THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

RAD-sequencing for estimating GRM-based heritability in the wild

OPEN ACCESS

DR. ERWAN QUEMERE (Orcid ID : 0000-0002-3880-1933)

RAD-sequencing for estimating GRM-based heritability in the wild:
a case study in roe deer

Gervais L[1,2], Perrier C[3], Bernard M[4,5], Merlet J[1], Pemberton J[6], Pujol B[2,7]*, Quéméré E[1]*.

[1]CEFS, INRA, Université de Toulouse, Castanet-Tolosan Cedex F-31326, France

[2]Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université Fédérale de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 31062 Toulouse, France

[3]CEFE UMR 5175, CNRS, route de Mende, 34290 Montpellier, France

[4]SIGENAE, INRA, Jouy-en-Josas, France

[5]GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[6]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK

[7]PSL Université Paris: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France.

*These authors contributed equally to this work as co-last author.

Corresponding authors: Laura Gervais, lauragervais51@gmail.com and Erwan Quéméré, erwan.quemere@inra.fr.
Phone: (+33) 561 285 497. Fax: (+33) 561 285 500.

Running head: RAD-Seq for estimating GRM-based heritability

Abstract

Estimating the evolutionary potential of quantitative traits and reliably predicting responses to selection in wild populations are important challenges in evolutionary biology. The genomic revolution has opened up opportunities for measuring relatedness among individuals with precision,

enabling pedigree-free estimation of trait heritabilities in wild populations. However, until now, most quantitative genetic studies based on a genomic relatedness matrix (GRM) have focused on long-term monitored populations for which traditional pedigrees were also available, and have often had access to knowledge of genome sequence and variability. Here, we investigated the potential of RAD-sequencing for estimating heritability in a free-ranging roe deer population for which no prior genomic resources were available. We propose a step-by-step analytical framework to optimize the quality and quantity of the genomic data and explore the impact of the SNP calling and filtering processes on the GRM structure and GRM-based heritability estimates. As expected, our results show that sequence coverage strongly affects the number of recovered loci, the genotyping error rate and the amount of missing data. Ultimately, this had little effect on heritability estimates and their standard errors, provided that the GRM was built from a minimum number of loci (above 7000). GRM-based heritability estimates thus appear robust to a moderate level of genotyping errors in the SNP dataset. We also showed that quality filters, such as the removal of low-frequency variants, affect the relatedness structure of the GRM, generating lower h^2 estimates. Our work illustrates the huge potential of RAD-sequencing for estimating GRM-based heritability in virtually any natural population.

Introduction

Estimating the evolutionary potential of quantitative traits and reliably predicting responses to selection in wild populations are important challenges in evolutionary biology. However, measuring the additive genetic variance and heritability of a trait, and the genetic correlation between a trait and fitness components (i.e. predicting the response to selection), require estimates of pairwise relatedness between individuals. For wild populations, these estimates have traditionally been based on a multigenerational pedigree. However, many species of research interest are hard to sample with the intensity and long-term effort required for pedigree construction (Pemberton,

2008). This has led to a strong taxonomic bias, with the great majority of heritability estimates available for vertebrates, especially birds and large mammals, because of their ease of monitoring (Postma, 2014).

Advances in high-throughput sequencing technology have opened up the possibility of getting access to the realized proportion of the genome that is shared among individuals (i.e. generating a Genomic Relatedness Matrix - GRM) in virtually any non-model species, with the potential to greatly expand the taxonomic coverage of quantitative genetic studies in the wild (Gienapp et al., 2017). While first applied to human data (Yang et al., 2011), GRM have recently been shown to provide heritability estimates which are similar to those from multigenerational pedigrees in free-ranging animal populations (Bérénos et al., 2014 on Soay sheep, Robinson et al., 2013 on great tits, Perrier et al., 2018 on blue tits). These studies have demonstrated that GRMs computed from a few hundreds of individuals and a few thousand loci may be enough to obtain heritability estimates with low standard errors when effective population size is relatively small and linkage disequilibrium (LD) high (Bérénos et al., 2014; Perrier et al., 2018; Stanton-Geddes, Yoder, Briskine, Young, & Tiffin, 2013).

The few studies that have attempted to estimate GRM-based heritabilities in the wild have all focused on long-term monitored populations for which pedigrees were already available. Moreover, the authors often had access to knowledge about both genome sequence and variability on the study species, or on a closely related species, facilitating the development of the genotyping SNP array. Whole genome sequencing is still prohibitive for many species in terms of cost, bioinformatic resources especially for organisms with large genomes, and required DNA quality (Ekblom & Wolf, 2014). Recently, a variety of Restriction site-Associated DNA sequencing (RAD-seq - Andrews et al 2016) methods and other similar approaches to sequencing a subset of the genome have been developed. These approaches are increasingly attractive because they are cost-effective, not

dependent on prior genomic information and SNP loci are discovered and genotyped in a single procedure (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013), making them applicable to a wide range of non-model species. RAD-seq approaches are also valued for their high level of flexibility, allowing optimization of the trade-off between the number of markers, the number of genotyped individuals and sequencing depth in an Illumina run using a wide variety of experimental designs and using different restriction enzymes (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012).

Despite the numerous benefits, the potential of RAD-seq for estimating GRM and quantitative genetic parameters in the wild, particularly in non-model species, has received little attention (but see Perrier et al., 2018). One reason is that long-term research projects willing to genotype thousands of individuals to estimate GRM often invest in developing a SNP chip. Another reason is that RAD-seq approaches are reputed to lead to high rates of missing data and allele dropout that might bias relatedness estimates and downstream biological inferences (Dodds et al., 2015; Gienapp et al., 2017). Furthermore, although easy-to-use bioinformatic pipelines are available for the *de novo* assembly of loci and SNP genotyping (J. M. Catchen et al., 2017; Eaton, 2014), little is known about the impact of sequencing strategy (e.g. marker read-depth coverage) and parameter choice during the SNP calling/filtering process on genotyping error/missing data rates, GRM structure and, ultimately, quantitative genetic estimates. Several recent studies have proposed methodologies to optimize the *de novo* assembly of markers and minimize error rates (Fountain, Pauli, Reid, Palsbøll, & Peery, 2016; Mastretta-Yanes et al., 2015; Paris, Stevens, & Catchen, 2017; Rochette & Catchen, 2017). Others have explored how relatedness estimators can circumvent the influence of genotyping errors (Attard, Beheregaray, & Möller, 2018) or tested the influence of the number of markers/samples on heritability estimates from whole genome data (Stanton-Geddes et al., 2013). However, to our knowledge, no study has explicitly explored the potential limits and pitfalls of RAD-seq for estimating quantitative genetic parameters in the wild.

In this study, we describe an analytical framework (see Figure 1) to estimate trait heritability from RAD sequencing data, using a free-ranging roe deer population (*Capreolus capreolus*, Linnaeus 1758) as a case study. We focus on the heritability of body mass, a trait closely associated with both survival and reproductive performance (Hamel, Gaillard, Festa-Bianchet, & Côté, 2009; Quéméré et al., 2018). There are three key parameters for estimating GRM-based heritability: the number of SNPs, the number of individuals and the sampling variance in relatedness (Visscher & Goddard, 2015). First, we provide guidance on the sampling and sequencing strategy. Using *in silico* simulation and double-digest RAD-sequencing (Peterson et al., 2012), we established a sequencing strategy that optimizes the balance between the number of loci and the number of individuals genotyped, and library/sequencing costs. The sequencing depth (i.e. average read depth per locus) may directly affect data quality, with potential impact on biological inferences (Sims et al 2014). We, therefore, explored how variable sequencing effort (coverage depth of 20x *versus* 60x) affects the rates of genotyping error and missing data, the accuracy of relatedness coefficients and, ultimately, heritability estimates. We then detail the different steps in the bioinformatic and analytical pipeline from the raw sequence data to the implementation of the GRM and the estimation of GRM-based heritability. Specifically, we explore how parameter choice during the *de novo* assembly of loci and the SNP data filtering may influence data quality (genotyping error rate) and quantity (number of informative loci), the structure of the GRM and GRM-based heritability. Lastly, we provide practical recommendations to minimize bias in the estimation of relatedness while maximizing the explanatory power of the GRM.

