Pedigree-Free Estimates of Heritability in the Wild: Promising Prospects for Selfing Populations

Laurene Gay[1]*, Mathieu Siol[1,2], Joelle Ronfort[1]

1 Diversity and Adaptation of Mediterranean Species, UMR AGAP 1334, Montpellier, France, **2** Genetics and Ecophysiology of Legume Species, UMR Agroecology 1347, Dijon, France

Abstract

Estimating the genetic variance available for traits informs us about a population's ability to evolve in response to novel selective challenges. In selfing species, theory predicts a loss of genetic diversity that could lead to an evolutionary dead-end, but empirical support remains scarce. Genetic variability in a trait is estimated by correlating the phenotypic resemblance with the proportion of the genome that two relatives share identical by descent ('realized relatedness'). The latter is traditionally predicted from pedigrees (Φ_A: expected value) but can also be estimated using molecular markers (average number of alleles shared). Nevertheless, evolutionary biologists, unlike animal breeders, remain cautious about using marker-based relatedness coefficients to study complex phenotypic traits in populations. In this paper, we review published results comparing five different pedigree-free methods and use simulations to test individual-based models (hereafter called animal models) using marker-based relatedness coefficients, with a special focus on the influence of mating systems. Our literature review confirms that Ritland's regression method is unreliable, but suggests that animal models with marker-based estimates of relatedness and genomic selection are promising and that more testing is required. Our simulations show that using molecular markers instead of pedigrees in animal models seriously worsens the estimation of heritability in outcrossing populations, unless a very large number of loci is available. In selfing populations the results are less biased. More generally, populations with high identity disequilibrium (consanguineous or bottlenecked populations) could be propitious for using marker-based animal models, but are also more likely to deviate from the standard assumptions of quantitative genetics models (non-additive variance).

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Laurene.Gay@supagro.inra.fr

Introduction

The genetic variance available for a trait in a population informs us about its potential ability to evolve in response to novel selective challenges [1]. This corresponds to one of Brookfield's definitions of the evolvability of a population: "a description of its current standing crop of genetic variability, and the consequence of the extent and nature of this variation for the population's ability to respond to current selective pressures" [2]. Additive genetic variance can be estimated from the resemblance between relatives by relating the phenotypic covariance of a quantitative trait with the proportion of the genome for which two relatives share genes identical by descent [1,3]. To achieve this, one would ideally like to know the actual proportion of loci controlling the trait that are identical by descent. This 'realised relatedness' is the outcome of a stochastic process (due to Mendelian segregation and linkage) with a variance that depends on genome size [4–6]. However, because causal loci are unknown, we traditionally use the expected value of identity by descent given the ancestry [7,8]. It can be deduced from a pedigree (hereafter called Φ_A), either in an experiment using specific relatedness classes (e.g. full sib-half sib design) or in a population with pedigree data ranging over several generations [9]. However, in wild populations, pedigree information is generally not available except for a few long term studies

[10–12]. An alternative solution is to estimate the genome-wide average of the realised relatedness between individuals using molecular markers [13].

With the ongoing rise of next generation sequencing, high density SNP panels become available and the realised proportion of the genome that two individuals share identical by descent can be estimated with increasing accuracy [14]. Several estimators of kinship (sometimes called coancestry) and relatedness (or relationship) coefficients have been proposed [14–21] and compared [16,22–24]. From these reviews, it appears that the relative performance of each method depends on the set of loci used, on allele frequency distributions, in particular minor allele frequency spectrum [25] and on the average relatedness between individuals in the population. Provided we can estimate it precisely, using the genome-wide average of the realised relatedness rather than resorting to its expected value (Φ_A) could improve the estimation of evolutionarily relevant parameters for quantitative traits (genetic variance, genetic correlations or selection gradients) [7].

Different methods are available to estimate quantitative genetic parameters from molecular marker data (hereafter called pedigree-free methods) [26]. Their respective reliability and suitability depending on the marker used and/or the population genetic structure remain unclear. While such methods based on genome-wide molecular information have received substantial attention by

animal breeders and human biologists [27,28], evolutionary biologists remain very cautious about how useful marker-based relatedness coefficients could be for studying complex phenotypic traits in populations lacking pedigree information [29].

Being able to accurately estimate the genetic variance of a trait is particularly important for inbred or selfing species that have been described as evolutionary dead ends due to the potential loss of genetic diversity [30,31]. Self-fertilization is common in angiosperms [32,33] and also occurs in hermaphrodite animals at a lower frequency [34]. Reduced genetic variation in highly selfing populations is frequently observed using molecular markers [35–37]. It can be explained by reduced effective population sizes accompanying increased homozygosity [38], enhanced genetic hitchhiking with selective sweeps and background selection caused by a reduced effective recombination rate [39,40] and frequent bottlenecks following the recurrent extinction - recolonisation events [41]. However, the effect of selfing on the genetic variation relevant for most adaptive change, i.e. quantitative genetic variation, is less clear. For a population with constant allelic frequencies, inbreeding is expected to increase the genetic variance of a trait to the point that when inbreeding is complete, the genetic variance in the population as a whole is doubled and appears as the between-lines component [1]. Simultaneously, inbreeding should reduce quantitative variation due to the fixation of alleles and to a reduced efficiency of selection in maintaining variants (see [42] and [43] for a review of theoretical arguments). Simulations have shown that if non-additive effects are important (dominance or epistasis), additive genetic variance increases with inbreeding, reaching a maximum for intermediate inbreeding coefficients (F) and then a declining towards zero at $F = 1$ [44,45]. In agreement with these theoretical expectations, some evidence for reduced within population genetic variance in highly inbreeding populations compared to outcrossing populations has been found [43,46]. More data are needed to compare the level and components of quantitative genetic variation between selfers and outcrossers and to further understand the consequences of self-fertilization on the adaptive potential of natural populations.

Besides direct effects on genetic variability, selfing generates a correlation in heterozygosity and/or homozygosity across loci, called identity disequilibrium [47]. This will broadly influence both the relatedness between individuals and the variance in relatedness in the population [48,49] and could improve marker-based estimates of the expected proportion of identity by descent [48]. Thereby, selfing may provide favourable conditions for pedigree-free quantitative genetics.

In this paper we review published results comparing several pedigree-free methods used to estimate quantitative genetics parameters for complex phenotypic traits in wild populations. We then report results from simulations aimed at further comparing the performance of individual-based models (hereafter called animal models) using pairwise relatedness predicted from the pedigree versus molecular markers, with a special focus on how mating systems affects the efficiency of these methods.

Materials and Methods

Review

As a preliminary step, we searched the literature (using keywords in Web of Science) looking for studies comparing methods for estimating quantitative genetic parameters using molecular markers, for example a pedigree-based animal model and another method, or a simulation study. Our review is based on the classification published by Garant & Kruuk [26], who identified three categories of methods that rely on molecular markers. (1) The 'Ritland' method estimates heritability as the covariance between pairwise phenotypic similarity and pairwise relatedness [50]. Alternatively (2), using a maximum likelihood approach, individuals can be classified into known classes of relatedness (for example sibs vs. unrelated) and analyzed in a mixture model [51]. Sibling groups can also be identified within one generation and analyzed in a classic quantitative genetics framework (analysis of variance) or using a more complex model (animal model, [52]). Finally (3), parentage assignment methods can help reconstruct a complete pedigree spanning several generations [13,53] and quantitative genetics parameters are then derived from an animal model [54]. We expanded Garant & Kruuk's classification to include two additional methods that are currently available: (4) an animal model method [3,12,55] directly using the full pairwise relatedness matrix estimated using molecular markers [56,57] and (5) a multilocus association method derived from genomic selection [27]. This last method was originally aimed at predicting individual breeding values using molecular markers, in order to accelerate and improve the response to artificial selection. Technically, it relies on multiple regressions with shrinkage, where the phenotype is explained by a set of markers (e.g. [58]).

