Received Date : 10-Oct-2014

Revised Date : 20-Jul-2015

Accepted Date : 22-Jul-2015

Article type : Original Article

Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity

Ana Vilas, Andrés Pérez-Figueroa, Humberto Quesada and Armando Caballero

Dep. Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, 36310 Vigo, Spain

CORRESPONDING AUTHOR:

Armando Caballero, Dpto. de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, 36310 Vigo, SPAIN

Tel: +34 986 812 568. Fax: +34 986 812 556

E-mail: armando@uvigo.es

Short title: Allelic diversity and adaptive potential

Keywords: number of alleles, gene diversity, heterozygosity, response to selection, synthetic populations

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.13334

Abstract

The adaptive potential of a population depends on the amount of additive genetic variance for quantitative traits of evolutionary importance. This variance is a direct function of the expected frequency of heterozygotes for the loci which affect the trait (QTL). It has been argued, but not demonstrated experimentally, that long-term response to selection is more dependent on QTL allelic diversity than on QTL heterozygosity. Conservation programs, aimed at preserving this variation, usually rely on neutral markers rather than on quantitative traits for making decisions on management. Here we address, both through simulation analyses and experimental studies with *Drosophila melanogaster*, the question of whether allelic diversity for neutral markers is a better indicator of a high adaptive potential than expected heterozygosity. In both experimental and simulation studies, we established synthetic populations for which either heterozygosity or allelic diversity was maximized using information from QTL (simulations) or unlinked neutral markers (simulations and experiment). The synthetic populations were selected for the quantitative trait in order to evaluate the evolutionary potential provided by the two optimization methods. Our results show that maximizing the number of alleles of a low number of markers implies higher responses to selection than maximizing their heterozygosity.

Introduction

It is universally recognized that the evolutionary potential of a population is proportional to its amount of genetic diversity, *i.e.* the variation in inheritable characteristics possessed by the population. Most traits of evolutionary or economic interest are quantitative in nature and genetic variation for these traits is measured by the additive genetic variance, which is proportional to the expected heterozygosity of the loci controlling the trait (Falconer & Mackay 1996). The number of alleles per locus (allelic diversity) is also a relevant measure of genetic diversity, being more sensitive to population bottlenecks than expected heterozygosity (Nei *et al.* 1975; Allendorf 1986; Cornuet & Luikart 1996; Luikart *et al.* 1998; Leberg 2002). Theoretical studies using biallelic locus models and analysing the selection response from populations with different initial census sizes (James 1970; Hill & Rasbash 1986) suggest that the response to selection heavily depends on the contribution of rare alleles. It has then been argued that although the short-term (immediate) response to selection is a direct

function of the expected heterozygosity, the number of allelic variants after a bottleneck mightbe the main factor responsible for the response to long-term selection or selection limits (Allendorf 1986; Allendorf *et al.* 2013; p. 289). Recent simulation studieshave in fact shown that gene frequency measures of variation such as heterozygosity are good predictors of short-term response to selection for the trait, whereas allelic number measures are better predictors of long-term and total response (Caballero & García-Dorado 2013). These predictions, however, have not been experimentally demonstrated.

When genetic variance components for traits of interest cannot be obtained, as it is very often the case, genetic variation for neutral molecular markers is used as a surrogate of variation for quantitative traits. The assessment of genetic variation through neutral molecular markers is current in use in population genetics and conservation biology studies (Toro *et al.* 2009; Kirk & Freeland 2011; Frankham *et al.* 2010; Allendorf *et al.* 2013). However, the relationship between neutral molecular variation and quantitative trait variation is controversial. The mean correlation between molecular heterozygosities and estimates of genetic variation for a range of quantitative traits across 71 studies was about 0.22 (Reed & Frankham 2001) whereas the correlation between variation at neutral markers and fitness traits was 0.43 in another meta-analysis of 34 studies (Reed & Frankham 2003). Moreover, the relationship between individual genetic diversity and fitness-related traits (heterozygosity-fitness correlation) is generally low in large populations at equilibrium (Coltman & Slate 2003; Slate *et al.* 2004; Chapman *et al.* 2009), but can be relevant in small populations or those recently affected by bottlenecks (Szulkin *et al.* 2010). In addition, direct associations between neutral marker variation and adaptive changes for quantitative traits have been experimentally found (Santos *et al.* 2012).

For neutral molecular markers, expected heterozygosity, also referred to as gene diversity (Nei 1973), is the most frequent used statistic to measure diversity. Maximization of the expected heterozygosity, equivalent to minimization of mean coancestry (kinship), is the most widely proposed criterion to maintain genetic diversity both in conservation programs (Ballou & Lacy 1995; Caballero & Toro 2000, 2002) and animal breeding programs (Toro & Pérez-Enciso 1990; Grundy *et al.* 1998; Meuwissen2007), as it implies the maximization of effective population size (Caballero & Toro 2000, 2002) and because of its direct relationship with the magnitude of

additive genetic variance and inbreeding depression. However, for multiallelic molecular markers, such as microsatellites, the number of different allelic types is also used to describe the genetic diversity of populations (Barker 2001; Foulley & Ollivier 2006; Caballero & Rodríguez-Ramilo 2010), to ascertain the occurrence of past population bottlenecks (Cornuet & Luikart 1996; Luikart *et al.* 1998) and to establish conservation priorities (Petit *et al.* 1998; Fernández *et al.* 2004; Simianer 2005; Medugorac *et al.* 2011). In addition, allelic heterogeneity is often overlooked in association studies, as standard genome wide association studies (GWAS) and linkage studies are based on bi-allelic markers (Nelson *et al.* 2013). However, sequencing technologies are allowing for this allelic heterogeneity to be unmasked (Flister *et al.* 2013).

Differences in allelic number rather than differences in allele frequencies have also been proposed as an alternative way to quantify genetic differentiation between populations (Jost 2008; but see, *e.g.*, Whitlock 2011; Wang 2012). In the context of subdivided populations, where rare allelic types, present in a given subpopulation, may arrive by migration into another subpopulation and contribute to its adaptation, allelic variation may play a fundamental role (Blanquart *et al.* 2012; Caballero & García-Dorado 2013). There is, therefore, an increased consensus in the idea that allelic diversity is of great importance for the evolutionary potential of populations (Slade & McCallum 1992; Petit *et al.* 1998; Wilson *et al.* 2009; Medugorac *et al.* 2011; Caballero & García-Dorado 2013) and a number of methods and computer tools have been recently developed to estimate and predict allelic numbers (Belkhir *et al.* 2006; Szpiech*et al.* 2008; Bashalkhanov *et al.* 2009), as well as to specifically retain the largest possible number of alleles in conservation programs (Fernández *et al.* 2004; Weiser *et al.* 2012).

