

## RESEARCH ARTICLE

# Automated genotyping of microsatellite loci from feces with high throughput sequences

Isabel Salado<sup>1\*</sup>, Alberto Fernández-Gil<sup>2</sup>, Carles Vilà<sup>1</sup>, Jennifer A. Leonard<sup>1\*</sup>

**1** Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain, **2** Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain

\* [jleonard@ebd.csic.es](mailto:jleonard@ebd.csic.es) (JAL); [isalado@ebd.csic.es](mailto:isalado@ebd.csic.es) (IS)



## Abstract

Ecological and conservation genetic studies often use noninvasive sampling, especially with elusive or endangered species. Because microsatellites are generally short in length, they can be amplified from low quality samples such as feces. Microsatellites are highly polymorphic so few markers are enough for reliable individual identification, kinship determination, or population characterization. However, the genotyping process from feces is expensive and time consuming. Given next-generation sequencing (NGS) and recent software developments, automated microsatellite genotyping from NGS data may now be possible. These software packages infer the genotypes directly from sequence reads, increasing throughput. Here we evaluate the performance of four software packages to genotype microsatellite loci from Iberian wolf (*Canis lupus*) feces using NGS. We initially combined 46 markers in a single multiplex reaction for the first time, of which 19 were included in the final analyses. Megasat was the software that provided genotypes with fewer errors. Coverage over 100X provided little additional information, but a relatively high number of PCR replicates were necessary to obtain a high quality genotype from highly unoptimized, multiplexed reactions (10 replicates for 18 of the 19 loci analyzed here). This could be reduced through optimization. The use of new bioinformatic tools and next-generation sequencing data to genotype these highly informative markers may increase throughput at a reasonable cost and with a smaller amount of laboratory work. Thus, high throughput sequencing approaches could facilitate the use of microsatellites with fecal DNA to address ecological and conservation questions.

## OPEN ACCESS

**Citation:** Salado I, Fernández-Gil A, Vilà C, Leonard JA (2021) Automated genotyping of microsatellite loci from feces with high throughput sequences. PLoS ONE 16(10): e0258906. <https://doi.org/10.1371/journal.pone.0258906>

**Editor:** Trishna Dutta, University of Goettingen, GERMANY

**Received:** June 7, 2021

**Accepted:** October 7, 2021

**Published:** October 25, 2021

**Copyright:** © 2021 Salado et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequencing files are available from the SRA database Bioproject PRJNA702144 accession numbers SAMN17926410 to SAMN17926439.

**Funding:** JAL- Junta de Andalucía Frontera grant P18-FR-5099 [www.juntadeandalucia.es](http://www.juntadeandalucia.es). IS was supported by a PhD fellowship “Ayudas para la formación de profesorado Universitario (FPU)” of the Spanish Ministry of Science, Innovation and Universities (FPU17/02584) [www.ciencia.gob.es/portal/site/MICINN/](http://www.ciencia.gob.es/portal/site/MICINN/). CSIC Open Access Publication Support Initiative through its Unit of Information

## Introduction

Microsatellites are neutral, codominant, highly polymorphic, and abundant genetic markers in eukaryotic genomes [1–3]. Microsatellites are motifs of one to six base pairs repeated in tandem. High polymorphism in these loci yields high information content from only a few markers [4]. These properties have made microsatellites a widely used tool in very different research areas, such as cancer diagnosis [5], human forensics [6], and conservation biology [7]. During the last twenty years, they have also been a powerful and commonly applied tool in molecular ecology for a variety of applications, including individual identification, kinship

Resources for Research (URICI) [www.csic.es](http://www.csic.es). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

determination, and population genetics [1–3, 8, 9]. Furthermore, because microsatellite loci are usually short in length, these PCR-based markers can be amplified from low quality DNA samples (e.g. feces or hair) in noninvasive genetic studies [10] as well as for studies involving ancient DNA or museum specimens [11–13].

Microsatellites are generally genotyped through electrophoresis, in which each targeted marker is amplified by polymerase chain reaction (PCR) and then fragment length is determined by capillary or gel electrophoresis [3]. This method is easy to implement, but the number of samples and loci that can be tested in a single run is limited [4]. In addition, this technology is becoming less common in labs, and more difficult to maintain where it remains. An additional problem is that genotypes are difficult to compare across laboratories, or even projects in the same laboratory, which impedes wider data utilization [4].

Next-generation DNA sequencing (NGS) has the potential to overcome some of these drawbacks. The massive parallel sequencing ability considerably enhances throughput, as hundreds to thousands of reads can be generated for many markers and samples simultaneously [14]. NGS may also increase the accuracy of genotyping, because it enables direct access to the nucleotide sequence of amplified microsatellites. Thus, alleles may be identified using the nucleotide sequence and the length of the microsatellite [14]. The potential of NGS to make microsatellite data more comparable across studies and laboratories would greatly enhance their utility. Genotyping microsatellites with NGS would take advantage of their useful traits such as high polymorphism, reduced ascertainment bias when markers are chosen because of the polymorphism at a given population, and high information content [4], while minimizing some of the problems associated with traditional electrophoresis typing such as lack of comparability of results across studies and low throughput.

Manual genotyping of the great number of sequences generated in a single NGS run is seriously time consuming. However, automatic genotyping may now be possible through bioinformatics pipelines. Several software packages have been specifically designed for genotyping amplicons using NGS data: AmpliSAS [15], Megasat [16], MicNeSs [17], and CHIIMP [18]. These packages examine sequences and apply user-definable decision rules to extract potential alleles while discarding amplification or sequencing artifacts (Table 1, for further details on algorithms or software workflow, see original references). Most of the packages have been tested in tissue samples only, although CHIIMP has been evaluated in fecal samples [18]. Non-invasive genetic samples might be a challenge for automatic genotyping software packages as their low DNA quantity and quality increases dropout and produces non-target amplification, increasing genotyping error rates.

Here we evaluate the ability of several available software packages to genotype microsatellite loci from NGS sequences generated from feces. We compared the performance and ease of use of four software packages for automatic microsatellite genotyping for multiplex-amplified microsatellite loci amplified from fecal DNA of wild Iberian gray wolves (*Canis lupus*). First, we summarize the characteristics of the programs. Second, we evaluated the genotyping success of these packages in terms of genotyping error rates. Third, we assessed the effect depth of coverage has on the genotyping success and estimated the number of reads and replicates needed to reach high confidence in the estimated genotypes.

## Materials and methods

### Materials and lab methods

We collected five apparently fresh Iberian gray wolf (*Canis lupus*) feces in the Cantabrian Mountains of Asturias, northern Spain. The feces samples were collected without any interaction with any animal. We put a small piece of each scat in a 50 ml falcon tube with 70%

**Table 1. Summary of program properties.** Description of programs compared in this study, including an outline of the workflow, default parameters, important features, and brief observations on the ease of use.