In order to compare additive genetic variances among traits or taxa, it is common practice to scale it with the total phenotypic variance (heritability) or with the trait mean (coefficient of additive genetic variation). As proposed by Houle [59] and recently confirmed by Hansen et al. [60] the additive genetic coefficient of variation is a better predictor of a population's ability to respond to selection and can be viewed as an accurate measure of evolvability. Nevertheless, heritability remains the most commonly reported measure of evolutionary potential (in particular in all studies using the Ritland method). In our review, we therefore compared heritabilities rather than coefficients of additive genetic variation. This can be misleading in the presence of a positive correlation between the additive variance and other components of phenotypic variance [60]. We also used heritabilities in our simulations in order to remain consistent. We argue that it is not problematic in our simulations as there is no inherent correlation between variance components.

The performance of a statistical inference can be evaluated by the bias, defined as $E(\hat{h}^2 - h^2)$, where $\hat{h}^2$ is the estimator and h^2 the parameter and the sampling error, $E(\hat{h}^2 - E(\hat{h}^2))^2$. When reviewing the literature for pedigree- free methods, we would ideally like to compare the bias and sampling error of heritability estimates obtained using pedigrees ($\hat{h}^2_{ped}$) or one of the marker-based methods ($\hat{h}^2_{marker}$). However, the review of empirical results only provides us with single estimates $\hat{h}^2_{ped}$ and $\hat{h}^2_{marker}$ to compare and we cannot say anything about the extent to which either or both are biased and which has smaller sampling error. We therefore only tested whether $\hat{h}^2_{ped}$ and $\hat{h}^2_{marker}$ significantly differed for each of the marker-based methods using linear models and also compared their standard errors. Most studies considered several traits that would be wrongly considered as independent. We took this into account by adding a random effect for study. For simulation studies, we report the bias and standard error of the estimators $\hat{h}^2_{ped}$ or $\hat{h}^2_{marker}$. Again, we tested for an influence of the method used (pedigree or marker-based methods) on the bias using linear mixed models with study as a random effect followed by posthoc tests (with Bonferroni correction) to identify significant pairwise differences. Studies using genomic selection methods generally evaluate the accuracy of the model as the correlation between the estimated breeding value and the real breeding value. We report these accuracies.

We tested whether the absolute value of the bias increases with heritability (measured by pedigree methods or simulated) and assessed the relationship between the bias and the number of markers available (or simulated) using a mixed linear model for each method, with study as a random effect.

Simulations

We simulate a population of constant size $N = 500$ individuals evolving for a number of generations. Individuals are diploid and unrelated at generation 0 (but they can share alleles identical by state). Generations are non-overlapping. We simulate the genotype of each individual at $L_M + L_{QTL}$ unlinked loci. The L_{QTL} loci are causal loci whereas the L_M loci are biallelic non-coding marker loci. Initial allele frequencies are drawn from the distribution expected at mutation – drift equilibrium using a Dirichlet distribution [61]. The phenotype of each individual is controlled by 500 loci (L_{QTL}) with five alleles each. Allelic effects are randomly drawn from a normal distribution for each allele. The phenotype of an individual is the sum of the two allelic effects at each of the L_{QTL} loci ($\alpha_{i,j}$ is the allelic effect of allele i at locus j) plus a random factor (ε_{ij}) drawn from a normal distribution representing the environmental effect:

$$P_{ij} = \sum \left(\alpha_{i,l} + \alpha_{j,l} \right) + \varepsilon_{ij} \qquad (1)$$

The variances of the distribution of allelic and environmental effects are adjusted to simulate heritabilities of 0.15, 0.3 and 0.6. The next generation is built either by random mating of pairs of individuals or with a set proportion of selfing ($S = 0.9$) to simulate inbred populations. The simulation program is written in C++ and runs in batch using a custom python script, with 20 replicates per parameter set. The code is available as supplementary material (Zipfile S1).

Estimation of Genetic Variance Using a Marker-based Animal Model –influence of Mating Regime and Manipulations on Relatedness Matrices

We simulated populations evolving for 10 generations and performed analyses at generation 10, using pedigree and marker-based animal models (method 4 in the review section). Every generation, the pedigree information (mother and father of each individual) was recorded and used to calculate the relatedness coefficient (Φ_A) between pairs of individuals. Genotypes at the L_M marker loci were used to estimate pairwise relatedness using the coefficient introduced by Loiselle et al. [20] (thereafter named $K_{i,j}$). It does not assume Hardy-Weinberg equilibrium and performs well, even in the presence of rare alleles [24,62]. To estimate heritability, we fitted a very simple linear model to the simulated phenotype: $y = Za + e$ where y is the phenotype, Z is a design matrix and a the vector of additive genetic effects; e is the vector of residual effect. The pedigree or the marker loci information were then used to specify a variance–covariance structure for the vector of additive genetic effects a, shaped as $2.\Phi_A.\sigma_A^2$ when using an animal model including the pedigree or as $2.K_{i,j}.\sigma_A^2$ when replacing the Φ_A matrix by a marker-based relatedness matrix. For selfing populations, we used $2.\Phi_A.\sigma_A^2/(1+F)$ and $2.K_{i,j}.\sigma_A^2/(1+F)$ to account for inbreeding [62], where F was the average inbreeding coefficient. F was either estimated using pedigrees or approximated as $S/(2-S)$ [63] for the marker-based method, where the pedigree is supposed unknown. We preliminarily ascertained that both values were highly similar ($F = 0.82$ for $S = 0.9$). We used restricted maximum likelihood to estimate the additive genetic variance and the standard errors with the program ASReml v3.0 [64]. We examined the effect of the number of marker loci by
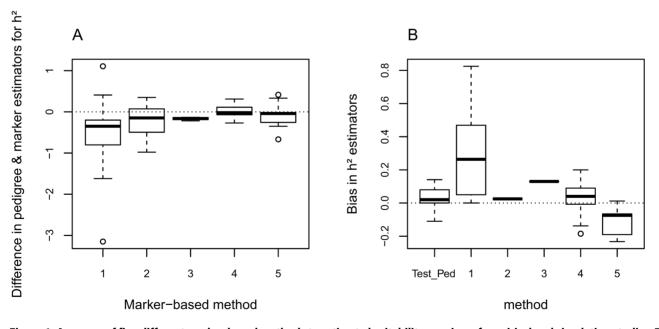

Figure 1. Accuracy of five different marker-based methods to estimate heritability – review of empirical and simulation studies. The efficiency was assessed in a review of 24 empirical studies (A) or 15 simulation studies (B), comparing heritability estimates using pedigree or one of the following methods: 1 - Ritland; 2 - relatedness classes; 3 - reconstructed pedigrees; 4 - marker-based animal model or 5 - genomic selection. Details of the number of studies for each method are given in Table 1. The bias was measured as $\hat{h}^2_{marker} - \hat{h}^2_{pedigree}$ in A and as $E(\hat{h}^2 - h^2)$ in B, where h^2 is the simulated parameter. The horizontal line shows the median bias for each method. The bottom and top of the box show the 25th and 75th percentiles. The vertical dashed lines show the maximum and minimum biases and the circles are outliers.
doi:10.1371/journal.pone.0066983.g001

Table 1. Summary of studies comparing estimates of quantitative genetics parameters using pedigree-free methods.

