

Article

Optimizing Sample Size for Population Genomic Study in a Global Invasive Lady Beetle, *Harmonia Axyridis*

Hongran Li ¹, Wanmei Qu ², John J. Obrycki ³, Ling Meng ¹, Xuguo Zhou ³ , Dong Chu ²  and Baoping Li ^{1,*} 

¹ College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China; Hongran.Li@uky.edu (H.L.); ml@njau.edu.cn (L.M.)

² Key Lab of Integrated Crop Pest Management of Shandong Province, College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao 266109, China; register1_kq@163.com (W.Q.); chinachudong@qau.edu.cn (D.C.)

³ Department of Entomology, University of Kentucky, Lexington, KY 40546, USA; john.obrycki@uky.edu (J.J.O.); xuguo Zhou@uky.edu (X.Z.)

* Correspondence: lbp@njau.edu.cn

Received: 23 April 2020; Accepted: 6 May 2020; Published: 9 May 2020



Abstract: Finding optimal sample sizes is critical for the accurate estimation of genetic diversity of large invasive populations. Based on previous studies, we hypothesized that a minimal sample size of 3–8 individuals is sufficient to dissect the population architecture of the harlequin lady beetle, *Harmonia axyridis*, a biological control agent and an invasive alien species. Here, equipped with a type IIB endonuclease restriction site-associated (2b-RAD) DNA sequencing approach, we identified 13,766 and 13,929 single nucleotide polymorphisms (SNPs), respectively, among native and invasive *H. axyridis* populations. With this information we simulated populations using a randomly selected 3000 SNPs and a subset of individuals. From this simulation we finally determined that six individuals is the minimum sample size required for the accurate estimation of intra- and inter-population genetic diversity within and across *H. axyridis* populations. Our findings provide an empirical advantage for population genomic studies of *H. axyridis* in particular and suggest useful tactics for similar studies on multicellular organisms in general.

Keywords: sample size; *Harmonia axyridis*; 2b-RAD; empirical investigation; population genomics; genetic diversity

1. Introduction

Using optimal sample sizes to accurately estimate genetic diversity of large natural populations is an imperative issue in the analysis of evolutionary processes [1]. Because larger samples per population than needed for accurate estimation, results in extra expense and wastes much time to analysis [2], while limited sample sizes will lead to significant errors in estimating the genetic diversity of species [3–5]. Still, most genetic studies of wild species sampled so many individuals per population, but using a small number of genetic markers, for example, microsatellites [5–8]. In practice, the number of microsatellite markers is often limited owing to the cost and time, and thus lowers the power in addressing phylogeographic questions [5,9,10]. Using microsatellites markers to analyze population genetics of invasive species could be a concern, since a recent study showed that low genetic diversity or genuine multimodality was observed [11].

Given these disadvantages of microsatellites marker in genetic analysis, a growing number of researches have recommended to use the genomic data over microsatellites for genetic study [12–14].

For example, a recent empirical study showed that genome-wide techniques can acquire large numbers of single nucleotide polymorphisms (SNPs), to obtain a finer population structure and stronger patterns of isolation-by-distance than microsatellites do with a smaller sample size [15]. It is gratifying that recent developments in restriction site-associated DNA sequencing (RAD-seq) techniques, could provide massive sequence data for efficient identification of SNPs at an unprecedented level [16–18]. Nevertheless, although the costs associated with next-generation sequencing have reduced substantially, RAD-seq remains a relatively costly approach due to the uncertainty of the number of individuals needed to be sampled in a given study. Therefore, a trade-off between the sample size and the number of molecular markers must be considered during the experimental design.

To our knowledge, the utilization of these new techniques has effectively addressed the establishment of an ideal sampling scheme for sample size determination [19–22]. For example, an empirical study via double digest RAD-seq technique showed that the genetic diversity and *Fst* in a plant species can be accurately estimated from six to eight individual plants [1]. A simulation analysis using a large number of SNPs has shown that sample size can be reduced to four to six individuals when estimating the genetic differentiation (*Fst*) [2]. Using a type IIB endonucleases RAD-seq method, Qu et al. [23] have provided the optimization of sampling schemes for an invasive whitefly species, *Bemisia tabaci* Gennadius, which showed that a sample size greater than four individuals has little impact on estimates of genetic diversity. However, the sampling space is limited, which might increase the overlaps between iterations [1].

The harlequin lady beetle, *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae), has emerged as an alien invasive species in many North America and European countries [24–28]. Inferences about introduction routes of this invasive species are needed to understand the fundamental eco-evolutionary aspects of colonization success or failure [29] and for preventing future invasions [30]. Using microsatellite loci data, the approximate Bayesian computation (ABC) methods with more traditional statistical approaches have been reconstructed the introduced routes of *H. axyridis* worldwide, as two bridgehead invasive populations were involved in North America and then served as the source populations for at least six independent introductions into other continents [31]. However, there is no research to characterize the population genomics of *H. axyridis* worldwide based on as many as SNP markers, which may provide novel insights in genetic structure and introduced routes. Indeed, the prerequisite is to determine effective sample size from individual populations to accurately estimate population genetic parameters. Following the results from the existing empirical studies in other organisms, we hypothesized that a minimal sample size of 3–8 individuals is sufficient to estimate genetic diversity within and across native and invasive *H. axyridis* populations, although other factors can also contribute to population genetic inferences [1,23].

To test this hypothesis, we conducted the experiments as follows: (1) surveyed SNPs among *H. axyridis* populations from both native and invasive ranges using type IIB endonuclease RAD sequencing (2b-RAD) method, (2) constructed simulated populations using a random subset of individuals, and finally (3) determined the minimal number of individuals to accurately estimate the several intra- and inter-population genetic diversity parameters, including number of effect alleles (*Ae*), observed heterozygosity (*Ho*), unbiased expected heterozygosity (*uHe*), and pairwise genetic differentiation (*Fst*) in *H. axyridis*; the ad hoc statistic ΔK was supplemented to determine the optimal sample sizes for intra- and inter-population genetic diversity running the STRUCTURE software [23,32].

2. Materials and Methods

2.1. *Harmonia Axyridis* Collection and DNA Extraction

A total of 20 *H. axyridis* females per population were collected from China (LNSY), its native range (41.83°N, 123.57°E) and Poland (PLKK), its invasive range (50.06°N, 19.93°W) for subsequent sequencing analysis. The specimens were preserved in 95% ethanol and stored at –80 °C until DNA extraction. Total DNA was extracted using a TIAMamp Micro DNA Kit (Tiangen Biotech (Beijing)

Co., Ltd., Beijing, China) following the manufacture's protocols. Extracted DNA was dissolved in DNAase-free water and DNA concentration was determined using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA integrity was assessed using 1.0% agarose gel electrophoresis.