Software	AmplISAS	Megasat	MicNeSs	CHIIMP
Operating system	· Web server	· Windows	· Linux	· Windows
	· Linux	· Mac OS* · Linux		· Mac OS* · Linux
GUI†	Yes	Yes	No	No
Programming language	Perl	Perl & R	Python 2.7	R
Workflow	1. <b>Demultiplexing</b> reads into amplicons based on primers sequences.	1. <b>Demultiplexing</b> reads into locus-specific files, based on primer and flanking sequences.	1. <b>Extraction</b> of the microsatellite with the largest repeat number from all reads of all individuals.	1. <b>Sample processing</b> classification of sequences assigning attributes and querying for potential PCR artifacts.
	2. <b>Clustering</b> potential alleles and artifacts are grouped together.	2. <b>Scoring</b> using sequence depth to obtain genotypes.	2. <b>Building</b> observed distribution of the repeat number for each individual.	2. <b>Genotype calling</b> filtering and genotyping only sequences that match locus attributes.
	3. <b>Filtering</b> separation of artifacts from alleles.	3. <b>Plotting</b> histograms of sequence length-frequency distributions for manual verification.	3. <b>Fitting</b> by least squares an equal mixture of two discretized asymmetric Gaussians used for all individuals.	3. <b>Summary and reporting</b> summary of genotypes and quality control tables and graphs.
Genotype calling (algorithm)	All identical reads added to the coverage of the unique variant ('dominant' sequence) and variant freq. calculated (if two highly sequenced variants and similar in sequence, 'subdominant' seq. considered). In clustering, variants are aligned to each other to find seq. errors, erroneous variants (artefacts) are identified and removed (filtering) and coverages added to the true ones, and a consensus sequence is created (allele assignment). Cluster only exact length/in frame can be user-defined.	Based on depth ratios of as many as four of the most common length variants among amplification products relative to the most common length variant (A1). Decision process considers the relative size (> or < A1) and difference in size of putative alleles relative to A1. If the sum of the two most common sequence length variants exceeds the min. read depth (default = 50), Megasat will score the genotypes. Decision variables are user-definable. Recommended to review genotype calls (depth vs size histogram plots).	Assigning a pair of asymmetric Gaussians from repeat number (that represent alleles, each characterized by four parameters: a mode, substitutions(s), a right and left variance) to each individual. Homozygotes have two identical distributions, while for heterozygotes distributions differ, even unequal number of substitutions are considered as different alleles.	Sequences that passes filters (locus' primer, repeat motif and length range) and exceed min. read depth will be genotyped. Only sequences accounting for at least a min. of the filtered reads are considered (5%). Potential stutters, artifacts or ambiguous sequences are excluded. After filters, if only one sequence remains, then sample labelled as homozygous; if two or more, heterozygous. Several quality control tables and graphs are generated for manually review.
Default parameters‡	· Substitution error rate (%) (clustering) = 1 (Illumina)	· No. mismatches (error tolerance to forward and reverse primers and flanking regions) = 2	· No. substitutions = 1	· Min. read depth = 500
	· Indel error rate (%) (clustering) = 0.001 (Illumina)		· Motif size = [2,5]	
	· Min. frequency respect to the dominant seq. (%) (subdominant seq.) (clustering) = 10–25 (Illumina)	· No. threads (multiprocessing) = 1	· Min. no. repeats = 4	· Min. no. repeats = 3
	· Min. amplicon depth (no. reads per amplicon) (filtering) = 100		· Min. read depth = 16	
	· Min. per-amplicon frequency (%) (filtering) = 3	· Min. read depth = 50	· Max. width of the distribution (upper limit for standard deviations) = 5	· Min. fraction retained of the total no. filtered reads (%) = 5
	· Min. chimera length (filtering) = 10		· Max. asymmetry of the distribution (ratio between right and left standard deviation) = 2.5	
· Max. no. alleles (filtering) = 10 (2, our study)				
Previous preprocessing steps§	· Demultiplexing by sample	· Demultiplexing by sample	· Demultiplexing by sample and locus	· Demultiplexing by sample
	· AmpliMERGE		· Format file conversion (fastq -> fasta)	
	· AmpliCLEAN		· Adapter trimming (cutadapt)	

(Continued)

Table 1. (Continued)

Software	AmplisAS	Megasat	MicNeSs	CHIIMP
<b>Input files (format)</b> §	· Primer file (TXT, CSV) · Sequences files (FASTA/FASTQ (R1 & R2 merged))	· Sequence files (FASTA/FASTQ) · Primer file & locus attributes (CSV)	· Sequence files (FASTA)	· Sequence files (FASTA/FASTQ) · Sample attributes (CSV) · Locus attributes (CSV) · known individuals (optional) (CSV) · Named alleles (optional)(CSV)
<b>Output files (format)</b>	· Clustered & filtered sequences (FASTA) · Summary of genotypes (XLS)	· Summary of genotypes (TXT) · Histograms (PDF)	· Summary of genotypes (CSV)¶	· Summary of genotypes (CSV) · Processed files & samples (CSV) · Histograms (PNG) · Allele sequences (FASTA) · Alignments (FASTA) · Alignments (PNG) · Report (HTML)
<b>URL</b>	<a href="http://evobiolab.biol.amu.edu.pl/amplisat/">http://evobiolab.biol.amu.edu.pl/amplisat/</a>	<a href="https://github.com/beiko-lab/MEGASAT">https://github.com/beiko-lab/MEGASAT</a>	<a href="https://bioinfo.mnhn.fr/abi/public/micness/">https://bioinfo.mnhn.fr/abi/public/micness/</a>	<a href="https://github.com/ShawHahnLab/chiimp">https://github.com/ShawHahnLab/chiimp</a>
<b>Reference</b>	15	16	17	18
<b>Available support for questions/bug reports/suggestions</b>	Yes, corresponding authors and Google forum available	Yes, corresponding authors	Not possible to contact with corresponding authors	Yes, corresponding authors
<b>Observations</b>	· Not intuitive output format. · AmpliCHECK did not work · Only possible to change minimum amplicon depth through the command line, not in the web server. · Last program version 1.0 (19-Nov-2018). Last documentation version (24-Jun-2018)	· GUI† does not work properly in Windows 10.  Last program version 1.0 (19-Apr-2017). Last documentation version (Dec-2015)	· It does not work with Python 3 · Genotype is not given in the standardize format (length).  Last program version 1.1 (07-Aug-2015). Last documentation version (11-Aug-2015)	· No need to know R language to run the program, an executable is available after installing required R packages.  Last program version 0.3.1 (31-Jan-2020). Last documentation version (10-Jul-2019)

†GUI: Graphical user interface;

‡ Most important parameters recommended by authors, most of them are user-definable;

§Pre-processing steps and input files used in this study, following the guidelines of authors;

¶ Alleles reported as (mode, substitutions);

\*Operating system not tested in this study.

<https://doi.org/10.1371/journal.pone.0258906.t001>

ethanol. We extracted DNA in an isolated, dedicated laboratory for low quality samples using the QIAmp DNA stool kit (Qiagen) with a negative control. Each extraction and negative was subject to six multiplex PCR reactions (i.e. technical replicates—each reaction was done six times from the same biological sample) including primers to amplify 46 autosomal microsatellites loci previously described in Canids [19–26] (S1 Table in S1 File). This was the first test of these markers as a single multiplex reaction. Reactions contained 1x Phusion Multiplex mix (Thermo Scientific), 0.8 mg/mL bovine serum albumin BSA, 0.05  $\mu$ M of each primer, and 2  $\mu$ L of extract in 20  $\mu$ L and were amplified in a touchdown reaction starting with 98°C for 1 min, then 10 touchdown cycles with 98°C for 10 sec, 67°C to 56°C for 30 sec and 72°C for 30 sec followed by 20 more cycles with the annealing temperature constant at 56°C, ending with a 10 min extension at 72°C and 3 min at 95°C to reduce the formation of primer chains. We checked all products against size standards on agarose gels, and all reactions that yielded product were bead cleaned and dual indexed in a second 12  $\mu$ L PCR with 1x Kapa HiFi mix

(Sigma-Aldrich) and 0.42  $\mu\text{M}$  of each primer (with tail including index and sequencing primer) for 30 sec at 98°C, 12 cycles of 98°C for 10 sec, 60°C for 20 sec and 72°C for 45 sec, and a final 3 min step at 95°C. We checked and quantified products on agarose gels using ImageLab v5.2.1 (BioRad), and pooled equimolar dilutions of all successful PCR reactions for sequencing on a MiSeq (Illumina) with 300 cycles of paired end sequencing at the Johns Hopkins Genetic Resources Core Facility.