Species	Method[1]	Marker[2]	Individuals	Traits	Kinship coef[3]	Bias[4]	Note	Variance in relatedness	Mating regime[5]	Ref
Bighorn sheep *Ovis canadensis*	1	32 SSR	110 to 150	15	LR - QG - W	−0.20	–	0.0093	Out	[99]
Yellow monkey-flower *Mimulus gattatus*	1	8 allozymes	665	7	R	−0.67	–	0.009 or 0.027 to 0.044 [100]	Mix	[101]
Turkey oak *Quercus laevis*	1	5 SSR	200	1	R	–	no variance of relatedness	0.0007	C+Out	[102]
Soay sheep *Ovis aries*	1	11 SSR	529	6	LR	−0.56	–	0.0005	Out	[79]
Capricorn silvereye *Zosterops lateralis chlorocephalus*	1	11 SSR	479	6	QG	−0.24	–	−0.001 to 0.008	Out	[65]
Shea tree *Vitellaria paradoxa*	1	12 SSR	200	5	LR - W	−0.70	–	0.004	Out	[103]
Yellow box *Eucalyptus melliodora*	1	6 SSR	259	1	LR	−0.21	bias relative to Doran and Matheson [104]	0.003	Mix	[105]
Algarrobo Blanco *Prosopis alba*	1	6 SSR	142	13	R	−0.84	–		Out	[106]
Monterey Pine *Pinus radiata*	1	8 SSR	355	1	R	0.41	–	0.009	Out	[107]
Japanese flounder *Paralichthys olivaceus*	1	7 SSR	134	20	LR - QG - R - W	0.29	bias relative to method3	0.0001	Out	[108]
Rainbow trout *Oncorhynchus mykiss*	1	16 SSR	628	3	R	1.70	bias relative to method3	0.002	Out	[109]
Yellow monkey-flower *Mimulus gattatus*	2	8 allozymes	665	7	–	−0.60	–		Mix	[101]
Salmon *Oncorhynchus tshawytscha*	2	1SLP[6]+1 SSR	170	2	–	0.27	bias relative to Heath et al. [110]		Out	[51]
Soay sheep *Ovis aries*	2	12	529	6	–	0.02	–		Out	[79]
Rainbow trout *Oncorhynchus mykiss*	2	16	628	3	–	−0.07	bias relative to method3		Out	[109]
Soay sheep *Ovis aries*	3	12	529	6	CERVUS	−0.17	–		Out	[79]
Rainbow trout *Oncorhynchus mykiss*	3	16	628	3	MCMC	–	no comparison		Out	[109]
Japanese flounder *Paralichthys olivaceus*	3	7 SSR	50 to 134	20	COLONY	–	no comparison		Out	[108]
Common sole *Solea solea*	3	10 SSR	2000	1	PAPA	–	no comparison	0.04	Out	[111]
Cattle - Angus	4	9323 SNP	379	1	SA	0.19	–		Out	[112]
Capricorn silvereye *Zosterops lateralis chlorocephalus*	4	11 SSR	479	6	QG	0.02	–		Out	[65]
Malaria parasite *Plasmodium falciparum*	4	335 SSR	185	8	SA	0.03	–		Mix	[113]
Common sole *Solea solea*	4	10 SSR	2000	1	SA	−0.12	bias relative to method3		Out	[111]
Mouse	4	10000 SNP	2200	4	IBS(i,j)	0.06	–		Out	[67]
Holstein cows	4	54001 SNP	517	3	GRM	−0.10	–		Out	[114]
Angus steers	4	41028 SNP	698	3	GRM	< −0.01	–		Out	[115]
Wheat	5	1279 SNP	599 lines	3	–	0.36	–		Self	[58]
Human	5	294831 SNP	3925	4	–	−0.30	–		Out	[25]
Mouse	5	10 656 SNP	1885	1	–	0.55	Correlation with simulated h^2		Out	[116]
Holstein cows	5	54001 SNP	1200	1	–	0.62	Correlation with simulated h^2		Out	[117]
Mouse	5	10946 SNP	1884	4	–	−0.31	–		Out	[73]

Table 1. Cont.

Species	Method[1]	Marker[2]	Individuals	Traits	Kinship coef[3]	Bias[4]	Note	Variance in relatedness	Mating regime[5]	Ref
Mouse	5	10000 SNP	2200	4	–	0.10	–		Out	[67]
Dairy cattle	5	47152 SNP	5217	6	–	–0.04	–		Out	[68]

[1]Method: 1 = Ritland's method; 2 = maximum likelihood on relatedness classes; 3 = sibship or pedigree reconstruction; 4 = animal model with marker-based relatedness matrix; 5 = genomic selection.

[2]SSR = microsatellite; SNP = single nucleotide substitution.

[3]Method used to estimate pairwise kinship coefficients using molecular markers: LR = Lynch & Ritland [16]; QG = Queller & Goodnight [15]; R = Ritland [21]; SA = % of shared alleles [118]; GRM = genomic relationship matrix [119]; IBS(ij) = probability of identity by state [120]. For method 3, this column indicates the method of pedigree reconstruction: MCMC (maximizes the pairwise likelihood ratios of being full siblings or unrelated [121]); CERVUS [122]; COLONY [123]; PAPA [124].

[4]The bias is defined as the difference between heritability estimated from the pedigree (Φ_A) and heritability estimated using molecular markers in one of the five methods presented $\hat{h}^2_{marker} - \hat{h}^2_{ped}$. When no information was available on $\hat{h}^2_{ped}$ estimated using pedigrees, the bias was measured relative to $\hat{h}^2$ estimated using method 3 (pedigree reconstruction), as indicated in the column "Note". When several traits were analyzed (column "traits"), we present the average bias.

[5]Mating regime: "Out" stands for outcrossing, "Mix" for mixed mating, "Self" for predominantly selfing and "C" for clonality.

[6]Single Locus Probe.

doi:10.1371/journal.pone.0066983.t001

letting L_M vary between 384, 1500, 3000 or 5000 SNPs. We also manipulated the marker-based relatedness matrix to summarize the information and examined how this influenced the bias in genetic variance estimates. It has become common practice in association studies (GWAS) to truncate the marker-based relatedness matrix by replacing any negative value by zero. We transformed our matrix accordingly.

As in the review section for simulation studies, we measured the bias in heritability as $E(\hat{h}^2 - h^2)$. We estimated the precision of the estimation using sampling errors, defined as $E(\hat{h}^2 - E(\hat{h}^2))^2$. We compared the bias and sampling errors in heritability estimators (pedigree versus molecular markers) using Wilcoxon signed rank tests. We tested for the effect of the mating regime, trait heritability and number of markers on the bias in heritability estimates using linear models.

The identity disequilibrium created by consanguineous matings could improve marker-based estimates of the expected proportion of identity by descent [48]. We verified this in our simulations by estimating the relatedness at the causal variants (500 QTLs) and examined how accurately it is predicted by the relatedness at the L_M marker loci, in selfing or outcrossing populations, using linear regressions. A slope close to one would indicate that the set of observed SNPs accurately predicts the relatedness at causal loci. Any deviation to one could be caused either by sampling error (due to the limited number of observed SNPs) or by rare alleles in the causal variants. All analyses were run in R version 2.15.1.

Results

Review of Published Results

We collected 39 papers comparing heritability estimates based on molecular data or pedigree information. Among those, 24 reported empirical results (Table 1) while 15 were based on simulations only (Table 2). Only seven studies reported heritability estimates for plant species, 16 for animals and one for a protist. Using linear mixed models, we found a significant difference between pedigree-based estimates and marker-based estimates for the Ritland method ($\chi^2 = 22.2$; $p = 2.5 \times 10^{-06}$), the relatedness classes ($\chi^2 = 5.9$; $p = 0.015$) and the reconstructed pedigrees ($\chi^2 = 29.8$; $p = 4.7 \times 10^{-08}$) but not for the animal model ($\chi^2 = 0.04$; $p = 0.836$) or genomic selection ($\chi^2 = 0.001$; $p = 0.977$). As shown by Figure 1.A, the difference between the heritability estimated using pedigree or molecular markers was lowest with method 4 (marker-based animal model, see also Figure S1). When a dataset was analyzed using Ritland's method and an animal model in parallel, the latter gave results closest to the estimates obtained using the pedigree [65]. In addition, pedigree-free methods seem to improve the precision of the estimation, except for the Ritland method and the genomic selection method (with Ritland: standard error increased by +0.23; p = 0.062 and +0.13 with genomic selection; p = 0.001).

Using the 15 simulation studies, we calculated the bias as $E(\hat{h}^2 - h^2)$ (Table 2). The mixed model highlighted a significant effect of the method on the bias (p = 0.005). Posthoc tests showed a significantly higher bias for the Ritland method (average 0.30; $n = 10$) than for the pedigree method (0.04; $n = 13$; p < 0.001), the marker-based animal models (0.03; $n = 15$; p = 0.049) or the genomic selection (−0.11; $n = 12$; p = 0.015). There was no significant difference between these three methods (p > 0.750). The number of simulation results testing methods 2 and 3 in the dataset we collected was insufficient to compare their biases. The average biases for each method are shown in Figure 1.B. We found no significant difference in standard error between these methods (p = 0.129).