2.2. Library Preparation and SNP Identification

Briefly, 2b-RAD libraries were constructed for each individual *H. axyridis* following Wang et al. [33]. Subsequently, genomic DNA was digested by 1 U BsaXI (New England Biolabs, cat. no. R0609), and short adapter sequences were ligated to the ends of the fragments. The ligation products were amplified in 50 μ L PCRs; only fragments starting with a sequence that can be hybridized by the selective sequence of the primer was efficiently amplified. PCR products were purified using a MinElute PCR Purification Kit and pooled for sequencing using the Illumina PE sequencing platform.

Raw sequence data were filtered as follows: (1) the reads with linker sequences were removed to obtain clean reads; and (2) reads with low-quality positions (>15% of nucleotide positions with a Phred score < 30) were deleted. In N bases greater than 8% and without restriction recognition sites, the filtered high-quality sequences were referred to as enzyme reads. Then mapping the enzyme Reads to the *H. axyridis* reference genome (http://bipaa.genouest.org/sp/harmonia_axyridis) was done using SOAP program (the parameter was set to: -r0-M4-v2) and the same reads were clustered into Unique Tags. Finally, the SNP-calling was performed using Maximum likelihood (ML) method [34]. To ensure the accuracy of SNP genotyping, the following filtering procedures were performed: (1) SNPs with a minor allele frequency (MAF) < 0.01 were deleted; (2) tags with more than 2 SNPs were deleted; (3) SNPs at each locus with 1 or 4 bases were deleted; and (4) SNPs that could be genotyped in more than 80% of the individuals were retained.

2.3. Population Genetic Analyses

The allelic data generated by 2b-RAD sequencing were used to build a neighbor-joining (NJ) phylogenetic tree between LNSY (China) and PLKK (Poland) populations. Phylogenetic reconstruction was carried out in treebest v.1.9.2, with 1000 bootstrap replicates [35]. ADMIXTURE (v.1.3.0) was carried out using the entire dataset to estimate the genetic ancestry of each sample [36]. This tool is based on a maximum likelihood approach, which provides an estimation of the number of genetic clusters and the proportion of derived alleles in one sample from each of the K populations. The program was run 10 times, varying the values of K from 2 to 10. A cross-validation test was performed to determine the optimal value of K. The results from 10 replicates of the selected K values were summarized into a single result and were then aligned and analyzed using pong, a network-graphical approach for analyzing and visualizing membership in latent clusters. The raw data have been deposited in the Sequence Read Archive (SRA) database under an accession number of SRP227109.

2.4. Construction of Simulated Populations

Before we proceeded to the optimal sample size, we carried out a power analysis to determine the minimum number of resampling replicates to ensure accurate estimation of genetic parameters following Qu et al. (2019) [23] with minor modifications. Specifically, GenALEx 6.5 was used to remove SNPs deviating from Hardy–Weinberg equilibrium for each population [37]. Then, SNPs that potentially under balancing and divergent selection were removed using the software BAYESCAN v. 2.1 with 20 pilot runs of 10,000 iterations, a burn-in of 50,000 iterations and a final run of 100,000 iterations [1]. To reduce the false positives, prior odds of the neutral model were set to 100 (i.e., the neutral model is 100 times more likely than the model with selection). After then, a total of 3000 SNPs (k) were randomly selected using data tools in Excel for each population. Considering that too limited sampling space would increase overlaps between iterations, we enlarged sampling space in the data set ($n = 20$) than that in previous study ($n = 10$) [23]. In detail, we constructed simulated data sets consisting of

different numbers of resampling replicates ($x = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$), each represented by the different sample sizes (number of individuals per population, $n = 2, 4, 6, 8, 10, 15$).

2.5. Optimizing Sample Size

GenALEx 6.5 was used to measure the genetic parameters, including the number of effective alleles (A_e), observed heterozygosity (H_o), and unbiased expected heterozygosity (uHe), for each replicate at each sample size (i.e., for each simulated population) [37]. To estimate the degree of genetic differentiation, i.e., pairwise genetic differentiation (F_{st}) among populations, a slightly different subsampling strategy, was used to resample the 3000 SNPs shared by LNSY and PLKK populations rather than LNSY or PLKK individually. In addition, we used an ad hoc statistic, ΔK , based on the rate of change in the log probability of data between successive K values, to evaluate the optimal number of replicates and sample sizes for population genomics analyses [32]. ΔK shows a clear peak at the optimal number of replicates and sample sizes.

GraphPad Prism 7.00 was used to measure the influence of sample sizes and replicates on intra- and inter-population genetic diversity parameters.

3. Results

3.1. Population Characterization and SNP Identification

Sequencing of the 2b-RAD libraries resulted in about 344.31 million raw reads from the 40 individuals. On average, 8.61 million reads with restriction site per individual were retained. After quality filtering, the percentage of high quality of reads was above 80% (305.57 million enzyme reads were generated) of the total reads in the libraries of the 40 individuals. Of the retained sequences, a total of 104.56 million (34.26%) enzyme reads aligned to the *H. axyridis* genome survey sequences (https://bipaa.genouest.org/sp/harmonia_axyridis/) (Supplementary Table S1). Of these, 1.67 million (57.92%) loci with minimum 3X and maximum 500X coverage were retained for SNP discovery (Table S1). After removing SNPs that significantly deviated from HWE (818 for population LNSY and 655 for population PLKK), we identified 13,766 and 13,929 polymorphic SNPs for further analysis (Supplementary Table S2). We did not detect any loci that were under selection for either population of *H. axyridis*, with the false discovery rate (FDR) set to 0.05. As such, no other loci were removed from subsequent analyses (Supplementary Table S2). The number of effective alleles (A_e) in the *H. axyridis* populations was 1.123 ± 0.001 SE (LNSY) and 1.124 ± 0.001 SE (PLKK). The expected heterozygosity (H_e) in the LNSY and PLKK populations was 0.087 ± 0.001 SE and 0.087 ± 0.001 SE. The observed heterozygosity (H_o) in the LNSY and PLKK populations were 0.076 ± 0.001 SE and 0.079 ± 0.001 SE, respectively. The F_{st} calculated from all detected SNPs distances between LNSY and PLKK populations was 0.036, indicating that no substantial genetic differentiation exists between the two populations.