One important question in the field of conservation genetics is what kind of diversity from neutral molecular markers should be optimized in order to keep the maximal evolutionary potential of a population. Caballero and García-Dorado (2013) performed simulations based on simulating different scenarios with different population sizes and mutation rates (generally focusing on subdivided populations) and looked for correlations between variability parameters in these highly variable population scenarios and responses to natural selection for a changing selection optima. They showed that

allelic diversity measures are more correlated to long-term and total response to selection than gene frequency diversity measures. Regarding single undivided populations, the results were clear forquantitative trait loci (QTL) variation. Although the same tendency was also found for neutral markers, the results were, however, much less obvious in this case. Thus, it remains to be shown if the above results also hold for populations of identical size but differing in their levels of genetic variability. In addition, it should also be seen whether these results hold when neutral marker variation is considered.

Many populations, both in nature and in conservation settings (zoos, botanic gardens, natural reserves, germplasm collections, etc.) are subdivided, and sometimes a decision has to be taken so as how to mix subpopulations to produce a synthetic population with the highest diversity. We use this practical setting to answer the following questions: (1) Can be demonstrated that for populations of the same size maximization of allelic diversity for unlinked neutral markers conveys a larger response to selection for a quantitative trait than maximization of their expected heterozygosity? (2) Can such a prediction be shown experimentally? (3) Are neutral molecular markers reliable for making conservation decisions regarding variation for quantitative traits?

To respond to these questions we used simulation studies and experimental data with *Drosophila melanogaster*. In both studies we established synthetic populations from a structured base population under different optimization criteria (maximization of expected heterozygosity or allelic diversity). In the simulation, optimization was implemented using information both from the loci controlling the trait (QTL) and from unlinked neutral markers. In the experiment, optimization was based on microsatellite markers. The synthetic populations were subjected to artificial selection for the quantitative trait (sternopleural bristle number in the experimental study) in order to compare the ability of the optimization methods to retain the evolutionary potential of the populations.

Materials and methods

Simulation study

The simulation procedure was carried out with the following steps. First, we simulated a large base population subdivided in eight subpopulations under four different historical scenarios. Second, from this subdivided base population, we built synthetic populations with maximal heterozygosity or maximal allelic diversity, both based on information from the loci controlling a quantitative trait (QTL) or from neutral molecular markers unlinked to the QTL. Finally, we subjected the synthetic populations to artificial selection for the quantitative trait.

Base population structure: The software quantiNemo (v1.0.4, Neuenschwander *et al.*, 2008), was used to simulate base population structures under four different scenarios, as shown in Figure 1. In scenarios A and B a large original population of size $N = 4000$ hermaphrodite individualswas first run for 13000 generations to reach sufficient genetic variation. This population was then subdivided into eight subpopulations of equal or different sizes, which were maintained separately for a further 50 generations. In scenarios C and D, two large ancestral populations were established instead of one, seven subpopulations arising from the first and the last one from the second. All populations were maintained throughout assuming random mating. 100 replicates of each scenario were simulated. Scenario A was used to simulate a range of different genetic parameters.

Quantitative trait and neutral markers: A quantitative trait subjected to stabilising selection was assumed to be controlled by 10 to 1000 unlinked multiallelic QTL. In addition, 10 to 1000 neutral multiallelic unlinked markers were simulated. All combinations of numbers of QTL and markers were run. For each locus, the initial allelic frequency distribution assumed for both the QTL and neutral markers was a uniform one for all the 256 possible alleles, the maximum number of alleles per locus defined in the software. New mutationsoccurred randomly to any of the possiblealleles (random mutation model of the software) with mutation rate generally $\mu = 5 \times 10^{-5}$ per locus and generation (values ten times smaller or larger were also simulated).

Allelic QTL effects on the quantitative trait were pre-assigned with the option of the quantiNemo software to cover a uniform distribution with valuesequally spaced between -6σ and 6σ, where $\sigma^2 = 0.625 / n$ and n is the number of QTL controlling the quantitative trait. This assured the same additive genetic variance for the different numbers of QTL considered. The genotypic value of each individual was obtained as the sum of the values of their allelic effects, and the phenotypic value was obtained by adding to the genotypic value an environmental deviation obtained from a normal distribution with mean zero and variance 0.8, constant along generations. The mean heritability of the hypothetical trait at generation 13,050 was about 0.20.

Individuals to be mated were chosen according to a probability proportional to their fitness. This was obtained from the function $w_i = \exp[-(p_i - \mathrm{Opt})^2 / 2V_S]$ (Turelli 1984), where w_i is the fitness of individual i, p_i is its phenotypic value, Opt is the optimum value for the quantitative trait, which was always assumed to be zero, and V_S is an inverse measure of the intensity of the stabilising selection. The value of V_S assumed was typically 100 but stronger selection (V_S = 1 or 20) was also considered.

Synthetic populations: We used the software METAPOP (v2.0.a1, Pérez-Figueroa *et al.* 2009) to obtain the optimal number of individuals contributed by each of the eight subpopulations in order to create a synthetic population of 100 individuals under different optimization criteria (Figure 2). These optimization criteria were aimed to maintain: the largest total expected heterozygosity (O_H) or the largest total number of alleles (O_A), either using information from the QTL or from the neutral markersof the eight subpopulations. METAPOP software applies a simulating annealing algorithm (Kirkpatrick *et al.* 1983) scanning the whole space of subpopulation contributions until the maximum value of the criterion is reached. Note that these optimizations are based on population genetic parameters rather than on individual genotypes. Therefore, the individuals contributed by a given subpopulation are chosen randomly from the subpopulation rather than according to their particular genotype. For each of the 100 simulation replicates of each scenario, optimization was repeated 100 times to obtain precise results.

Artificial selection for the quantitative trait: Once the optimized contributions from subpopulations were obtained, a C program was used to gather the genotypes (QTL and markers) of the 100 individuals contributed bythe subpopulations to form the synthetic pools. Then, artificial selection was performed on these 100 individuals by selecting the 20 with the largest phenotypic value for the quantitativetrait (a 50% proportion was also considered) (Figure 2). Random mating of selected parents was carried out and genotypic and phenotypic values of the progeny were obtained as explained above. Neither new mutation nor stabilising selection were considered during the period of artificial selection. The selective process was maintained for 25-100 generations depending on the number of QTL involved.