Table 2. Summary of simulation studies comparing estimates of quantitative genetics parameters using pedigree-free methods.

Method[1]	h²	Markers	Individuals	Kinship coef[2]	Bias	Ref
1	0.25	32 SNP	2000 (sib families)	R	0.01	[125]
1	0 to 1	2 to 30	100 to 1000 (sib families)	R	0.05	[126]
1	0.1	10 to 100 - 6 alleles	500	R	0.32	[127]
1	0.5	10 to 100 - 6 alleles	500	R	0.43	[127]
2	0 to 1	2 to 30; 5 to 20 alleles	100 to 1000	–	0.03	[126]
3	0.25	11 SSR	1955	CERVUS - COLONY	0 to 013	[128]
4	0.4 or 0.6	400 SNP	240 inbred lines	L	0.08	[129]
4	0.5	2000 to 5000 SNP	1000	UAR	−0.10	[14]
4	0.33	9000 SNP	1000	SA	0.12	[112]
4	0.33	1000 SNP	1000	SA	0.01	[112]
4	0.1 or 0.3	5 to 100 QTLs	100 (sib families)	SA (on QTLs)	0.86 (corr)	[130]
4	0.2	20 to 1200 SNP	20	SA	0.63 (corr)	[131]
5	0.8	294831 SNP	3925	–	0.01	[25]
5	0.5	1000 SNP	2200	–	0.85 (corr)	[132]
5	0.7/0.3/0.1	6000 SNP	5865	–	0.89 (corr)	[116]
5	0.2/0.5/0.9	2000 SNP	1000	–	0.58 (corr)	[117]
5	0.1	396 SNP	1000	–	−0.07	[133]
5	0.5	396 SNP	1000	–	−0.20	[133]
5	0.5	5000 SNP	1000 (family structure)	–	0.79 (corr)	[134]
ped	0.1		500	pedigree	0.05	[127]
ped	0.5		500	pedigree	0.14	[127]
ped	0.4 or 0.6		240 inbred lines	pedigree	0.03	[129]
ped	0.33		1000	pedigree	−0.11	[112]
ped	0.1 or 0.3		100 (sib families)	pedigree	0.52 (corr)	[130]
ped	0.2		20	pedigree	0.52 (corr)	[131]

[1]The column "Method" is the same as described for Table 1, with the additional category "Ped" that stands for pedigree-based animal models. Numbers in italics indicate that we report the correlation with the simulated heritability (corr) rather than the bias.
[2]Method used to estimate pairwise kinship coefficients using molecular markers: L = Loiselle [20]; UAR = raw unified additive relationship (estimator of the genome-wide relationship between individuals) [14]; SA = % of shared alleles [135]; pedigree = Φ_A.
doi:10.1371/journal.pone.0066983.t002

The average value of the bias in heritability tended to increase for traits with low heritabilities (Figure S2.A and S2.B, p<0.0001) and this effect was significant for all methods except the marker-based animal model (p = 0.743). We detected a significant but small negative effect of the number of microsatellites or SNP used on the average value of the bias with the animal model method (p = 0.012 and p = 0.048) (Figure S2.C and S2.D).

Simulation Results - Heritability Estimation Using a Marker-based Animal Model

Simulation results confirmed that using pedigree information in an animal model provides accurate estimates of heritability for both outcrossing and selfing populations, even with low heritabilities (Figure 2.A). We analysed the bias in heritability estimates using Wilcoxon signed rank tests and found that replacing the pedigree-based relatedness matrix by a marker-based relatedness matrix strongly worsens the estimation of heritability for outcrossing populations (average bias = −0.014 with pedigree and 0.174 with markers; $V_{Wilcoxon} = 9$; p = 4.1×10^{-10}; Figure 2.B) but not for selfing populations (average bias = −0.016 with pedigree and −0.015 with markers; $V_{Wilcoxon} = 738$; p = 0.270). The bias in outcrossing populations was reduced when using truncated marker-based relatedness coefficients (0.072; Figure 2.C). Results

were the same when testing the effect of the method on the bias using linear models. In addition, if the marker-based method seemed more biased than the pedigree method, the later had higher sampling error (0.15 for the pedigree method and 0.08 for marker-based method) and the difference was significant in outcrossing populations (0.16 versus 0.01; $V_{Wilcoxon} = 1422$; p = 4.1×10^{-10}) but not in selfing populations (0.14 for both methods; $V_{Wilcoxon} = 1028$; p = 0.408).

In outcrossing populations, the absolute value of the bias with the marker-based method increased with higher simulated heritabilities ($F_{53,1} = 155.0$; p<10^{-16}), but it should be pointed out that with low heritabilities (0.15) the model sometimes failed to converge or estimated an additive variance not significantly different from zero.

Selfing decreased the bias in marker-based heritability estimates. With a heritability of 0.3, for example, the bias was more than seven times larger in outcrossing compared to selfing populations ($F_{36,1} = 57.2$; p = 6.10^{-9}). The distribution of pairwise relatedness in selfing populations may explain such a better performance in estimating heritability [48,49]. Indeed, as shown on Figure 3, pairwise relatedness coefficients have a higher mean and a larger variance in selfing compared to outcrossing populations. This effect extends beyond the simple influence of population size (Figure S3). The higher performance of the marker
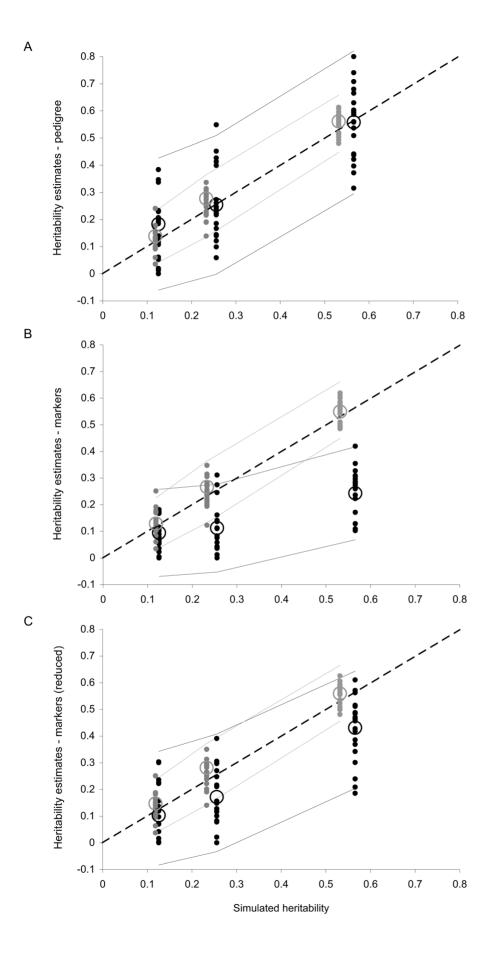
A

B

C

Simulated heritability

Figure 2. Simulation results testing the accuracy of pedigree or marker-based methods to estimate heritability. This figure shows the correlation between the heritability simulated and heritability estimates obtained using pedigree-based animal models (A), marker-based animal models (B) or marker-based relatedness coefficients truncated before the analysis (C). Each dot stands for a simulated population, with 90% selfing (in grey) or complete outcrossing (in black). Circles stand for means across 20 replicates and solid lines show the 95% confidence intervals, as estimated by Asreml (and averaged across replicates). The dashed lines represent $y = x$.
doi:10.1371/journal.pone.0066983.g002

based method under selfing can also be related to the fact that the relatedness at causal loci is more closely correlated with the relatedness at a set of marker loci in selfing than in outcrossing populations (Figure 4). This is in agreement with results by Szulkin *et al.* [48] on inbreeding coefficients and is caused by high identity disequilibrium in selfing populations.