3.2. Population Genetic Structure

The resultant neighbor-joining (NJ) tree of these ladybug populations featured two main clusters, namely LNSY and PLKK (Figure 1). The analysis of population structure estimated the K value was two, indicating that the uppermost hierarchical level detected by STRUCTURE was two distinct genetic clusters (Figure 2).

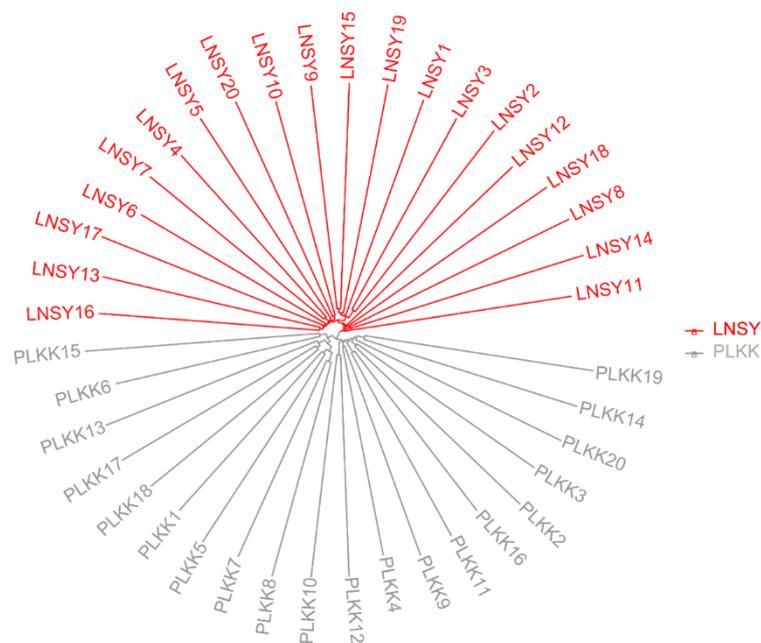


Figure 1. Neighbor-joining phylogram of the LNSY and PLKK populations of *H. axyridis*. The different colors represent the individuals from different populations.

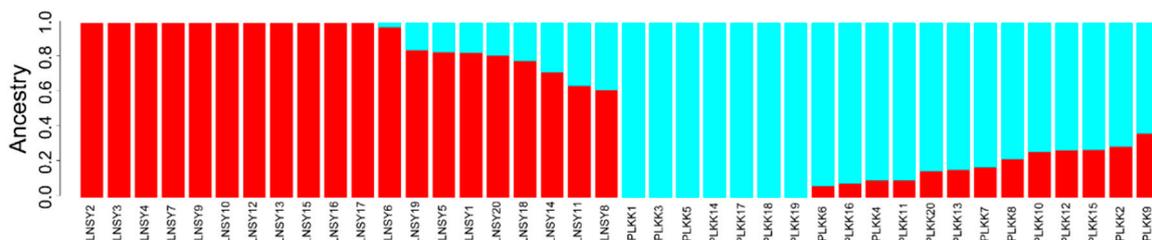


Figure 2. Admixture for LNSY and PLKK populations of *H. axyridis* (K = 2). Each bar represents an individual from each of the collection locations (X axis). Individual admixture coefficients are represented in each column (Y axis).

3.3. Depicting Sample Sizes for Intrapopulation Genetic Diversity

We assessed the impact of increasing sample sizes for intrapopulation genetic diversity estimates by resampling 2, 4, 6, 8, 10, and 15 individuals from empirical data sets obtained for the two *H. axyridis* populations. At first, accurate estimates of population genetic parameters in two populations were acquired in our simulations with only $x = 30$ resampling replicates (Figures 3 and 4). In detail, when we fixed the number of individuals (n) to three and the number of SNPs (k) to 3000, there was no statistical difference for the mean values of A_e , H_o and uHe even when the number of replicates was set to $x = 30$ [LNSY: $A_e = 1.1047$, 95%CI (1.1015, 1.1079); $H_o = 0.0767$, 95% CI (0.0738, 0.0796); $uHe = 0.0802$, 95% CI (0.0775, 0.0829). PLKK: $A_e = 1.1112$, 95%CI (1.1091, 1.1133); $H_o = 0.0803$, 95% CI (0.0786, 0.0820); $uHe = 0.0845$, 95% CI (0.0832, 0.0858)] or $x = 100$ [LNSY: $A_e = 1.1006$, 95%CI (1.0970, 1.1042); $H_o = 0.0731$, 95% CI (0.0703, 0.0759); $uHe = 0.0771$, 95% CI (0.0745, 0.0797). PLKK: $A_e = 1.1121$, 95%CI (1.1089, 1.1153); $H_o = 0.0803$, 95% CI (0.0781, 0.0825); $uHe = 0.0848$, 95% CI (0.0825, 0.0871)]. At the same time, the ΔK line chart showed a peak at $x = 30$ (Supplementary Table S3).

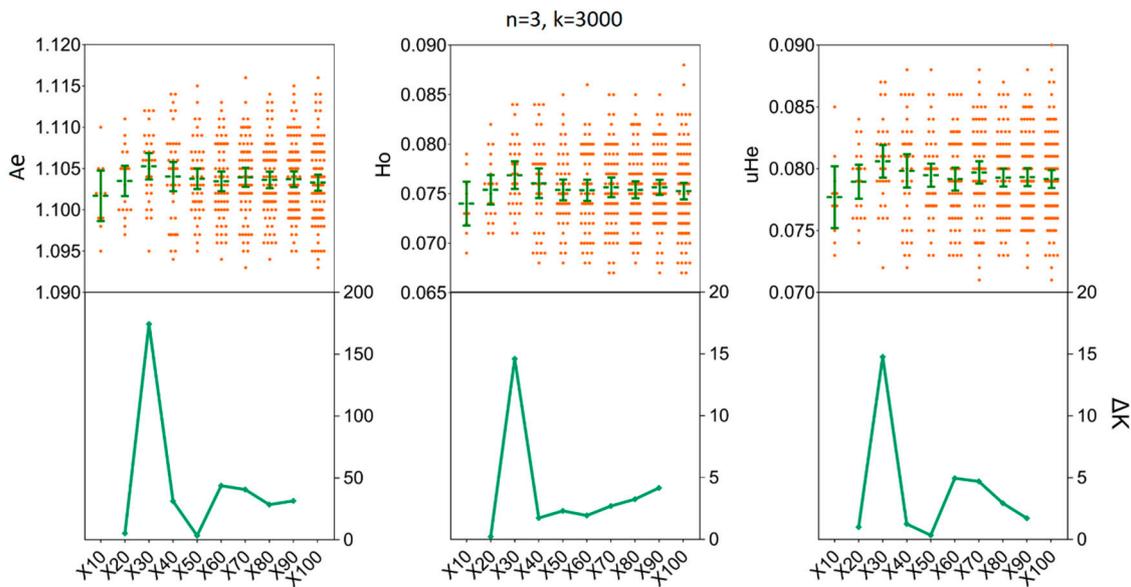


Figure 3. The minimum number of resampling replicates (x) required to estimate the genetic diversity in LNSY population. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y -axis) shows a clear peak at the optimal replicates (x). Ae , number of effective alleles; H_o , observed heterozygosity; uHe , unbiased expected heterozygosity.