2. Methods

2.1. Sampling design

The study focused on a roe deer population inhabiting a heterogeneous agricultural landscape in southwestern France (43°13′N, 0°52′E). We selected a set of 250 individuals which were caught using large scale drives during winter, between 2002 and 2016 at seven sampling sites, separated by a few

kilometers, and with a variable proportion of woodland *vs.* open areas (e.g. crops, meadows, pasture) (Figure S1). Adult roe deer generally occupy a single home range across their entire adult life, whereas around 40% of juveniles disperse (Debeffe et al., 2012). Our hierarchical sampling scheme is thus expected to maximize variance in relatedness by including potentially closely related individuals within sampling sites, individuals with intermediate relatedness among sampling sites and unrelated individuals that immigrated from outside the study area. At each capture, animals were sexed, weighed and assigned to an age class: juveniles (< 1 year old), yearling (1-2 years) or adults (> 2 years). When individuals were captured in successive years, we used only data from the first capture. Direct field observations of females equipped with Very High Frequency (VHF) transmitters allowed us to identify mother-offspring relationships (N= 8 pairs included in further analysis). All applicable institutional and European guidelines for the care and use of animals were followed. All the procedures involving animals were approved by the Ethical Committee 115 of Toulouse and authorized by the French Ministry in charge of ethical evaluation (n° APAFIS#7880-2016120209523619v5).

2.2. Sequencing strategy

DNA was extracted from skin samples using DNeasy^{TM} Tissue Kit (Qiagen). SNP genotyping was performed using double-digest restriction site associated DNA sequencing (ddRAD-seq, Peterson et al., 2012). This method was chosen because it offers greater flexibility than traditional RAD-seq, with the possibility of optimising the number of loci sequenced by testing different combinations of restriction enzymes and fragment size selection. This optimization process was carried out by performing *in-silico* digestions of the red deer genome (*Cervus elaphus*) (Bana et al., 2018) (divergence time from roe deer 7.7-9.6 million years, Gilbert et al., 2006) using the R package Sim-RAD (Lepais & Weir, 2014). We tested various experimental designs and finally retained the two enzymes *EcoR1* and *MsPI* and size-selected libraries of 270-330 bp insert-size that produced around

35,000 fragments in the red deer genome. Using this design, we expected to sequence 96 individuals with an average of 60X coverage in a single Illumina HiSeq 2500 lane (based on average read counts of 200 M reads per lane). To ensure that this design provides at least 15,000 polymorphic loci (i.e. the minimum required to capture the additive genetic variance of traits in Soay sheep Bérénos et al., 2014), we performed a pilot study on a set of 12 samples sequenced using an Illumina MiSeq lane (V2; Paired-end reads, 2x150 bp) . We then applied our protocol to the whole dataset of 250 samples which were multiplexed in equimolar proportions in groups of 24 individuals and sequenced on Illumina HiSeq 2500 (V4; Paired-end reads of 2x125 bp). Library preparation and Illumina sequencing was performed at the NERC Biomolecular Analysis Facility-Edinburgh (Genepool, Edinburgh, UK). To evaluate the reliability of the genotyping process and optimize the *de novo* loci assembly (see below), seven individuals were repeat-processed, either at the library preparation step (3 pairs of 'library replicates' individuals) or at the sequencing step (4 pairs of 'sequencing replicates').

2.3. De novo reconstruction of loci and error quantification

Raw sequences were first inspected with FASTQC (Andrews, 2010) for quality control. Reads were then demultiplexed (i.e. assigned to each sample) and trimmed to 117 bp using *process_radtags* (part of the *Stacks* 1.35 pipeline, Catchen et al. 2013) without allowing any mismatch in the barcode sequence. The '*de novo map'* pipeline of *Stacks* was used to build loci *de novo* (without a reference genome) and call single nucleotide polymorphisms (SNPs). *De novo* loci assembly in *Stacks* is governed by three main parameters that influence the balance between the number of loci, genotyping errors and missing data. The first two parameters affect the way loci are built at the individual level: -*m* is the minimum number of reads to form a stack (allele) and -*M* is the maximum number of mismatches allowed between alleles. A high value of –*m* may cause allele dropout, while a low value may generate false alleles (Catchen et al., 2013). A high value of –*M* may generate false

heterozygotes (due to the erroneous combination of different homozygous loci), while a low value may erroneously generate homozygous loci. The third parameter *-n* is the maximum number of mismatches allowed between homologous loci across all samples to build the population catalog.

The choice of parameter values is specific to each study since it depends on the biology of the study species (inherent polymorphism, ploidy level) and the experimental design (restriction enzyme used, number of samples multiplexed, sequencing coverage) (Paris et al., 2017). Following Mastretta-Yanes et al., 2015, we carried out a preliminary analysis to identify the parameter values that maximized the number of loci recovered while minimizing genotyping error rates in our case study. To guide our choice, we estimated two error rates between the seven pairs of technical replicates: (1) the *locus error rate* (LE_R) corresponding to the number of loci present in only one of the two replicates, divided by the total number of loci being compared (i.e. missing data at the locus level) and (2) the *allele error rate* (AE_R), calculated as the number of incongruent genotypes between the two replicates, divided by the number of common loci. We first explored the influence of each parameter (-m, -M and –n) one-by-one (while holding the two others fixed) (see Text S1 in supplementary material). On this basis, we tested four sets of parameters corresponding to different ways to deal with the trade-off between data quantity and quality: (S1) parameters designed to maximize the number of markers ('MaxLoci', m=2, M=2, n=1), (S2) parameters designed to minimize the error rates ('MinError', m=11, M=2, n=1), (S3) parameters designed to lead to a low error rate with an intermediate number of markers ('Intermediate', m=7, M=2, n=1) and (S4) parameters by default ('Default' hereafter, m=3, M=2, n=0). This preliminary analysis was restricted to loci with a single SNP that was shared by at least 80% of individuals. We retained the *Stacks* model that provided the optimal balance between number of loci and genotyping errors (AE_R) and used it on the full dataset (250 samples).

2.4. Quality filtering

Once the SNP calling process is completed, additional filtering steps are regularly performed to remove loci and/or individuals with too many missing data (loci and individual missingness rates respectively) and persistent genotyping errors (removing rare variants and loci that deviate from Hardy-Weinberg equilibrium). Here, we explored the impact of these quality filters on genotyping and locus error rates, number of loci and GRM structure (overall variance and mother-offspring relationship coefficients). We successively applied three filters to the SNP dataset by testing different threshold values using VCFtools (Danecek et al. 2011): (1) the "HWE filter" removes loci that deviate from Hardy-Weinberg Equilibrium (p-value<0.05) because the population is known to be panmictic (Coulon et al. 2006, see Text S1), (2) the "LM filter" (Loci Missingness) discards loci with a missingness rate above 10%, 20%, 30% or 40% and (3) the "MAF" retains loci with Minimum Allele Frequency greater than 1%, 5% or 10%. We retained one SNP per locus (to minimize linkage disequilibrium between markers) and removed individuals with more than 50% missing data. All analyses were performed on the dataset built with the S3 'Intermediate' *Stacks* model described above which offered the best compromise between the number of loci and genotyping error rate, in other words which maximized the number of markers and minimized allele error rates (AE_R) (see results below).