Figure 5 highlights that when a larger number of loci is used to estimate the matrix of pairwise relatedness, heritability estimates become more accurate, most probably because the prediction of Φ_A is improved. The linear model showed that the absolute value of the bias decreases significantly with the number of markers ($F_{78,1} = 11.9$; p = 0.001). If marker-based relatedness coefficients are truncated, the effect becomes non-significant ($F_{78,1} = 0.7$; p = 0.389) and a lower number of markers is required to get an accurate estimate of heritability. In selfing populations, there is no relationship between the bias and the number of marker used ($F_{78,1} = 0.8$; p = 0.371). 384 markers seem already sufficient to estimate heritability reasonably well. Surprisingly, the sampling error of the estimates did not decrease with the number of loci.

Discussion

Pedigree-free Methods to Estimate Quantitative Genetics Parameters in the Wild: What have we Learnt in Nearly 20 Years?

Reviewing empirical and simulation studies of quantitative genetics in wild populations using marker-based estimates of relatedness confirms that it is extremely difficult to derive reliable estimates for quantitative genetic parameters in wild populations using Ritland's pairwise regression model, as suggested by several authors [26,65,66]. Nevertheless, being a pioneer, the Ritland method played a significant role in stimulating the development of further marker-based methods to estimate quantitative genetics parameters in wild populations. Despite performing slightly better than the Ritland method, the relationship classes method (method 2) requires a known family structure with only two classes of relatedness and is therefore of restricted use [52]. Methods 3 and 4 both use the statistical machinery of the animal model (mixed model) after using molecular markers to reconstruct the pedigree (method 3) or the
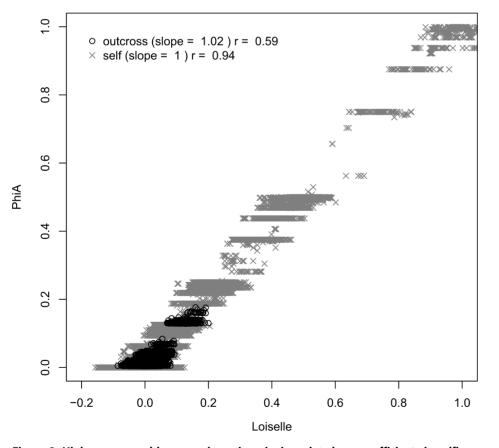

Figure 3. Higher mean and larger variance in pairwise relatedness coefficients in selfing compared to outcrossing populations. Regression between pairwise Loiselle coefficients estimated using 1500 SNP and Φ_A. The population comprised 500 individuals with 90% selfing (grey crosses) or complete outcrossing (black circles). The legend indicates the slope of the regression of Φ_A against Loiselle and the correlation coefficient r. The variance in relatedness was 0.0026 in the outcrossing population and 0.0108 in the selfing population (within the range of variances observed in wild populations, see Table 1 and [23]).
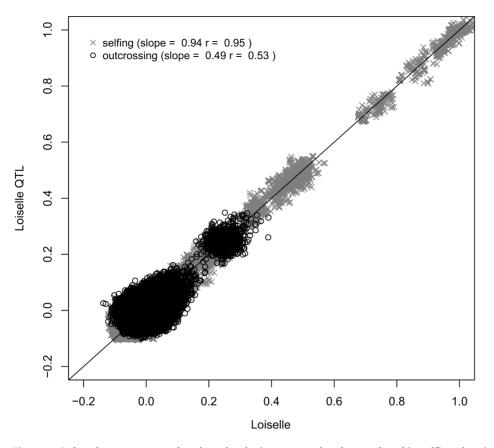doi:10.1371/journal.pone.0066983.g003

Figure 4. Relatednesses at causal and marker loci are more closely correlated in selfing than in outcrossing populations. Regression between pairwise Loiselle coefficients estimated using 1500 SNPs and pairwise Loiselle coefficients estimated using the allele frequency at QTLs determining the phenotypic trait. Outcrossing populations are shown in black and selfing populations (selfing rate 90%) in grey. The legend indicates the slope of the regression and the correlation coefficient r. The slope is expected to be close to one if the relatedness at causal loci is accurately predicted by the relatedness at observed SNPs.

doi:10.1371/journal.pone.0066983.g004

relatedness matrix (method 4). But empirical and simulation results suggest that method 4 performs best. Finally, genomic selection (method 5) has become extremely popular among breeders, and even though their main purpose is to predict breeding values with the highest accuracy, some of these studies report estimates for additive genetic variance [27]. Nevertheless, the lowest biases are found in studies using samples with family structure [67,68] and simulation studies seem more encouraging than actual empirical results (Tables 1 and 2). Besides, despite using a colossal number of SNPs for genomic selection (and not only those showing significant association with the phenotype – GWAS [69]), the SNPs still often explain only a small proportion of heritability (missing heritability) [25,70,71]. Insufficient linkage disequilibrium between causal variants and genotyped SNPs and low minor allele frequency of causal variants might be involved [25]. Currently, models combining pedigree and molecular markers are being developed to partition the additive genetic variance into genomic and "remaining polygenic" components [68,72,73]. These results highlight that for traits with highly polygenic determinism, the classic infinitesimal model [74,75], as implemented in the animal model, might still perform best [27].

Promising Prospects for Marker-based Animal Models?

Our review showed that animal models including marker-based relatedness matrices (method 4) offer promising prospects.

Nevertheless, our simulations show that using molecular markers instead of pedigrees seriously worsens the estimation of quantitative genetics parameters in outcrossing populations, even if an increased number of loci and truncated relatedness coefficients improved the result. Conversely, our simulations in selfing populations suggest that pedigree-free methods are successful. We discuss several arguments that could help explain this contrast, in the light of our simulation results and the literature available about inbreeding.

Firstly, we sampled individuals from a single generation in our simulations (non-overlapping generations, i.e. annual populations). Such a sampling design constrains the variance in pairwise relatedness in outcrossing populations (no parent-offspring relatedness and a probability of $1/N$ to be full-sibs) but not so much in selfing populations (full-sibs from selfing events have $\Phi_A > 0.5$). More generally, the variance in relatedness in a population is strongly affected by its size and mating system (see Figure 3, Figure S3 and [49]). Variances in relatedness are generally low in large outcrossing populations (see [23] and the column "variance in relatedness" in Table 1) but this constrains pedigree methods just as much as marker-based methods, and require collecting data over a large number of generations.

A more serious issue is that uncertainty in relatedness estimates is ignored when estimating heritability using this pedigree-free method and this could be a problem if uncertainty is high. Previous studies have shown that marker-based estimates of
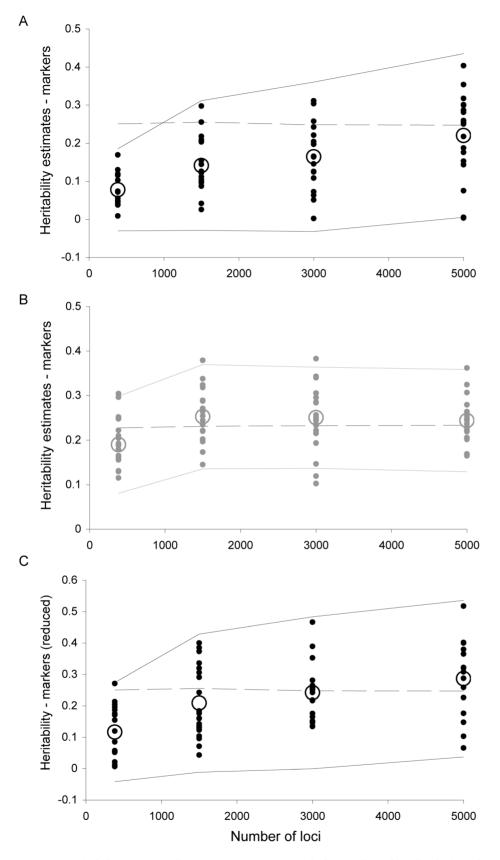
Figure 5. Heritability estimates become more accurate with the number of marker loci used to estimate relatedness. Influence of the number of loci used to estimate pairwise relatedness coefficients (Loiselle coefficients) on the bias in heritability estimates, when using a marker-based animal model. Each dot stands for a simulated population of 500 individuals, with complete outcrossing (panel A, in black) or 90% selfing (panel B, in grey). Panel C shows the results when marker-based relatedness coefficients are truncated before the analysis. Large circles stands for the

inbreeding coefficients are improved by non-random mating because it generates identity disequilibrium [48]. We observe the same thing here for pairwise relatednesses (i.e. the inbreeding coefficient of a hypothetical offspring) (Figure 3 and 4). In outcrossing populations, it remains that a very large set of markers is required to estimate the 'realised relatedness' without too much uncertainty [16,22] and to capture the variance within relatedness classes more accurately than when using the pedigree (Φ_A) [4,5,28,69].

