation to final combined multilocus estimates (Miyamoto et al. 2008; Pruett and Winker 2008). To our knowledge, no method has been applied to calculate threshold-based minimum sample sizes for individual markers, but this information could improve the efficiency of ecological, evolutionary or conservation studies (including long-term genetic monitoring programs) by aiding in the process of marker set selection.

The sufficiency of sample has important implications for the accuracy and precision of genetic estimates, but it is difficult to assess empirically (Fitzpatrick 2009; Alex Buerkle and Gompert 2013; Chao and Jost 2015). In fact, the minimum sample size is marker-, species-, and even population-dependent so it should be addressed through pilot studies, but these are often expensive and time-consuming (Taberlet and Luikart 1999). Alternatively, the performance of genetic markers can be supervised by exploring how cumulative curves approach final estimates obtained from a large sample of a reference population (e.g., Miyamoto et al. 2008). Different measures can be used to characterize the approximation of subsample estimates to final estimates, such as the root mean square error of estimates (Miyamoto et al. 2008; Pruett and Winker 2008) or the successive slopes of the accumulation curve (Chao et al. 2013). Here we adapt a method originally derived for diversity accumulation curves (Ariño et al. 2008) to calculate the minimum sample size required for each marker to estimate AR and H_E. This method could be routinely performed in reference populations to test the suitability of molecular markers to address ecological and conservation questions, and so inform marker set choice and sampling design. In this study, we complement this approach with the calculation of diversity profile curves as proposed in Chao and Jost (2015).

Similarly, the presence of excessive relatives in the sample can also bias population inferences. All natural populations contain relatives, so including relatives is necessary for representative sampling. Unfortunately, knowing the exact proportion of relatives of each class in a wild population is practically impossible. Therefore, it is difficult to assess whether a sample, even with known or inferred genealogical relationships among individuals, represents the population from which it was drawn (Waples and Anderson 2017). In samples with an excess of relatives, alleles present in large (or small) families might be over- (or under-) represented, thus leading to inaccurate estimation

of population allelic frequencies (Jourdan-Pineau et al. 2012). An excess (compared to random sampling) of relatives in the sample is a frequent problem when tissue sampling is performed among early stage individuals in iteroparous species with overlapping generations, a scenario in which the aggregation of single cohort relatives (especially full sibs) is common in many taxa (Goldberg and Waits 2010). Estimates obtained from such samples may not be representative of the whole population, which can sometimes lead to biased conclusions (Anderson and Dunham 2008; Goldberg and Waits 2010; Rodríguez-Ramilo and Wang 2012; Rodríguez-Ramilo et al. 2014). It has been suggested that removing siblings from the samples can reduce bias in unsupervised Bayesian clustering programs such as STRUCTURE (Anderson and Dunham 2008; Rodríguez-Ramilo and Wang 2012), although this approach might often be counter-productive in certain circumstances (Waples and Anderson 2017). However, the effect of removing full sibs from genetic samples on genetic diversity indexes (such as AR and H_E) and in commonly employed tests of genotypic proportions such as Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) has not been explored in wild populations.

Here we introduce a method for calculating the minimum sample size required to assess the genetic diversity at each individual marker in a dataset, and explore the effect of full sibs on genetic diversity characterization. We used specifically optimized microsatellite markers to score multilocus genotypes for three co-distributed pond-breeding amphibians: the Iberian treefrog (*Hyla molleri*), the Natterjack toad (*Bufo calamita*) and the Iberian green waterfrog (*Pelophylax perezi*). These three species are iteroparous, with overlapping generations, and molecular protocols are required to obtain information about their demography, mating system and genetic structure. We estimated several genetic diversity indexes in 17-21 populations per species and assessed the effect of the presence of full sibs in the samples by comparing results including or excluding full sibs. We also used large samples (n = 77-96 individuals) in a reference population where the three species co-occur to explore the effect of sample size on single-locus AR and H_E estimates and to calculate minimum required sample sizes for each marker. We discuss the benefits of this approach for establishing efficient sampling design protocols in conservation genetics studies.

Materials and methods

Tissue sampling

Between 2010 and 2015 we collected larval tissue samples of *H. molleri* (*n* = 547), *B. calamita* (*n* = 652) and *P. perezi* (*n* = 516) in 17-21 localities per species along both slopes of Sierra de Guadarrama, in the Iberian Central System, encompassing different habitat types and with altitudes ranging between 875 and 1720 m.a.s.l. (see Table 1 and Fig. 1). In one of the localities (Valdemanco) we collected 77 to 96 tadpoles of each of the three species. In the remaining locations, 19 to 36 tadpoles per locality were collected (Table 1). For each species in each locality, we used nets to sample larvae from the same year cohort. Surveys were performed uniformly throughout the water surface and samples included individuals of different body sizes, to minimize potential sampling biases arising from the aggregative behavior of full sib tadpoles. Small tadpoles were euthanized and preserved in absolute ethanol. In the case of large tadpoles, tail tips were clipped and stored in absolute ethanol for subsequent DNA extraction, and larvae were released back in the same pond of capture.

DNA extraction and genotyping

Two enriched partial genomic libraries, one for *H. molleri* and another for *B. calamita*, were prepared at the Sequencing Genotyping Facility, Cornell Life Sciences Core Laboratory Center (CLC) (New York, NY) following the method described in Gutiérrez-Rodríguez & Martínez-Solano (2013). They were generated from DNA of one tadpole of *H. molleri* collected in Arzila, Portugal (40.20º N, 8.65º W) and one adult male of *B. calamita* collected in Valdemanco, central Spain (40.85º N, 3.64º W). From each of the two libraries, 60 loci containing microsatellite motifs (30 trimers and 30 tetramers) between 5 and 12 repetitions long were selected for further screening. Although some tri-nucleotides might be under selection, we don't expect that it would dramatically affect our results, except if selection was very strong, which is highly unlikely. This notion was further supported by the similar polymorphism and diversity profiles shown by the tri- and tetra-nucleotide loci in this work (see Table 2 and Supplementary Information), and also by other demographic analyses performed with different subsets of loci (data not shown). For DNA purification, optimization of multiplex reactions, genotyping

and allele scoring, we followed the methods described in Sánchez-Montes et al. (2016). Final sets of markers consisted of 18 and 16 newly developed microsatellite loci for *H. molleri* and *B. calamita*, respectively (see Supplementary Material), and 15 previously optimized markers for *P. perezi* (Sánchez-Montes et al. 2016). These sets of markers were used to genotype the samples of each species. We selected a subsample for repetition of the DNA amplification process (between 3.7% and 17.8% of the sample in each species) to check for consistency of genotype calling.

Characterization of genetic diversity and effect of full sibs

For characterization of genetic diversity, allelic richness (AR), observed (H_O) and expected (H_E) heterozygosity and F_{IS} were calculated for each locus in each population using GENALEX 6.5 (Peakall and Smouse 2006). Tests for departures from HWE and evidence of LD were performed with GENEPOP v.4.3 (Raymond and Rousset 1995; Rousset 2008), with 10,000 dememorisation steps, 1,000 batches and 10,000 iterations per batch. The Bonferroni sequential correction was applied to account for multiple testing (Rice 1989). The presence of null alleles was assessed with MICRO-CHECKER v.2.2.3 (Van Oosterhout et al. 2004). We calculated the information content of the markers by means of their informativeness for genetic relationship (R Info) using the software KININFOR (Wang 2006). The other information indexes calculated by the program were highly correlated with R Info in the three species (data not shown). Sibship analyses were performed in COLONY (Jones and Wang 2010b) to identify full sibs in each locality and to infer mistyping rates due to allele dropout and false allele scoring. All analyses for genetic diversity characterization were conducted both on the original genotype data (referred to as the complete samples) and on the data after excluding all but one of the identified full sibs in every full sib family from each population (referred to as the reduced samples).

Effect of sample size

We explored the effect of sample size on the estimates of AR and H_E for each locus in the locality with the highest total sample size for the three species (i.e., the complete samples from locality Valdemanco, see Table 1). In order to compute approximate 95% confidence intervals for final AR and H_E estimates (i.e., for the estimates obtained with the complete samples), we randomly produced

10,000 bootstrap samples for each locus, each with the same number of individuals as the whole sample. We calculated Simpson's complementary diversity index in PAST v. 3 (Hammer et al. 2001); this index is identical to H_E ($H_E \equiv 1\text{-}D_S$, where D_S is Simpson's dominance). We also quantified the rate of approximation to final AR and H_E obtained by the molecular markers with increasing sample size, using 10,000 jackknifed subsamples of one, two, three...*n* individuals of the total sample, and obtained AR and H_E accumulation curves for each locus. We used diversity functions in ESTIMATES (v.9.1.0, Colwell 2013, http://purl.oclc.org/estimates); this software provides expected *S* that is identical to AR, and $1/D_S$ that we converted to H_E.

We used R (R Development Core Team 2008) to inspect the accumulation curves looking for asymptotic stabilization of AR and H_E (see Supplementary Figures S1-3). Our criterion for defining "sufficient samples" was to minimize a Type-II (β) error (Snedecor and Cochran 1989) by selecting the first point along the section of curve that would persistently exceed the lower bound of the (bootstrapped) confidence interval of the final estimate, while no further points would consistently fall below. We summarized in boxplots the observed minimal sample sizes for each locus necessary to approximate final estimates of AR and H_E. For comparison with our results, we also obtained empirical and Chao's diversity profiles for each marker for values $0 \leq q \geq 3$, by adapting the R script in Chao and Jost (2015, Appendix 8). The parameter q defines the sensitivity of the diversity estimate to the rarest categories in the sample, and most of the variation in the diversity profile is expected to be comprised within the interval q = [0,3] (Chao and Jost 2015). The empirical profile at q = 0 corresponds to AR measured as the total number of alleles (analogous to species richness 'S' in Chao and Jost 2015), and at q = 2 it approximates the Simpson's diversity index (Chao, Ma and Hsieh 2015), which is complementary to H_E, as stated above.

Results

Characterization of genetic diversity and effect of full sibs

Almost all microsatellite markers were polymorphic in nearly all sampled populations (see Supplementary Tables S4-S6). The only exception was *Hmol3.7*, which was monomorphic in all

populations except for CAN (see Supplementary Table S4). This marker might result more informative at larger-scale studies and so we decided to describe it within the multiplex, although neither average F_{IS}, nor minimum sample sizes could be calculated for this locus. Genetic diversity measures obtained with the reduced samples were very similar to those obtained with the complete samples (see Supplementary Tables S4-6), although F_{IS} estimates changed slightly (see Table 2 and Supplementary Appendix S1). F_{IS} and the allelic dropout rate (inferred from COLONY analyses) were highly correlated in the three species (*H. molleri*: Spearman's rho = 0.57, p = 0.015; *B. calamita*: rho = 0.85, p < 0.001; *P. perezi*: rho = 0.70, p = 0.005) although the trend was clearer in *B. calamita*, which showed the highest variance in the values of both F_{IS} and allelic dropout rate (see Supplementary Figure S7). However, F_{IS} was not correlated with false allele rate in any of the three species.

Four markers of the *H. molleri* set and five markers of the *P. perezi* set showed significant deviations from HWE in more than one population in the complete samples (Table 2). However, after removing full sibs from the samples, no locus departed from HWE in more than one population (out of 20 and 17 total localities of *H. molleri* and *P. perezi*, respectively, see Table 2 and Supplementary Tables S4 and S6). Only one marker (out of 18 total loci) in the *H. molleri* set (*Hmol3.15*) and four loci (out of 15 total loci) in the *P. perezi* set (*Pper4.7*, *Pper3.1*, *Pper3.23* and *Pper4.24*) showed evidence of null alleles in more than one population (three populations at most), and these effects mostly remained after removing full sibs from the sample (see Supplementary Tables S4 and S6, respectively). In contrast, almost all loci in the *B. calamita* set were found to be out of HWE in some populations when using the complete samples. Five of them (*Bcal4.6*, *Bcal4.14*, *Bcal4.2*, *Bcal3.26* and *Bcal3.19*, out of 16 total loci) still showed departures from HWE in 4-11 populations after removing full sibs from the samples (Table 2 and Supplementary Table S5). According to MICRO-CHECKER results, these five loci, as well as *Bcal4.21*, showed evidence of null alleles in many populations (see Supplementary Table S5).