2.5. Inference of GRM-based heritability

GRMs were computed using Genome-wide Complex Traits Analysis (GCTA) (Yang et al 2011) from identity by state (IBS) SNP relationships. We used the Unified Additive Relationship (UAR) estimator that has been shown to provide accurate quantitative genetic estimates similar to pedigree-based inference (Bérénos et al., 2014; Perrier et al., 2018). At each locus, relatedness was scaled by the expected heterozygosity 2pq (Yang et al. 2010, 2011). GRMs were then fitted in a mixed-linear

model with REstricted Maximum-Likelihood (REML) approach to estimate the amount of phenotypic variance in body mass that was explained by additive genetic variation (SNPs). We used the REML method because it is known to perform as well as Bayesian inference (with MCMC) for Gaussian traits like body mass while being far less computationally intensive (Wilson et al., 2010). Univariate animal models for body mass were run on 243 individuals (7 missing data) in ASREML-R v3.00 (VSN International, Hemel Hempstead, UK) as follows:

$$y_i = \mu + age_i + sex_i + age_i * sex_i + a_i + e_i$$

where y is the studied phenotypic trait for an individual i, μ is the overall population mean of y, a_i is the breeding value (i.e. effects of i's genotype relative to μ) and e_i the residual term. Because body mass is known to vary between sexes and age-classes (Gaillard, Delorme, & Jullien, 1993), individual's sex and age-class as well as the interaction between them (age_i*sex_i) were included in the model as fixed effects. Narrow-sense heritability of body mass was estimated as $h^2 = V_A / V_P$ where V_A is the additive genetic variance, V_R is the residual variance and V_P the total phenotypic variance ($V_{A+}V_R$). A Likelihood Ratio Test (LRT) (Meyer & Hill, 1992) was applied to test the statistical significance of h^2, *i.e.* by comparing the likelihood of the full model with and without an additive genetic effect fitted. Residuals and fitted values were visually inspected to verify assumptions of normality and homogeneity of variance (*i.e.* histogram of residuals, Q-Q plot of residuals, and scatter plot of residuals *versus* fitted values, Bolker et al., 2009). We investigated the influence of SNP calling (parameters in *Stacks*) and quality filtering (LM and MAF filter) on GRM-based heritability estimates. First, we explored the influence of loci missingness rate (LM ranging from 10% to 40%) by fixing the MAF to 1%. Then, we evaluated the impact of MAF (1%, 5% or 10%) by fixing the LM ratio to 20%. These analyses were carried out on SNP datasets called with the four sets of parameters in Stacks ('Max Loci', 'Min Loci', 'Intermediate' and 'Default'). Since these four models generated a variable number of informative loci, and because estimates of heritability increase up to an asymptote with the number of SNPs used to build the GRM (Bérénos et al., 2014; Perrier et al.,

2018), to be comparable, the h^2 value must be estimated using GRMs built from the same number of SNPs. Here, we fixed the number of SNPs to the minimum obtained in all analyses and performed fifty resampling iterations.

2.6. Influence of sequencing coverage

Genotyping error rates are assumed to be highly sensitive to sequencing coverage, particularly in *de novo*-assembled data sets (Catchen et al., 2013; Fountain et al., 2016). To explore the effect of sequencing coverage, we randomly sampled 20% of the raw reads of the initial dataset (using Seqtk: https://github.com/lh3/seqtk) so as to obtain an expected average read depth of 12 reads per locus (hereafter 'low-coverage dataset'). We then calculated genotyping error rates (LE_R and AE_R) between pair of replicates as we did for the full dataset and used the same sensitivity analysis described above to select the parameters in *Stacks* that offered the best compromise between data quality and quantity. Then, we evaluated how the filters affected the number of polymorphic loci, relationship coefficients, variance in relatedness of the GRM and heritability estimates.

3. Results

3.1. Optimization of the de novo loci assembly of loci and SNP discovery

An average of 6.9 million reads per individual was obtained after demultiplexing. The number of polymorphic loci recovered by *Stacks* in the exploratory analysis ranged from 14,536 for the 'MinError' parameter set to 21,681 loci for 'MaxLoci', with a median depth coverage per individual and per locus of between 39.1 and 88.4 reads. Median locus error rate (LE_R) across the seven pairs of replicates varied between 3.1% for 'MaxLoci' and 5.1% for 'default' (Figure 2). Median allele error rates (AE_R) ranged from 1.1% for 'MinError' to 3.2% for 'MaxLoci'. We retained the parameters that

offered the optimal balance between data quantity and quality for further analysis: 'MinError' and 'Intermediate' models generated similar low genotyping error rates (AE_R), but the 'intermediate' model yielded more loci (17,013). When loci assembly was attempted using a subset of our data (20% of the reads), we actually obtained a median depth coverage of 20x. In this case, loci and allele error rates were two to three times higher than when using the original dataset (Median LE_R ranging from 4.8% to 16%; Median AE_R from 1.8% and 7.6%, Figure S2). The number of loci recovered also dropped sharply from 17,534 polymorphic loci for 'MaxLoci' to 2,092 for the 'MinError' model.

3.2.   Quality filtering: impact on genotyping error rates, number of loci and relatedness coefficients

High-coverage dataset

HWE filter: Once the *de novo* assembly of loci had been completed on the full dataset (250 samples, using the 'intermediate' *Stacks* model), we obtained 96,773 loci totaling 154,540 SNPs. When removing loci that deviated from HWE, we retained 83,893 SNPs (Figure 3a). The 'HWE filter' had very little impact on loci and allele error rates (Figure 3b, c), but led to a substantial decrease in the number of SNPs (-13%), of the mother-offspring relatedness coefficients (median relatedness of 0.24 *versus* 0.31 before filtering), and of the off-diagonal variance in relatedness of the GRM (V_{GRM} = 0.6 x 10^{-3} *versus* 0.8 x 10^{-3} for the filtered and unfiltered datasets, respectively) (Figure 3d).

Loci missingness filter (LM): When we applied the 'loci missingness' filter (LM), we observed a two to three-fold decrease in the number of loci: between 51,355 and 64,400 SNPs were excluded when removing loci typed in less than 60% (LM = 40%) and 90% (LM = 10%) of individuals respectively. Mother-offspring relatedness (median relatedness ranging from r = 0.26-0.29) and variance in relatedness did not change markedly (V_{GRM} ranging from 0.6-0.7 x 10^{-3}), but the remaining loci had significantly lower error rates, particularly with a LM = 10% (LE_R ~ 0. 3%, AE_R ~ 0.2%).

MAF filter: Removing rare variants only affected locus error rate slightly, but this resulted in a lower allele error rate, particularly when using high LM: AE_R fell from 1.9% to 0.6% when using LM 40% and MAF 10% filters. MAF filtering greatly improved mother-offspring relatedness estimates which approximated the expected value of 0.5 (figure 3e). We got more accurate estimates of mother-offspring relatedness with MAF = 10% (r=0.49) than with MAF = 1% (r=0.45). However, this stringent MAF threshold also led to a >50% decrease in the number of markers (to a final number ranging between 8434 and 12,528 SNPs). MAF filtering had also a marked influence on the GRM off-diagonal variance in relatedness (V_{GRM}) which substantially increased from ~0.8 x 10^{-3} to ~1.6 x 10^{-3} when using a MAF = 1% filter and to ~2 x 10^{-3} with MAF = 10%. However, the MAF threshold had little effect on pairwise relatedness coefficients *per se*, although the highest coefficients tended to increase when using a stringent MAF = 10% filter (Figure 4a).