It is worth pointing out that in our simulations, we used the complete pedigree (no missing individual), exempt from errors. Even in domestic animals, pedigree errors have been estimated to range between 1% and 10% [76]. In wild populations, the requirement of large sampling efforts and the occurrence of extra-pair paternity will inevitably scale down the pedigree information available for the analysis and might favour the use of alternative methods for a wide range of populations.

Finally, our simulations do not consider non-genetic sources of resemblance between relatives (common environment, maternal effects…). Such non-genetic effects are expected to overestimate heritability quite markedly [77]. Nevertheless, without transplantation or cross-fostering experiment, these effects are difficult to account for with any of the methods currently available (including pedigree methods, see [78]), in particular if there are few different classes of relatedness in the dataset [79].

Promising Prospects for Selfing Populations

Stebbins [30] suggested that extreme selfing is an "evolutionary dead-end", because it reduces genetic diversity within populations and may thereby lower their adaptive potential compared to outcrossing populations. Until recently, accurate comparisons of the levels of additive variance between selfers and outcossers were impaired by methodology because most studies on quantitative genetic variation in plants involved analyses of variance between and within families [80], and the pollinations required to derive paternal families were rarely performed (e.g. in 13 out of 37 studies reported by Charlesworth & Charlesworth, [43]), particularly in selfing species. Animal models coupled with marker-based related-ness information offer an inclusive, conceptually simple and flexible framework to quantify additive variance in plant and animal populations by taking advantage of the recombined genotypes produced by rare outcrossing events in their wild environment [69].

Yet, in selfing populations, inbreeding generates additional variance components specifically associated with dominance effects of alleles when autozygous (identical by descent) (detailed in equation (4) in Shaw et al. [81]). The covariance between the genotypic values of two individuals X and Y becomes [80,82]:

$$Cov(G_X, G_Y) = 2\Phi_A V_A + 2\delta_{\ddot{X}+\ddot{Y}} V_D + 2\left(\gamma_{\ddot{X}Y} + \gamma_{X\ddot{Y}}\right)$$
$$\text{cov}(A,D) + \delta_{\ddot{X}\ddot{Y}} V_{DI} \tag{2}$$
$$+ \left(\Delta_{\ddot{X}.\ddot{Y}} - F_X F_Y\right) H* + (2\Phi_A)^2 V_{AA}$$

where V_A and V_D are the additive and dominance variances, $cov(A,D)$ is the covariance between the additive effect of alleles and their autozygous dominance deviation, V_{DI} is the total variance due to autozygous dominance effects, $H*$ is the inbreeding depression and V_{AA} the variance due to additive-additive epistatic effects (other epistatic variances are neglected here). Each variance

component is preceded by a probability measure ($2\Phi_A, 2\delta_{\ddot{X}+\ddot{Y}} \ldots$), a function of the identity of alleles by descent (described in further details by Cockerham [83] and Harris [82]). This equation highlights the complexity of quantitative genetics in partially inbred populations. In our simulations, we only focused on V_A and neglected the effect of directional dominance because we were interested in methods providing comparable estimates of additive variance in selfing and outcrossing populations. These additional variance components could be included in the mixed model ($y = Za + \ldots + e$), keeping in mind that their estimation is not straightforward without a complex experimental design [84]. Besides, dominance effects due to heterozygote effects (d_{ij} in Harris [82]) are expected to contribute little to genetic variance in highly selfing populations where homozygosity is very high [80]. Therefore, neglecting V_D in selfing populations is not as inaccurate as in outcrossing populations, where it is rarely examined. Theoretical models also predict that inbreeding depression resulting from deleterious recessive alleles should be purged with selfing [85,86]. Some empirical data support this prediction [87,88], but the opposite has also been reported [89], in agreement with the prediction that mildly deleterious mutations could accumulate in selfing populations [90]. Finally, with partial selfing the estimate of V_A using the animal model can be inflated by variance due to additive-additive epistasis, directly proportional to Φ_A^2. Importantly, this is also true for outcrossing species, even if we expect lower effects [91]. We are therefore confident that additive variance estimated using marker-based animal models in selfing populations should be comparable with estimates in outcrossing populations, in spite of inbreeding.

It remains that such estimates of additive variance may not be sufficient to predict the response to selection (R) when there is directional dominance and epistasis. Simulations suggest that evolution in partially selfing populations can strongly differ from the predictions obtained using the breeders' equation ($R = h^2 S$ where S is the selection differential), even if the later accounts for inbreeding depression [81]. Methods have been suggested to extend the breeder's equation to selfing populations [92–96], but predictions are difficult when all individuals do not share the same level of inbreeding, as expected in most natural populations.

Conclusion

Our literature review highlighted that more testing is required for the most promising marker-based methods: animal models including a marker-based matrix of relatedness and genomic selection. Our simulations of the animal model showed that estimates in selfing populations are as accurate when using molecular markers or pedigrees, thanks to their high identity disequilibrium. It is undeniable that a very large set of molecular markers is required in large random mating populations, but recent advances in next generation sequencing technologies provide encouraging prospects, even for non-model species [97,98]. More generally, populations with high identity disequi-librium (consanguineous or bottlenecked populations) could promote the use of marker-based animal models, but at the same time are more likely to deviate from the standard assumptions of quantitative genetics models (e.g. non-additive variance).

Supporting Information

Figure S1 Correlation between heritability estimates obtained using pedigree-based animal models or one of the five marker-based method. Each dot stands for an empirical result and the colour indicates the method (Ritland in black; relatedness classes in grey; pedigree reconstruction in pink; animal model in green and genomic selection in blue). The dashed lines represent $y = x$.
(DOC)

Figure S2 Slight increase in the bias with lower heritabilities or lower number of markers. The panels A and B show the relationship between the bias in heritability estimate and the value of heritability for the empirical data (in A – bias expressed as $\hat{h}^2_{marker}$ - $\hat{h}^2_{pedigree}$) or the simulation data (in B – bias expressed as $E(\hat{h}^2 - h^2)$). The panels C and D show the relationship between the bias in heritability estimate and the number of microsatellites or SNPs used. As explained in the text, the bias did not systematically increase for traits with low heritabilities but was more variable. Surprisingly, we found no overall relationship between the bias and the number of markers used in empirical data.
(DOC)

Figure S3 Influence of population size and mating regime on the variance in pairwise relatedness in a population. Selfing could improve marker-based estimation of heritability because it affects the structure and the variance of pairwise relatedness, as has been shown for the inbreeding coefficient. We assessed the effect of the reduced effective population size in selfing populations ($Ne = N/(1+ F)$) by simulating populations with varying census size ($N = 50$, 100, 250, 500) and different mating regimes (outcrossing in black and 90% selfing in grey). Error bars stand for the standard error estimated from 10 replicated simulations. This figure confirms that pairwise relatedness coefficients have a higher mean and a larger variance in selfing compared to outcrossing populations and that this effect extends beyond the simple influence of population size (e.g. large excess of variance in a selfed population of $N = 50$ compared to an outcrossed population of $N = 100$). The identity disequilibrium created by selfing might explain such higher variance in relatedness.
(DOC)

Table S1 Full data literature review.
(XLS)

Zipfile S1 Simulation program in C++.
(ZIP)

Author Contributions

Conceived and designed the experiments: LG MS JR. Performed the experiments: LG MS. Analyzed the data: LG. Contributed reagents/materials/analysis tools: LG MS JR. Wrote the paper: LG MS JR.