Experimental study

An experimental setting analogous to that followed in the simulation study was carried out with *Drosophila melanogaster* (Figure 3). First, a number of experimental lines were established starting from a large base population, in order to create a structured population. This structured population was analysed for nine microsatellite markers and, using this information, synthetic populations were created maximizing the expected heterozygosity or the allelic diversity for the markers. Artificial selection was then carried out for increased and decreased bristle number for 8 generations. This experimental designis explained in what follows.

Base population and culture conditions: A large population, found in 2006 and maintained with about 3000 individuals (Ávila *et al.* 2011), was used for the experiment. Flies were reared in a culture medium composed of 1 L water, 100 g brewer's yeast, 100 g sucrose, 12 g agar, 2.5 g NaCl, and 5 mL propionic acid, and were handled at room temperature under CO_2 anaesthesia. All cultures were incubated in a chamber at 25 ± 1 °C, 65 ± 5 % relative humidity, and maintained under continuous lighting. Virgin males and females were used for mating across the entire experiment.

Experimental lines: Twenty lines (lines L1, L2, ... L20) were established from the large base population, each line being generated with 50 males and 50 females taken at random from the large base population. These lines were maintained isolated with mass mating and constant population size (50 males and 50 females) for 58 generations (Figure 3).

Analysis of diversity using microsatellite loci: Genetic diversity was analysed using nine microsatellite loci (Ávila *et al.* 2011). Details of the protocol followed is given in Online Supporting information 1. At generation 50 of the lines´s maintenance, 40 individuals sampled per line as well as 100 individuals sampled fromthe large base population were genotyped for the microsatellites. The software Genepop (v4.2, Rousset 2008) was used to estimate the heterozygosity (*H*), number of alleles (*A*), and F_{ST} (Weir & Cockerham 1984). The effective population size of the lines (N_e) was estimated from $F_{ST} = 1 - (1 - 1/2N_e)^t$ (Falconer & Mackay 1986), where *t* is the generation number. The maximum likelihood approach implemented in the software ML_NullFreq (Kalinowski & Taper 2006) was used toaccount for genotypic errors and the presence of null alleles.

Creation of a complex structured population using the experimental lines: The 20lines were used to create a complex population structure. For this, some lines were assumed to be pooled in larger groups (Figure 3). Flies were not actually crossed but pools were made with the genotypic data, which is equivalent to assume crossing followed by panmixia. In this grouping, eight lines remained independent (L1 to L8), lines 9-16 were pooled in pairs (L9-10, L11-12, L13-14, and L15-16), and lines 17-20 were pooled in a single large subpopulation. Thus, a complex structuredpopulation was composed of 8 small subpopulations, 4 subpopulations twice as large, one four times as large, and the large base population.

Synthetic populations: Using the microsatellite data from the structured population, synthetic populations of *N* = 100 individuals (50 males and 50 females) were created using, as before, the software METAPOP (v2.0.a1, Pérez-Figueroa *et al.* 2009). The optimal contributions (*i.e.* number of individuals) from each subpopulation were found to maximize, as before, the expected heterozygosity in the synthetic populations (O_H), or the total number of alleles (O_A). Apopulation analogous to the synthetic ones wasalso established where 100 individuals were randomly taken from the original large base population (BP). Actual individuals to create the synthetic populations were obtained from generation 58 of the lines or pools of lines.With the optimized set of contributions obtained by O_H, 100 individuals (50 males and 50 females) were randomly sampled from the pertinent subpopulations, and the same procedure (with the same contributions) was repeated four times to produce four replicates. The same was done for the O_A contributions. Four replicates were also obtained of the population BP by

randomly sampling four groups of 100 individuals (50 males and 50 females) from the original base population.

Artificial selection: Once the synthetic populations (and BP) were established, they were subjected to upward and downward artificial selection for sternopleural bristles. Every generation, 100 flies (50 males and 50 females) from each population, replicate and direction of selection were evaluated for the number of sternopleural bristles in both sternopleural plates. The 10 females and the 10 males with the highest (upward selection) or lowest (downward selection) number of bristles were selected to be the parents of the next generation. Mass mating of the selected parents was carried out in single bottles. Selection was continued for 8 generations. Average heterozygosity and numbers of alleles were estimated for the 9 microsatellite loci at generation 0 of selection (50 individuals per replicate and population type) and generation 8 (40 individuals for each replicate, population type and direction of selection). In total, the experiment implied the genotyping of 2,460 individuals for 9 microsatellites.

Results

Simulation study

Synthetic populations were obtained with contributions from the different subpopulations simulated under the two optimization methods (O_H and O_A; see Supporting Information Table A1 for examples of average contributions from each subpopulation). The average values of heterozygosity (*H*) and number of alleles (*A*) for the synthetic populations obtained from optimizations based on QTL or on markers are shown in Tables 1 and 2, respectively. Two main patterns emerge from these results.

First, we observe that when the number of loci is small (10 QTL or 10 markers; cases a-i in Tables 1 and 2), the synthetic populations obtained by the method O_A were,for all scenarios and parameters, those with the largest number of alleles for the set of loci optimized, as expected. Likewise, the synthetic populations obtained by the method O_H were those with the largest values of heterozygosity. Therefore, the optimization methods achieved their objectives. When QTL data was used for the optimization, the method which gave the highest response to artificial selection was

O_A (Table 1). Remarkably, the same result was observed when neutral markers unlinked to the QTL were used for optimization (Table 2; Supporting Information Figure A1), clearly showing that maximal allele diversity for markers leads to a higher adaptive potential for quantitative traits than maximal marker heterozygosity.

Second, as the number of QTL or markers was increased in the simulations (cases j-o in Tables 1 and 2, respectively), both optimizations (O_H and O_A) tended, however, to lead to the same average values of *H* and *A*. For example, for 500-1000 QTL or markers, the optimizations led to identical average values of *H* and *A* in the synthetic populations and, therefore, selection produced identical responsesfor the two optimization methods (see Tables 1 and 2, and Supporting Information Fig. A2-A3).This can be explained by the decreasingly lower variation in the average values of *H* and *A* for the synthetic populations as the number of loci was increased (Supporting Information Fig. A4). Thus, when a large number of QTL or markers was assumed it was not possible to test, with our approach, whether allelic diversity retains higher adaptive potential than expected heterozygosity.