Low-coverage dataset

In comparison, the 'low-coverage dataset' provided 70,956 loci totaling 111,803 SNPs (i.e. a 27% decrease compared to the full coverage dataset) using the S4 'Default' stack model (optimal model for this dataset). The number of loci fell to between 3,493 loci (LM = 10%, MAF = 10%) and 11,441 (LM = 40%, MAF = 1%) after the filtering procedure (Figure 3b). The LM filter substantially reduced locus error rate (from 19.6% to ~3.5%). However, across all filtering options, allele error rate remained 10 times higher than in the high coverage dataset. As in the full dataset, variance in relatedness markedly increased when using the MAF filter to reach values between 1.4 x 10^{-3} (LM=40%, MAF=1%) and 2.3 x 10^{-3} (LM=10%, MAF=10%) while median mother-offspring relatedness rose from 0.17-0.20 to 0.41-0.49. We noted that applying a MAF=10% allowed to get relatively accurate relationship coefficients (median ~ 0.497-0.498) very similar to those obtained using the high coverage dataset after the same filtering process (figure 3e). Modifying the MAF threshold from 1 to 10% led to more changes in relationships than with the high-coverage dataset (Pearson r =

0.969, P<10^{-16} for low coverage dataset *versus* r = 0.994, P<10^{-16} for high coverage dataset) and relatedness coefficients >0.1 tended to be higher with MAF = 10% than with MAF = 1% (Figure S3a). Lastly, the full dataset tended to generate higher relatedness coefficients than that the low-coverage dataset (Wilcoxon signed-rank test, P<10^{-7}, Figure S3b), probably because of a much lower genotyping error.

3.3. GRM structure and GRM-based heritability

The GRM mostly included unrelated individuals (Figure 4b in yellow). However, on closer inspection, the histogram of relatedness indicated the presence of related individuals made up of parent-offspring or full-sib links and a number of half-sib-like links with relatedness around ~0.25. Estimated heritability of body mass (h^2) ranged from 0.59 (se=0.14) ('Default' *stacks* model, LM = 10%, MAF = 10%) to 0.70 (se = 0.15) ('Default' model, LM = 30%, MAF = 1%) (see Table S1 for the full list of h^2). The lowest h^2 values were obtained using MAF = 10% for fixed LM (figure 5a) and LM = 10% for fixed MAF (Figure S4a) which correponds to the filtering parameters that yielded the smallest number of SNPs. As expected, h^2 gradually increased with the number of markers up to an asymptote (here around 7000-8000 SNPs) (Figure S5). However, even when the number of loci used to build the GRM was fixed, h^2 tended to be lower with a stringent MAF threshold (MAF = 10%) (Figure 5b).

When we built the GRM from the low-coverage dataset using the same filtering procedure as with the full dataset (HWE, LM = 20% and MAF = 1%), we obtained h^2 estimates that ranged from 0.39 (se = 0.13) (S2 'MinError' *Stacks* models, 212 individuals, 599 loci) to 0.67 (se=0.16) (S1 'MaxLoci' *Stacks* model, 240 individuals, 10,876 loci) (Table S2). In this case, the 'MinError' *Stacks* model with a very high –*m* value (11) was not suitable as it removed too many loci and individuals (individuals with

>50% missing loci). However, the three other models provided h^2 estimates that were similar to those produced from the full dataset, despite a lower number of loci (5,302 for S3 'Intermediate', 8,646 for S4 'default') and higher genotyping error rate (see above).

4. Discussion

Our results illustrate how RAD-sequencing data can be used to obtain relatively robust estimates of heritability in a free-ranging population for which no genomic resources are available. Our estimates of body mass heritability are consistent with those obtained from a pedigree on other roe deer populations (h^2=0.43-0.65, Quéméré et al., 2018) and other closely related deer species (e.g. 0.54-0.68 in white-tailed deer,Williams, Krueger, & Harmel, 1994). We showed that this process is not straightforward and that critical decisions must be taken during the sampling design and the computational process to optimize the quality and quantity of the genomic data. We showed that the choice of sequencing depth and of the values for bioinformatic parameters (calling and filtering genomic markers) markedly affect the rates of missing data and genotyping error, ultimately impacting the accuracy of genome-wide relatedness estimates. However, in our case study, these decisions had little impact on h^2 estimates, provided that the GRMs were built from a minimum number of loci (above 7000-8000 loci). This suggests that GRM-based heritability estimates are relatively robust to genotyping error and missing data in the SNP dataset. One important exception is the removal of rare variants that led to lower h^2 estimates when a high MAF threshold was used (MAF = 10%). We hereafter discuss the implications of our results and provide some practical recommendations for each step of the analytical process.

4.1 Sequencing design: target a high average sequencing coverage

The sequencing strategy, and especially read depth, may have a significant impact on the quality and quantity of the genomic data, especially when loci are built *de novo* i.e. without a reference genome (Catchen et al., 2013). The *in-silico* digestion of a related reference genome (red deer) followed by a pilot study on a small subset of samples enabled us to optimize the genotyping of around 70,000 loci with a mean coverage of 60 reads per locus (60x). Targeting such high sequencing coverage has multiple benefits. First, it allows maximization of the number of genotyped individuals. Indeed, it is frequent to observe among-sample variation in DNA quality and quantity during the RAD library preparation that may translate into variation in sequencing depth. When using a low average sequencing depth, individuals that are not sufficiently covered may be filtered out because they do not share enough loci with other individuals. Second, using a high average sequencing depth allows optimization of the number of loci and considerably reduces the rate of missing data and genotyping error. In our case, we obtained very low error rates with the 60x coverage dataset (LE_R and AE_R ~ 2-3%, before filtering), well below the values obtained when genotyping was performed on a subset of our sequencing data (coverage ~20x, LE_R > 5% and AE_R ~2-8%). These error rates were also lower than those obtained by Mastretta-Yanes et al., 2015 *in Betula alpina* with a coverage ~ 10x (where LE_R >10% and AE_R > 5% regardless of assembly parameters). At the end of the SNP calling and filtering optimization procedure (using a MAF=1% and a LM=20%), we obtained nearly twice the number of SNPs with the 60x-coverage than with the 20x-coverage dataset (15,930 *versus* 8,809) with far fewer missing data (4.7% *versus* 7%). Several genotype imputation methods have been developed for handling missing data (Das et al., 2016; Nielsen, Paul, Albrechtsen, & Song, 2011), but incorrect SNP calling and allelic dropout are hard to detect and may have a strong impact on downstream population genetics and biological inferences (e.g. parentage assignment in Fountain et al.,2016; demographic inference in Shafer et al., 2017). In the past, because of budget limitations, sequencing coverage was often traded-off against increasing the number of individuals and loci genotyped (Mastretta-Yanes et al., 2015). However, the continuous improvement of Next

Generation Sequencing technology has led to a rapid decrease in sequencing costs so that it is now possible to genotype tens of thousands of loci with >50x coverage for less than 50$ per sample (including library preparation cost). Hence, we strongly encourage targeting a high sequencing depth to maximize the quality and completeness of genomic datasets.

4.2. Impact of SNP genotyping and filtering on GRM and GRM-based heritability

Calling SNPs from RAD-seq data without using a reference genome to guide the assembly of loci is not a simple task since it requires the researcher to select the values for the bioinformatic parameters that determine how closely the sequences must match and the minimum coverage to identify true loci and alleles (Catchen et al., 2013; Paris et al., 2017). Our work illustrates the importance of performing a sensitivity analysis based on technical replicates to help select the optimal parameters and understand the limitations of the dataset (Flanagan, Forester, Latch, Aitken, & Hoban, 2018; Mastretta-Yanes et al., 2015). In our case, the high sequencing depth allowed us to design an assembly model with a relatively high minimum coverage threshold (7 reads) to recover allelic variants. This enabled us to eliminate most sequencing errors (reduced to <2%) while retaining a high number of loci with a minimum of missing data. However, for this dataset, genotyping error rates remained low (<5%) irrespective of the assembly parameters so that the SNP calling procedure ultimately had little impact on the mean and standard error of the body mass heritability estimate.