A few pairs of loci showed evidence of linkage disequilibrium (LD) across some populations in the complete datasets after applying the Bonferroni correction. One, three and three pairs of loci were in LD in more than 20% of the populations in the *H. molleri*, *B. calamita* and *P. perezi* sets respectively. The most widespread disequilibrium involved markers *Bcal4.20* and *Bcal3.26*, which were in LD in 14 populations. However, none of these disequilibria remained significant in the reduced samples (data not shown).

Effect of sample size

Minimum sample sizes required for approaching final estimates of AR and H_E in each locus are summarized in the boxplots of Fig. 2. Median values ranged between 50 and 80 individuals for characterization of AR in each species, while less than 20 individuals were sufficient to estimate H_E. Minimum sample sizes required for estimation of H_E were highly correlated with marker polymorphism, measured as AR, in the three species (*H. molleri*: Spearman's rho = 0.79, $p < 0.001$; *B. calamita*: rho = 0.67, p = 0.005; *P. perezi*: rho = 0.74, p = 0.002, see Fig. 2). In contrast, minimum sample sizes required for estimation of AR were negatively correlated with marker AR, although only significantly in the case of *B. calamita* (*H. molleri*: rho = -0.28, p = 0.270; *B. calamita*: rho = -0.61, p = 0.013; *P. perezi*: rho = -0.09, p = 0.738, see Fig. 2). Loci in the three marker sets showed different diversity profiles (see Supplementary Figures S4-6). The least polymorphic loci in each set showed flat profiles, but the most polymorphic loci showed a more or less decreasing function along the range of q. Profiles obtained following Chao's correction for sampling bias were very similar to empirical profiles in most cases, although some highly polymorphic loci showed some differences at q = 0, like *Hmol4.8* (15 observed alleles vs. 23 alleles estimated by Chao's correction), *Bcal4.26* (42 vs. 45) or *Pper4.7* (20 vs. 28, see Supplementary Figures S4-6).

Discussion

A thorough empirical assessment of marker polymorphism and performance is a key step to evaluate their adequacy for genetic diversity characterization and therefore to inform marker set choice for future studies (Matson et al. 2008; Queirós et al. 2015). The moderate to high polymorphism observed in our marker sets (Table 2) suggests that a high power of resolution could be obtained by combining a subset of the most polymorphic markers in a single (or two) multiplex reaction(s), which might be useful e.g. for management purposes (Cornuet and Luikart 1996; Holleley and Geerts 2009; Harrison et al. 2013; Queirós et al. 2015). However. in studies including genetically impoverished regions, for instance near range borders (Rowe et al. 1999; Edenhamn et al. 2000; Allentoft et al. 2009), more loci could be necessary to obtain similar power of resolution, and these loci could be selected from each

set after testing their degree of polymorphism in the area of interest. Marker set composition should therefore be informed before addressing the study, to guarantee unbiased comparison among populations (i.e., using the same marker set for all sampling localities) while also avoiding problems caused by insufficient marker information. Mistyping rates are also essential to assess the practical utility of newly developed markers, but this information is often overlooked (Pompanon et al. 2005; Lampa et al. 2013). Inferred error rates in our markers rarely exceeded 0.05, except for the six markers of *B. calamita* in which we also detected evidence of null alleles (Supplementary Tables S1-3). These markers showed dropout rates between 0.09 and 0.32 (Supplementary Table S2). In all three species, dropout rates inferred by COLONY were highly correlated with F_{IS}, but this trend was more obvious in the case of *B. calamita* than in *H. molleri* and *P. perezi*, because larger variance was observed in the former species (Supplementary Fig. S7). These results highlight the usefulness of pedigree reconstruction in COLONY for the estimation of error rates since they are in agreement with HWE tests, which are based on F_{IS} (Waples 2015).

Our analyses of marker genotypes across many populations allowed assessing the effect of sampling full sibs on estimates of genetic diversity, which may be problematic when pedigree information is not available (Allendorf and Phelps 1981; Goldberg and Waits 2010). We identified full sibs in each population after reconstruction of one- or two-generation pedigrees (Jones and Wang 2010a) and found that samples from some localities were mostly composed of full sibs (see Table 1), thus potentially misleading some downstream analyses (Anderson and Dunham 2008; Jourdan-Pineau et al. 2012; Rodríguez-Ramilo and Wang 2012). However, removing all relatives from the sample is not always a good solution, because the degree of nonrandomness (with respect to sibship frequency) in empirical samples is unknown (Waples and Anderson 2017). More theoretical work, coupled with empirical data, is needed to derive guidelines about how best to account for this factor. Here we report some preliminary conclusions drawn from both theoretical (see Supplementary Appendix S1) and empirical work, with consistent results across species and populations.

The presence of full sibs in our samples did not significantly affect estimates of genetic diversity (AR, H_O and H_E), although there were slight variations in F_{IS} estimates (Table 2). Theoretically, full sibs in the sample are expected to affect the genotype distributions (see Supplementary Appendix S1). For this reason, F_{IS}, HWE and LD are most affected, although the pattern of change is complex and dependent on the mating system (Goldberg and Waits 2010). As

expected, tests for HWE and LD were strongly affected by the presence of full sibs in the samples (Waples 2015), and most initially inferred disequilibria were no longer significant after removing full sibs (Table 2). While this could also be caused by the lower statistical power in some reduced samples due to reduced sample sizes, some consistent departures from HWE were still detected in many reduced samples of *B. calamita* (see Supplementary Tables S4-6). Five loci (*Bcal 4.6*, *Bcal4.14*, *Bcal4.2*, *Bcal3.26* and *Bcal3.19*) departed from the expected HWE in more than 15% of populations in the reduced samples. Disequilibria in these five loci, as well as in *Bcal4.21*, were probably due to the presence of null alleles, as indicated by analyses with MICRO-CHECKER (Supplementary Table S5). These six markers are highly informative and can be useful in some analyses accounting for genotyping errors (such as sibship analyses in COLONY), but otherwise they should only be used when downstream analyses are robust to violation of HWE assumptions. Altogether, these results suggest that genetic diversity indexes (AR, H_O, H_E) are not affected by the presence of close relatives in the sample, at least in the absence of strongly unbalanced data structure (i.e. when there are not very large families combined with unrelated individuals in the same sample), such as in our case (see also Waples and Anderson 2017). In contrast, the presence of close relatives in the sample strongly affects the results of tests of HWE and LD, especially in small samples/populations.

On the other hand, accounting for the minimum sample size required for genetic diversity characterization is crucial for the accuracy of results and the efficient design of monitoring programs (Wang 2002). Here we have adapted methods based on diversity accumulation curves (Ariño et al. 1996; Ariño et al. 2008) by observing the rate at which jackknifed subsamples approach the confidence interval of bootstrapped replicates of the entire dataset and can no longer be statistically separated from each other at a pre-specified significance level (see Supplementary Figs. S1-3). Our threshold criterion was useful for defining a realistic minimum sample size in most markers, although it was dependent on the width of the 95% confidence interval (CI) of final estimates. As a consequence, in the case of markers with very narrow 95% CI, large sample sizes were required to reach the lower bound of the 95% CI. This resulted in an artificially inflated minimum sample size for AR estimation in some markers (see, for example, *Hmol3.3*, *Bcal3.19* or *Pper3.24* in Supplementary Figs. S1-3). Conversely, for some indexes with a very wide 95% CI, inferred minimum sample sizes were artificially low (e.g., H_E curves for *Hmol3.9*, *Bcal3.22* or *Pper3.22* in Supplementary Figs. S1-3). Too wide (or narrow) 95% CIs in highly (or very little) polymorphic loci probably caused the negative relationship

between AR and the minimum sample size for AR estimation (Fig. 2). These problems associated with the calculation of minimum sample sizes are in essence caused by the potentially continuous diversity measurement profile (i.e., the parameter *q*), in which choosing a particular threshold value (e.g. choosing between AR or H_E to characterize genetic diversity) necessarily involves some degree of arbitrariness.

Furthermore, although our total sample sizes in Valdemanco can be considered large enough to characterize genetic diversity in pond-breeding amphibian populations, our final estimates cannot be taken as actual population parameters. As a consequence, these minimum sample sizes cannot be regarded as generally applicable to other systems. Rather, our goal is double: to encourage the general use of a simple method to explore the rate of approximation to final genetic diversity estimates with cumulative sample size (such as those applied in Miyamoto et al. 2008, Pruett and Winker 2008, Hale et al. 2012, Chao and Jost 2015, or in this paper), and to empirically calculate minimum sample size. Our method could be easily adapted to sequential sampling schemes where additional individuals are genotyped, and their alleles added to the pool at each step. Thus, additional sampling is no longer necessary when the added individual(s) do not significantly improve the estimates of AR and H_E. This way, minimum sample sizes can be defined when required (e.g., for the design of sampling protocols). Nevertheless, since AR and H_E are two particular cases of the continuous diversity measurement, we also followed Chao and Jost (2015)'s proposal of reporting the continuous diversity profile at the most relevant values of *q*. As expected, the most polymorphic loci in our datasets also showed more rare alleles and, as a consequence, their diversity profile varied through the range of *q*. In contrast, the profiles of the least polymorphic loci were largely flat (Supplementary Figs. S4-6). This is in agreement with the observed positive correlation between marker polymorphism and the minimum sample size required for H_E estimation (Fig. 2). Empirical profiles were markedly similar to Chao's profiles in most markers, suggesting that our empirical accumulation curves of AR and H_E did not dramatically underestimate diversity (Supplementary Figs. S4-6). However, some differences arose between both profiles in some markers with alleles at low frequencies, like *Hmol*4.8, *Bcal*4.2, *Bcal*4.26 or *Pper*4.7 (Supplementary Figs. S4-6), which concordantly showed wide 95% CIs in their corresponding accumulation curves for AR estimation (Supplementary Figs. S1-3). Highly polymorphic loci are usually associated with rare alleles, and therefore higher sample sizes are required to estimate AR (but not necessary H_E) with this markers. These results support the

usefulness of our method for reliable minimum sample size calculation and also for detecting possible diversity underestimations caused by loci with rare alleles.

Our results highlight that the presence of full sibs can slightly alter F_{IS} estimates and affect tests of HWE and LD. We proved that some disequilibria are no longer significant after removing full sibs from the samples, therefore allowing detection of truly problematic markers (e. g. those presenting null alleles). The minimum sample size is dependent on the marker(s) selected and should also be assessed in each case for the configuration of the final marker set (Harrison et al. 2013). The required sample size for genetic diversity characterization can be optimized from an exhaustively sampled population by means of accumulation curves and some threshold criterion. This methodology is easy to apply to any empirical dataset and can be readily used to help design sampling protocols for genetic monitoring studies. These two aspects are basic for the efficient design of ecological studies aiming to obtain reliable and comparable inferences about demography and genetic diversity distribution in non-model species.

Funding

This work was supported by *Ministerio de Ciencia e Innovación*, Spain, and FEDER (grant number CGL2008-04271-C02-01/BOS); and *Ministerio de Economía y Competitividad*, Spain, and FEDER (grant number CGL2011-28300) to IMS, who was also supported by funding from the Spanish Severo Ochoa Program (SEV-2012-0262). G. Sánchez-Montes is funded by a predoctoral grant provided by the *Asociación de Amigos de la Universidad de Navarra* and benefited from funding from the *Programa de ayudas de movilidad de la Asociación de Amigos de la Universidad de Navarra*.

Acknowledgements

We thank M. Peñalver, L. San José, J. Gutiérrez, E. Iranzo, L. Carrera, C. Valero, M. Rojo, G. Rodríguez, J. Agüera, A. Sabalza, N. Escribano, R. Goñi, R. Santiso and I. Miqueleiz for help during fieldwork. R. Waples, W.B. Sherwin and an anonymous reviewer provided valuable comments on a

previous version of this manuscript. Authorizations for animal tissue sampling were provided by *Consejería de Medio Ambiente y Ordenación del Territorio. Comunidad de Madrid (Spain)*.

Data accessibility

Sequences of contigs containing newly developed microsatellite loci were deposited in the NCBI GenBank (accession numbers to be added upon manuscript acceptance). The dryad archive (doi:XXXXXXX) contains microsatellite genotype data for the three species.

References

Alex Buerkle C, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? Mol Ecol 22:3028–3035. doi: 10.1111/mec.12105

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. Can J Fish Aquat Sci 38:1507–1514.