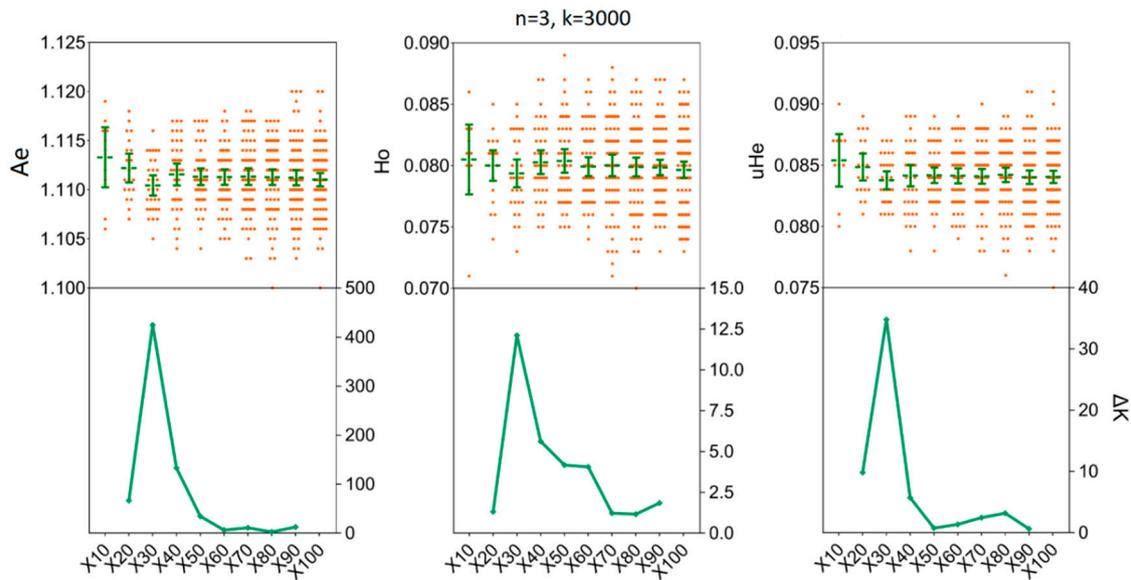


Figure 4. The minimum number of resampling replicates (x) required to estimate the genetic diversity in PLKK population. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y -axis) shows a clear peak at the optimal replicates (x). Ae , number of effective alleles; H_o , observed heterozygosity; uHe , unbiased expected heterozygosity.

Our simulations allowed us to determine the minimum sample size of *H. axyridis* required to ensure that the sample precisely reflects the genetic diversity of the empirical data sets. In the LNSY population, increasing sample sizes above four ($n \geq 4$) individuals appears to have little impact on the mean Ae , H_o and uHe when 3000 SNPs were selected for $n = 4$ [$Ae = 1.1071$, 95%CI (1.1056, 1.1085); $H_o = 0.0764$, 95% CI (0.0750, 0.0777); $uHe = 0.0805$, 95% CI (0.0793, 0.0817)] and for $n = 15$ [$Ae = 1.1097$, 95%CI (1.1093, 1.1102); $H_o = 0.0758$, 95% CI (0.0754, 0.0762); $uHe = 0.0807$, 95% CI (0.0804, 0.0811)] (Figure 5). Also, at $n = 4$, the ΔK line chart showed a clear peak (Figure 5). For the PLKK population, sample sizes above six individuals appear to have little impact on the mean H_o , when 3000 SNPs

were considered. The mean values of H_o for $n = 6$ was 0.0795 [95% CI (0.0788, 0.0802)] and for $n = 15$ was 0.0795 [95% CI (0.0792, 0.0799)]. Simultaneously, the ΔK line chart showed a clear peak at $n = 6$ (Figure 6). For Ae and uHe parameters, a small sample size ($n = 4$) with 3000 SNPs was sufficient to recover the genetic diversity found in PLKK populations [for $n = 4$: $Ae = 1.1149$, 95% CI (1.1139, 1.1159); $uHe = 0.0848$, 95% CI (0.0842, 0.0855)]; for $n = 15$, $Ae = 1.1185$, 95% CI (1.1182, 1.1189) and $H_o = 0.0849$, 95% CI (0.0847, 0.0852)] (Table S4). The ΔK line chart shows a clear peak at $n = 4$, separately (Figure 6).