To get more insight into the impact of the number of QTL or markers in the optimizations, Figure 4 shows the average contributions from the subpopulations to the synthetic populations for Scenario A. When the number of loci was large (either of QTL or of markers) the contributionfrom subpopulation 1 (the largest subpopulation) tended to be about 70%,whereas that from each of the other small subpopulations was about 4%. This was also the case for a low number of markers when the number of alleles was maximized (O_A). However, when heterozygosity was maximized (O_H), the contributions were only close to the optimal ones when the number of loci was large.

Thus, the responses to selection from synthetic populations based on optimizations with neutral markers were similar to those based on QTL when marker allelic diversity was maximized for any number of markers. This is shown in Figure 5, which gives results based on 1000 QTL and a variable number of neutral markers. However, the responses were substantially smaller when optimizations were based on heterozygosity for a low number of neutral markers (Figure 5). This suggests that, when the number of markers is small, optimization through allelic diversity is more effective than optimization through expected heterozygosity.

Experimental study

After 50 generations, the experimental linesmaintained with population size N = 100 had an average heterozygosity for microsatellite markers of H = 0.41 ± 0.02 and an average number of alleles of A = 2.7 ± 0.07 (4.6 in the whole setof lines). The corresponding values in the base population were H = 0.53 and A = 3.9. The estimate of the fixation index for the lines at generation 50 was F_{ST} = 0.29 ± 0.05, from which an estimated average effective line size of N_e = 71.9, was obtained. For the 9 microsatellite markers examined there was no evidence for scoring errors due to stuttering, large allele dropout, or null alleles, and subsequently no correction of allele frequencies was necessary.

The heterozygosities and numbers of alleles in the subpopulations as well as the optimal contributions from these to the synthetic populations are shown in Table 3. Note that methods O_H and O_A produced synthetic populations composed of individuals from several, coincident or not, subpopulations, and no individuals from the base population, despite that this latter was also included in the optimizations. The average heterozygosity and number of alleles in the synthetic populations before selection are shown at the bottomof Table 3 (values for each of the microsatellite loci and replicate are given in Supporting Information Table A2). The optimization methods achieved their objectives. The synthetic populations obtained by the method O_H were those with the largest average heterozygosity for markers, while those obtained by the method O_A were those with the largest average number of alleles. Average values of H and A in the populations formed by individuals from the base population (BP) were intermediate between those for O_H and O_A (see Table 3).

The average number of bristles per replicate and generation of selection for each of the optimization methods is shown in Supporting Information Figure A5. Estimates of the realized heritability, obtained from the linear regression of the cumulative response to selection on the cumulative selection differential across generations were 0.3 for upward selection and 0.2 for downward selection (Supporting Information Figure A6). The heterozygosities and numbers of alleles remaining in the selected lines at generation 8 are given in Table A3.

Figure 6 shows the average cumulative divergence across generations for upward and downward selection. In both directions of selection the method involving the largest response to selection, consistently over generations, was O_A, followed by O_H. The base population (BP) showed the lowest response.Therefore, maximization of marker allelic number led to higher selection responses than maximization of marker heterozygosity.

Discussion

It has beenarguedthat whereas the expected heterozygosity is the main factor responsible for the short-term response to selection, allelic diversity is the main factor responsible for the long-term response (Allendorf 1986; Allendorf *et al.* 2013). This has been deduced from the results of the response to selection after bottlenecks (James 1970), as these have a much larger impact on allelic diversity than on heterozygosity. However, a formal experimental proof of this statement was missing. Note that bottlenecks also reduce, although only a little, the expected heterozygosity. Therefore, the selection response occurred after bottlenecks of different magnitudes cannot be used on its own to investigate the relative importance of one or the other type of variation on the response. Here we bypassed this problem by creating synthetic populations of the same size but either with higher heterozygosity or higher allelic diversity (see Supporting Information Figure A7 for a detailed explanation of the consequences of the two optimization methods). The second issue addressed is whether neutral molecular variation is an appropriate surrogate of quantitative genetic variation in the context of preserving variation in conservation settings. Previous simulation studies showed that, for single undivided populations, gene frequency diversity measures for QTL, such as expected heterozygosity, are the best predictors of short-term adaptation, whereas allelic diversity measures are the best predictors of long-term and total adaptation (Caballero & García-Dorado 2013). Results were not so clear for neutral marker variation however (see Fig. 1 of Caballero & García-Dorado 2013). Here, our objective was to demonstrate, both through simulation and empirically, thatmarker allelic diversity in a single population is abetter indicator of higher adaptive potential than marker expected heterozygosity.

In the simulation study we considered a variable number of QTL and markers and found that, when the number of loci is large (larger than about 50), all optimizations lead to the same results on average. Thus, synthetic populations with different numbers of alleles or different heterozygosities could not be obtained with this optimization approach, and the hypothesis to be addressed could not be tested. When the number of loci is small, nevertheless, it could be shown that optimizations based on maximization of the number of alleles implied a higher adaptive potential for quantitative traits than maximization of expected heterozygosity. Our results therefore support the idea that allelic diversity is the main source of variation with regard to long-term response to selection (Allendorf 1986; Medugorac *et al.* 2011; Weiser *et al.* 2012; Caballero & García-Dorado 2013). We did not find, however, that a higher heterozygosity is associated with a larger short-term selection response. Instead, both O_A and O_H gaveabout the same responses at the start of the selection process (Supporting Information Figures A1-A3) or with even a slighter higher response for O_A (see generation 5 responses in Figure 5). The simulations of Caballero & García-Dorado (2013) showed, in contrast, a stronger correlation of short-term response with QTL heterozygosity than with number of QTL alleles. Although no optimizations were performed to increase either *H* or *A* in that study, there was a large variation in *H* and *A* values between populations when pooling all demographic scenarios assumed, because these involved a range of population sizes (100 to 1000) and mutation rates (0.0001-0.0004). This large variation between *H* or *A* values in the simulated populations allowed for detecting significant correlations between these parameters and short or long-term responses. In contrast, the present simulations were based on single synthetic populations of the same size, obtained from a common base with a single demographic and genetic history, and optimized to have the largest possible *H* or *A*. The achieved differences in *H* and *A*between synthetic populations obtained through O_H and O_A were generally small (Tables 1-2). This is perhaps the reason why it was not possible to observe any substantial differences in the short-term responses achieved. An alternative explanation might be that the scenario of a cross between lines followed by panmixia and selection, as assumed in this study, may imply a certain advantage for allelic diversity also at the short term, as any new allele supplied bya contributing line may imply a larger response from the start.