Allentoft ME, Siegismund HR, Briggs L, Andersen LW (2009) Microsatellite analysis of the natterjack toad (*Bufo calamita*) in Denmark: populations are islands in a fragmented landscape. Conserv Genet 10:15–28.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. Mol Ecol Resour 8:1219–1229. doi: 10.1111/j.1755-0998.2008.02355.x

Ariño AH, Belascoáin C, Jordana R (1996) Determination of minimal sampling for soil fauna by asymptotic biodiversity accumulation. In: *XII International Colloquium on Soil Zoology*. Dublin, Ireland. 21-26 July 1996.

Ariño AH, Belascoáin C, Jordana R (2008) Optimal sampling for complexity in soil ecosystems. In: Minai AA, Bar-Yam Y (eds) Unifying themes in complex systems IV. Springer, Berlin Heidelberg, pp 222–230.

Arntzen JW, Recuero E, Canestrelli D, Martínez-Solano I (2013) How complex is the *Bufo bufo* species group? Mol Phylogenet Evol 69:1203–1208. doi: 10.1016/j.ympev.2013.07.012

Banks MA, Rashbrook VK, Calavetta MJ, et al (2000) Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Can J Fish Aquat Sci 57:915–927.

Broquet T, Petit EJ (2009) Molecular estimation of dispersal for ecology and population genetics. Annu Rev Ecol Evol Syst 40:193–216. doi: 10.1146/annurev.ecolsys.110308.120324

Buckley D (2009) Toward an organismal, integrative, and iterative phylogeography. BioEssays 31:784–793. doi: 10.1002/bies.200800162

Chao A, Jost L (2015) Estimating diversity and entropy profiles via discovery rates of new species. Methods Ecol Evol 6:873–882. doi: 10.1111/2041-210X.12349

Chao A, Ma KH, Hsieh TC (2015) The online program SpadeR: Species-richness prediction and diversity estimation in R. Program and User's Guide.

Chao A, Wang YT, Jost L (2013) Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. Methods Ecol Evol 4:1091–1100. doi: 10.1111/2041-210X.12108

Colwell RK (2013) EstimateS: Statistical estimation of species richness and shared species from samples.

Cornuet J-M, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

Edenhamn P, Höggren M, Carlson A (2000) Genetic diversity and fitness in peripheral and central populations of the European tree frog *Hyla arborea*. Hereditas 133:115–122.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinose* (L. Skeels)] endemic to Morocco. Theor Appl Genet 92:832–839.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. Nat Rev Genet 7:745–758. doi: 10.1038/nrg1904

Fahey AL, Ricklefs RE, Dewoody JA (2014) DNA-based approaches for evaluating historical demography in terrestrial vertebrates. Biol J Linn Soc 112:367–386. doi: 10.1111/bij.12259

Fitzpatrick BM (2009) Power and sample size for nested analysis of molecular variance. Mol Ecol 18:3961–3966.

Foulley JL, Ollivier L (2006) Estimating allelic richness and its diversity. Livest Sci 101:150–158. doi: 10.1016/j.livprodsci.2005.10.021

Gallardo C, Correa C, Morales P, et al (2012) Validation of a cheap and simple nondestructive method for obtaining AFLPs and DNA sequences (mitochondrial and nuclear) in amphibians. Mol Ecol Resour 12:1090–1096. doi: 10.1111/1755-0998.12013

Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. Mol Ecol Resour 10:304–313. doi: 10.1111/j.1755-0998.2009.02755.x

Guichoux E, Lagache L, Wagner S, et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611. doi: 10.1111/j.1755-0998.2011.03014.x

Gutiérrez-Rodríguez J, Martínez-Solano I (2013) Isolation and characterization of sixteen polymorphic microsatellite loci in the Western Spadefoot, *Pelobates cultripes* (Anura: Pelobatidae) via 454 pyrosequencing. Conserv Genet Resour 5:981–984. doi: 10.1007/s12686-013-9948-y

Habel JC, Husemann M, Finger A, et al (2014) The relevance of time series in molecular ecology and conservation biology. Biol Rev 89:484–492. doi: 10.1111/brv.12068

Habel JC, Zachos FE, Dapporto L, et al (2015) Population genetics revisited – towards a multidisciplinary research field. Biol J Linn Soc 115:1–12. doi: 10.1111/bij.12481

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:e45170. doi: 10.1371/journal.pone.0045170

Hamilton MB (2009) Population genetics. Wiley-Blackwell, Chinchester, West Sussex.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron 4:9pp.

Harrison HB, Saenz-Agudelo P, Planes S, et al (2013) Relative accuracy of three common methods of parentage analysis in natural populations. Mol Ecol 22:1158–1170. doi: 10.1111/mec.12138

Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans

and optimizes multiplex PCR. Biotechniques 46:511–517. doi: 10.2144/000113156

Jones OR, Wang J (2010a) Molecular marker-based pedigrees for animal conservation biologists. Anim Conserv 13:26–34. doi: 10.1111/j.1469-1795.2009.00324.x

Jones OR, Wang J (2010b) COLONY: A program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour 10:551–555. doi: 10.1111/j.1755-0998.2009.02787.x

Jourdan-Pineau H, Folly J, Crochet P-A, David P (2012) Testing the influence of family structure and outbreeding depression on heterozygosity-fitness correlations in small populations. Evolution (N Y) 66:3624–3631. doi: 10.5061/dryad.bb33v

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierachical sampling designs. Conserv Genet 5:539–543.

Lampa S, Henle K, Klenke R, et al (2013) How to overcome genotyping errors in non-invasive genetic mark-recapture population size estimation - A review of available methods illustrated by a case study. J Wildl Manage 77:1490–1511. doi: 10.1002/jwmg.604

Leberg PL (2002) Estimating allelic richnes: effects of sample size and bottlenecks. Mol Ecol 11:2445–2449. doi: 10.1046/j.1365-294X.2002.01612.x

Luikart G, Ryman N, Tallmon DA, et al (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. Conserv Genet 11:355–373. doi: 10.1007/s10592-010-0050-7

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. Trends Ecol Evol 26:448–456. doi: 10.1016/j.tree.2011.05.007

Matson SE, Camara MD, Eichert W, Banks MA (2008) P-LOCI: a computer program for choosing the most efficient set of loci for parentage assignment. Mol Ecol Resour 8:765–768. doi: 10.1111/j.1755-0998.2008.02128.x

Miyamoto N, Fernández-Manjarrés JF, Morand-Prieur M-E, et al (2008) What sampling is needed for reliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)? Ann For Sci 65:403. doi: 10.1051/forest:2008014

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for consevation on the basis of genetic markers. Conserv Biol 12:844–855. doi: 10.1046/j.1523-1739.1998.96489.x

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859. doi: 10.1038/nrg1707

Pruett CL, Winker K (2008) The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. J avian Biol 39:252–256. doi: 10.1111/j.2008.0908-8857.04094.x

Queirós J, Godinho R, Lopes S, et al (2015) Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. Mol Ecol Resour 15:747–760.

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Rice WR (1989) Analyzing tables of statistical tests. Evolution (N Y) 43:223–225.

Rodríguez-Ramilo ST, Toro MA, Wang J, Fernández J (2014) Improving the inference of population

genetic structure in the presence of related individuals. Genet Res 96:e003. doi: 10.1017/S0016672314000068

Rodríguez-Ramilo ST, Wang J (2012) The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. Mol Ecol Resour 12:873–884. doi: 10.1111/j.1755-0998.2012.03156.x

Rousset F (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Rowe G, Beebee TJC, Burke T (1999) Microsatellite heterozygosity, fitness, and demography in natterjack toads *Bufo calamita*. Anim Conserv 2:85–92.

Sánchez-Montes G, Recuero E, Gutiérrez-Rodríguez J, et al (2016) Species assignment in the *Pelophylax ridibundus* x *P. perezi* hybridogenetic complex based on 16 newly characterised microsatellite markers. Herpetol J 26:101–110.

Scherer RD, Muths E, Noon BR, Oyler-McCance SJ (2012) The genetic structure of a relict population of wood frogs. Conserv Genet 13:1521–1530. doi: 10.1007/s10592-012-0395-1

Segelbacher G, Cushman SA, Epperson BK, et al (2010) Applications of landscape genetics in conservation biology: concepts and challenges. Conserv Genet 11:375–385. doi: 10.1007/s10592-009-0044-5

Snedecor GW, Cochran WG (1989) Statistical Methods, 8th edn. Iowa State College Press.

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. Biol J Linn Soc 68:41–53.

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends Ecol Evol 19:489–496. doi: 10.1016/j.tree.2004.07.003

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x

Wang J (2006) Informativeness of genetic markers for pairwise relationship and relatedness inference. Theor Popul Biol 70:300–321.

Wang J (2002) An estimator for pairwise relatedness using molecular markers. Genetics 160:1203–1215.

Waples RS (2015) Testing for Hardy-Weinberg proportions: have we lost the plot? J Hered 106:1–19. doi: 10.1093/jhered/esu062

Waples RS, Anderson EC (2017) Purging putative siblings from population genetic data sets: a cautionary view. Mol Ecol. doi: 10.1111/mec.14022.

Table 1. List of localities included in the present study. For each locality, the abbreviation (Abr), geographic coordinates and sample sizes for each species including (and excluding) full sibs are displayed.

Locality	Abr	Coordinates	*H. molleri*	*B. calamita*	*P. perezi*
Alameda del Valle	ALA	40.91º N 3.85º W	-	24 (13)	-
Arcones	ARC	41.13º N 3.73º W	30 (27)	-	19 (14)
Berrocal	BRC	41.06º N 3.98º W	-	30 (6)	-
Boceguillas	BOC	41.31º N 3.66º W	-	20 (1)	-
Bustarviejo	BUS	40.85º N 3.68º W	30 (29)	28 (19)	30 (17)
Cabanillas de la Sierra[a]	CAB	40.85º N 3.65º W	22 (19)	30 (26)	20 (20) / 27 (20) / 30 (15)
Cerceda	CER	40.72º N 3.96º W	20 (16)	30 (14)	23 (18)
Collado Hermoso	HER	41.05º N 3.93º W	23 (7)	-	32 (28)
Colmenar Viejo	COL	40.69º N 3.83º W	21 (18)	30 (7)	-
Dehesa de Roblellano	ROB	40.86º N 3.63º W	30 (20)	36 (33)	23 (4)
El Berrueco	BER	40.93º N 3.57º W	21 (18)	29 (3)	20 (8)
Fuenterrebollo	FUE	41.33º N 3.93º W	20 (12)	-	20 (10)
Gargantilla del Lozoya	GAR	40.95º N 3.72º W	-	30 (27)	-
Gascones	GAS	41.01º N 3.65º W	21 (19)	-	-
La Pradera de Navalhorno	PRA	40.88º N 4.03º W	22 (9)	30 (11)	23 (19)
Lozoyuela	LOZ	40.92º N 3.65º W	-	28 (17)	-
Medianillos	MED	40.76º N 3.68º W	21 (9)	-	25 (20)
Muñoveros	MUN	41.20º N 3.95º W	-	32 (16)	-
Navafría	NAV	41.06º N 3.83º W	-	30 (10)	-
Navalafuente	NVL	40.81º N 3.68º W	-	30 (5)	-
Puerto de Canencia	CAN	40.87º N 3.76º W	25 (22)	28 (26)	22 (19)
Puerto de La Morcuera	MOR	40.84º N 3.83º W	30 (24)	20 (11)	22 (15)
Puerto del Medio Celemín	CEL	40.88º N 3.66º W	-	30 (21)	-
Rascafría	RAS	40.85º N 3.91º W	20 (18)	-	22 (20)
Santo Tomé del Puerto	STO	41.19º N 3.59º W	-	30 (8)	21 (17)
Sauquillo de Cabezas	SAU	41.19º N 4.06º W	20 (12)	-	22 (10)
Soto del Real	SOT	40.76º N 3.80º W	20 (18)	30 (14)	-
Torrecaballeros	TOR	41.00º N 4.02º W	34 (28)	-	-
Turrubuelo	TUR	41.32º N 3.59º W	21 (19)	-	21 (15)
Valdemanco	VAL	40.85º N 3.64º W	96 (88)	77 (27)	94 (58)

[a]In Cabanillas de La Sierra, three samples of *P. perezi* were obtained in different years (2010 / 2013 / 2014).

Table 2. Mean (and standard deviation) of several indexes averaged across all sampled populations for every marker in each species. For those measures affected by the presence of full sibs in the sample (F_{IS}, HW), estimates obtained in the reduced samples are also displayed for comparison. AR = allelic richness, H_O and H_E = observed and expected heterozygosity. HW: number of populations in which significant departures from Hardy-Weinberg equilibrium were detected in each marker.