Additional quality control filtering steps are generally performed to further clean the dataset of uninformative markers and statistical artefacts based on loci missingness ratio, minor allele frequency and/or loci deviating from Hardy-Weinberg proportions (Huang & Knowles, 2014). The relevance of such filters is increasingly questioned, particularly when the final purpose is to estimate genomic relationships (Eynard, Windig, Leroy, Van Binsbergen, & Calus, 2015) or detect loci

potentially under selection (Benestan et al., 2016; Roesti, Salzburger, & Berner, 2012; Waples, 2014). We explored the impact of each of these filters on the number of loci, error rates, GRM structure and GRM-based heritability. We found that the "HWE filter" had little impact on error rates, but led to a substantial decrease in the number SNPs and in the variance in relatedness of the GRM, two parameters that are critical for the estimation of GRM-based heritability (Bérénos et al., 2014). This filtering step is applied to remove sequencing or SNP calling error when the population is assumed to be panmictic. However, genotyping errors generally lead to only slight departures from HW proportions (Cox and Kraft 2006) in contrast to factors such as selection, age structure or non-random sampling that are often neglected (see Waples et al 2014 for a review). One can question the relevance of a filter that has little effect on the quality of the dataset but removes numerous loci that are potentially important for downstream analyses. In contrast, the "Loci missingness" filter (LM) appeared to have a significant impact on locus error rate that was reduced five-fold when excluding loci shared by less than 80% of individuals. Filtering unshared loci thus appears important to remove erroneous loci that are based on artefactual sequences during *de novo loci* assembly.

Lastly, as expected, we found that removing rare variants decreased allele error rate, particularly when a high MAF threshold was used (MAF < 10%). However, such a stringent MAF threshold also significantly reduced the number of loci (figure 3a) and may lead to bias in the allele frequency spectrum. One of the major advantages of RAD-seq genotyping is that it captures rare variants that are often not covered by SNP chips in appropriate proportions (Eynard et al., 2015). Previous work has shown that imposing a strict MAF filter may significantly affect the estimate of relatedness coefficients (Eynard et al. 2015), measures of genomic differentiation among populations (F_{ST}, Hendricks et al., 2018) and may lead to inaccurate demographic inference (Nielsen et al., 2011). Removing rare alleles from data sets may also impede our ability to detect fine scale patterns of connectivity and local adaptation (O'Leary, Puritz, Willis, Hollenbeck, & Portnoy, 2018). Our results

suggest that the MAF threshold has a direct impact on the GRM structure: applying a stringent MAF filter led to higher, and probably more accurate, pairwise relatedness coefficients among the most closely related individuals with mother-offspring relatedness very close to expectation (i.e. 0.5). More broadly, more stringent MAF filtering increased the variance in relatedness of the GRM, probably because most remaining genotyping errors were eliminated. However, the MAF threshold also had an impact on the h^2 estimate: h^2 tended to decrease when increasing the MAF threshold, even when the number of SNPs used to build the GRMs was fixed. This suggests that rare variants of functional importance that might play a key role in segregating weakly related individuals were excluded by mistake in an attempt to remove genotyping errors. This is in agreement with the recent work of Maroulie et al. (2017), who showed that low-frequency variants may individually have greater influence than common variants on adult height in humans.

Together, our results suggest that it is preferable to start with a high sequencing depth and the corresponding SNP calling procedure to obtain a high quality genomic dataset with few missing data rather than applying a stringent filtering procedure to offset low starting coverage (Catchen et al., 2013). A primary reason is that strict MAF or LM filters may lead to a sharp decline in the number of markers and so in our ability to capture the genetic variance of traits. The second is that it is particularly difficult to discriminate sequencing errors from low-frequency variants that may contribute to trait heritability.

4.3. Limits

While encouraging, our study also identified several limits to this approach. A first concern is that while the number of samples used here appears sufficient to detect high heritability (h^2>0.50), it is probably not adequate to detect low heritability values or to estimate genetic covariance between

traits (Perrier et al., 2018). We performed a power analysis using the GCTA-GREML power calculator (Visscher et al., 2014) (Figure S6). Given our number of samples (N=243) and the observed variance of the GRM ($V_{GRM\ =}\ 1.5 \times 10^{-3}$), there is only a 38% chance of detecting a heritability that is lower or equal to 0.25. In our case, a minimum of 500 samples would be required to generate a power of 92% for detecting values of heritability this low. Another consequence is that our standard errors (SE) are relatively large (~0.15) and this might have masked slight differences in h^2 estimates related to SNP calling and filtering setting. It should be noted, however, that our SE are similar to those obtained when body mass h^2 is estimated from a pedigree (SE=0.11-0.24 in Bighorn sheep, Réale et al., 1999; SE=0.16 in roe deer, Quéméré et al., 2018).

Another important issue is that the variance of the GRM is inversely proportional to the sampling variance of the estimate of h^2 (Visscher & Goddard, 2015). Scientists must design their sampling scheme to capture the maximum variance in relatedness in their study population which is partly dependent on the social and spatial structure of the population (Flanagan et al., 2018). In our case study, the GRM mainly included distantly related individuals, reflecting the complex dynamics of the study population which has a high turn-over due to heavy hunting pressure and high dispersal rates. The GRM also captured family structure generated by highly related individuals establishing their home range close to each other within sampling sites. Values for V_{GRM} (1.5×10^{-3}) are within the same range as those reported in Soay sheep (1.3×10^{-3}) (Bérénos et al., 2014) and blue tits (4×10^{-3}) (Perrier C. pers. obs), but one hundred times higher than between unrelated humans (Vinkhuyzen, Wray, Yang, Goddard, & Visscher, 2013) (2×10^{-5}).

Another caveat is the extent of linkage disequilibrium (LD) that dictates how SNP loci tag causal loci and, thus, capture the genetic variance (Bérénos et al., 2014). Since the study population colonized the region only a few decades ago, we expected a low historical effective population size and, thus, a relatively high LD, as observed in Soay sheep (Bérénos et al., 2014). In accordance, we showed that a relatively modest number of markers (around 8000) appears to be enough to capture the genetic variance of the population. Hence, closely-linked SNPs that provide redundant information should be removed. However, without access to a reference genome sequence to physically map loci, we had no *a priori* knowledge of the genetic distance between SNPs. Estimating LD between each pair of SNPs could be an alternative option to prune statistically linked SNPs, however, this is computationally intensive. Moreover, pruning SNPs based solely on LD values, without any knowledge on their physical genomic distance, might be particularly risky since patterns of LD may also reflect the past demographic history of the population and/or effects of selection. Here, in order to reduce the aforementioned redundancy, we retained only one SNP per locus, but we did not filter for additional LD. Given the relatively modest number of RAD loci compared to the genome size, we also hypothesize a moderate representation of pairs of loci with high LD. Furthermore, Yang et al., (2015) showed that there is only a limited bias in h^2 due to heterogeneity in LD across the genome. Nevertheless, further analysis may be required to explore how h^2 estimates may be biased by LD among SNPs in our study system. In particular, simulation of both genomic (mimicking a RAD-sequencing procedure with various coverage/genotyping error rates) and phenotypic data can help evaluate the impact of sequencing design, SNP calling and filtering setting on the accuracy of h^2 estimates in the ideal case where the true heritability is known.

Lastly, a downside of pedigree-free estimation of quantitative genetic parameters is that partitioning of other variance components, such as maternal effects, is not possible without additional information. Here, heritabilities were estimated and interpreted for comparative purposes only,

hence, we deliberately did not account for potential confounding environmental sources of similarity among individuals (e.g. maternal effect, habitat type, cohort effect). Hence, it is very likely that our h^2 estimates are biased upwards. This issue may be partly resolved by the recent development of methods to infer relationships between pairs of individuals (e.g. parent-offspring, full sibs, half-sibs) from SNP data (K. R. Andrews et al., 2018; Huisman, 2017).