### Bioinformatics tools

We independently genotyped microsatellites from the same underlying NGS data with four software packages: AmpliSAS (v1.0) [15], Megasat (v1.0) [16], MicNeSs (v1.0) [17], and CHIIMP (v0.2.2) [18]. CHIIMP was installed and run in Windows 10; MicNeSs, Megasat and AmpliSAS were run in Linux (Ubuntu 18.04.1); and AmpliSAS was also run on the web server (Table 1).

Some pre-processing steps were necessary before performing automatic genotyping (Table 1). Following the AmpliSAT guidelines, we used AmpliMERGE and AmpliCLEAN before AmpliSAS itself to merge paired-end reads and remove reads that did not contain any primer sequences. For MicNeSs, demultiplexing by locus was performed using primer sequences in the “separate by barcode” function in Geneious Prime v2019.1.3 (Biomatters, Auckland, NZ). FASTQ to FASTA format file conversion was also needed for MicNeSs. For CHIIMP, adapters were trimmed using cutadapt [27].

Before the genotyping process, we selected the best microsatellite loci (19 loci), mostly dinucleotides, with alleles ranging in size from 56–172 bp, among all of those initially included for sequencing (46 loci). We discarded microsatellite loci which failed in PCR amplification (i.e. those that did not amplify or only amplified a non-target sequence) in most samples (more than two thirds), loci with very low read depth of the target sequence (less than 16 reads per PCR replicate) in most samples, and compound microsatellites. Compound microsatellites were discarded because not all assessed programs could handle that kind of data—MicNeSs considers only pure repeats (not compound or interrupted).

We used default settings as much as possible to perform genotyping (Table 1). The critical parameter to standardize when comparing programs was the minimum read depth (number of reads) required to call a genotype. Analyses were performed twice for each software: first with the default minimum depth (AmpliSAS = 100 reads; Megasat = 50 reads; MicNeSs = 16 reads; CHIIMP = 500 reads), then with a minimum depth of 16 reads, the lowest default of any of the programs, established by MicNeSs. We used this same, minimum number of reads (16) to make results comparable among software packages. MicNeSs’ genotyping algorithm calculates a Gaussian distribution of the number of repeats that represents each allele, reporting the allele as (*mode*, *s*); being *mode* the mode of the distribution (i.e. largest repeat number) and *s* the number of substitutions in the repeat array [17] (Table 1). To allow comparison with other software packages, we converted genotype results from MicNeSs to allele length (which is the output of the other programs) by considering the number of repeats, the repeat motif and the length of the flanking regions.

### Genotyping error rates

Genotyping of microsatellite loci from fecal DNA is particularly prone to errors [10, 28, 29]. We considered as genotyping error any allelic difference between the genotype obtained from the noninvasive sample using the genotyping software and a reference genotype [30]. Reference genotypes were obtained through consensus of replicates of loci individually scored by the same person (IS) using the lengths graphs function in Geneious Prime (v2019.1.3) on the

same NGS data. A second step in the genotyping process was to obtain a reference genotype for all samples and loci. Reference genotypes from manual allele calling of all data were estimated following the criteria: (1) allele must be present at least in 2 PCRs for heterozygotes and 3 PCRs for homozygotes, and (2) homozygotes must not have same second allele in more than 20% of PCRs (Supplementary material, S1 Appendix in [S1 File](#)). In just two cases the reference consensus genotype was ambiguous and so were not considered for downstream analyses. We considered allelic dropout (ADO) as the failure to amplify one allele in heterozygous individuals [31, 32] and false allele (FA) as an allele-like artifact that is generated by PCR [33]. Equations (2) and (4) from [31] were used to calculate proportions of ADO and FA, respectively.

We defined the following variables for each locus and each software. *Proportion genotyped* was the number of genotypes estimated by a software pipeline ( $N_{genotypes}$ ) divided by the total number of available samples and PCR replicates for the considered locus (excluding the ambiguous genotypes in the reference, see above). *Genotyping success* was the number of genotypes that coincided with the reference ( $N_{successful}$ ) divided by the total number of genotypes inferred by the software per locus across all replicates and samples. *Proportion ADO* was the number of heterozygote genotypes for which only one of the two alleles could be genotyped ( $N_{ADO}$ ) divided by the number of heterozygote genotypes in the reference. *Proportion FA* was the number of genotypes including a false allele ( $N_{FA}$ ) divided by the total number of reference genotypes.

## Statistical analyses

A multidimensional scaling analysis (MDS) was carried out to visualize the similarity between the genotypes generated by the different programs and the reference genotype. A symmetric dissimilarity matrix was obtained from the comparison of consensus genotypes per locus and sample generated by each software when using a minimum read depth of 16 for all programs. We assigned a distance value of 0.5 to ADO and FA (one of the two alleles erroneous), while a distance value of 0 to successful genotypes. Distance matrix included the mean distances values across samples and locus per software (S2 Table in [S1 File](#)). Then we performed a two-dimensional ratio MDS, which does not transform input dissimilarities, using the `mds` function from the package `smacof` [34] in R version 3.5.2.

We assessed the effect of software and coverage on *Proportion genotyped* and *Genotyping success* when using a minimum read depth of 16 with a generalized linear mixed model (GLMM) with a binomial distribution and using the function `glmer` from the `lme4` R package [35]. Two models were established *a priori* with *Proportion genotyped* and *Genotyping success* as dependent variables. *Software*, *Locus* and *Mean read depth* were included as fixed effects while *Sample* was considered as a random factor. For this analysis, *Proportion genotyped* and *Genotyping success* were calculated for each software, locus and sample, excluding the ambiguous genotypes in the reference (see above). Weights argument were included in each model as the number of trials used to generate each proportion. *Software* and *Locus* were factor variables referring to the genotyping programs and the microsatellite loci. *Mean read depth* was calculated as the mean number of reads of the six PCR replicates per sample and locus (S3 Table in [S1 File](#)). *Mean read depth* was centered at the mean and scaled by the standard deviation. The significance of each variable was tested using a Chi-squared test with `drop1` function in `lme4`, which compares the likelihood of the full model including and excluding the variables of interest. When significant differences were found, we also performed pairwise comparisons with Tukey's post hoc test among programs using `emmeans` package [36].

We wrote Python, Bash and R scripts (Supplementary material, S2 Appendix in [S1 File](#)) to calculate the proportion of correct genotypes for the best performing software when varying

the coverage for each PCR product and locus. We used a coverage of 6, 10, 20, 40, 80, 100, 150 and 200 reads. Simulations were run on a dataset with the best quality fecal sample using eleven heterozygous loci with > 200 reads. First, we performed a random subsampling without replacement of reads for each locus and PCR product. Second, those datasets were analyzed using the best performing software. Finally, we evaluated the genotyping success in relation to the number of reads. Simulations were performed with 100 simulation replicates for each locus and PCR product.