References

1. Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Harlow, Essex, UK: Longmans Green.
2. Brookfield JFY (2009) Evolution and evolvability: celebrating Darwin 200. Biology Letters 5: 44–46.
3. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer.
4. Visscher PM (2009) Whole genome approaches to quantitative genetics. Genetica 136: 351–358.
5. Hill WG (1993) Variation in genetic identity within kinships. Heredity 71: 652–653.
6. Cockerham CC, Weir BS (1983) Variance of actual inbreeding. Theoretical Population Biology 23: 85–109.
7. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, et al. (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLOS Genetics 2: 316–325.
8. Guo S-W (1998) Variation in genetic identity among relatives. Human Heredity 46: 61–70.
9. Wright S (1922) Coefficients of inbreeding and relationship. The American Naturalist 56: 330–338.
10. Nussey DH, Wilson AJ, Brommer JE (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20: 831–844.
11. Kruuk LEB, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. Proceedings of the Royal Society B: Biological Sciences 275: 593–596.
12. Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'Animal Model'. Philosophical Transactions: Biological Sciences 359: 873–890.
13. Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology & Evolution 18: 503–511.
14. Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. Nature Review Genetics 11: 800–805.
15. Queller DC, Goodnight KF (1989) Estimating relatedness using molecular markers. Evolution 43: 258–275.
16. Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152: 1753–1766.
17. Wang J (2002) An estimator for pairwise relatedness using molecular markers. Genetics 160: 1203–1215.
18. Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. Human Heredity 43: 45–52.
19. Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83: 145–154.
20. Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany 82: 1420–1425.
21. Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genetical Research 67: 175–185.
22. Van De Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. Molecular Ecology 10: 1539–1549.
23. Csillery K, Johnson T, Beraldi D, Clutton-Brock T, Coltman D, et al. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics 173: 2091–2101.
24. Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology 13: 921–935.
25. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42: 565–569.
26. Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters in wild populations. Molecular Ecology 14: 1843–1859.
27. Sillanpää MJ (2011) On statistical methods for estimating heritability in wild populations. Molecular Ecology 20: 1324–1332.
28. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era - concepts and misconceptions. Nature Review Genetics 9: 255–266.
29. Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: The evolutionary quantitative genetics of wild animal populations. Annual Review of Ecology, Evolution, and Systematics 39: 525–548.
30. Stebbins GL (1957) Self fertilization and population variability in the higher plants. American Naturalist 91: 337–354.
31. Takebayashi N, Morrell PL (2001) Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. American Journal of Botany 88: 1143–1150.
32. Igic B, Kohn JR (2006) The distribution of plant mating systems: study bias against obligately outcrossing species. Evolution 60: 1098–1103.
33. Vogler DW, Kalisz S (2001) Sex among the flowers: The distribution of plant mating systems. Evolution 55: 202–204.
34. Jarne P, Auld JR (2006) Animals mix it up too: The distribution of self-fertilization among hermaphroditic animals. Evolution 60: 1816–1824.

35. Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 358: 1051–1070.

36. Hamrick JL, Godt MJW (1997) Allozyme diversity in cultivated crops. Crop Science 37: 26–30.

37. Glemin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proceedings of the Royal Society B-Biological Sciences 273: 3011–3019.

38. Pollak E (1987) On the theory of partially inbreeding finite populations.1. Partial selfing. Genetics 117: 353–360.

39. Hedrick PW (1980) Hitchhiking - a comparison of linkage and partial selfing. Genetics 94: 791–808.

40. Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

41. Schoen DJ, Brown AHD (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proceedings of the National Academy of Sciences of the United States of America 88: 4494–4497.

42. Crow J, Kimura M (1970) Introduction to Theoretical Population Genetics. New York: Harper and Row.

43. Charlesworth D, Charlesworth B (1995) Quantitative genetics in plants - the effect of the breeding system on genetic variability. Evolution 49: 911–920.

44. Van Buskirk J, Willi Y (2006) The change in quantitative genetic variation with inbreeding. Evolution 60: 2428–2434.

45. Barton NH, Turelli M (2004) Effects of genetic drift on variance components under a general model of epistasis. Evolution 58: 2111–2132.

46. Geber MA, Griffen LR (2003) Inheritance and natural selection on functional traits. International Journal of Plant Sciences 164: S21–S42.

47. Weir BS, Cockerham CC (1973) Mixed self and random mating at two loci. Genetical Research 21: 247–262.

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

49. Reid JM, Keller LF (2010) Correlated inbreeding among relatives: occurence, magniture and implications. Evolution 64: 973–985.

50. Ritland K (2000) Detecting Inheritance with Inferred Relatedness in Nature. Adaptive genetic variation in the wild. USA: Oxford University Press.

51. Mousseau TA, Ritland K, Heath DD (1998) A novel method for estimating heritability using molecular markers. Heredity 80: 218–224.

52. Thomas SC, Hill WG (2000) Estimating quantitative genetic parameters using sibships reconstructed from marker data. Genetics 155: 1961–1972.

53. Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. Molecular Ecology Resources 10: 6–30.

54. Milner JM, Pemberton JM, Brotherstone S, Albon SD (2000) Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach. Journal of Evolutionary Biology 13: 804–813.

55. Henderson CR (1984) Applications of Linear Models in Animal Breeding. Guelph, Canada: University of Guelph.

56. Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136: 245–257.

57. Strandén I, Garrick DJ (2009) Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. Journal of Dairy Science 92: 2971–2975.

58. Crossa J, Campos Gdl, Perez P, Gianola D, Burgueno J, et al. (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186: 713–724.

59. Houle D (1992) Comparing evolvability and variability of quantitative traits. Genetics 130: 195–204.

60. Hansen T, Pélabon C, Houle D (2011) Heritability is not evolvability. Evolutionary Biology 38: 258–277.

61. Ewens WJ (2004) Mathematical Population Genetics. New York: Springer-Verlag.

62. Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. Molecular Ecology 12: 1577–1588.

63. Nordborg M, Donnelly P (1997) The coalescent process with selfing. Genetics 146: 1185–1195.

64. Gilmore AR, Gogel BJ, Cullis BR, Thompson R (2006) Asreml User Guide Release 2.0. Hemel Hempsted: VSN International Ltd.

65. Frentiu FD, Clegg SM, Chittock J, Burke T, Blows MW, et al. (2008) Pedigree-free animal models: the relatedness matrix reloaded. Proceedings of the Royal Society B: Biological Sciences 275: 639–647.

66. Pemberton JM (2008) Wild pedigrees: the way forward. Proceedings of the Royal Society B: Biological Sciences 275: 613–621.

67. Lee S, Goddard M, Visscher P, van der Werf J (2010) Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. Genetics Selection Evolution 42: 22.

68. Jensen J, Su G, Madsen P (2012) Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. BMC Genetics 13: 44.

69. Hill WG (2012) Quantitative genetics in the genomics era. Current genomics 13: 196–206.

70. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.

71. Maher B (2008) The case of the missing heritability. Nature 456: 18–21.

72. Christensen O, Lund M (2010) Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42: 2.

73. Legarra A, Robert-Granié C, Manfredi E, Elsen J-M (2008) Performance of genomic selection in mice. Genetics 180: 611–618.

74. Bulmer MG (1980) The mathematical theory of quantitative genetics. Oxford, UK: Oxford University Press.

75. Fisher RA (1918) The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh: Earth Sciences 52: 399–433.

76. Leroy G, Danchin-Burge C, Palhiere I, Baumung R, Fritz S, et al. (2012) An ABC estimate of pedigree error rate: application in dog, sheep and cattle breeds. Animal Genetics 43: 309–314.

77. Stopher KV, Walling CA, Morris A, Guinness FE, Clutton-Brock TH, et al. (2012) Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in red deer. Evolution 66: 2411–2426.

78. Kruuk LEB, Hadfield JD (2007) How to separate genetic and environmental causes of similarity between relatives. Journal of Evolutionary Biology 20: 1890–1903.

79. Thomas SC, Coltman DW, Pemberton JM (2002) The use of marker-based relationship information to estimate the heritability of body weight in a natural population: a cautionary tale. Journal of Evolutionary Biology 15: 92–99.