4.4. Conclusion and prospects

Our study illustrates the huge potential of genomic-based relatedness for estimating quantitative genetic parameters in free-ranging populations (see also Bérénos et al., 2014; Malenfant, Davis, Richardson, Lunn, & Coltman, 2018; Perrier et al., 2018). Here, we showed that robust heritability estimates can be obtained from RAD-sequencing data in populations or species for which no genomic resources are available. This opens up new and exciting avenues in evolutionary biology, for example, by providing the opportunity to explore how the evolutionary potential of morphological, behavioural or life-history traits varies across space or time in virtually any species (Gienapp et al., 2017). This also paves the way towards more comparative/community quantitative genetic studies, for example, to explore how the heritability of traits varies among populations across environmental gradients (Martinez-Padilla et al 2017) or to better understand constraints on the evolution of phenotypic variation in several species interacting within a given ecosystem (Whitham et al., 2006).

References

Andrews, K. R., Adams, J. R., Cassirer, E. F., Plowright, R. K., Gardner, C., Dwire, M., … Waits, L. P. (2018). A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RAD seq data. *Molecular Ecology Resources*.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81.

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*.

Attard, C. R., Beheregaray, L. B., & Möller, L. M. (2018). Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias. *Molecular Ecology Resources*, *18*(3), 381–390.

Bana, N. Á., Nyiri, A., Nagy, J., Frank, K., Nagy, T., Stéger, V., … Horn, P. (2018). The red deer Cervus elaphus genome CerEla1. 0: sequencing, annotating, genes, and chromosomes. *Molecular Genetics and Genomics*, *293*(3), 665–684.

Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J., Baums, I. B., … Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, *25*(13), 2967–2977.

Bérénos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology*, *23*(14), 3434–3451.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi:10.1111/mec.12354

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365.

Coulon, A., Guillot, G., Cosson, J. F., Angibault, J. M. A., Aulagnier, S., Cargnelutti, B., … Hewison, A. J. M. (2006). Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, *15*(6), 1669–1679. doi:10.1111/j.1365-294X.2006.02861.x

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., … McGue, M. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284.

Debeffe, L., Morellet, N., Cargnelutti, B., Lourtet, B., Bon, R., Gaillard, J. M., … Fryxell, J. (2012). Condition-dependent natal dispersal in a large herbivore: heavier animals show a greater propensity to disperse and travel further. *J Anim Ecol*. doi:10.1111/j.1365-2656.2012.02014.x

Dodds, K. G., McEwan, J. C., Brauning, R., Anderson, R. M., Stijn, T. C., Kristjánsson, T., & Clarke, S. M. (2015). Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genomics*, *16*(1), 1047.

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849.

Ekblom, R., & Wolf, J. B. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042.

Eynard, S. E., Windig, J. J., Leroy, G., Van Binsbergen, R., & Calus, M. P. (2015). The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genetics*, *16*(1), 24.

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, *11*(7), 1035–1052.

Fountain, E. D., Pauli, J. N., Reid, B. N., Palsbøll, P. J., & Peery, M. Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Molecular Ecology Resources*, *16*(4), 966–978.

Gaillard, J. M., Delorme, D., & Jullien, J. M. (1993). Effects of cohort, sex, and birth date on body development of roe deer (Capreolus capreolus) fawns. *Oecologia*, *94*(1), 57–61.

Gienapp, P., Fior, S., Guillaume, F., Lasky, J. R., Sork, V. L., & Csilléry, K. (2017). Genomic quantitative genetics to study evolution in the wild. *Trends in Ecology & Evolution*.

Gilbert, C., Ropiquet, A., & Hassanin, A. (2006). Mitochondrial and nuclear phylogenies of Cervidae (Mammalia, Ruminantia): systematics, morphology, and biogeography. *Molecular Phylogenetics and Evolution*, *40*(1), 101–117.

Hamel, S., Gaillard, J.-M., Festa-Bianchet, M., & Côté, S. D. (2009). Individual quality, early-life conditions, and reproductive success in contrasted populations of large herbivores. *Ecology*, *90*(7), 1981–1995.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., … Koop, B. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197–1211.

Huang, H., & Knowles, L. L. (2014). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, *65*(3), 357–365.

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, *17*(5), 1009–1024.

Lepais, O., & Weir, J. T. (2014). Sim RAD: an R package for simulation-based prediction of the number of loci expected in RAD seq and similar genotyping by sequencing approaches. *Molecular Ecology Resources*, *14*(6), 1314–1321.

Malenfant, R. M., Davis, C. S., Richardson, E. S., Lunn, N. J., & Coltman, D. W. (2018). Heritability of body size in the polar bears of Western Hudson Bay. *Molecular Ecology Resources*.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., … Highland, H. M. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, *542*(7640), 186–190.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*(1), 28–41.

Meyer, K., & Hill, W. G. (1992). Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. *Journal of Animal Breeding and Genetics*, *109*(1–6), 264–280.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443.

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*(16), 3193–3206.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373.

Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proc Biol Sci*, *275*(1635), 613–21. doi:10.1098/rspb.2007.1531

Perrier, C., Delahaie, B., & Charmantier, A. (2018). Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. *Molecular Ecology Resources*.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest radseq: An inexpensive method for de novo snp discovery and genotyping in model and non-model species. *Plos One*, *7*(5), e37135.

Postma, E. (2014). Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates. *Quantitative Genetics in the Wild*, 16–33.

Quéméré, E., Gaillard, J. M., Galan, M., Vanpé, C., David, I., Pellerin, M., … Pemberton, J. M. (2018). Between-population differences in the genetic and maternal components of body mass in roe deer. *BMC Evolutionary Biology*, *18*(1), 39.

Réale, D., Festa-Bianchet, M., & Jorgenson, J. T. (1999). Heritability of body mass varies with age and season in wild bighorn sheep. *Heredity*, *83*(5), 526–532.

Robinson, M. R., Santure, A. W., DeCauwer, I., Sheldon, B. C., & Slate, J. (2013). Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, *22*(15), 3963–3980.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640.

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*(1), 94.

Shafer, A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917.

Stanton-Geddes, J., Yoder, J. B., Briskine, R., Young, N. D., & Tiffin, P. (2013). Estimating heritability using genomic data. *Methods in Ecology and Evolution*, *4*(12), 1151–1158.

Vinkhuyzen, A. A., Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2013). Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annual Review of Genetics*, *47*, 75–95.

Visscher, P. M., & Goddard, M. E. (2015). A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics*, *199*(1), 223–232.

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A., Chen, G.-B., Lee, S. H., Wray, N. R., … Yang, J. (2014). Statistical power to detect genetic (co) variance of complex traits using SNP data in unrelated samples. *PLoS Genetics*, *10*(4), e1004269.

Waples, R. S. (2014). Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity*, *106*(1), 1–19.

Whitham, T. G., Bailey, J. K., Schweitzer, J. A., Shuster, S. M., Bangert, R. K., LeRoy, C. J., … Potts, B. M. (2006). A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics*, *7*(7), 510.

Williams, J. D., Krueger, W. F., & Harmel, D. H. (1994). Heritabilities for antler characteristics and body weight in yearling white-tailed deer. *Heredity*, *73*(1), 78–83.

Wilson, A. J., Reale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., … Nussey, D. H. (2010). An ecologist's guide to the animal model. *J Anim Ecol*, *79*(1), 13–26. doi:10.1111/j.1365-2656.2009.01639.x

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., … Hayes, M. G. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, *43*(6), 519.