We also calculated the number of PCR replicates needed to reach high confidence in a multilocus genotype for the best performing software taking into account the proportion of failed amplifications. As commonly done in noninvasive studies, we considered that a successful genotype must be replicated at least twice if heterozygote and at least three times if homozygote [37]. We first calculated the average proportion of amplifications that resulted in successful genotypes for homozygous and heterozygous loci when using the software with fewer genotyping errors. We then used these proportions to calculate the probability of genotyping success for a single locus after a varying number of replicates using the cumulative probability function of a binomial distribution (probability of obtaining at least 2/3 correct genotypes for a heterozygous/homozygous locus). Finally, we used this value to estimate the probability of correctly genotyping at least 16, 17, 18 and 19 of 19 loci.

## Results

### Sequencing results

A total of 807,084 reads were obtained from the six amplification replicates of the five fecal samples analyzed (mean  $\pm$  se: 26903  $\pm$  42567 reads per PCR replicate). There was high variance in coverage per locus (from mean  $\pm$  se: 1  $\pm$  0 (locus 403) to 4154  $\pm$  2165 reads (locus u213), S1a Fig in S1 File). The best 19 autosomal microsatellite loci were selected for downstream analyses (S1 Table and S1b Fig in S1 File).

### Comparison of software performance: Genotyping error rates

Software pipelines differed in the proportion of genotypes obtained as output (*Proportion genotyped*) and in the proportion of correct genotypes among them (*Genotyping success*) (Tables 2 and 3). *Proportion genotyped* increased when the minimum read depth was decreased to 16 reads in all programs (except for MicNeSs, for which 16 reads is the default minimum read threshold). However, *Genotyping success* declined with this lower depth threshold. With a standardization of 16 reads, *Proportion genotyped* was generally high for all programs, while *Genotyping success* varied widely (Table 3). Megasat was the software with the highest proportion of correct genotypes, although MicNeSs obtained a greater number of genotypes as output.

**Table 2. Genotyping error rates among software packages with default settings.** Default minimum read depth; AmpliSAS, 100 reads; Megasat, 50 reads; CHIIMP, 500 reads; MicNeSs, 16 reads. *Proportion genotyped*, the number of genotypes estimated by each software ( $N_{genotypes}$ ) divided by the total number of reference genotypes per locus across all replicates and samples. *Genotyping success*, the number of genotypes that coincide with the consensus ( $N_{successful}$ ) divided by the total number of genotypes estimated by the software per locus across all replicates and samples ( $N_{genotypes}$ ). *Proportion ADO*, the number of heterozygote genotypes for which only one of the two alleles could be genotyped ( $N_{ADO}$ ) divided by the number of heterozygote genotypes in the reference. *Proportion FA*, the number of genotypes including a false allele ( $N_{FA}$ ) divided by the total number of reference genotypes. Values are shown as mean  $\pm$  standard error per locus.

Software	$N_{genotypes}$	$N_{successful}$	$N_{ADO}$	$N_{FA}$	Proportion genotyped	Genotyping success	Proportion ADO	Proportion FA
AmpliSAS	12 $\pm$ 1	6 $\pm$ 1	0	6 $\pm$ 1	0.40 $\pm$ 0.04	0.47 $\pm$ 0.09	0	0.53 $\pm$ 0.09
CHIIMP	5 $\pm$ 1	4 $\pm$ 1	0	2 $\pm$ 1	0.16 $\pm$ 0.02	0.69 $\pm$ 0.10	0.01 $\pm$ 0.01	0.28 $\pm$ 0.11
Megasat	15 $\pm$ 2	12 $\pm$ 2	2 $\pm$ 1	2 $\pm$ 1	0.50 $\pm$ 0.05	0.74 $\pm$ 0.08	0.11 $\pm$ 0.04	0.15 $\pm$ 0.07
MicNeSs	27 $\pm$ 1	16 $\pm$ 2	3 $\pm$ 1	8 $\pm$ 2	0.91 $\pm$ 0.02	0.59 $\pm$ 0.06	0.19 $\pm$ 0.04	0.31 $\pm$ 0.07

<https://doi.org/10.1371/journal.pone.0258906.t002>

**Table 3. Genotyping error rates among software packages using a minimum read depth of 16.** *Proportion genotyped*, the number of genotypes estimated by each software ( $N_{genotypes}$ ) divided by the total number of reference genotypes per locus across all replicates and samples. *Genotyping success*, the number of genotypes that coincide with the reference ( $N_{successful}$ ) divided by the total number of genotypes estimated by the software per locus across all replicates and samples ( $N_{genotypes}$ ). *Proportion ADO*, the number of heterozygote genotypes for which only one of the two alleles could be genotyped ( $N_{ADO}$ ) divided by the number of heterozygote genotypes in the reference. *Proportion FA*, the number of genotypes including a false allele ( $N_{FA}$ ) divided by the total number of reference genotypes. Values are shown as mean  $\pm$  standard error per locus.

Software	Ngenotypes	Nsuccessful	NADO	NFA	Proportion genotyped	Genotyping success	Proportion ADO	Proportion FA
AmpliSAS	25 $\pm$ 1	9 $\pm$ 2	0	16 $\pm$ 2	0.84 $\pm$ 0.04	0.36 $\pm$ 0.06	0	0.64 $\pm$ 0.06
CHIIMP	25 $\pm$ 1	11 $\pm$ 2	1 $\pm$ 0	12 $\pm$ 2	0.84 $\pm$ 0.03	0.46 $\pm$ 0.07	0.07 $\pm$ 0.02	0.49 $\pm$ 0.06
Megasat	22 $\pm$ 1	16 $\pm$ 2	3 $\pm$ 1	2 $\pm$ 1	0.74 $\pm$ 0.04	0.72 $\pm$ 0.06	0.22 $\pm$ 0.05	0.12 $\pm$ 0.04
MicNeSs	27 $\pm$ 1	16 $\pm$ 2	3 $\pm$ 1	8 $\pm$ 2	0.91 $\pm$ 0.02	0.59 $\pm$ 0.06	0.19 $\pm$ 0.04	0.31 $\pm$ 0.07

<https://doi.org/10.1371/journal.pone.0258906.t003>

Megasat had a slightly higher proportion of ADO compared to the other programs. AmpliSAS and CHIIMP showed a very high proportion of FA, with these being present in half or more of the genotypes (Table 3). The comparison of consensus genotypes of each software showed Megasat results were closest to the reference (Fig 1), and that the errors in the genotypes generated from different software occurred from distinct source of bias, as each of them is distant from the others in the plot.

A GLMM analysis revealed that *Proportion genotyped* was influenced by *Software* (LRT = 72.36,  $p < 0.001$ ,  $N = 371$ ), *Locus* (LRT = 139.83,  $p < 0.001$ ,  $N = 371$ ) and *Mean read depth* (LRT = 18.82;  $p < 0.001$ ,  $N = 371$ ). *Genotyping success* was also influenced by the *Software* (LRT = 172.66,  $p < 0.001$ ,  $N = 367$ ) and *Locus* (LRT = 299.08,  $p < 0.001$ ,  $N = 367$ ), but not by *Mean read depth* (LRT = 0.58,  $p = 0.448$ ,  $N = 367$ ). Post-hoc tests revealed significant pairwise differences among all software packages for *Genotyping success* and for *Proportion genotyped*, except between AmpliSAS and CHIIMP for *Proportion genotyped* (S4 Table in S1 File). Post-hoc tests also showed that, when using the minimum read depth of 16 reads, MicNeSs was the software with the highest *Proportion genotyped*, but Megasat was the software with the highest *Genotyping success* (S4 Table in S1 File). These results suggested that Megasat offered the most consistent results with regard to the reference genotypes generated manually and this program was chosen for subsequent analyses.