The fact that optimization methods of allelic diversity based on just 10 unlinked markers gave the same overall result as the optimization based on 1000 QTL (or on hundreds of markers; Figures 4-5) suggests that allelic diversity from a few neutral markers can be rather effective in terms of taking decisions in conservation settings.This supportsthe use of the marker allele number for optimization as a surrogate of quantitative variation. This capacity of neutral markers to predict the evolutionary potential for quantitative traits can be ascribed to the information they convey on the population history and structure rather than to a direct association with the genetic variation responsible for adaptation (Caballero & García-Dorado 2013). These remarkable results obtained with few markers were not extended, however, to maximizationof the expected heterozygosity. In fact, the correlations between *H* for QTL and *H* for markers were generally small and not significant (0.075, 0.063, 0.12, 0.073 for scenarios A to D, respectively). This is in line with the results of several studies which have shown that the correlation between molecular variability for neutral markers and genetic variation for quantitative traits is, at most, moderate (Reed & Frankham 2001, 2003; Chapman *et al.* 2009) and difficult to interpret (Grueber *et al.* 2008). In contrast, the corresponding correlations for number of alleles of QTL and markers were much larger and significant for all scenarios (0.30, 0.26, 0.21 and 0.28 for scenarios A to D respectively). This suggests that this variability measure is more sensitive to previous demographic changes (Allendorf 1986; Cornuet & Luikart 1996; Luikart *et al.* 1998; Leberg 2002), producing and maintaining a more homogeneous and stronger genomic signal.

Note that we do not intend to give the message that the use of a few markers is adequate for any conservation purpose. Rather, our conclusion is that, if only a few markers are available, allelic diversity should be used rather than heterozygosity. Although our results showed that optimizations tended to the optimal ones with large numbers of markers, the current situation is that most studies of conservation are carried out with the use of few genetic markers despite that high density markers are increasingly available in genetic studies. As an illustration, we reviewed (using the Thomson Reuter Web of Science database) the number of publications in the journal Conservation Genetics in the last 5 years (2010-2015) making use of molecular markers. We found that 529 studies used microsatellite data, 153 used mDNA data, and other markers wereused much less often (33 with AFLP, 26 with allozymes, 20 with RAD, 13 with SNP, 7 with RAPD, and 2 with DArT). Considering only the year 2014, the tendency was the same (91with microsatellites, 24 with mDNA, 4 with SNP, 3 with RAD and 1 with AFLP). An analogous five year review (2010-2015) for papers published in the

journal Molecular Ecology which include the term "conservation genetics" found that 211 studies used microsatellite data, 40 mDNA, 28 RAD, 26 SNP data, 6 AFLP and 3 allozymes. Thus, the great majority of conservation genetics studies use microsatellite markers. We also reviewed, among the 89 studies using microsatellites published in Conservation Genetics in 2014, the number of loci used in each study (Figure 7). The average number of loci used was 11.96 and the median was 11.0. This suggests that most conservation studies, at least for non-model species, are based on relatively few multiallelic markers. The results found in this work are, therefore, of applicability to these studies. It should be taken into account that biallelic markers such as SNP or RAD sequencing may take over microsatellites in the near future. However, multiallelic diversity measures can also be applied to SNP markers, either as such (Pérez-Figueroa *et al.* 2012; de Cara *et al.* 2013), which is then equivalent to the proportion of polymorphic loci, or as short haplotypes involving a certain number of SNPs (Pérez-Figueroa *et al.* 2012).

Our experimental evaluation with *Drosophila melanogaster* allowed us to reach the same conclusions as for the simulation study. The synthetic populations obtained with the method O_A using 9 markers produced the highest average response, from the initial generations, followed by those obtained with O_H. The synthetic populations obtained from the base population alone yielded the lowest response for both directions of selection. Overall realized heritabilities (0.3 for upward selection and 0.2 for downward selection) were within the range of estimates obtained for morphological traits in Drosophila (Roff & Mousseau 1987), and with realized heritabilities for other synthetic populations (López-Fanjul & Hill 1973).

In the experiment we could compare the response to selection from the base population alone (BP) with that from synthetic populations obtained exclusively from the 20 isolated lines started from the large base population (O_H and O_A; Table 3). Population genetics theory (Wright 1952) indicates that the total additive genetic variance for an additive neutral quantitative trait in a group of lines started from a base population is 1 + F times the variance in the latter, where F is the inbreeding coefficient of the lines. In our experiment, the experimental lines showed a differentiation of F_{ST} = 0.29. Thus, we could expect that the response to selection from a mixture of lines (although not all lines) would produce up to 29% more response to selection than that obtained from the base population. In agreement with this, the total response to selection with O_A was 25.3% larger than that with BP, and that for O_H was 17.7% larger.

We conclude that, whenthe number of markers available is small, the optimization of the number of alleles is the best method to create synthetic populations with the largest adaptive potential. Our results also support the use of neutral markers to make optimization decisions in conservation.

Acknowledgements

We thank Raquel Sampedro and Nieves Santamaría for technical assistance, and Aurora García-Dorado, Beatriz Villanueva, Jesús Fernández, Miguel Toro and three anonymous reviewers for useful comments on the manuscript. This work was funded by Ministerio de Economía y Competitividad (CGL2012-39861-C02-01), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (RZ, 2012-00011-C02-01), Xunta de Galicia (GPC2013-011) and Fondos Feder: "*Unha maneira de facer Europa*".

References

Allendorf FW (1986) Genetic drift and loss of alleles versus heterozygosity. *Zoo Biology*, **5**, 181-190.

Allendorf FW, Luikart GH, Aitken SN (2013) *Conservation and the Genetics of Populations.John Wiley and Sons*, Chichester, West Sussex, UK.

Ávila V, Fernández J, Quesada H, Caballero A (2011) An experimental evaluation with *Drosophila melanogaster* of a novel dynamic system for the management of subdivided populations in conservation programs. *Heredity*, **106**, 765-774.

Ballou JD, Lacy RC (1995) Identifying genetically important individuals for management of genetic diversity in pedigreed populations. In Ballou JD, Gilpin M, Foose TJ (eds.), *Population Management for Survival and Recovery*. Columbia University Press, New York.

Barker JSF (2001) Conservation and management of genetic diversity: a domestic animal perspective. *Canadian Journal of Forest Research*, **31**, 588-595.