Data accessibility

Raw data was deposited at NCBI *SRA* database under accession PRJNA533008, the Stacks output (VCF file) generated from the high-coverage dataset using the optimal "Intermediate" model and script procedures for quality filtering and heritability analysis is available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.n6d4mm5

Conflict of interest

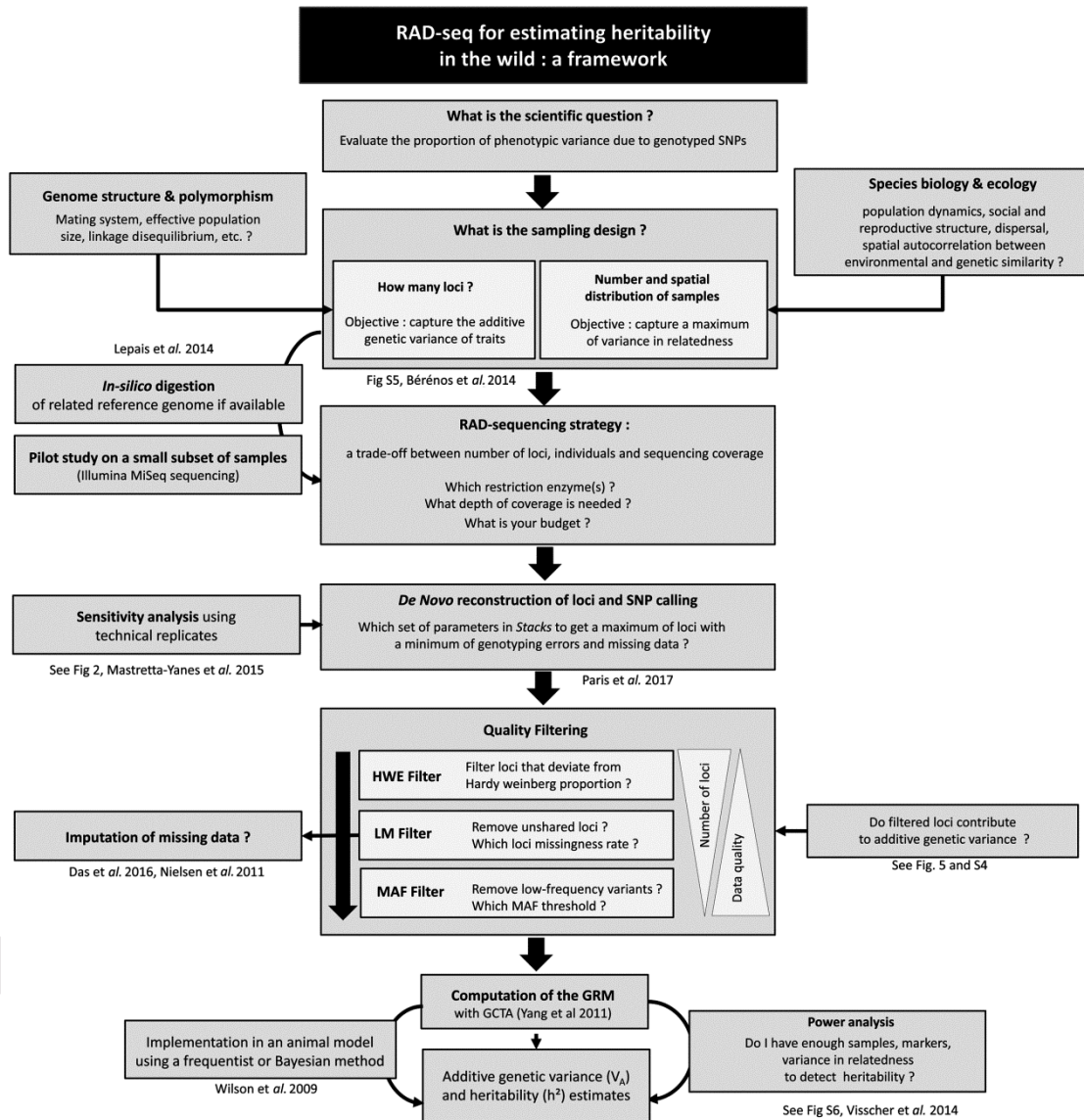The authors declare no conflict of interest

Author Contributions

LG, EQ, BP, JP designed the study and interpreted the data; JM and EQ carried out the DNA extraction and quality controls; LG and MB performed the analyses; LG and EQ led the writing, all authors contributed to the manuscript and gave final approval for publication.

Figures (with captions)

Figure 1. Practical framework with steps for estimating a GRM-based heritability from RAD-seq data. Throughout the process (sampling design, SNP calling, quality filtering), there are trade-offs between data quality (choice of sequencing coverage, genotyping error/ missing data rates) and quantity (number of samples and loci).


RAD-seq for estimating heritability in the wild : a framework

What is the scientific question ?
Evaluate the proportion of phenotypic variance due to genotyped SNPs

Genome structure & polymorphism
Mating system, effective population size, linkage disequilibrium, etc. ?

Species biology & ecology
population dynamics, social and reproductive structure, dispersal, spatial autocorrelation between environmental and genetic similarity ?

What is the sampling design ?

How many loci ?
Objective : capture the additive genetic variance of traits

Number and spatial distribution of samples
Objective : capture a maximum of variance in relatedness

Lepais et al. 2014

Fig S5, Bérénos et al. 2014

***In-silico* digestion**
of related reference genome if available

Pilot study on a small subset of samples
(Illumina MiSeq sequencing)

RAD-sequencing strategy :
a trade-off between number of loci, individuals and sequencing coverage

Which restriction enzyme(s) ?
What depth of coverage is needed ?
What is your budget ?

Sensitivity analysis using technical replicates

See Fig 2, Mastretta-Yanes et al. 2015

***De Novo* reconstruction of loci and SNP calling**
Which set of parameters in *Stacks* to get a maximum of loci with a minimum of genotyping errors and missing data ?

Paris et al. 2017

Quality Filtering

HWE Filter — Filter loci that deviate from Hardy weinberg proportion ?

Imputation of missing data ?

LM Filter — Remove unshared loci ? Which loci missingness rate ?

MAF Filter — Remove low-frequency variants ? Which MAF threshold ?

Number of loci

Data quality

Do filtered loci contribute to additive genetic variance ?
See Fig. 5 and S4

Das et al. 2016, Nielsen et al. 2011

Computation of the GRM
with GCTA (Yang et al 2011)

Implementation in an animal model using a frequentist or Bayesian method

Wilson et al. 2009

Additive genetic variance (V_A) and heritability (h^2) estimates

Power analysis
Do I have enough samples, markers, variance in relatedness to detect heritability ?