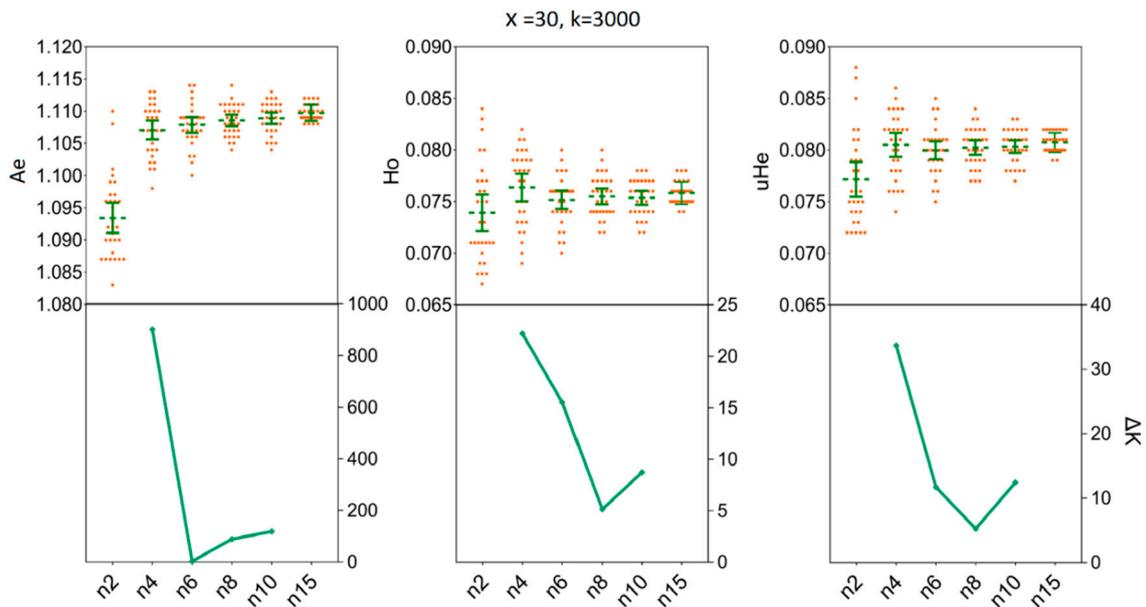


Figure 5. The minimum sample size (n) required to estimate the genetic diversity in LNSY population. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y-axis) show a clear peak at the minimum sample sizes (n). Ae , number of effective alleles; H_o , observed heterozygosity; uHe , unbiased expected heterozygosity.

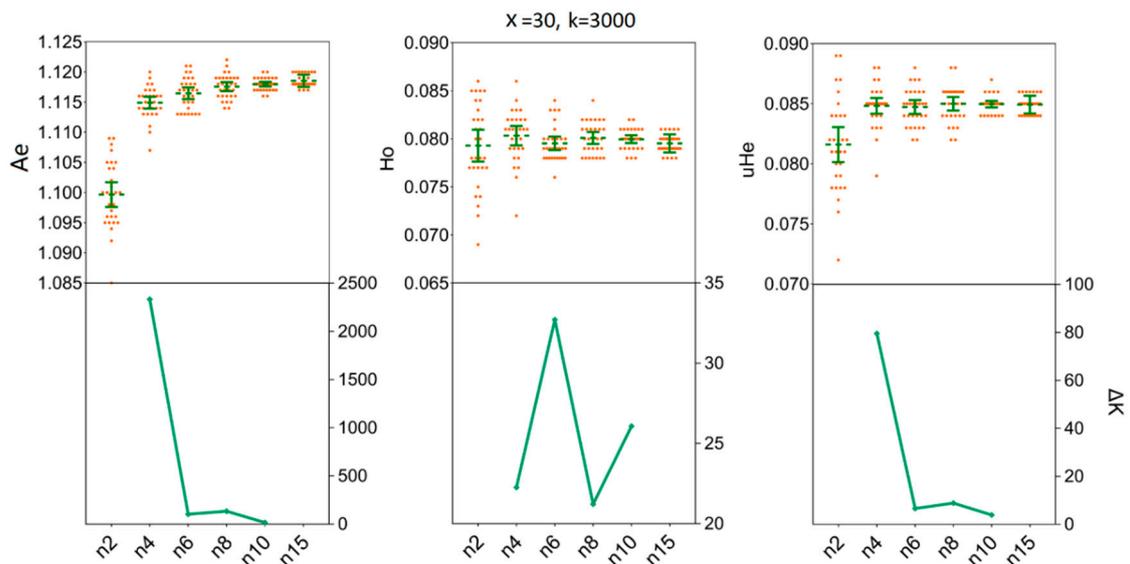


Figure 6. The minimum sample size (n) required to estimate the genetic diversity in PLKK population. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y-axis) shows a clear peak at the minimum sample sizes (n). Ae , number of effective alleles; H_o , observed heterozygosity; uHe , unbiased expected heterozygosity.

3.4. Determination of the Sample Sizes for Interpopulation Genetic Diversity

To estimate the degree of population genetic differentiation, our results showed that compared to $x = 100$, there was no statistical difference for the mean values of F_{st} when we set the number of replicates to $x = 60$; the number of individuals (n) were fixed to three and the number of SNPs were fixed to 3000. For instance, the mean values of F_{st} for $x = 60$ were 0.0398 [95% CI (0.0265, 0.0531)] and for $x = 100$ were 0.0347 [95% CI (0.0286, 0.0408)] (Supplementary Table S3). At the same time, the ΔK line chart showed a peak at $x = 60$ (Figure 7). Furthermore, increasing sample size above six ($n \geq 6$) did not decrease the sample size needed to recover the genetic differentiation of individual populations of *H. axyridis* based on 60 replicates (Figure 8). The mean values of F_{st} for $n = 6$ were 0.0370 [95% CI (0.0345, 0.0396)] and for $n = 15$ were 0.0394 [95% CI (0.0388, 0.0400)]. Similarly, the ΔK line chart showed a peak at $n = 6$ (Supplementary Table S4).

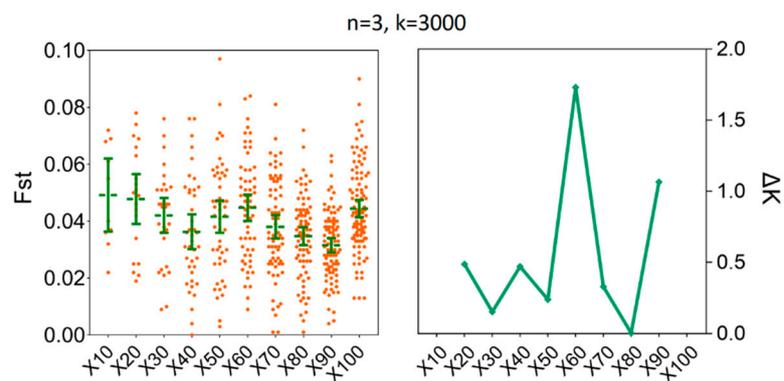


Figure 7. The minimum number of resampling replicates (x) required to estimate F_{st} between LNSY and PLKK populations. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y-axis) shows a clear peak at the optimal replicates (x). F_{st} , pairwise genetic differentiation.