Bashalkhanov S, Pandey M, Rajora OP (2009) A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC genetics*, **10**, 84.

Bataillon TM, David JL, Schoen DJ (1996) Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics*, **144**, 409-417.

Belkhir K, Dawson KJ, Bonhomme F (2006) A comparison of rarefaction and Bayesian methods for predicting the allelic richness of future samples on the basis of currently available samples. *Journal of Heredity*, **97**, 483-492.

Blanquart F, Gandon S, Nuismer S (2012) The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology*, **25**, 1351-1363.

Caballero A, García-Dorado A (2013) Allelic diversity and its implications for the rate of adaptation. *Genetics*, **195**, 1373-1384.

Caballero A, Rodríguez-Ramilo ST (2010) A new method for the partition of allelic diversity within and between subpopulations. *Conservation Genetics*, **11**, 2219-2229.

Caballero A, Toro MA (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research*, **75**, 331-343.

Caballero A, Toro MA (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, **3**, 289-299.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009). A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology*, **18**, 2746-2765.

Coltman DW, Slate J (2003) Microsatellite measures of inbreeding: a meta-analysis. *Evolution*, **57**, 971-983.

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

de Cara MAR, Fernández J, Toro MA, Villanueva B (2011) Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, **128**, 456-464.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, Ed 4. Longmans Green, Harlow, Essex, UK.

Fernández J, Toro MA, Caballero A (2004) Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, **18**, 1-10.

Flister MJ, Tsaih SW, O'Meara CC, Endres B, Hoffman MJ, Geurts AM, Dwinell MR, Lazar J, Jacob HJ, Moreno C (2013) Identifying multiple causative genes at a single GWAS locus. *Genome Research*, **23**, 1996-2002.

Foulley JL, Ollivier L (2006) Estimating allelic richness and its diversity. *Livestock Science*, **101**, 150-158.

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Grueber CE, Wallis GP, Jamieson IG (2008) Heterozygosity-fitness correlations and their relevance to studies on inbreeding in threatened populations. *Molecular Ecology*, **17**, 3978-3984.

Grundy B, Villanueva B, Woolliams JA (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetical Research*, **72**,

159-168.

Hill WG, Rasbash J (1986) Models of long term artificial selection in finite populations. *Genetical Research*, **48**, 41-50.

James JW (1970) The founder effect and response to artificial selection. *Genetical Research*, **16**, 241-250.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015-4026.

Kalinowski ST, Taper ML (2006) Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*, **7**, 991-995.

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutralmarkers in molecular ecology. *International Journal of Molecular Sciences*, **12**, 3966-3988.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science*, **220**, 671-680.

Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology*, **11**: 2445-2449.

López-Fanjul C, Hill WG (1973) Genetic differences between populations of *Drosophila melanogaster* for a quantitative trait: II. Wild and laboratory populations. *Genetical Research,***22**, 69-78.

Luikart G, Allendorf F, Cornuet JM, Sherwin W (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238-247.

Medugorac I, Veit-Kensch CE, Ramljak J, Brka M, Markovic B, Stojanovic S, Bytyqi H, Kochoski L, Kume K, Grünenfelder HP, Bennewitz J, Förster M (2011) Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and Evolution*, **1**, 408-420.

Meuwissen THE (2007) Operation of conservation schemes. In Oldenbroek, K. (Ed.), *Utilisation and Conservation of Farm Animal Genetic Resources*.Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 167-193.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA*, **70**, 3321-3323.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

Nelson RM, Pettersson ME, Carlborg Ö (2013) A century after Fisher: time for a new paradigm in quantitative genetics. *Trends in Genetics*, **29**, 669-676.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic

architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552-1553.

Pérez-Figueroa A, Saura M, Fernández J, Toro MA, Caballero A (2009) METAPOP—a software for the management and analysis of subdivided populations in conservation programs. *Conservation Genetics*, **10,** 1097-1099.

Pérez-Figueroa A, Rodríguez-Ramilo ST, Caballero A (2012) Analysis and management of subdivided populations with METAPOP. In *Data Production and Analysis in Population Genomics: Methods and Protocols* (Ed. François Pompanon and Aurélie Bonin). Humana Press. Chap. 15: 261-276.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844-855.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, **55**, 1095-1103.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity.*Conservation Biology*, **17**, 230-237.

Roff DA, Mousseau TA (1987) Quantitative genetics and fitness: lessons from Drosophila. *Heredity*, **58**, 103-118.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Santos J, Pascual M, Simões P, Fragata I, Lima M, Kelle B, Santos M, Marques A, Rose MR, Matos M (2012) From nature to the laboratory: the impact of founder effects on adaptation. *Journal of Evolutionary Biology*, **25**, 2607-2622.

Simianer H (2005) Using expected allele number as objective function to design between and within breed conservation of farm animal biodiversity. *Journal of Animal Breeding and Genetics*, **122**, 177-187.

Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, **132**, 861-862.

Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity,* **93**, 255-265.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498-2504.

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: a time for reappraisal. *Evolution*, **64**, 1202-1217.

Toro MA, Fernández J, Caballero A (2009) Molecular characterization of breeds and its use in conservation. *Livestock Science*, **120**, 174-195.

Toro M, Pérez-Enciso M (1990) Optimization of selection response. *Genetics Selection Evolution*, **22**, 93-107.

Turelli M (1984) Heritable genetic variation via mutation-selection balance: Lerch´s zeta meets the abdominal bristle. *Theoretical Population Biology*, **25**,138-193.

Wang J (2012) On the measures of genetic differentiation among populations. *Genetics Research*, **94**, 275-289.

Wilson A, Arcese P, Keller LF, Pruett CL, Winker K, Patten MA, Chan Y (2009) The contribution to island populations to in situ genetic conservation. *Conservation Genetics*, **10**, 419-430.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Weiser EL, Grueber CE, Jamieson IG (2012) AlleleRetain: a program to assess management options for conserving allelic diversity in small, isolated populations. *Molecular Ecology Resources*, **12**, 1161-1167.

Whitlock MC (2011) G_{ST}' and *D* do not replace F_{ST}. *Molecular Ecology*, **20**, 1083-1091.

Wright S (1952) The theoretical variance within and among subdivisions of a population that is in a steady state. *Genetics*, **37**, 312-21.

Data accessibility

Genotypic data for genetic markers is available in Dryad (doi:10.5061/dryad.15v00). All software and scripts necessary to run the simulations are available in Github (https://github.com/anpefi/Vilas_et_al.15).