See Fig S6, Visscher et al. 2014

Figure 2. Effect of the different sets of parameters in *Stacks* on allele and locus error rates for the high-coverage dataset. For each *Stacks* profile and each pair or technical replicates, we computed the locus error rate (LE_R) corresponding to the number of loci present in only one of the two replicates, divided by the total number of loci being compared (i.e. missing data at the locus level) and the allele error rate (AE_R), calculated as the number of incongruent genotypes between the two replicates, divided by the number of common loci. For each *Stacks* model, we reported the number of loci recovered and the LE_R and AE_R distribution across the 7 replicate pairs.
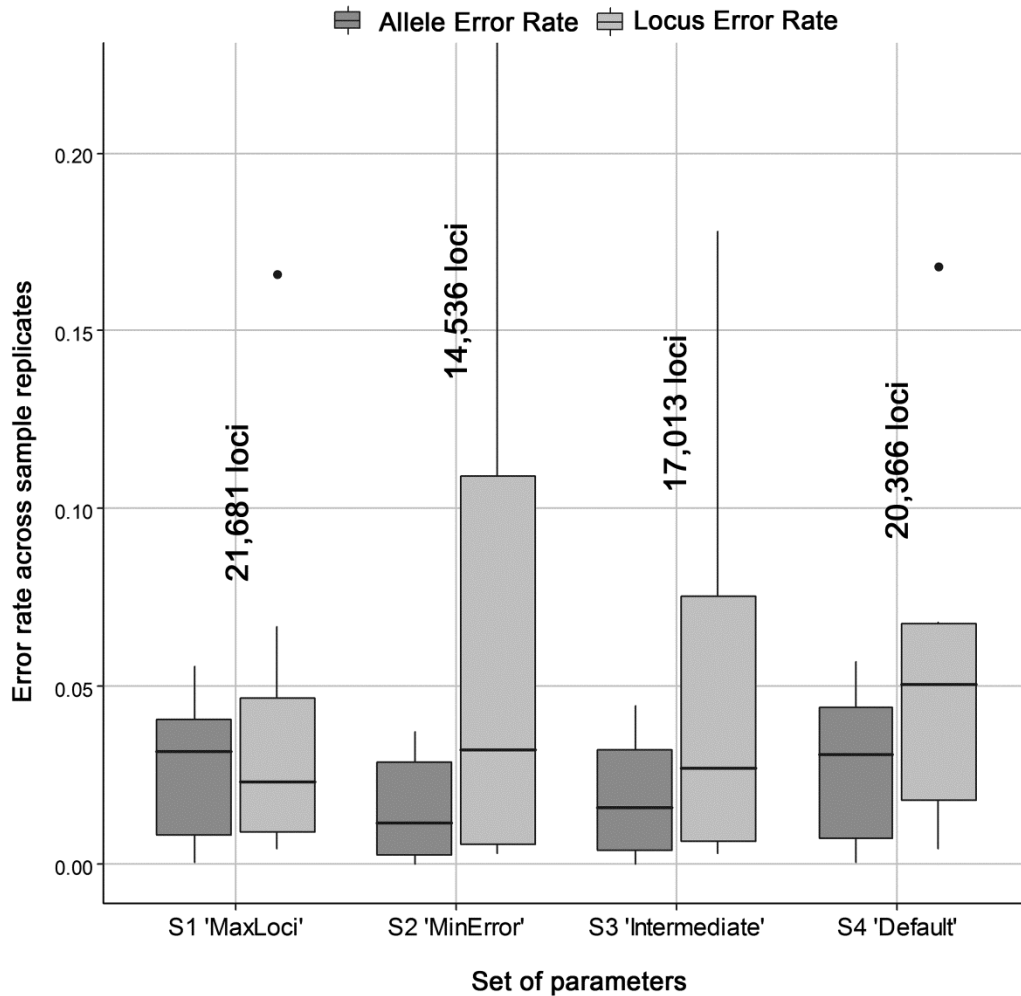
Figure 3. Impact of quality filtering on number of markers, genotyping error rates and GRM. Evolution of the (a) number of SNPs, (b) median locus error rate, (c) median allele error rates (across replicates), (d) the variance in relatedness of the GRM (V_{GRM}) and (e) mother-offspring relatedness, as a function of the filters successively applied for the low-coverage (in grey) and high-coverage (in black) dataset: We first removed loci that deviated from Hardy-Weinberg equilibrium (HWE 0.05- p<0.05) (step 1). We then discarded loci with missingness rate LM >10%, 20%, 30% or 40% (step 2). Lastly, we kept loci with a minimum allele frequency (MAF) greater than 1%, 5% or 10% (step 3). The vertical dashed line in (e) indicated the expected relatedness for mother-offspring (0.5).
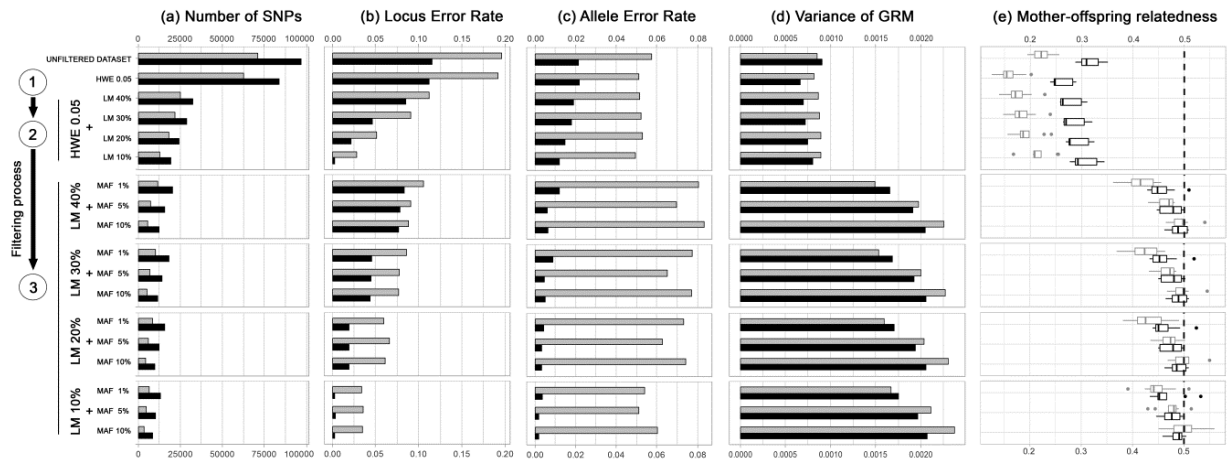
Figure 4. Genomic relatedness. (a) Scatter plot of genome-wide relatedness matrices (GRM) computed before and after filtering SNPS with a MAF < 1% (green dots) or MAF < 10% (blue dots) threshold (from the high-coverage dataset): The highest relatedness coefficients increased when applying a stringent MAF (10%). (b) Histogram of pairwise genome wide relatedness (using MAF=1% and LM=20%). Unrelated individuals were represented in yellow, half-sib-like links in red, and parent-offspring or full-sib relationships in grey.
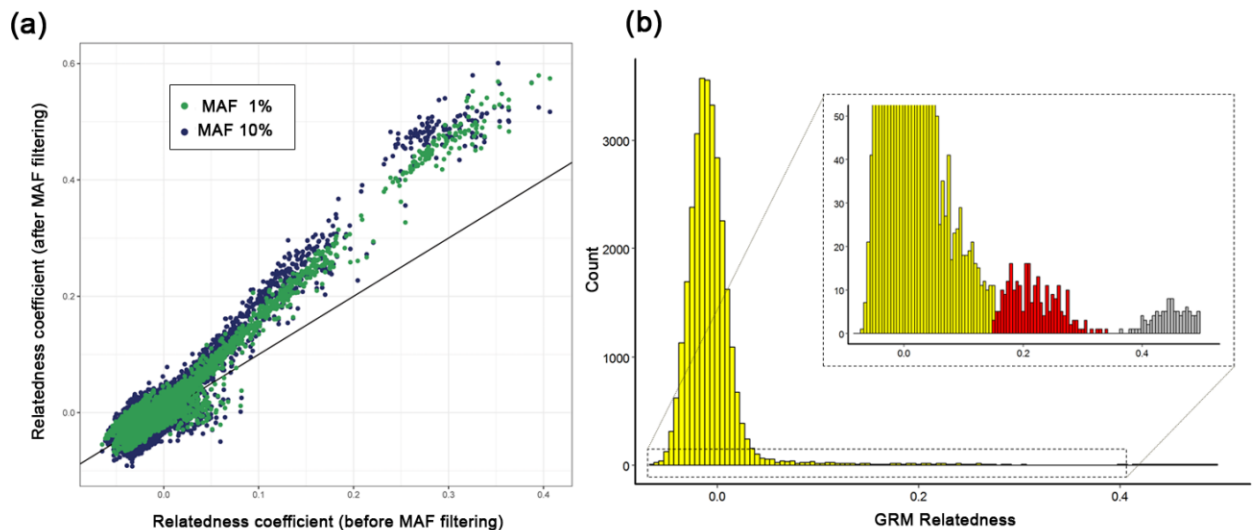
Figure 5. Sensitivity of estimates for *h²* of body mass to SNP calling (*Stacks* model) and MAF threshold (high-coverage dataset): (a) raw *h²* estimates with standard error (error bars) obtained using GRMs computed with all SNPs generated by a given SNP calling (*Stacks* model) and filtering method (MAF = 1%, 5% or 10% and LM = 20%). For each analysis, we reported the number of SNPs used to build the GRMs; (b) *h²* estimates obtained by fixing the number of SNPs to 8,567 SNPs for all GRMs and resampling the SNP data 50 times. Boxplot and whiskers indicate median *h²* and variation across the 50 resampled datasets.