		Complete samples					Reduced samples	
Species	**Marker**	**AR**	H_O	H_E	F_{IS}	**HW**	F_{IS}	**HW**
H. molleri	*Hmol3.7*	1.05 (0.22)	0 (0.01)	0 (0.01)	-	0	-	0
	Hmol3.28	3.95 (1.1)	0.71 (0.15)	0.61 (0.11)	-0.17 (0.23)	2	-0.18 (0.23)	0
	Hmol4.2	2.75 (0.64)	0.5 (0.12)	0.45 (0.1)	-0.12 (0.15)	0	-0.14 (0.18)	0
	Hmol3.9	2.95 (0.94)	0.32 (0.19)	0.31 (0.16)	-0.01 (0.22)	0	-0.02 (0.25)	0
	Hmol3.3	3.05 (0.6)	0.37 (0.15)	0.35 (0.12)	-0.04 (0.17)	0	-0.02 (0.23)	0
	Hmol4.12	10.8 (3.09)	0.86 (0.12)	0.81 (0.1)	-0.06 (0.12)	1	-0.05 (0.13)	0
	Hmol4.16	8.5 (2.69)	0.83 (0.11)	0.78 (0.08)	-0.05 (0.1)	0	-0.09 (0.12)	0
	Hmol4.1	7.6 (2.19)	0.79 (0.09)	0.75 (0.07)	-0.06 (0.09)	1	-0.06 (0.1)	0
	Hmol4.9	4.3 (1.08)	0.65 (0.12)	0.6 (0.08)	-0.09 (0.17)	0	-0.05 (0.18)	0
	Hmol4.10	9.05 (2.93)	0.86 (0.09)	0.81 (0.06)	-0.06 (0.07)	2	-0.05 (0.08)	0
	Hmol3.22	6.3 (1.26)	0.8 (0.12)	0.75 (0.07)	-0.07 (0.13)	1	-0.06 (0.15)	0
	Hmol4.22	2.05 (0.39)	0.34 (0.18)	0.3 (0.15)	-0.1 (0.18)	0	-0.1 (0.17)	0
	Hmol3.15	3.85 (0.67)	0.61 (0.14)	0.59 (0.06)	-0.02 (0.2)	0	-0.03 (0.22)	0
	Hmol4.27	3.3 (0.73)	0.55 (0.2)	0.57 (0.11)	0.06 (0.26)	2	0.05 (0.26)	0
	Hmol3.8	4.05 (1.05)	0.57 (0.15)	0.55 (0.13)	-0.04 (0.2)	1	-0.07 (0.21)	1
	Hmol4.11	2.15 (0.49)	0.27 (0.15)	0.27 (0.14)	-0.02 (0.21)	0	-0.01 (0.23)	0
	Hmol4.8	10.25 (3.18)	0.88 (0.09)	0.82 (0.06)	-0.07 (0.1)	1	-0.07 (0.11)	0
	Hmol4.29	10.35 (3.33)	0.86 (0.12)	0.83 (0.07)	-0.04 (0.13)	3	-0.06 (0.11)	0
B. calamita	*Bcal4.21*	7.43 (2.09)	0.58 (0.17)	0.75 (0.07)	0.22 (0.21)	12	0.19 (0.19)	1
	Bcal4.20	16.48 (5.65)	0.96 (0.04)	0.89 (0.04)	-0.08 (0.08)	10	-0.11 (0.22)	0
	Bcal4.8	15.57 (5.9)	0.89 (0.12)	0.86 (0.08)	-0.04 (0.11)	5	-0.08 (0.25)	0
	Bcal4.29	7.81 (1.72)	0.86 (0.08)	0.8 (0.07)	-0.09 (0.16)	2	-0.13 (0.24)	0
	Bcal4.16	4.38 (1.07)	0.61 (0.12)	0.57 (0.11)	-0.09 (0.14)	0	-0.05 (0.21)	0
	Bcal4.18	7.05 (1.32)	0.85 (0.07)	0.79 (0.05)	-0.07 (0.12)	2	-0.1 (0.26)	0
	Bcal4.3	9.76 (3.22)	0.82 (0.1)	0.81 (0.09)	-0.02 (0.09)	6	0 (0.1)	0
	Bcal4.6	7.05 (1.63)	0.65 (0.16)	0.77 (0.1)	0.15 (0.24)	11	0.17 (0.25)	4
	Bcal4.14	9.05 (2.69)	0.57 (0.18)	0.82 (0.05)	0.3 (0.24)	18	0.28 (0.36)	9
	Bcal4.2	16.81 (7.15)	0.71 (0.19)	0.88 (0.05)	0.2 (0.23)	20	0.11 (0.42)	11
	Bcal3.26	12.76 (4.6)	0.63 (0.18)	0.85 (0.08)	0.25 (0.23)	17	0.23 (0.36)	11
	Bcal4.24	8.95 (2.56)	0.87 (0.1)	0.82 (0.05)	-0.07 (0.13)	4	-0.11 (0.24)	0
	Bcal3.4	5.38 (1.56)	0.7 (0.19)	0.67 (0.18)	-0.06 (0.11)	2	-0.08 (0.16)	0
	Bcal3.29	4.24 (1.37)	0.44 (0.14)	0.48 (0.13)	0.08 (0.21)	3	0.03 (0.31)	0
	Bcal3.19	6.57 (1.96)	0.43 (0.18)	0.73 (0.13)	0.42 (0.21)	15	0.37 (0.34)	8
	Bcal4.26	21.14 (9.67)	0.95 (0.07)	0.9 (0.05)	-0.06 (0.09)	12	-0.09 (0.23)	0
P. perezi	*Pper4.25*	13.16 (5.11)	0.88 (0.09)	0.86 (0.08)	-0.03 (0.09)	2	-0.06 (0.13)	0
	Pper4.15	8.79 (2.64)	0.81 (0.13)	0.8 (0.08)	-0.01 (0.11)	0	-0.05 (0.11)	0
	Pper4.28	4 (1.63)	0.55 (0.17)	0.52 (0.12)	-0.06 (0.21)	0	-0.07 (0.2)	0
	Pper3.9	6.21 (1.55)	0.71 (0.14)	0.69 (0.1)	-0.02 (0.13)	0	-0.05 (0.1)	0
	Pper4.5	3.11 (0.46)	0.64 (0.09)	0.63 (0.04)	-0.02 (0.14)	0	-0.03 (0.16)	0
	Pper4.16	7.95 (2.3)	0.81 (0.1)	0.79 (0.06)	-0.03 (0.12)	0	-0.01 (0.16)	0
	Pper3.24	6.21 (1.65)	0.77 (0.17)	0.74 (0.12)	-0.04 (0.16)	1	-0.06 (0.17)	1
	Pper4.20	2.05 (0.23)	0.39 (0.16)	0.4 (0.12)	0.05 (0.32)	1	0.06 (0.32)	0
	Pper3.22	3.68 (1.16)	0.44 (0.12)	0.42 (0.11)	-0.05 (0.11)	0	-0.05 (0.12)	0
	Pper4.13	9.58 (3.61)	0.82 (0.13)	0.81 (0.13)	-0.02 (0.08)	1	-0.04 (0.15)	0
	Pper4.7	11.63 (4.76)	0.83 (0.21)	0.84 (0.08)	0.03 (0.22)	5	0.02 (0.21)	1
	Pper3.1	5.74 (1.79)	0.7 (0.15)	0.72 (0.07)	0.02 (0.2)	4	0.01 (0.24)	1
	Pper4.29	6.05 (1.9)	0.76 (0.18)	0.67 (0.14)	-0.13 (0.11)	1	-0.14 (0.12)	0
	Pper3.23	4.89 (1.05)	0.67 (0.16)	0.67 (0.07)	-0.01 (0.23)	2	0.03 (0.27)	1
	Pper4.24	9.21 (2.8)	0.82 (0.16)	0.81 (0.09)	-0.01 (0.18)	2	0 (0.2)	1

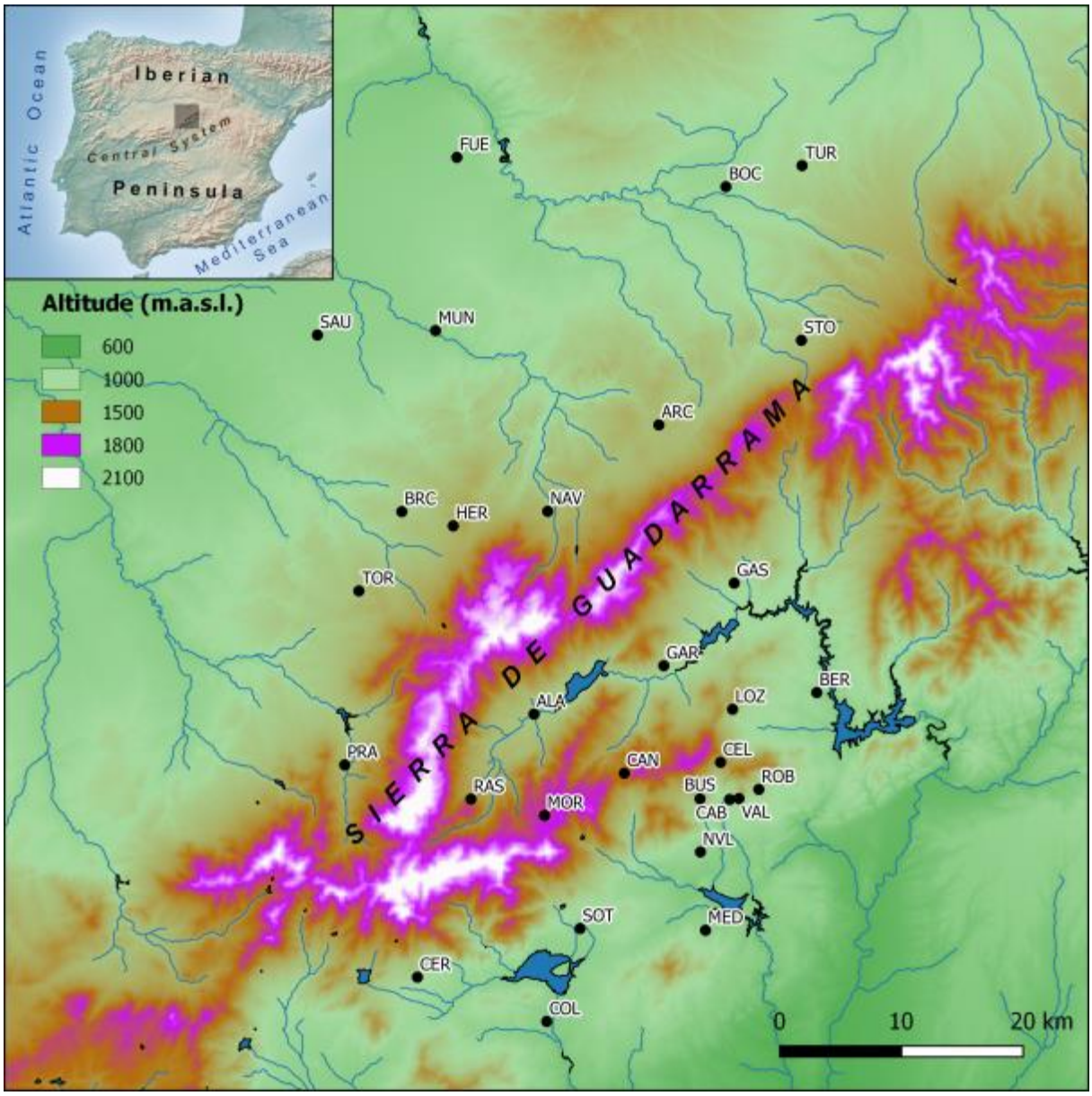

Figure 1. Topographic map showing location of the study area in the Iberian Peninsula and sampling localities. See Table 1 for abbreviations.