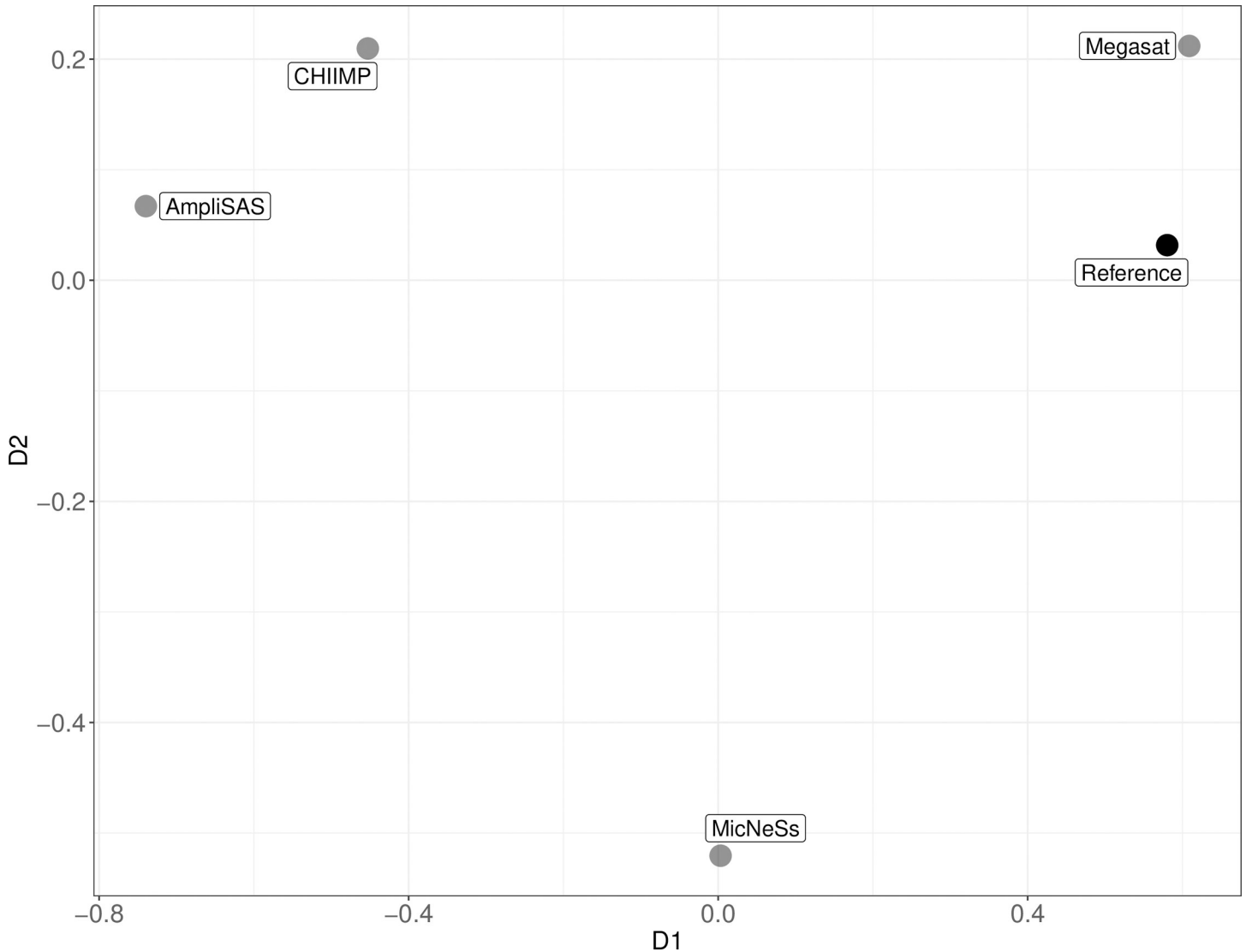
### Effect of coverage and number of replicates on genotyping success

For Megasat, the most reliable software in our analyses, subsampling reads from the genotypes with the largest amounts of data showed that more than about 100 reads per PCR led only to small increases in *Genotyping success* (Fig 2). With 150 reads *Genotyping success* was greater than 0.9 (Fig 2).

Average rates of genotyping success per PCR in Megasat were 0.46 and 0.66 for heterozygotes and homozygotes, respectively, due to the high ADO rate (Table 3), which can only be detected in heterozygotes. In order to obtain the three positive replicates required to confirm a homozygous genotype or two for heterozygotes, at least seven PCR replicates were needed to assess a homozygote, and eight for a heterozygote, with a probability of 0.95 (Fig 3A). Therefore, it is more difficult to assess a correct heterozygous genotype than a homozygous one.

Taking locus dropout into account, we estimated that to obtain a correct genotype with a probability of 0.95 in 18 of 19 homozygous loci in multiplex, we would need at least eight PCR replicates (Fig 3B). For heterozygotes, we would need at least ten replicates (Fig 3C). This number of replicates is panel-specific, and could be decreased if protocols and marker selection were optimized to reduce locus dropout.