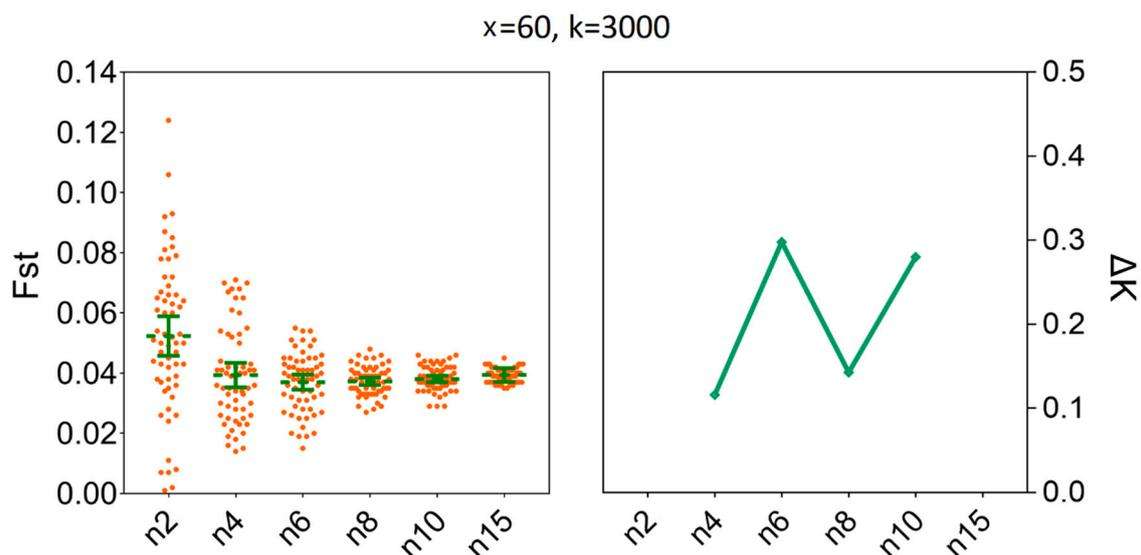


Figure 8. The minimum sample size (n) required to estimate F_{st} between LNSY and PLKK populations. Bars represent sample means and 95% confidence intervals of the means. The ΔK (Y-axis) shows a clear peak at the minimum sample sizes (n). F_{st} , pairwise genetic differentiation.

4. Discussion

In this study, we conducted a rigorous empirical determination of the optimal sample size in an invasive ladybeetle, *H. axyridis*, which confirmed our hypothesis that a minimal sample size of 3–8 individuals is sufficient to estimate genetic diversity within and across native and invasive populations. The next step in our studies will be to sample global populations of *H. axyridis* to investigate the genetic diversity. To our knowledge, only one published study has examined a reliable number of individuals to address invasive population genomics using SNP markers [23], while several studies have investigated the impact of sample size based on microsatellite markers [5,8,38–41].

Sample size is an important study design factor that influences population genomic studies [1,15,23,42,43]. There are generally two types of sampling that occur in invasive genomic studies. First, there is process variance, due to variations in genetic metrics caused by the number of individuals introduced, the diversity and differentiation of the source population(s), multiple introductions, genetic drift, and natural dispersal [44–46]. Second, there is sampling variance, caused by variation in allele frequencies when a subset of individuals (the sample) is drawn from the population [1,23,44].

Our results indicate that, in general, even small sample sizes are likely to be sufficient, which is similar to previous studies [1,15,23,42,43]. Specifically, we found that four to six individuals were enough to calculate within-population genetic diversity estimates using RADseq that provides a large number of SNPs. These results are inconsistent to previous study of another invasive whitefly *B. tabaci*, in which 3–4 individuals were required to recover within-population genetic diversity parameters (Table 1). The exact reason for this phenomenon is unclear here, but may be associated with the invasive process variance of two invasive species (e.g., bottleneck effects and founding effects occurred in different periods for these species, causing different extent of gene flow, etc.). Another interpretation may be our enlarged sampling space in the data set ($n = 20$) than that in previous study ($n = 10$) [23], which could deliberately decrease overlaps between iterations. Actual and recommended sample sizes for evaluating a population have varied widely. For example, previous studies have used a larger data set ($n = 25$ –30 or 30 individuals) from each population to identify the optimal sample size [43,47], and 35 individuals were evaluated in previous study [1]. In our study, we removed SNP markers showing deviation from Hardy–Weinberg equilibrium and those under selection for each population of *H. axyridis* to estimate genetic parameters [1], which should not be overlooked to explain the different results with Qu et al. [23].

Table 1. The optimal sample sizes required for different genetic parameters.

Species Analyzed	Genetic Diversity			Genetic Differentiation	References
	Ae	Ho	uHe	(Fst)	
<i>Bemisia tabaci</i> MED Q1 clade	3	3–4	3	4	Qu et al. 2019
<i>Bemisia tabaci</i> MED Q2 clade	3–4	4	3–4	3	Qu et al. 2019
<i>Amphirrhox longifolia</i>	2	2	6–8	2	Nazareno et al. 2017
<i>Harmonia axyridis</i>	4	4–6	4	6	The present study

Additionally, our results showed that sample sizes of six individuals per population, provide accurate estimates of *Fst*, when a large number of polymorphic SNPs are employed. This minimum sample size in *H. axyridis* is larger than two individuals for an Amazonian plant species, *Amphirrhox longifolia* (Violaceae) and three to four individuals for *B. tabaci* based on empirical analysis [1,23] (Table 1). However, in a simulation study, the population sample size reported can be as small as $n = 4$ –6 to measure *Fst* metrics when using a large number of SNPs (>1000) [2]. It is worth noting that 3000 SNPs as described in Qu et al. 2019 [23] is sufficient for reliable estimation of genetic diversity parameters, which is greater than the number used in previous studies to estimate *Fst* (>1000 or ≥ 1500 SNPs) [1,2]. However, a different study used 23,057 SNPs to determine optimal sample size

for the Galapagos tortoise [42], while another empirical simulation study employed approximately 14,000 SNPs [43]. Thus, we confirmed that this resampling method is effective and robust, but it may be necessary to assess the appropriate sample size for each invasive species prior to the characterization of their invasion genetics. The main reason is these estimates can be affected by many factors in evolutionary process including the bottleneck effect, founder effect, and bridgehead effect [46].

In present study, we selected native (Asian) and invasive (European) populations of *H. axyridis* to conduct a sample size optimization analysis. The SNP data support the presence of two highly divergent lineages of *H. axyridis* populations. Lineage 1 is distributed in China and Lineage 2 generally in Poland, which suggests a long-standing barrier to gene flow between these geographic regions (Figures 1 and 2). Our results revealed that similar sample size could accurately estimate the genetic metrics of the two populations, which indicate that the optimal sample size of *H. axyridis* is not dramatically affected by the invasion process.