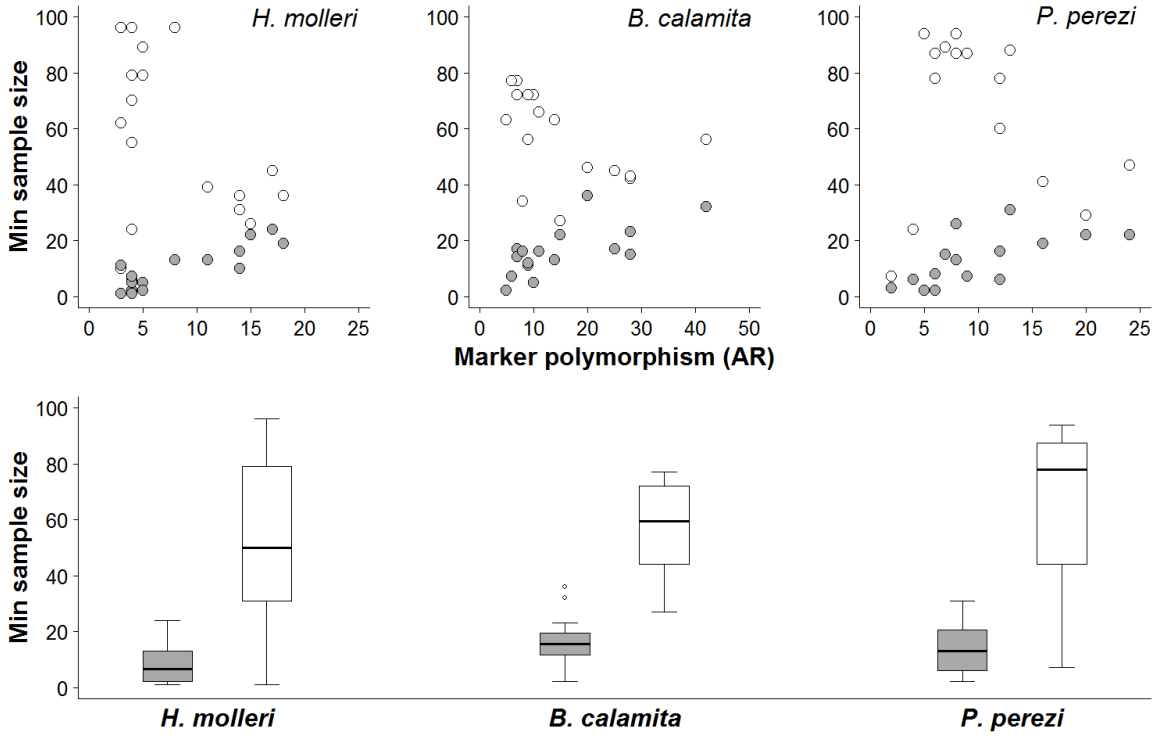

Figure 2. Minimum sample sizes (i.e., minimum number of individuals) required to obtain final estimates of H_E (grey) and AR (white) in the complete samples from Valdemanco. Scatterplots in the top panel show the minimum sample sizes (y-axis) required to estimate each parameter for each marker individually (grey dots: H_E, white dots: AR), while each marker is represented in the x-axis by the polymorphism (AR) shown in Valdemanco. Minimum sample sizes required for estimation of H_E were highly correlated with marker polymorphism, measured as AR, in the three species (*H. molleri*: Spearman's rho = 0.79, p < 0.001; *B. calamita*: rho = 0.67, p = 0.005; *P. perezi*: rho = 0.74, p = 0.002). In contrast, minimum sample sizes required for estimation of AR were negatively correlated with marker AR, although only significantly in the case of *B. calamita* (*H. molleri*: rho = -0.28, p = 0.270; *B. calamita*: rho = -0.61, p = 0.013; *P. perezi*: rho = -0.09, p = 0.738). Boxplots summarize the minimum sample sizes for the marker set of each species (bottom panel).

Table S1. Characterization of the *H. molleri* microsatellite set, with multiplex combinations, primer sequences, repeated motifs and observed allele size ranges (in base pairs). Annealing temperature was 60ºC in all cases. The mean (and standard deviation) percentage of missing data, allelic dropout and false allele scoring rates across all sample populations are shown for each marker. R Info: Informativeness for relationship. GB: GenBank accession numbers.

Locus	Multiplex reaction	Primer sequences	Repeated motif	Size range (bp)	Missing data (%)	Allele dropout	False alleles	R Info	GB
Hmol3.7	1	5' GAAGGAAGGGCATTAAGAGGATG 3' 5' TCCTCTGGATTAACTCAGTAGGG 3'	$(ACT)_7$	140 - 149	0.23 (1.02)	-	-	5.23E-09	
Hmol3.28	1	5' TGTACCAGAGCTTCTCCACTTAG 3' 5' CCTACATTGGTCAGGATTAGGTAC 3'	$(AAT)_{10}$	188 - 203	0.5 (1.63)	0 (0.01)	0.06 (0.06)	0.02	
Hmol4.2	1	5' GCCGAAACGTAACTCTATGTACC 3' 5' TGACTTGCACTGGGACTTTAAAC 3'	$(ACAT)_6$	283 - 311	1.87 (3.75)	0.01 (0.02)	0.01 (0.02)	0.01	
Hmol3.9	1	5' AACACAATCACAGTTAGCTTCCC 3' 5' GTTGTCTAGAAGCAGAGTACCAC 3'	$(ACT)_7$	442 - 451	0.56 (1.39)	0.03 (0.07)	0.01 (0.02)	0.00	
Hmol3.3	2	5' AATAGGACTGAAAGGAACAACGC 3' 5' AAGTGATCTGATCGGCTACTTTG 3'	$(AAT)_5$	136 - 145	0.23 (1.02)	0.02 (0.04)	0.01 (0.02)	0.00	
Hmol4.12	2	5' CTAAGTCATCTAGTGGTCCCTGG 3' 5' TTTACAAATGCGACGTTTCAACC 3'	$(AGAT)_8$	228 - 344	2.22 (4.29)	0.01 (0.02)	0.04 (0.05)	0.07	
Hmol4.16	2	5' ATTTACTCAGGGAATGTGCATCC 3' 5' TCATGCTAACTGTGTTTATGTTGC 3'	$(AGAT)_9$	147 - 235	0.24 (1.06)	0 (0.02)	0.03 (0.03)	0.05	
Hmol4.1	2	5' TGCAATGTATCTATTAGCCTCCAC 3' 5' GCCCATTTAAGCATACAGTCTAGC 3'	$(AGAT)_9$	236 - 292	1.66 (2.67)	0.01 (0.04)	0.04 (0.04)	0.04	
Hmol4.9	3	5' GGACAACGTTCTGCAAGTTAATC 3' 5' TGTCTCTTCATGTTGGTGTGATC 3'	$(AGAT)_{10}$	165 - 221	0.45 (1.23)	0 (0.01)	0.01 (0.02)	0.02	
Hmol4.10	3	5' TATTGCCCATATCCTCCCTTCTC 3' 5' ATGACATCACCTCATCAGCCAG 3'	$(AGAT)_{10}$	103 - 175	0.39 (1.23)	0 (0.01)	0.02 (0.03)	0.06	
Hmol3.22	3	5' GACATCCATCATTCACATCCCTG 3' 5' TTCTGCCTTCTCTTCCCATAGAC 3'	$(AAT)_{10}$	294 - 324	0.84 (1.74)	0.01 (0.02)	0.04 (0.04)	0.04	
Hmol4.22	4	5' GCTTCATCACCACTTAACCTGAG 3' 5' TGGACATGATCAGAGACCATTAC 3'	$(AAAC)_6$	236 - 244	0.73 (3.05)	0.01 (0.05)	0.03 (0.05)	0.00	
Hmol3.15	4	5' TTTGTCTAGTGTCAGCCCTCTAG 3' 5' AGCATACAGTGGCATATTTCAGC 3'	$(AAG)_5$	161 - 169	0 (0)	0.02 (0.03)	0.02 (0.03)	0.02	
Hmol4.27	4	5' GACGTCAATACCAAGTACGCTAG 3' 5' GTAAGTCAAGGGCCCTGAAGTC 3'	$(AGAT)_6$	204 - 220	1.21 (2.16)	0.06 (0.09)	0.04 (0.05)	0.02	
Hmol3.8	4	5' ATAGTCTTATGCTTGTTGGGCTG 3' 5' TATGGGAAACTGCACCACTCTTC 3'	$(ACT)_{12}$	258 - 279	1.36 (5.09)	0.03 (0.07)	0.04 (0.05)	0.02	

Hmol4.11	5	5' TTAAGCCTGAATGTATGGAATTGG 3' 5' TTTCGAGCATATTGATCCCTCCC 3'	$(AGAT)_{10}$	276 - 292	2.38 (3.44)	0.04 (0.08)	0.03 (0.03)	0.00
Hmol4.8	5	5' GTTGTGCTGACCTTGAAAGTATTG 3' 5' CTAGGCTTGATAATGGCAGTGTG 3'	$(AGAT)_{10}$	384 - 441	2.49 (3.64)	0.01 (0.01)	0.02 (0.03)	0.07
Hmol4.29	5	5' CTTTCCTTGGCTTCTTTATGCAC 3' 5' GTATGTGAGCTCTTTACTGCCTG 3'	$(AGAT)_{6}$	356 - 461	3.58 (6.22)	0.02 (0.07)	0.04 (0.05)	0.07

Table S2. Characterization of the *B. calamita* microsatellite set, with multiplex combinations, primer sequences, repeated motifs and observed allele size ranges (in base pairs). Annealing temperature was 60ºC in all cases. The mean (and standard deviation, SD) percentage of missing data, allelic dropout and false allele scoring rates across all sample populations are shown for each marker. R Info: Informativeness for relationship. GB: GenBank accession numbers.

Locus	Multiplex reaction	Primer sequences	Repeated motif	Size range (bp)	Missing data (%)	Allele dropout	False alleles	R Info	GB
Bcal4.21	3	5' CACAGAAGGACAGTAGTTAGACG 3' 5' AGATCTGCTGGTTTACAAAGTGG 3'	$(AGAT)_9$	80 - 128	2.61 (3.53)	0.13 (0.12)	0.02 (0.03)	0.04	
Bcal4.20	3	5' TGAGCAAATCCTCCAAACATGAG 3' 5' TTTGGCCTTTCAACCTTAATCCC 3'	$(AAAG)_{10}$	238 - 314	1.3 (2.58)	0 (0)	0.03 (0.05)	0.09	
Bcal4.8	2	5' GACATCTGTTTGCGTTTCATTGG 3' 5' GCTAGTGTCATTTACTACAACAGC 3'	$(AGAT)_8$	362 - 448	0.38 (1.03)	0.01 (0.04)	0.03 (0.05)	0.08	
Bcal4.29	2	5' ATGTTGAATGCTAAGCCGAAATG 3' 5' ACATACCTTCATTTGGCTGTGAG 3'	$(AGAT)_{10}$	122 - 174	0.16 (0.73)	0.01 (0.02)	0.03 (0.04)	0.05	
Bcal4.16	2	5' GATAGCCCTCCATTCTAGTCTCC 3' 5' ATGGTTATGAACAGACATGCAAC 3'	$(AAAT)_5$	164 - 184	0 (0)	0.01 (0.01)	0.01 (0.02)	0.02	
Bcal4.18	3	5' CTGGAAAGGTCATTGATTCAGGG 3' 5' AGACCCTGTGTAGTCATATACCC 3'	$(AGAT)_8$	178 - 214	0.16 (0.73)	0.01 (0.01)	0.01 (0.03)	0.04	
Bcal4.3	2	5' AACAACCACCAGAACTAACATGG 3' 5' TGACGCAGATATGTATACAGTTGG 3'	$(AGAT)_6$	305 - 357	0 (0)	0.01 (0.01)	0.02 (0.03)	0.06	
Bcal4.6	1	5' AGGGTGTCTGAATACTTTCCGTC 3' 5' TTGACAAAGGCCTCATTGAGAAG 3'	$(AGAT)_{10}$	145 - 181	1.68 (2.39)	0.09 (0.09)	0.01 (0.01)	0.05	
Bcal4.14	1	5' TTACTTAGGCCCTGAACAGTGTC 3' 5' AATTGGCAATGATCAACGGTTTG 3'	$(AGAT)_8$	426 - 476	5.05 (5.27)	0.21 (0.19)	0.03 (0.05)	0.06	
Bcal4.2	1	5' GACTGTTTCCTGGATGTGAATTTC 3' 5' ACAAGGATGATTACTTTGAGCAGG 3'	$(AGAT)_9$	311 - 592	5.16 (5.3)	0.16 (0.17)	0.05 (0.09)	0.10	
Bcal3.26	2	5' GTGTATGGGCATCTTTAGAATGAG 3' 5' TATCTGCCACTTTGAACGGTTTC 3'	$(AAT)_7$	270 - 323	5.5 (5.96)	0.17 (0.14)	0.02 (0.04)	0.08	
Bcal4.24	3	5' ATCAGGAGCCACTAGTACTGAAC 3' 5' ATGCCAGATGACACTACTCTTGG 3'	$(AGAT)_7$	302 - 358	1.1 (1.7)	0.01 (0.02)	0.03 (0.05)	0.05	
Bcal3.4	3	5' TGACTATGGTGGGAAGGGTTAAG 3' 5' AGGAAATTCTGGGACTCTGAGG 3'	$(AAC)_8$	130 - 154	0.16 (0.73)	0 (0.01)	0.02 (0.03)	0.03	
Bcal3.29	1	5' GCCAGGAATACTTCTTCACTCTG 3' 5' TATCTGTTtGTTGATGGCAGACC 3'	$(ACT)_7$	222 - 240	1.54 (3.47)	0.06 (0.11)	0.02 (0.03)	0.01	
Bcal3.19	1	5' GCCATCCAATCCACAATCTCATC 3' 5' ACCATTCCATACTTTGTGTGACG 3'	$(ACT)_9$	234 - 270	9.32 (5.46)	0.32 (0.22)	0.02 (0.03)	0.04	

Bcal4.26	1	5' CGGATCTAACCTTCATGTAACCAC 3' 5' AGAAAGTCTAGCTACACCTTTGG 3'	$(AGAT)_8$	155 - 375	1.46 (2.86)	0 (0)	0.03 (0.04)	0.10

Table S3. Characterization of the *P. perezi* microsatellite set, with multiplex combinations, primer sequences, repeated motifs and observed allele size ranges (in base pairs). Annealing temperature was 60ºC in all cases. The mean (and standard deviation, SD) percentage of missing data, allelic dropout and false allele scoring rates across all sample populations are shown for each marker. R Info: Informativeness for relationship. Primer sequences, repeated motifs and GenBank accession numbers (GB) from Sánchez-Montes et al. (2016).