Table 1. Simulated mean heterozygosities (*H*) and mean numbers of alleles (*A*) in the synthetic populations obtained maximizing *H* (O_H) or maximizing *A* (O_A) using QTL data, and mean response to selection for the quantitative trait after 25 generations (60 for case i). #QTL: number of loci controlling the quantitative trait. All cases refer to Scenario A, except cases b, c and d which refer to Scenarios B, C and D, respectively. Mutation rate per locus and generation $\mu = 5 \times 10^{-5}$ except for cases g (5×10^{-6}) and h (5×10^{-4}). Intensity of stabilising selection $V_S = 100$ except for cases e (1) and f (20). Proportion selected 20% except for case i (50%). Standard errors ranged between 0.0001 and 0.004 for *H*, between 0.0001 and 0.1 for *A*, and between 0.004 and 0.20 for the response. The largest mean value across methods for each scenario is presented in bold face for clarity.

Optimizations based on QTL		*H*		*A*		*Response*	
Case	#QTL	O_H	O_A	O_H	O_A	O_H	O_A
a	10	**0.18**	0.17	2.4	**3.1**	2.9	**4.1**
b	10	**0.26**	0.23	2.3	**3.1**	2.6	**3.4**
c	10	**0.62**	0.56	5.4	**5.7**	6.5	**6.9**
d	10	**0.61**	0.50	4.9	**5.2**	6.0	**6.2**
e	10	**0.17**	0.14	1.5	**1.6**	0.1	**0.2**
f	10	**0.19**	0.16	1.9	**2.4**	0.9	**1.2**
g	10	**0.09**	0.07	1.4	**1.6**	0.9	**1.1**
h	10	**0.72**	0.70	14.1	**16.4**	13.2	**14.4**
i	10	**0.18**	0.17	2.4	**3.1**	3.1	**4.2**
j	20	0.22	0.22	3.2	**3.8**	5.3	**6.2**
k	50	0.28	**0.29**	4.3	**4.6**	8.5	**8.6**
l	100	0.31	0.31	4.7	**4.9**	8.4	8.4
m	200	0.32	0.32	4.9	**5.0**	6.1	6.1
n	500	0.32	0.32	4.9	**5.0**	3.1	3.1
o	1000	0.32	0.32	5.0	5.0	1.7	1.7

Table 2. Simulated mean heterozygosities (*H*) and mean numbers of alleles (*A*) in the synthetic populations obtained maximizing *H* (O_H) or maximizing A (O_A) using neutral marker data, and mean response to selection for the quantitative trait after 25 generations (60 for case i). #NM: number of neutral markers. All cases refer to Scenario A, except cases b, c and d which refer to Scenarios B, C and D, respectively. Mutation rate per locus and generation $\mu = 5 \times 10^{-5}$ except for cases g (5×10^{-6}) and h (5×10^{-4}). Intensity of stabilising selection $V_S = 100$ except for cases e (1) and f (20). Proportion selected 20% except for case i (50%). Standard errors ranged between 0.0001 and 0.004 for *H*, between 0.0001 and 0.1 for *A*, and between 0.004 and 0.20 for the response. Cases a-i assumed 10 QTL whereas cases j-n show averages over simulations with all different numbers of QTL (10 to 1000). The largest mean value across methods for each scenario is presented in bold face for clarity.

Optimizations based on markers		*H*		*A*		*Response*	
Case	#NM	O_H	O_A	O_H	O_A	O_H	O_A
a	10	**0.30**	0.29	3.7	**4.3**	2.9	**3.9**
b	10	**0.45**	0.42	3.9	**4.5**	2.4	**2.8**
c	10	**0.71**	0.68	7.6	**7.9**	6.2	**6.5**
d	10	**0.70**	0.64	6.8	**7.1**	5.8	**5.9**
e	10	**0.40**	0.37	3.4	**3.9**	0.12	**0.13**
f	10	**0.44**	0.41	3.7	**4.3**	0.8	**1.0**
g	10	**0.25**	0.22	1.8	**2.0**	0.6	**0.8**
h	10	**0.86**	0.85	20.7	**21.7**	14.4	**14.7**
i	10	**0.30**	0.29	3.7	**4.3**	3.0	**4.5**
j	20	0.29	**0.30**	3.9	**4.3**	5.0	**5.3**
k	50	0.29	**0.29**	4.1	**4.3**	5.2	**5.4**
l	100	0.26	0.26	4.1	**4.2**	5.4	**5.5**
m	200	0.29	0.29	4.1	**4.2**	5.4	5.4
n	500	0.29	0.29	4.2	4.2	5.4	5.4
o	1000	0.29	0.29	4.2	4.2	5.3	5.3

Table 3. Experimental average expected heterozygosity (*H*), number of alleles (*A*) and contributions from the subpopulations to the synthetic populations in the *Drosophila* experiment, averaged over 9 microsatellite loci and four replicates, under two optimization criteria: maximization of expected heterozygosity (O_H) and maximization of number of alleles (O_A). Apopulation analogous to the synthetic oneswas established taking all individuals from the large base population (BP). The two last rows show the average *H* and *A* in the synthetic and BP populations created, where the largest values of *H* and *A* are marked in bold face for clarity. Standard errors ranged between 0.005 and 0.013 for *H* and between 0.028 and 0.116 for *A*.

			Contributions to syntheti		
Subpop.	*H*	*A*	O_H	O_A	BP
L1	0.30	2.00	-	-	-
L2	0.47	2.67	27	-	-
L3	0.43	2.67	1	-	-
L4	0.44	2.33	-	-	-
L5	0.41	2.78	9	-	-
L6	0.42	3.00	-	8	-
L7	0.50	3.11	-	21	-
L8	0.52	3.22	43	2	-
L9-10	0.36	2.94	13	20	-
L11-12	0.37	2.78	1	4	-
L13-14	0.46	2.44	-	-	-
L15-16	0.40	2.72	6	21	-
L17-20	0.37	2.67	-	24	-
Base Pop.	0.53	3.89	-	-	100
Synthetic Populations					
H			**0.58**	0.53	0.56
A			3.58	**3.67**	3.61

FIGURE LEGENDS

Figure 1. Schematic representation of the structure of the base populations simulated for the different scenarios considered.