5. Conclusions

Our results showed that a sample size greater than six individuals ($n \geq 6$) has little impact on the estimates of genetic diversity within *H. axyridis* populations. Accurate estimates of *Fst* can also be easily obtained at a small simple size (six individuals). The findings demonstrated that SNP markers can accurately estimate the genetic diversity of *H. axyridis* populations, even when small numbers of individuals are sampled, which provides a starting point for future genome-wide population studies.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2075-4450/11/5/290/s1>, Table S1: Sequencing details of *H. axyridis* in LNSY and PLKK populations, Table S2: SNP data of *H. axyridis* in LNSY and PLKK populations, Table S3: Mean and 95% confidence interval (CI) of estimated parameters in resampling replicates in LNSY and PLKK populations, Table S4: Mean and 95% confidence interval (CI) of estimated parameters in resampling individuals in LNSY and PLKK populations.

Author Contributions: Conceptualization, B.L.; methodology, W.Q. and H.L.; software, H.L.; validation, B.L., L.M. and X.Z.; formal analysis, H.L.; investigation, H.L.; resources, H.L.; data curation, H.L.; writing—original draft preparation, H.L.; writing—review and editing, J.J.O., X.Z., D.C. and L.M.; visualization, W.Q.; supervision, B.L.; project administration, B.L.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the National Natural Science Foundation of China, grant number NSFC-31570389, granted to B.P.L.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nazareno, A.G.; Bemmels, J.B.; Dick, C.W.; Lohmann, L.G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Mol. Ecol. Res.* **2017**, *17*, 1136–1147. [[CrossRef](#)] [[PubMed](#)]
2. Willing, E.M.; Dreyer, C.; Oosterhout, C.V. Estimates of genetic differentiation measured by *FST* do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **2012**, *7*, e42649. [[CrossRef](#)] [[PubMed](#)]
3. Swatdipong, A.; Primmer, C.; Vasemagi, A. Historical and recent genetic bottlenecks in European grayling, *Thymallus thymallus*. *Conserv. Genet.* **2010**, *11*, 279–292. [[CrossRef](#)]
4. Nazareno, A.G.; Jump, A.S. Species-genetic diversity correlations in habitat fragmentation can be biased by small sample sizes. *Mol. Ecol.* **2012**, *21*, 2847–2849. [[CrossRef](#)]
5. Hobas, S.; Gaggiotti, O.; Consortium, C.; Bertorelle, G. Sample planning optimization tool for conservation and population genetics (SPOTG): A software for choosing the appropriate number of markers and samples. *Methods Ecol. Evol.* **2013**, *4*, 299–303.
6. Paetkau, D.; Slade, R.; Burden, M.; Estoup, A. Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. *Mol. Ecol.* **2004**, *13*, 55–65. [[CrossRef](#)]
7. Ward, S.M.; Jasieniuk, M. Review: Sampling weedy and invasive plant populations for genetic diversity analysis. *Weed Sci.* **2009**, *57*, 593–602. [[CrossRef](#)]

8. Hale, M.L.; Burg, T.M.; Steeves, T.E. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE* **2012**, *7*, e45170. [[CrossRef](#)]
9. Landguth, E.L.; Fedy, B.C.; Oyeler-McCance, S.J.; Garey, A.L.; Emel, S.L.; Mumma, M.; Wanger, H.H.; Fortin, M.J.; Cushman, S.A. Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Mol. Ecol. Res.* **2012**, *12*, 276–284. [[CrossRef](#)]
10. Peery, M.Z.; Kirby, R.; Reid, B.N.; Stoelting, R.; Doucet-B  er, E.; Robinson, S.; V  squez-Carrillo, C.; Pauli, J.N.; Palsb  ll, P.J. Reliability of genetic bottleneck tests for detecting recent population declines. *Mol. Ecol.* **2012**, *21*, 3403–3418. [[CrossRef](#)]
11. Lombaert, E.; Guillemaud, T.; Deleury, E. Biases of STRUCTURE software when exploring introduction routes of invasive species. *Heredity* **2018**, *120*, 485–499. [[CrossRef](#)] [[PubMed](#)]
12. Frankham, R.; Ballou, J.D.; Ralls, K.; Eldridge, M.D.B.; Dudash, M.R.; Fenster, C.B.; Lacy, R.C.; Sunnucks, P. *Genetic Management of Fragmented Animal and Plant Populations*; Oxford University Press: Oxford, UK, 2017.
13. Saura, M.; Fern  ndez, A.; Rodr  guez, M.C. Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient iberian pigs. *PLoS ONE* **2013**, *8*, e78314. [[CrossRef](#)] [[PubMed](#)]
14. Toro, M.A.; Villanueva, B.; Fern  ndez, J. Genomics applied to management strategies in conservation programs. *Livest. Sci.* **2014**, *166*, 48–53. [[CrossRef](#)]
15. Jeffries, D.L.; Copp, G.H.; Handley, L.L.; Ols  n, H.K.; Sayer, C.D.; H  nfling, B. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Biol. Invasions* **2016**, *25*, 2997–3018. [[CrossRef](#)] [[PubMed](#)]
16. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [[CrossRef](#)] [[PubMed](#)]
17. Andrews, K.R.; Good, J.M.; Miller, M.R.; Luikart, G.; Hohenlohe, P.A. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **2016**, *17*, 81–92. [[CrossRef](#)]
18. Van Tassell, C.P.; Smith, T.P.; Matukumalli, L.K.; Taylor, J.F.; Schnabel, R.D.; Taylor Lawley, C.; Haudenschild, C.D.; Moore, S.S.; Warren, W.C.; Sonstegard, T.S. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* **2008**, *5*, 247–252. [[CrossRef](#)]
19. Deagle, B.E.; Faux, C.; Kawaguchi, S.; Meyer, B.; Jarman, S.N. Antarctic krill population genomics: Apparent panmixia, but genome complexity and large population size muddy the water. *Mol. Ecol.* **2015**, *24*, 4943–4959. [[CrossRef](#)]
20. Martin, C.H.; Crawford, J.E.; Turner, B.J.; Turner, B.J.; Simons, L.H. Diabolical survival in Death Valley: Recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proc. Royal Soc. B* **2016**, *283*, 20152334. [[CrossRef](#)]
21. Ozerov, M.Y.; Gross, R.; Bruneaux, M.; V  h  , J.P.; Burimski, O.; Pukk, L.; Vasem  gi, A. Genome-wide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. *Mol. Ecol.* **2016**, *25*, 1275–1293. [[CrossRef](#)]
22. Vera, M.; D  ez-del-Molino, D.; Garc  a-Mar  n, J.L. Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Mol. Ecol.* **2016**, *25*, 1089–1105. [[CrossRef](#)] [[PubMed](#)]
23. Qu, W.M.; Liang, N.; Wu, Z.K.; Zhao, Y.G.; Chu, D. Minimum sample sizes for invasion genomics: Empirical investigation in an invasive whitefly. *Ecol. Evol.* **2020**, *10*, 38–49. [[CrossRef](#)] [[PubMed](#)]
24. Koch, R.L. The multicolored Asian lady beetle, *Harmonia axyridis*: A review of its biology, uses in biological control, and non-target impacts. *J. Insect Sci.* **2003**, *3*, 1–16. [[CrossRef](#)]
25. Brown, P.M.J.; Thomas, C.E.; Lombaert, E.; Jeffries, D.L.; Estoup, A.; Handley, L.J.L. The global spread of *Harmonia axyridis* (Coleoptera: Coccinellidae): Distribution, dispersal and routes of invasion. *Biocontrol* **2011**, *56*, 623–641. [[CrossRef](#)]
26. Van Lenteren, J.C. The state of commercial augmentative biological control: Plenty of natural enemies, but a frustrating lack of uptake. *BioControl* **2012**, *57*, 1–20. [[CrossRef](#)]
27. Sloggett, J.J. *Harmonia axyridis* invasions: Deducing evolutionary causes and consequence. *Entomol. Sci.* **2012**, *15*, 261–273. [[CrossRef](#)]
28. Roy, H.E.; Brown, P.M.J.; Adriaens, T.; Zhao, Z.H. The harlequin ladybird, *Harmonia axyridis*: Global perspectives on invasion history and ecology. *Biol. Invasions* **2016**, *18*, 997–1044. [[CrossRef](#)]