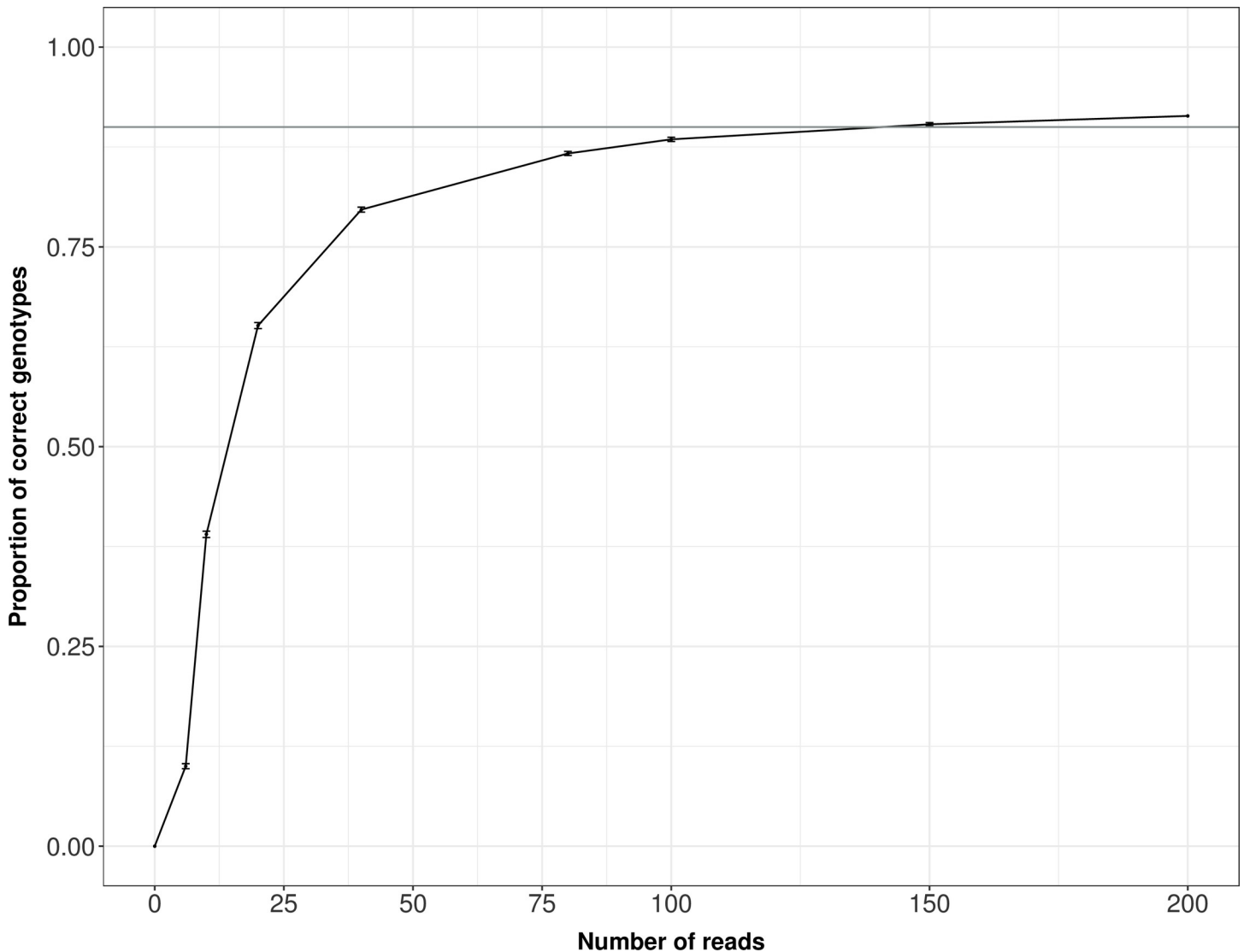


**Fig 1. Multidimensional Scaling (MDS) of genotyping programs.** Axes show distance values of distance matrix. Distance matrix was obtained from the comparison of consensus genotypes of each software and establishing minimum read depth of 16 (S2 Table in S1 File). The closer two programs are, the more similar their results are. Model statistics: two-dimensions ratio MDS using majorization, stress-1 value (normalized): 0.061, number of iterations: 17. *Reference* indicates consensus genotypes used as reference for comparison.

<https://doi.org/10.1371/journal.pone.0258906.g001>

## Discussion

In our evaluation of four different software packages to genotype microsatellites using NGS sequence data, we found differences in performance in terms of proportion of genotypes obtained (*Proportion genotyped*) as well as proportion of correct genotypes (*Genotyping success*). These differences might be due to differences in software design, algorithms and/or filters. While Megasat, MicNeSs and CHIIMP were especially designed for detecting and genotyping microsatellites, AmpliSAS was designed to genotype MHC markers. As MHC markers are generally polygenic in vertebrates [38, 39], AmpliSAS might be biased to call a higher number of heterozygotes, yielding a very low rate of dropout, but a very high rate of false alleles. All packages called substantially higher rates of false alleles than have been reported using traditional microsatellite genotyping of feces via gel electrophoresis [28, 32, 40].

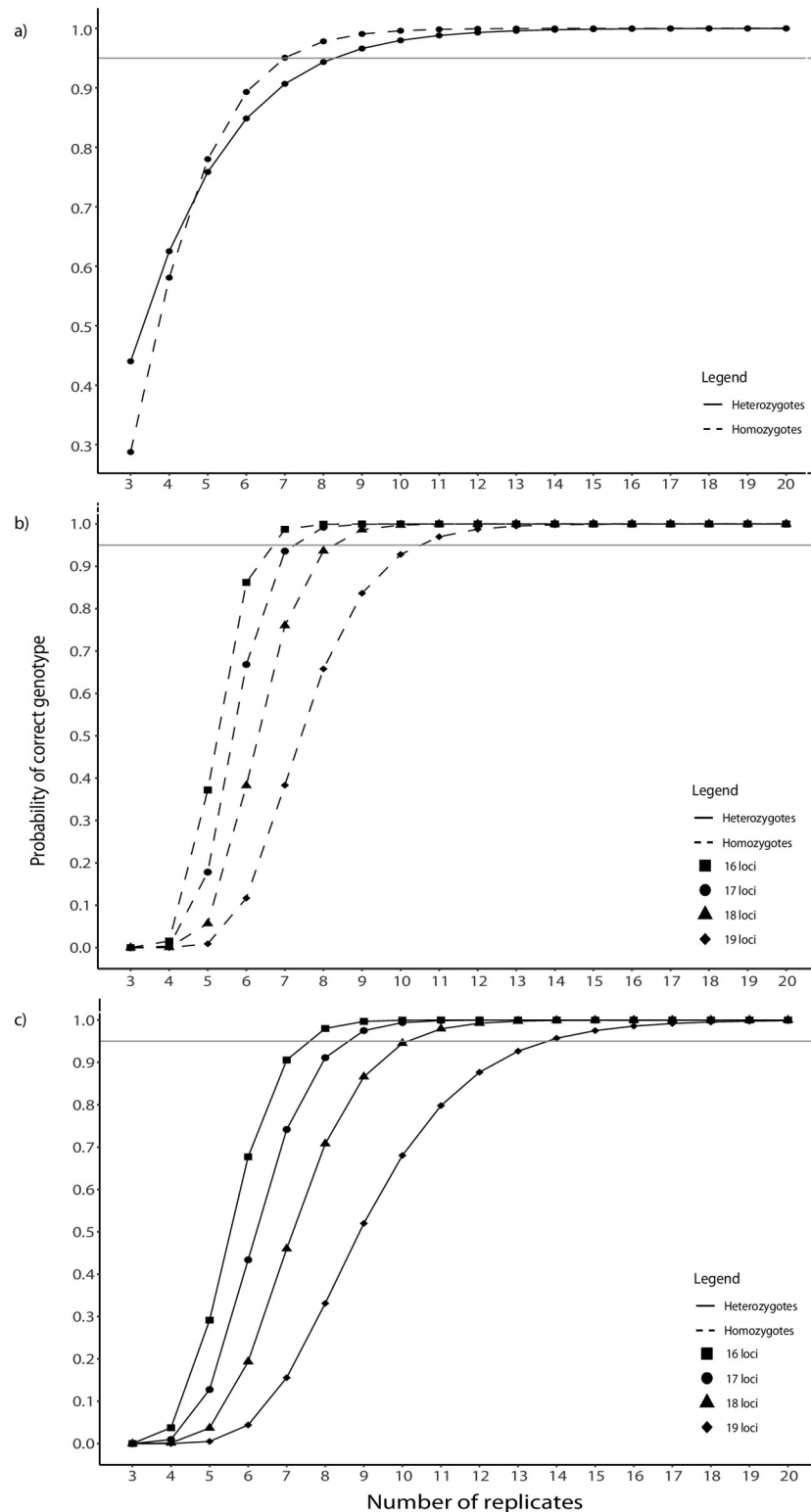


**Fig 2. Relation between genotyping success and read depth per locus in Megasat.** Proportion of correct genotypes using Megasat when varying the sequencing coverage. With 150 reads *Genotyping success* was greater than 0.9 (horizontal line). Simulations were performed with 100 random draws of a given number of reads for each locus and PCR replicate. The average of 100 random draws is represented; error bars indicate standard errors.

<https://doi.org/10.1371/journal.pone.0258906.g002>

It is not clear if this is an artifact of the sequencing or of the downstream analyses, and should be further explored in future studies.

We attempted to type a large number of markers, but less than half of them were included in the final panel. Locus selection and sample quality play a key role in genotyping success (see also [41]). In this case we employed loci that our group had previously used for other studies on wolves using gel electrophoresis (e.g. [13, 42, 43]). Not surprisingly, many loci that had previously worked with traditional methods failed in this project. It could be because they were not compatible with the multiplex (the multiplex used in this study was very much larger than we ever used for gel electrophoresis due to the limits imposed by few available fluorescent dyes as compared to sequences), or that some loci sequenced better than others. Most of the loci considered here were di-nucleotide repeat microsatellites. It is possible that other, perhaps longer, motifs will sequence better because of reduced strand slippage during PCR [14]. When



**Fig 3. Relation between genotyping success and number of replicates using Megasat. Single locus (a):** although the probability of genotyping success is higher for homozygotes (0.66; dashed line) than for heterozygotes (0.46; solid line), seven independent replicates are required to determine a homozygous genotype in a noninvasive sample, while eight were for a heterozygote. **Multiple loci (b, c):** Probability of obtaining the correct genotype for multiple homozygous (b, dashed lines) and heterozygous (c, solid lines) loci with different number of replicates. Considering a total of 19 loci,

probabilities were calculated to obtain a correct genotype for at least 16 (squares), 17 (circles), 18 (triangles) and 19 (diamonds) of the 19 loci. Horizontal line marks probability of 0.95.