Locus	Multiplex reaction	Primer sequences	Repeated motif	Size range (bp)	Missing data (%)	Allele dropout	False alleles	R Info	GB
Pper4.25	1	5' TCCCTTCTAGTGCTGTAACTTCG 3' 5' AGTTCATCTGCAGTTCCTACATG 3'	$(AGAT)_8$	183 - 403	0.58 (1.48)	0.01 (0.02)	0.05 (0.05)	0.09	KT166015
Pper4.15	1	5' ACATATTGTGCTGCTCCATCAAG 3' 5' AATTTCTTCAGTGCTGTCATGTC 3'	$(AGAT)_8$	177 - 249	0.06 (0.24)	0.01 (0.02)	0.03 (0.04)	0.06	KT166016
Pper4.28	1	5' CATGTACAGCTGACTTTAGAGCC 3' 5' TTCTTTCCAATTTGAGACTCGGG 3'	$(AAGG)_5$	200 - 260	0.06 (0.24)	0.04 (0.1)	0.04 (0.04)	0.02	KT166017
Pper3.9	1	5' CAACATATCTTCCCGAATGAGGC 3' 5' GTTTCTCTCAGTCTAGTTGGTGC 3'	$(AAG)_6$	191 - 262	0.06 (0.24)	0.02 (0.03)	0.03 (0.03)	0.03	KT166018
Pper4.5	2	5' TGTGCGCTATCCTCTGTAGTTAG 3' 5' TGAATCCTGGCATTGTCATCTTG 3'	$(AAAC)_6$	148 - 164	0.16 (0.72)	0.03 (0.06)	0.04 (0.05)	0.02	KT166019
Pper4.16	2	5' AGAGCAGATATACCACACTCCAG 3' 5' ACCTCAAGCATTTATAGACCAGC 3'	$(AGAT)_9$	140 - 192	0.22 (0.74)	0.01 (0.02)	0.02 (0.04)	0.05	KT166020
Pper3.24	2	5' ATGTGGAGACTATCAGCAGACAG 3' 5' CAAGTCTTGACTGTTCATACCGG 3'	$(AAC)_7$	248 - 278	1.18 (2.81)	0.02 (0.06)	0.05 (0.06)	0.04	KT166021
Pper4.20	3	5' TCTTAGCAGTGACAGATGTGAAC 3' 5' TCTTAGTGCAGATTAGGGACCTG 3'	$(AAGT)_6$	220 - 228	0 (0)	0.06 (0.18)	0.02 (0.05)	0.01	KT166022
Pper3.22	3	5' ACTGTCATCTGGTCTGGTATCAC 3' 5' ACACTAATTGTCCTCCTGTAGAAC 3'	$(ACT)_9$	358 - 382	0.42 (1.28)	0.01 (0.03)	0.03 (0.05)	0.01	KT166023
Pper4.13	3	5' AGAGACCATATATCGGAGCCATC 3' 5' TGGCAAATCACTCCACTTAACAG 3'	$(AGAT)_{10}$	425 - 513	0.42 (1.28)	0.01 (0.02)	0.06 (0.07)	0.06	KT166024
Pper4.7	4	5' TACCTCTTCTGCTGATCTCTTGG 3' 5' AAGCAATTTATCAAGCAGGAGGG 3'	$(AGAT)_9$	280 - 364	1.42 (2.8)	0.05 (0.14)	0.02 (0.04)	0.08	KT166025
Pper3.1	4	5' TTGCCAGCAGAAGAGAACATTAC 3' 5' TCTCACAGACATCGCATTTGATC 3'	$(AGG)_9$	337 - 376	0.49 (1.35)	0.06 (0.13)	0.04 (0.06)	0.04	KT166026
Pper4.29	5	5' CTGTGCTACGAGGATTGTAATGG 3' 5' TTCATTCTCTGTGTCGTGAATGC 3'	$(AAAG)_7$	313 - 357	0.34 (1.03)	0 (0.02)	0.02 (0.03)	0.04	KT166028
Pper3.23	5	5' ACTTGTATCATCTTTCTCTGCGC 3' 5' TTTCTGCCCAATTCTACTACTGC 3'	$(ACT)_6$	154 - 196	0.34 (1.09)	0.03 (0.06)	0.03 (0.04)	0.03	KT166029
Pper4.24	5	5' TTTCCCTATTGCCTATGAACTGC 3' 5' AGTGCTATGGTTGGGATTTGAAC 3'	$(AGAT)_{10}$	195 - 339	0.67 (1.62)	0.05 (0.07)	0.05 (0.05)	0.07	KT166030

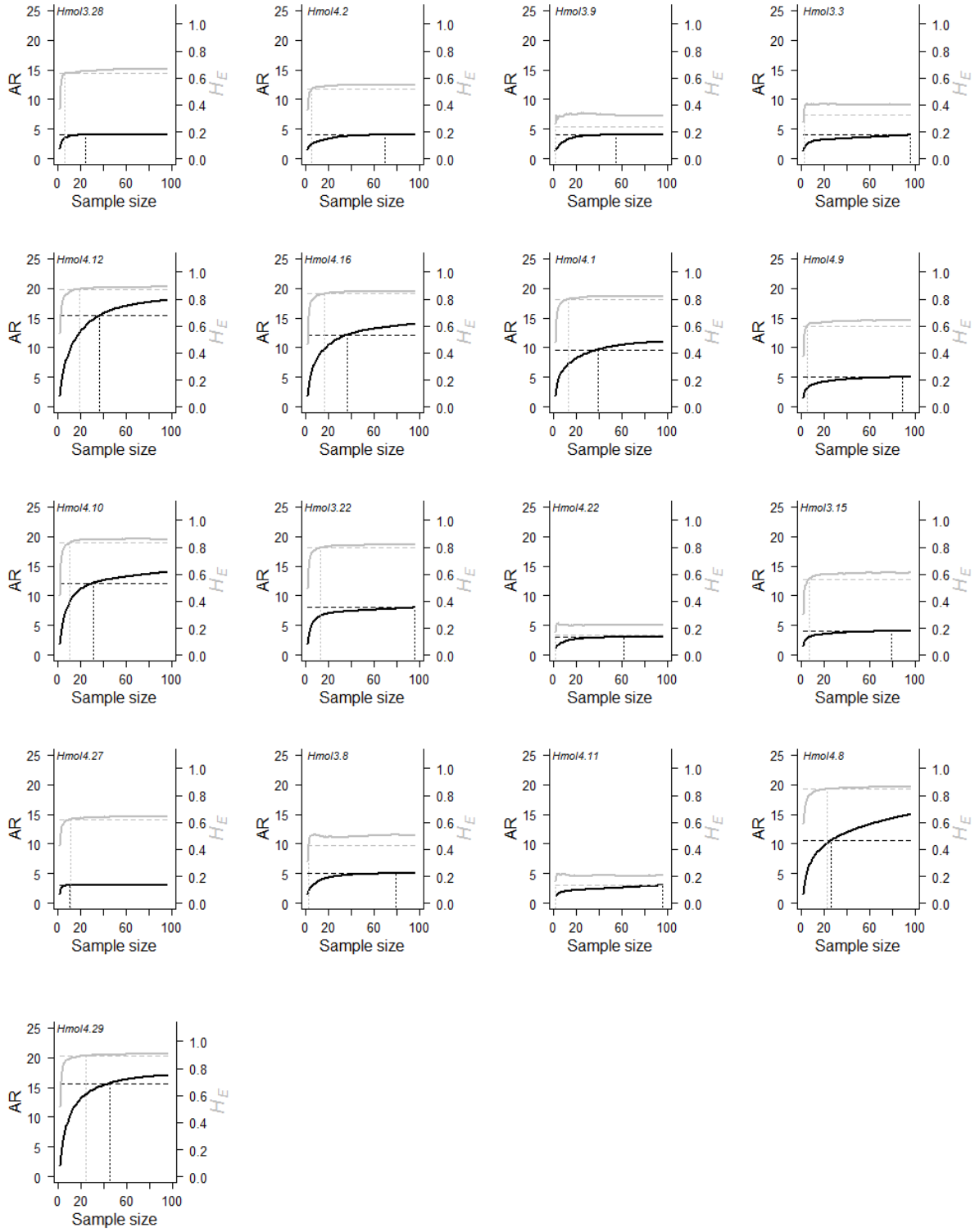

Figure S1. Accumulation curves of AR (dark lines) and H_E (grey lines) as a function of sample size (measured as number of individuals) for each marker in the *H. molleri* marker set. Jackknifed curves were calculated from the complete samples in Valdemanco. Vertical dotted lines show the minimum sample size at which the lower bound of the 95% confidence interval of each final estimate (shown as horizontal dashed lines) is reached.