80. Holland JB, Nyquist WE, Cervantes-Martínez CT (2010) Estimating and Interpreting Heritability for Plant Breeding: An Update. Plant Breeding Reviews: John Wiley & Sons, Inc. pp. 9–112.

81. Shaw RG, Byers DL, Shaw FH (1998) Genetic components of variation in *Nemophila menziesii* undergoing inbreeding: Morphology and flowering time. Genetics 150: 1649–1661.

82. Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50: 1319–1348.

83. Cockerham CC (1971) Higher order probability functions of identity of alleles by descent. Genetics 69: 235–246.

84. Goldringer I, Brabant P, Gallais A (1997) Estimation of additive and epistatic genetic variances for agronomic traits in a population of doubled-haploid lines of wheat. Heredity 79: 60–71.

85. Lande R, Schemske DW (1985) The evolution of self- fertilization and inbreeding depression in plants. 1. Genetics models. Evolution 39: 24–40.

86. Wright S (1969) Evolution and the Genetics of Populations. Vol. II. The Theory of Gene Frequencies. Chicago: University of Chicago Press.

87. Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics 10: 783–796.

88. Husband BC, Schemske DW (1996) Evolution of the magnitude and timing of inbreeding depression in plants. Evolution 50: 54–70.

89. Byers DL, Waller DM (1999) Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. Annual Review of Ecology and Systematics 30: 479–513.

90. Glémin S, Ronfort J, Bataillon T (2003) Patterns of inbreeding depression and architecture of the load in subdivided populations. Genetics 165: 2193–2212.

91. Holland JB (2010) Epistasis and Plant Breeding. Plant Breeding Reviews: John Wiley & Sons, Inc. pp. 27–92.

92. Kelly JK, Williamson S (2000) Predicting response to selection on a quantitative trait: A comparison between models for mixed-mating populations. Journal of Theoretical Biology 207: 37–56.

93. Pederson DG (1969) The prediction of selection response in a self-fertilizing species. Australian Journal of Biological Sciences 22: 117–130.

94. Cockerham CC, Matzinger DF (1985) Selection response based on selfed progenies. Crop Science 25: 483–488.

95. Wright, A J, Cockerham, C C (1985) Selection with partial selfing. I: Mass selection. Bethesda, MD, USA: Genetics Society of America.

96. Kelly JK (1999) Response to selection in partially self-fertilizing populations. I. Selection on a single trait. Evolution 53: 336–349.

97. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. Plos One 7.

98. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. Plos One 3.

99. Coltman DW (2005) Testing marker-based estimates of heritability in the wild. Molecular Ecology 14: 2593–2599.

100. Ritland K, Ritland C (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. Evolution 50: 1074–1082.

101. van Kleunen M, Ritland K (2005) Estimating heritabilities and genetic correlations with marker-based methods: An experimental test in *Mimulus guttatus*. Journal of Heredity 96: 368–375.

102. Klaper R, Ritland K, Mousseau TA, Hunter MD (2001) Heritability of phenolics in *Quercus laevis* inferred using molecular markers. Journal of Heredity 92: 421–426.

103. Bouvet JM, Kelly B, Sanou H, Allal F (2008) Comparison of marker- and pedigree-based methods for estimating heritability in an agroforestry population of *Vitellaria paradoxa* CF Gaertn. (shea tree). Genetic Resources and Crop Evolution 55: 1291–1301.

104. Doran JC, Matheson AC (1994) Genetic parameters and expected gains from selection for monoterpene yields in Petford *Eucalyptus camaldulensis*. New Forests 8: 155–167.

105. Andrew RL, Peakall R, Wallis IR, Wood JT, Knight EJ, et al. (2005) Marker-based quantitative genetics in the wild?: The heritability and genetic correlation of chemical defenses in eucalyptus. Genetics 171: 1989–1998.

106. Bessega C, Saidman BO, Darquier MR, Ewens M, Sanchez L, et al. (2009) Consistency between marker- and genealogy-based heritability estimates in an experimental stand of *Prosopis alba* (Leguminosae). American Journal of Botany 96: 458–465.

107. Kumar S, Richardson TE (2005) Inferring relatedness and heritability using molecular markers in radiata pine. Molecular Breeding 15: 55–64.

108. Shikano T (2008) Estimation of quantitative genetic parameters using marker-inferred relatedness in Japanese flounder: A case study of upward bias. Journal of Heredity 99: 94–104.

109. Wilson AJ, McDonald G, Moghadam HK, Herbinger CM, Ferguson MM (2003) Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. Genetics Research 81: 145–156.

110. Heath DD, Devlin RH, Heath JW, Iwama GK (1994) Genetic, environmental and interaction effects on the incidence of jacking in *Oncorhynchus tshawytscha* (chinook salmon). Heredity 72: 146–154.

111. Blonk RJW, Komen H, Kamstra A, van Arendonk JAM (2010) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184: 213–219.

112. Hayes BJ, Goddard ME (2008) Technical note: Prediction of breeding values using marker-derived relationship matrices. Journal of Animal Sciences 86: 2089–2092.

113. Anderson TJC, Williams JT, Nair S, Sudimack D, Barends M, et al. (2010) Inferred relatedness and heritability in malaria parasites. Proceedings of the Royal Society B: Biological Sciences 277: 2531–2540.

114. Veerkamp RF, Mulder HA, Thompson R, Calus MPL (2011) Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. Journal of Dairy Science 94: 4189–4197.

115. Rolf M, Taylor J, Schnabel R, McKay S, McClure M, et al. (2010) Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. BMC Genetics 11: 24.

116. Usai MG, Goddard ME, Hayes BJ (2009) LASSO with cross-validation for genomic selection. Genetics Research 91: 427–436.

117. Goddard ME, Hayes BJ, Meuwissen THE (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of Animal Breeding and Genetics 128: 409–421.

118. Toro M, Barragán C, Óvilo C, Rodrigañez J, Rodriguez C, et al. (2002) Estimation of coancestry in Iberian pigs using molecular markers. Conservation Genetics 3: 309–320.

119. VanRaden PM (2008) Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414–4423.

120. Oliehoek PA, Windig JJ, van Arendonk JAM, Bijma P (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics 173: 483–496.

121. Smith BR, Herbinger CM, Merry HR (2001) Accurate partition of individuals into full-sib families from genetic data without parental information. Genetics 158: 1329–1338.

122. Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7: 639–655.

123. Wang J (2004) Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963–1979.

124. Duchesne P, Godbout M-H, Bernatchez L (2002) PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. Molecular Ecology Notes 2: 191–193.

125. Ritland K (1996) A marker-based method for inferences about quantitative inheritance in natural populations. Evolution 50: 1062–1073.

126. Thomas SC, Pemberton JM, Hill WG (2000) Estimating variance components in natural populations using inferred relationships. Heredity 84: 427–436.

127. Rodriguez-Ramilo ST, Toro MA, Caballero A, Fernandez J (2007) The accuracy of a heritability estimator using molecular information. Conservation Genetics 8: 1189–1198.

128. DiBattista JD, Feldheim KA, Garant D, Gruber SH, Hendry AP (2009) Evolutionary potential of a large marine vertebrate: quantitative genetic parameters in a wild population. Evolution 63: 1051–1067.

129. Yu J, Zhang Z, Zhu C, Tabanao DA, Pressoir G, et al. (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. The Plant Genome 2: 63–77.

130. Nejati-Javaremi A, Smith C, Gibson JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. Journal of Animal Science 75: 1738–1745.

131. Villanueva B, Pong-Wong R, Fernández J, Toro MA (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. Journal of Animal Science 83: 1747–1752.

132. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

133. Calus MPL, Veerkamp RF (2007) Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. Journal of Animal Breeding and Genetics 124: 362–368.

134. Zhang Z, Liu J, Ding X, Bijma P, de Koning D-J, et al. (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS ONE 5: e12648.

135. Sørensen A, Pong-Wong R, Windig J, Woolliams J (2002) Precision of methods for calculating identity-by-descent matrices using multiple markers. Genetics Selection Evolution 34: 1–23.