<https://doi.org/10.1371/journal.pone.0258906.g003>

selecting loci for NGS it is important not only that they amplify in the target organism, and in multiplex if that strategy is used, but also that they are small enough for the largest alleles to be sequenced through. We sequenced on the Illumina MiSeq platform, so our limit was 300 bp.

Sample quality has an important impact on the rate of both allelic dropout and false alleles [44]. In this study all samples were feces, which are generally low-quality samples. However, there is still a lot of variation in quality across fecal samples. This variation could result from the age of the feces when they were collected, weather, or environmental conditions [45–47]. The standard method to overcome the higher rate of genotyping errors associated with the low quality of DNA extracted from feces in the wild has been replications [28, 32]. In some genomic studies higher coverage per locus is used to support the authenticity of a genotype. However, in studies such as this, where it is PCR products and not genomic DNA that is being sequenced, increasing coverage to very high levels should not be considered a replacement for PCR replications. Our results showed that 100 reads per locus maximized information content from the PCR product (Fig 2).

Once a minimum coverage per locus is reached, our results emphasize the importance of including multiple PCR replicates in the analyses, in a similar way to the traditional multi-tube approach [32]. In our case, a minimum of 8 to 10 replicates would be required to ensure reliable genotyping of 18 out of 19 loci (Fig 3). If high reliability at all 19 loci was needed, 11 to 14 replicates could be needed (Fig 3). However, not all research questions have the same requirements, and the number of replicates for a given study could be planned taking into consideration the goals of the study. While robust pedigree reconstructions will require accurate multilocus genotypes, studies about population structure that depend on allele frequencies may need many fewer replicates. The number of replicates required will also depend on the specific panel of loci, target organism, sample type, and level of optimization. Panels can be optimized to lower locus dropout, which would reduce the number of replicates necessary to obtain a genotype. This was the first test of this multiplex reaction, so the large number of replicates necessary for high confidence genotypes could likely be reduced with optimization.

In our analyses, Megasat was the software with the best performance. We tested both the command line and Graphical User Interface (GUI) for Windows versions. We had substantial trouble with the GUI, and chose to do all analyses using the command line. This could be an obstacle to some users, but the command line version is simple, and no advanced programming skills are required to run the program, although a basic knowledge of UNIX commands is. Of the four programs, only AmpliSAS had an operational and online GUI (Table 1), although occasionally the web server had issues refreshing the dataset or some user-defined parameters. Until GUIs are fully implemented in next versions of these programs, we recommend running them through the command line. For technical issues, besides the support provided by the authors, it would be very useful to include a troubleshooting section in the documentations of the programs and a forum for discussion between authors and users in order to promote new versions of the programs with updated documentations and an active support community. AmpliSAS is the only program that has a forum while CHIIMP is the only software which continues releasing new versions with updated documentation.

## Conclusions

Since the advent of high throughput sequencing technologies, a debate has arisen among conservation geneticists about the use of single nucleotide polymorphisms (SNPs) instead of microsatellites. However, the high information content of microsatellites and their applicability in less

known systems (where polymorphic loci are not known in advance) make microsatellites useful markers to address a wide range of ecological questions [4]. The primary threat to the long-term utility of these loci is the clearly dated nature of the equipment on which they are generally genotyped, and the lack of comparability between genotypes generated in different studies [48]. The sequencing of these loci on high throughput platforms appears to cure both of these ills. In order to apply this new technology to these markers, protocols for each step of the process from the field to the genotype need to be re-evaluated and made compatible. The number of sequences that come out of NGS sequencers is far beyond what can be visually inspected, as traditional microsatellite genotypes were. In our tests of how to best genotype this data, we find that a relatively low coverage of a reasonable number of replicates can yield automated, high confidence genotypes from fecal DNA. Given the capacity of NGS machines, and automated genotyping, the necessary level of replications should be feasible and should facilitate the higher throughput analysis of feces from wildlife for ecological and conservation studies.

## Supporting information

**S1 File.**  
(ZIP)

## Acknowledgments

We are grateful to A. Marín Alguacil for his bioinformatic support in the analyses. Logistical support was provided by Irene Quintanilla, Anna Cornellas, Laboratorio de Ecología Molecular (LEM-EBD) and by the infrastructures offered by Doñana's Singular Scientific-Technical Infrastructure (ICTS-EBD). We thank two anonymous reviewers for their helpful comments.

## Author Contributions

**Conceptualization:** Jennifer A. Leonard.

**Formal analysis:** Isabel Salado, Carles Vilà, Jennifer A. Leonard.

**Investigation:** Alberto Fernández-Gil.

**Project administration:** Jennifer A. Leonard.

**Resources:** Jennifer A. Leonard.

**Writing – original draft:** Isabel Salado, Jennifer A. Leonard.

**Writing – review & editing:** Alberto Fernández-Gil, Carles Vilà.

## References

1. Jarne P, Lagoda P. J. L. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*. 1996; 11: 424–429. [https://doi.org/10.1016/0169-5347\(96\)10049-5](https://doi.org/10.1016/0169-5347(96)10049-5) PMID: 21237902
2. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*. 2004; 5: 435–445. <https://doi.org/10.1038/nrg1348> PMID: 15153996
3. Vieira MLC, Santini L, Diniz AL, Munhoz C. de F. Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*. 2016; 39: 312–328. <https://doi.org/10.1590/1678-4685-GMB-2016-0027> PMID: 27561112
4. Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, et al. Current trends in microsatellite genotyping. *Molecular Ecology Resources*. 2011; 11: 591–611. <https://doi.org/10.1111/j.1755-0998.2011.03014.x> PMID: 21565126
5. Sidransky D. Emerging molecular markers of cancer. *Nature Reviews Cancer*. 2002; 2: 210–219. <https://doi.org/10.1038/nrc755> PMID: 11990857