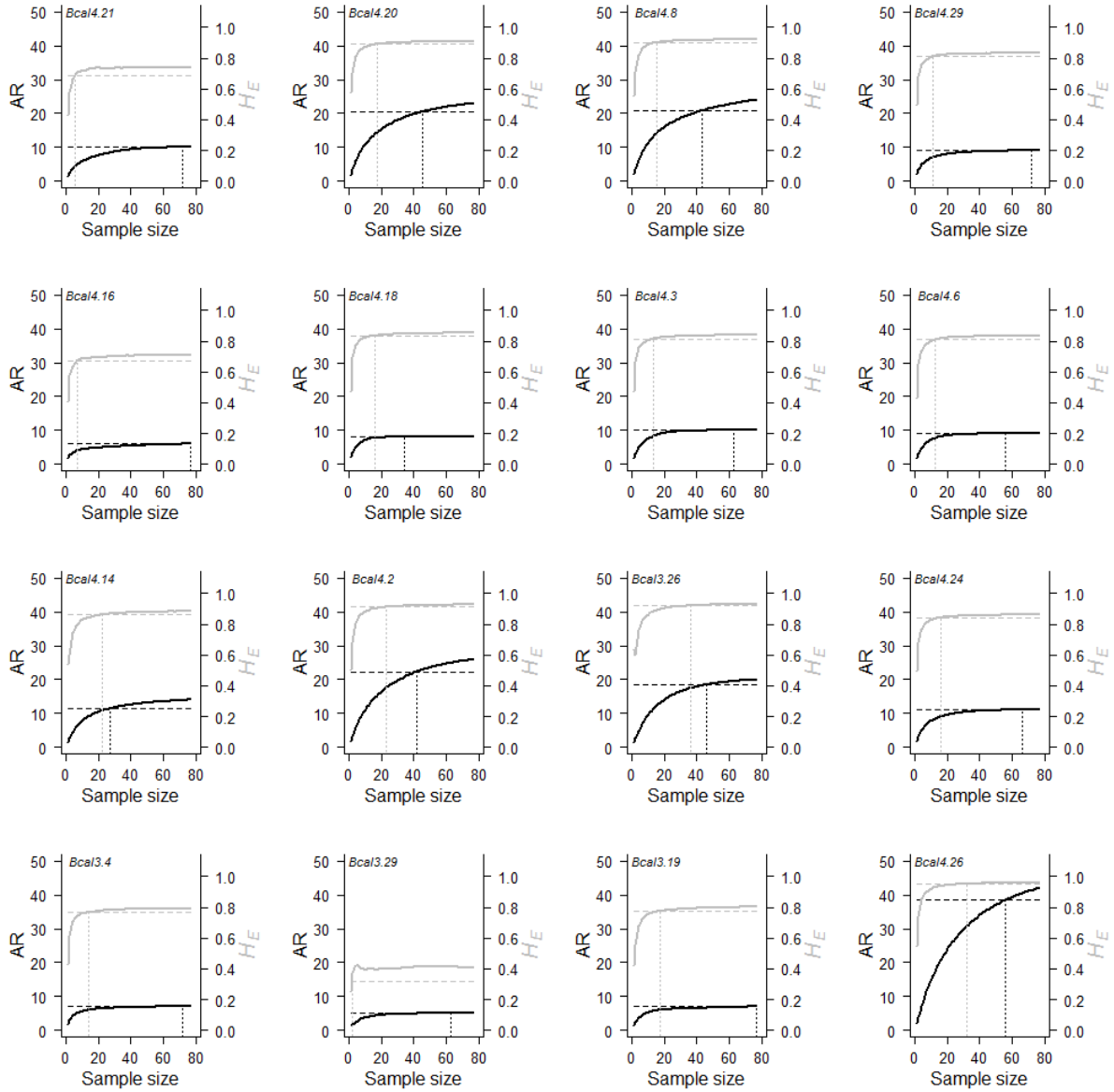

Figure S2. Accumulation curves of AR (dark lines) and H_E (grey lines) as a function of sample size (measured as number of individuals) for each marker in the *B. calamita* marker set. Jackknifed curves were calculated from the complete samples in Valdemanco. Vertical dotted lines show the minimum sample size at which the lower bound of the 95% confidence interval of each final estimate (shown as horizontal dashed lines) is reached.

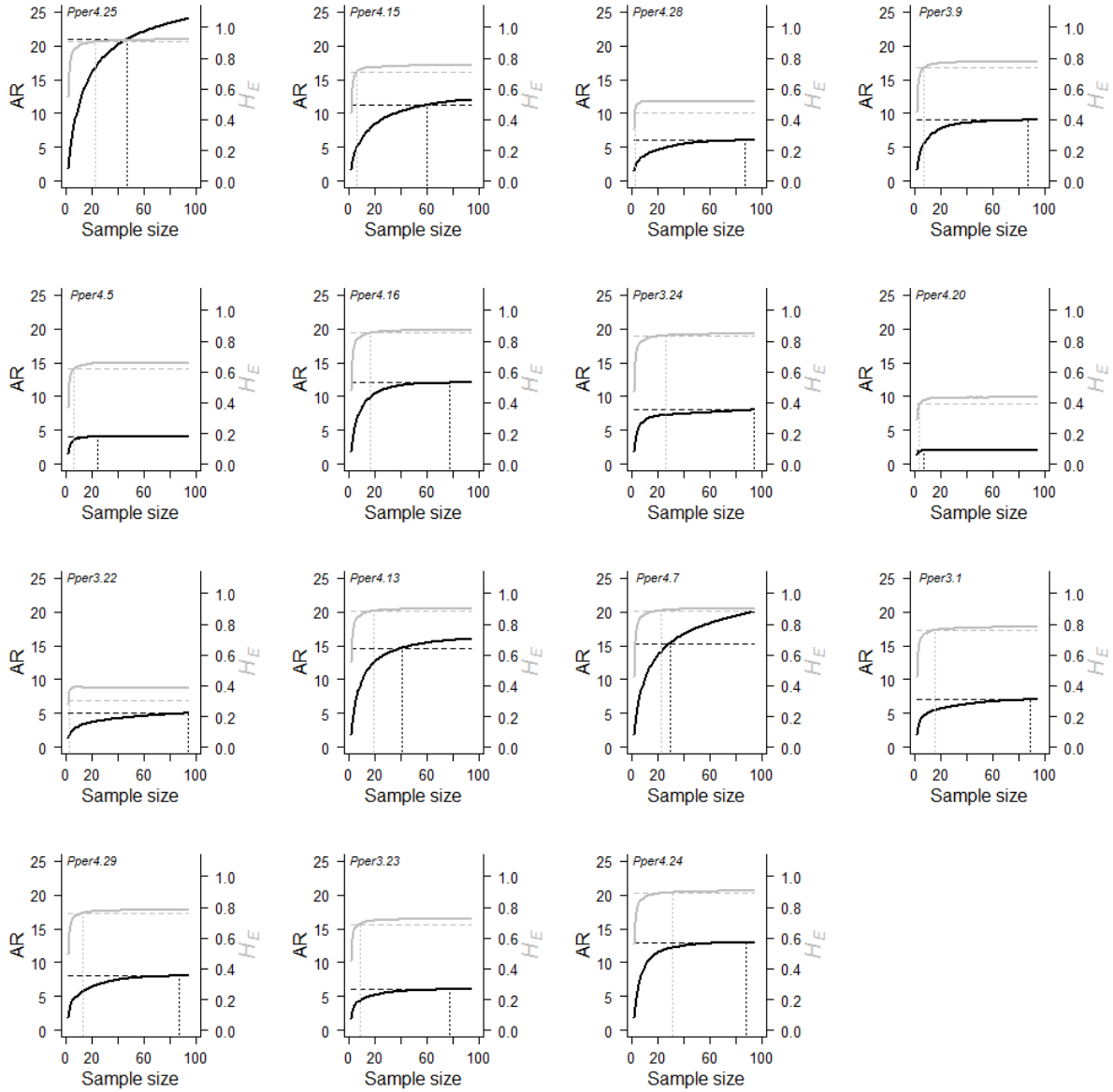

Figure S3. Accumulation curves of AR (dark lines) and H_E (grey lines) as a function of sample size (measured as number of individuals) for each marker in the *P. perezi* marker set. Jackknifed curves were calculated from the complete samples in Valdemanco. Vertical dotted lines show the minimum sample size at which the lower bound of the 95% confidence interval of each final estimate (shown as horizontal dashed lines) is reached.

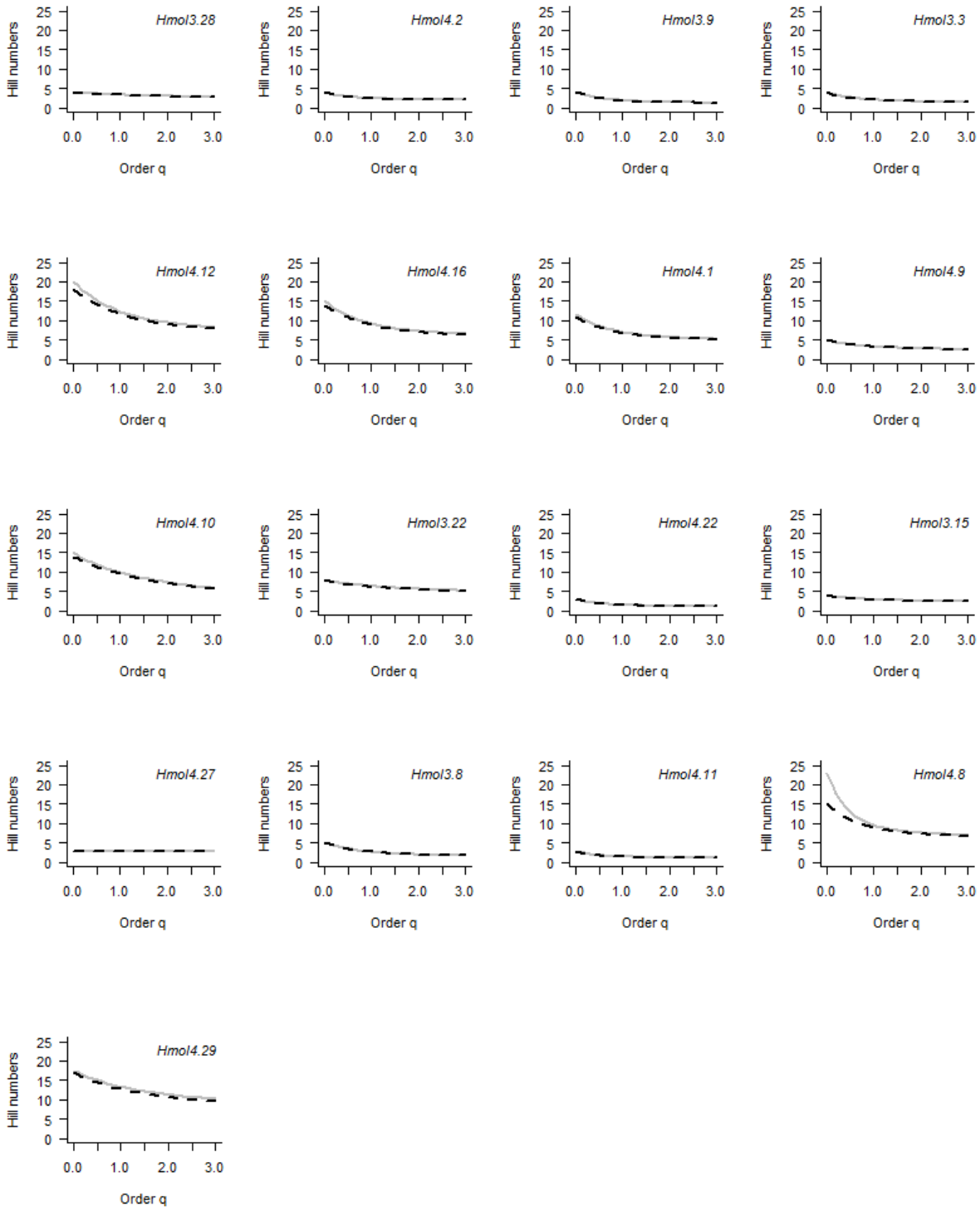

Figure S4. Empirical (dotted line) and Chao and Jost (2015) profile (grey solid line) for each marker in the *H. molleri* set.

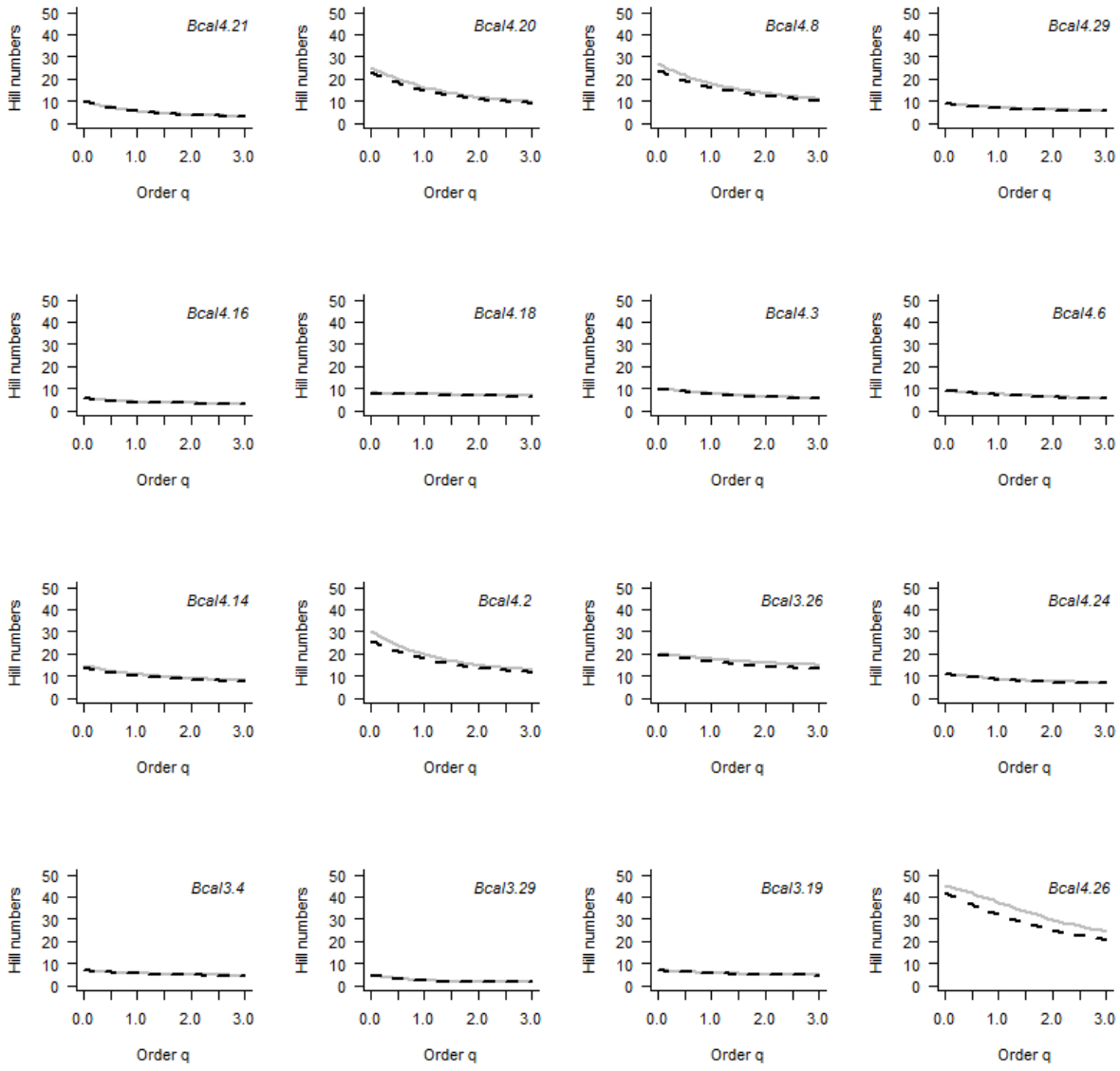

Figure S5. Empirical (dotted line) and Chao and Jost (2015) profile (grey solid line) for each marker in the *B. calamita* set.