Figure 2. Example of the establishment of a synthetic population in the simulations. In this case only five subpopulations contributed to create a synthetic population of 100 individuals. The lower part of the figure illustrates the artificial selection process which was simulated by mating the 20 or 50 individuals with the highest phenotypic value for a quantitative trait in each generation.

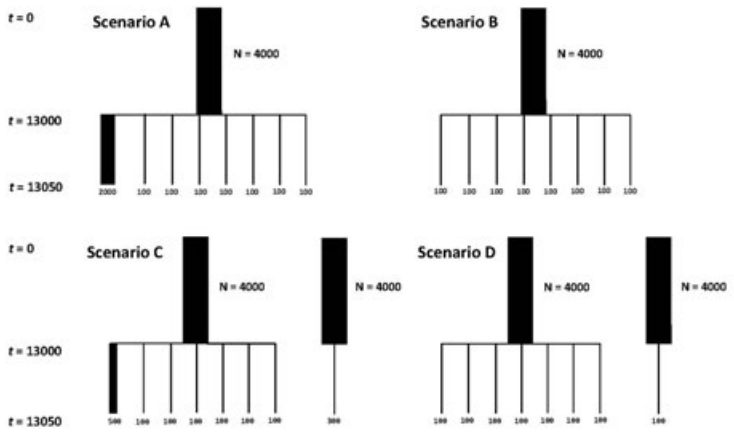
Figure 3. Experimental design with *Drosophila*. Twenty experimental lines of size N = 100 were established from a large base population and maintained for 58 generations to generate a structured population. Synthetic populations of size N = 100 were obtained with contributions from the structured population to maximize the expected heterozygosity or the number of alleles. The synthetic populations were artificially selected for increased and decreased sternopleural bristle number for 8 generations.

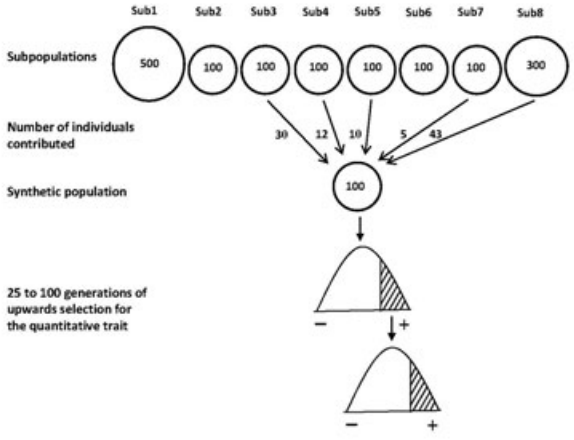
Figure 4. Average simulated contributions from subpopulation 1 and each of the other seven subpopulations to the synthetic population for Scenario A, assuming different numbers of loci (QTL or markers) in the simulations. In the case of the optimizations based on markers results are averages over simulations assuming all numbers of QTL (10 to 1000). Optimizations are obtained maximizing H (O_H) or maximizing A (O_A).

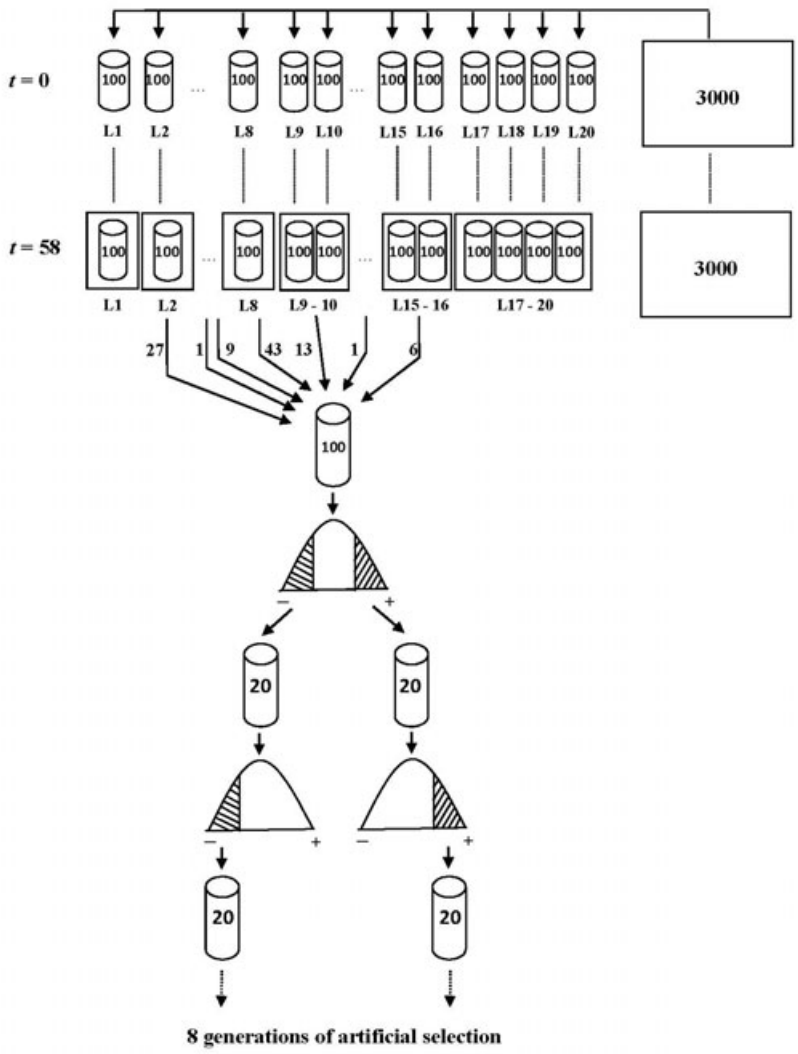
Figure 5. Average responses to selection at different generations from synthetic populations obtained through optimization based on different numbers of markers (lines) or optimization based on QTL (dots). Results refer to Scenario A with 1000 QTL. Optimizations are obtained maximizing H (O_H) or maximizing A (O_A).

Figure 6. Mean cumulative response to selection for upward selection (upper graph) and downward selection (lower graph) across 8 generations of artificial selection for sternopleural bristle number in the synthetic populations. Results are averaged over four replicates and error bars represent one standard error at each side of the mean. Synthetic populations were obtained maximizingthetotal expected heterozygosity (O_H) or the total number of alleles (O_A). Method BP refers to populations obtained taking individuals from the base population.

Figure 7. Distribution of the number of microsatellite loci used in the 89 studies published in the Journal Conservation Genetics in 2014. The average number of loci is 11.96 and the median 11.0.







8 generations of artificial selection