29. Keller, S.R.; Taylor, D.R. History, chance and adaptation during biological invasion: Separating stochastic phenotypic evolution from response to selection. *Ecol. Lett.* **2008**, *11*, 852–866. [[CrossRef](#)]
30. Simberloff, D.; Martin, J.L.; Genovesi, P.; Maris, V.; Wardle, D.A.; Aronson, J.; Franck, C.; Galil, B.; García-Berthou, E.; Pascal, M.; et al. Impacts of biological invasions: What's what and the way forward. *Trends Ecol. Evol.* **2013**, *28*, 58–66. [[CrossRef](#)]
31. Lombaert, E.; Guillemaud, T.; Lundgren, J.; Koch, R.; Facon, B.; Grez, A.; Loomans, A.; Malausa, T.; Nedved, O.; Rhule, E.; et al. Complementarity of statistical treatments to reconstruct worldwide routes of invasion: The case of the Asian ladybird *Harmonia axyridis*. *Mol. Ecol.* **2014**, *23*, 5979–5997. [[CrossRef](#)]
32. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
33. Wang, S.; Meyer, E.; McKay, J.K.; Matz, M.V. 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nat. Methods* **2012**, *9*, 808. [[CrossRef](#)] [[PubMed](#)]
34. Fu, X.T.; Dou, J.Z.; Mao, J.X.; Su, H.L.; Jiao, W.Q.; Zhang, L.L.; Hu, X.L.; Huang, X.T.; Wang, S.; Bao, Z.M. RADtyping: An integrated package for accurate de novo codominant and dominant RAD genotyping in mapping populations. *PLoS ONE* **2013**, *8*, e79960. [[CrossRef](#)] [[PubMed](#)]
35. Ruan, J.; Li, H.; Chen, Z.Z.; Coghlan, A.; Coin, L.J.M.; Guo, Y.R.; Hériché, J.; Hu, Y.F.; Kristiansen, K.; Li, R.Q.; et al. TreeFam: 2008 Update. *Nucleic Acids Res.* **2008**, *36*, 735–740. [[CrossRef](#)]
36. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)]
37. Peakall, R.; Smouse, P.E. Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]
38. Luikart, G.; Cornuet, J.M. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv. Biol.* **1998**, *12*, 228–237. [[CrossRef](#)]
39. Koskinen, M.T.; Hirvonen, H.; Landry, P.A.; Primmer, C.R. The benefits of increasing the number of microsatellites utilized in genetic populations studies: An empirical perspective. *Hereditas* **2004**, *141*, 61–67. [[CrossRef](#)]
40. Kalinowski, S.T. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* **2005**, *94*, 33–36. [[CrossRef](#)]
41. González-Ramos, J.; Agell, G.; Uriz, M.J. Microsatellites from sponge genomes: The number necessary for detecting genetic structure in Hemimycale columella populations. *Aquat. Biol.* **2015**, *24*, 25–34. [[CrossRef](#)]
42. Gaughran, S.J.; Quinzin, M.C.; Miller, M.J.; Garrick, R.C.; Edwards, D.L.; Russello, M.A.; Poulakakis, N.; Ciofi, C.; Beheregaray, L.B.; Caccone, A. Data from: Theory, practice, and conservation in the age of genomics: The Galápagos giant tortoise as a case study. *Evol. Appl.* **2017**, *7*, 1084–1093.
43. Flesch, E.P.; Rotella, J.J.; Thomson, J.M.; Graves, T.A.; Garrott, R.A. Evaluating sample size to estimate genetic management metrics in the genomics era. *Mol. Ecol. Res.* **2018**, *18*, 1077–1091. [[CrossRef](#)] [[PubMed](#)]
44. Holsinger, K.; Eir, B.S. Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST} . *Nat. Rev. Genet.* **2009**, *10*, 639–650. [[CrossRef](#)] [[PubMed](#)]
45. Lallias, D.; Boudry, P.; Batista, F.M.; Beaumont, A.; King, J.W.; Turner, J.R.; Lapègue, S. Invasion genetics of the Pacific oyster *Crassostrea gigas*, in the British Isles inferred from microsatellite and mitochondrial markers. *Biol. Invasions* **2015**, *17*, 1–15. [[CrossRef](#)]
46. Chu, D.; Qu, W.M.; Guo, L. Invasion genetics of alien insect pests in China: Research progress and future prospects. *J. Integr. Agric.* **2018**, *18*, 748–757. [[CrossRef](#)]
47. Hoban, S.; Schlarbaum, S. Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. *Biol. Conserv.* **2014**, *177*, 90–99. [[CrossRef](#)]