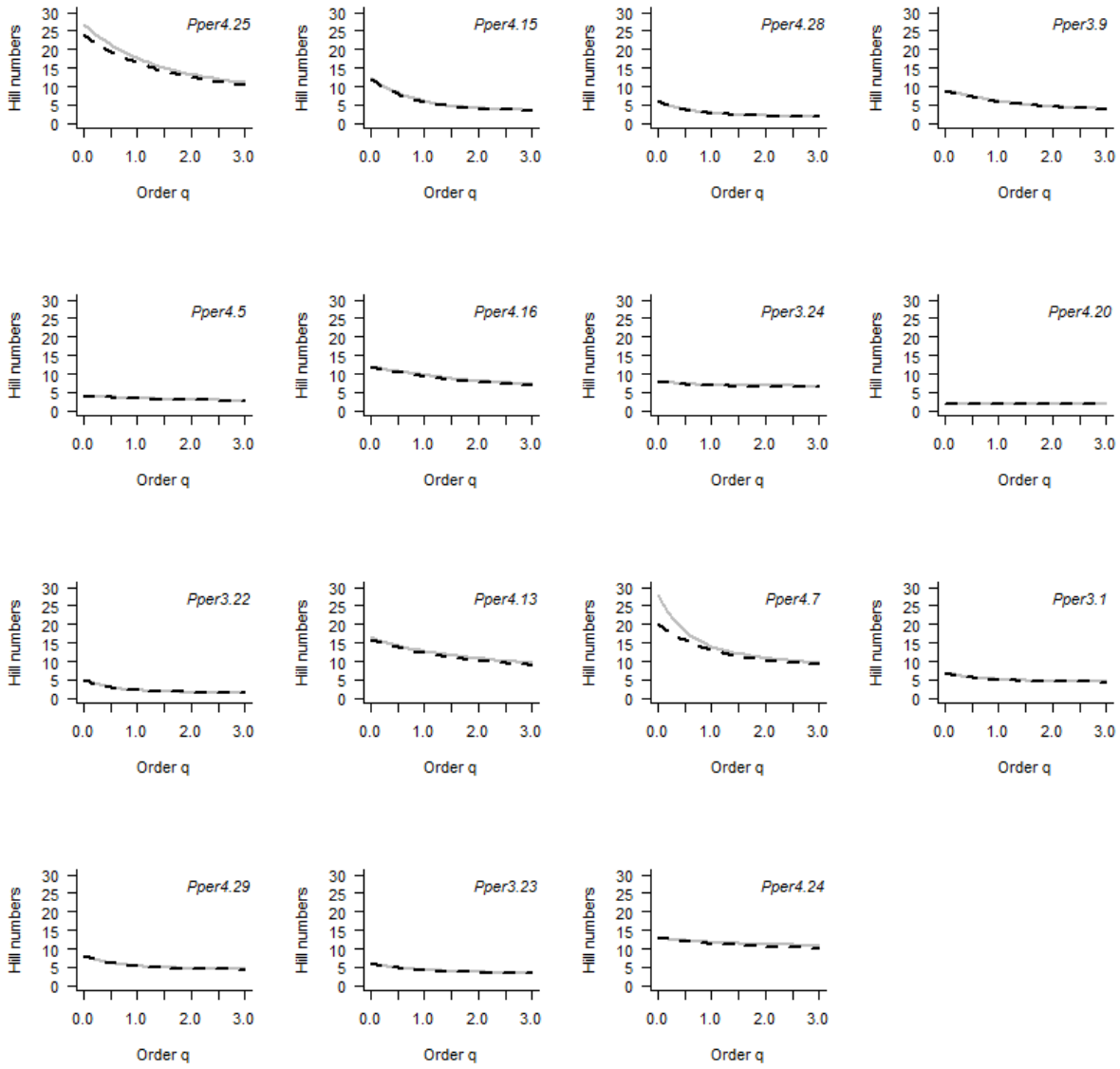

Figure S6. Empirical (dotted line) and Chao and Jost (2015) profile (grey solid line) for each marker in the *P. perezi* set.

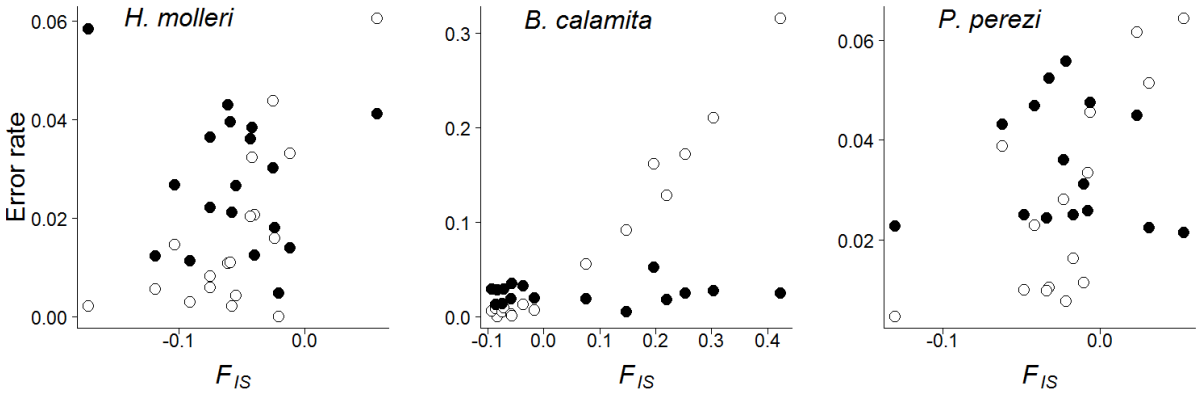

Figure S7. Relationship between F_{IS} and error rate estimates (empty dots: allelic dropout rate, solid dots: false allele rate) obtained from sibship analyses for each marker in the three species. Note the difference in axis scales in the *B. calamita* graph.

Appendix S1: Effect of sampling excessive close relatives on F_{IS} and deviation from HWE

Wright's (1931) F_{IS} is the traditional and most popular statistic used in measuring the distribution of genetic variation within and among individuals in a population. For a population at Hardy-Weinberg equilibrium (HWE), homologous allelic copies are independently distributed within and between individuals. In such a situation, $F_{IS} = 0$. For a population with subdivision (e.g. in social groups) or with close relative mating, the two allelic copies within an individual are more probable to be identical in state than those in different individuals. In such a situation, the observed homozygosity is higher than that expected if the population is at HWE, leading to $F_{IS} > 0$ (since $F_{IS} = 1 - \frac{H_O}{H_E}$, where H_O and H_E are the observed and expected heterozygosity, respectively. Nei 1977). In contrast, admixture and hybridization lead to $F_{IS} < 0$.

The F_{IS} of a population is usually unknown, and is estimated by the marker or pedigree data of a sample of individuals drawn from the population. Here we show analytically that sampling too many close relatives would lead to a reduced F_{IS} estimate. For a large population at HWE in which $F_{IS} = 0$, a sample from it can yield a negative F_{IS} estimate if it contains excessive close relatives. These predictions are true no matter whether pedigree or marker data are used in the estimation.

Denoting the probabilities of identity by descent (PIBD) for two homologous genes drawn at random from an individual and between two individuals in a population by α and β, respectively, we have

$$F_{IS} = \frac{\alpha - \beta}{1 - \beta}, \quad (1)$$

by definition (Cockerham 1969, eqn 41; Weir 1996, p.176). If a random sample (random with regard to genealogy) is taken from the population, then unbiased estimates of α, β, and thus F_{IS} estimates would be obtained. However, if too many (excessive) close relatives, such as full or half siblings, are included in a sample, the PIBD between individuals in the population would be overestimated, from the true value β to β', while the estimated PIBD within individuals would remain unbiased as α. As a result, F_{IS} would be expected to be decreased to

$$F'_{IS} = \frac{\alpha - \beta'}{1 - \beta'} \quad (2)$$

Equation (2) implies that $F'_{IS} < F_{IS}$, because $\beta' > \beta$. The larger the increase in PIBD between sampled individuals, β', due to the inclusion of a greater proportion of close relatives, the smaller will be F'_{IS} relative to F_{IS}.

For illustration, let's consider some numerical examples for a dioecious diploid species in a large random mating population. It is expected that two homologous genes at an autosomal locus are identical by descent with probabilities 0, 0, ¼ when they are in a single individual, in two unrelated individuals, and two full siblings respectively. In a random sample of individuals taken from the population, the estimated PIBDs are expected to be $\alpha = 0$, $\beta = 0$, and thus the estimated F_{IS} is also

expected to be zero. In an inadequately drawn sample of individuals with a proportion of δ full-sib pairs, the estimated PIBDs are expected to be $\alpha = 0$, $\beta = (1-\delta)\times 0 + \frac{\delta}{4} = \frac{\delta}{4}$, and the estimated F_{IS} is expected to be $\frac{0-\frac{\delta}{4}}{1-\frac{\delta}{4}} = -\delta/(4-\delta)$. Suppose a sample has n=50 individuals, with 10 individuals taken from full sib family X, 20 individuals from full sib family Y, and the remaining 20 individuals from 20 different and unrelated families. The estimated PIBDs are expected to be $\alpha = 0$, $\beta = \frac{10\times\frac{9}{2}+20\times\frac{19}{2}+}{50\times\frac{49}{2}}\times\frac{1}{4} = 0.048$, and the estimated F_{IS} is expected to be $\frac{0-0.048}{1-0.048} = -0.0504$.

As it can be seen from the examples, the inclusion of an excessive proportion of relatives (in this case, full siblings) in a sample causes a reduction in the estimated F_{IS}. Conversely, including an excessively low proportion of full sibs in the sample (relative to the true proportion in the population) results in an artificially inflated estimate of F_{IS}. Depending on the values of α and β, this bias may lead in some cases to false inferences of negative inbreeding (and the false conclusion that the population is affected by admixture (hybridization) or/and avoids close relative matings) or positive inbreeding (false conclusion of positive assortative mating or population subdivision). For the same reason, removing all but one of the full sibs in every full sib family in the sample does not always eliminate the bias caused by unrepresentative proportion of relatives in the sample. In fact, it could lead to an underrepresentation of relatives in the sample and thus to the opposite bias, with $F'_{IS} > F_{IS}$.

As a result, excessive close relatives in a sample cause an apparent decrease in observed homozygotes and an apparent increase in observed heterozygotes at each locus (i.e., higher H_O/H_E ratio), and they also cause nonrandom associations between alleles in different loci. This leads to an increase in statistically significant deviations from HWE across loci and evidences of LD, which disappear when the excess of relatives is removed.

Reference

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.

Cockerham C C (1969) Variance of gene frequencies. *Evolution*, 72-84.

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of human genetics*, *41*(2), 225-233.

Weir BS (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc., Inc., Sunderland, MA, USA.

.