6. Fordyce S, Ávila-Arcos M, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen F, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques*. 2011; 51: 127–133. <https://doi.org/10.2144/000113721> PMID: 21806557
7. Allendorf FW, Luikart G, Aitken SN. Conservation and the genetics of Populations. 2nd ed. West Sussex, UK: Wiley-Blackwell; 2013.
8. Queller DC, Strassmann JE, Hughes CR. Microsatellites and kinship. *Trends in Ecology & Evolution*. 1993; 8: 285–288. [https://doi.org/10.1016/0169-5347\(93\)90256-O](https://doi.org/10.1016/0169-5347(93)90256-O) PMID: 21236170
9. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*. 2006; 9: 615–629. <https://doi.org/10.1111/j.1461-0248.2006.00889.x> PMID: 16643306
10. Taberlet P, Waits LP, & Luikart G. Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution*. 1999; 14: 323–327. [https://doi.org/10.1016/s0169-5347\(99\)01637-7](https://doi.org/10.1016/s0169-5347(99)01637-7) PMID: 10407432
11. Wandeler P, Hoeck PEA, Keller LF. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*. 2007; 22: 634–642. <https://doi.org/10.1016/j.tree.2007.08.017> PMID: 17988758
12. Leonard JA. Ancient DNA applications for wildlife conservation. *Molecular Ecology*. 2008; 17: 4186–4196. <https://doi.org/10.1111/j.1365-294X.2008.03891.x> PMID: 18684136
13. Koblmüller S, Nord M, Wayne RK, & Leonard JA. Origin and status of the Great Lakes wolf. *Molecular Ecology*. 2009; 18: 2313–2326. <https://doi.org/10.1111/j.1365-294X.2009.04176.x> PMID: 19366404
14. De Barba M, Miquel C, Lobréaux S, Quenette PY, Swenson JE, Taberlet P. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*. 2017; 17: 492–507. <https://doi.org/10.1111/1755-0998.12594> PMID: 27505280
15. Sebastian A, Herdegen M, Migalska M, Radwan J. Amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*. 2016; 16: 498–510. <https://doi.org/10.1111/1755-0998.12453> PMID: 26257385
16. Zhan L, Paterson IG, Fraser BA, Watson B, Bradbury IR, Ravindran Nadukkalam et al. MEGASAT: automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources*. 2017; 17: 247–256. <https://doi.org/10.1111/1755-0998.12561> PMID: 27333119
17. Suez M, Behdenna A, Brouillet S, Graça P, Higuete D, Achaz G. MicNeSs: Genotyping microsatellite loci from a collection of (NGS) reads. *Molecular Ecology Resources*. 2016; 16: 524–533. <https://doi.org/10.1111/1755-0998.12467> PMID: 26400716
18. Barbian HJ, Connell AJ, Avitto AN, Russell RM, Smith AG, Gundlapally MS, et al. CHIIMP: An automated high-throughput microsatellite genotyping platform reveals greater allelic diversity in wild chimpanzees. *Ecology and Evolution*. 2018; 8: 7946–7963. <https://doi.org/10.1002/ece3.4302> PMID: 30250675
19. Ostrander EA, Sprague GF, Rine J. Identification and Characterization of Dinucleotide Repeat (CA)<sub>n</sub> Markers for Genetic Mapping in Dog. *Genomics*. 1993; 16: 207–213. <https://doi.org/10.1006/geno.1993.1160> PMID: 8486359
20. Shibuya H, Collins BK, Huang TH, Johnson GS. A polymorphic (AGGAAT)<sub>n</sub> tandem repeat in an intron of the canine von Willebrand factor gene. *Animal Genetics*. 1994; 25: 122. <https://doi.org/10.1111/j.1365-2052.1994.tb00094.x> PMID: 8010530
21. Ostrander EA, Mapa FA, Yee M, Rine J. One hundred and one new simple sequence repeat-based markers for the canine genome. *Mammalian Genome*. 1995; 6: 192–195. <https://doi.org/10.1007/BF00293011> PMID: 7749226
22. Fredholm M, Winter AK. Variation of short tandem repeats within and between species belonging to the Canidae family. *Mammalian Genome*. 1995; 6: 11–18. <https://doi.org/10.1007/BF00350887> PMID: 7719020
23. Francisco LV, Langsten AA, Mellersh CS, Neal CL, Ostrander EA. A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome*. 1996; 7: 359–362. <https://doi.org/10.1007/s003359900104> PMID: 8661717
24. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, et al. A Linkage Map of the Canine Genome. *Genomics*. 1997; 46: 326–336. <https://doi.org/10.1006/geno.1997.5098> PMID: 9441735
25. Jouquand S, Priat C, Hitte C, Lachaume P, Andre C, Galibert F. Identification and characterization of a set of 100 tri- and dinucleotide microsatellites in the canine genome. *Animal Genetics*. 2000; 31: 266–272. <https://doi.org/10.1046/j.1365-2052.2000.00642.x> PMID: 11086536
26. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, Holmes NG et al. Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Research*. 2001; 11: 1784–1795. <https://doi.org/10.1101/gr.189401> PMID: 11591656

27. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*. 2011; 17: 10–12. <https://doi.org/10.14806/ej.17.1.200>
28. Frantz AC, Pope LC, Carpenter PJ, Roper TJ, Wilson GJ, Delahay RJ, et al. Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*. 2003; 12: 1649–1661. <https://doi.org/10.1046/j.1365-294x.2003.01848.x> PMID: 12755892
29. Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 2004; 13: 3261–3273. <https://doi.org/10.1111/j.1365-294X.2004.02346.x> PMID: 15487987
30. Broquet T, Ménard N, Petit E. Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*. 2007; 8: 249–260. <https://doi.org/10.1007/s10592-006-9146-5>
31. Broquet T, Petit E. Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* 2004; 13: 3601–3608. <https://doi.org/10.1111/j.1365-294X.2004.02352.x> PMID: 15488016
32. Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N et al. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 1996; 24: 3189–3194. <https://doi.org/10.1093/nar/24.16.3189> PMID: 8774899
33. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. *Nature Reviews. Genetics* 2005; 6: 847–859. <https://doi.org/10.1038/nrg1707> PMID: 16304600
34. Leeuw J de, Mair P. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software* 2009; 31: 1–30.
35. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015; 67: 1–48. <https://doi.org/10.18637/jss.v067.i01>
36. Lenth R. emmeans: Estimated Marginal Means, aka Least-Square Means. R package version 1.3.5. 2019. Available from <https://cran.r-project.org/package=emmeans>
37. Forcina G, Vallet D, Le Gouar PJ, Bernardo-Madrid R, Illera G, Molina-Vacas G et al. From groups to communities in western lowland gorillas. *Proceedings of the Royal Society B: Biological Sciences*. 2019; 286: 2018–2019. <https://doi.org/10.1098/rspb.2018.2019> PMID: 30963928
38. Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. *Immunogenetics*. 2005; 56: 683–695. <https://doi.org/10.1007/s00251-004-0717-7> PMID: 15605248
39. Ujvari B, Belov K. Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *International Journal of Molecular Sciences* 2011; 12: 5168–5186. <https://doi.org/10.3390/ijms12085168> PMID: 21954351
40. Miller CR, Joyce P, Waits LP. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*. 2002; 160: 357–366. <https://doi.org/10.1093/genetics/160.1.357> PMID: 11805071
41. Sarabia C, Salado I, Cornellas A, Fernández-Gil A, Vilà C, Leonard JA. Towards high-throughput analyses of fecal samples from wildlife. *Animal Biodiversity and Conservation*. 2020; 43: 271–283. <https://doi.org/10.32800/abc.2020.43.0271>
42. Musiani M, Leonard JA, Cluff HD, Gates CC, Mariani S, Paquet PC et al. Differentiation of tundra/taiga and boreal coniferous forest wolves: genetics, coat colour and association with migratory caribou. *Molecular Ecology*. 2007; 16: 4149–4170. <https://doi.org/10.1111/j.1365-294X.2007.03458.x> PMID: 17725575
43. Muñoz-Fuentes V, Darimont CT, Paquet PC, Leonard JA. The genetic legacy of extirpation and re-colonization in Vancouver Island wolves. *Conservation Genetics*. 2010; 11: 547–556. <https://doi.org/10.1007/s10592-009-9974-1>
44. Taberlet P, Luikart G. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*. 1999; 68: 41–55. <https://doi.org/10.1111/j.1095-8312.1999.tb01157.x>
45. Santini A, Lucchini V, Fabbri E, Randi E. Ageing and environmental factors affect PCR success in wolf (*Canis lupus*) excremental DNA samples. *Molecular Ecology Notes*. 2007; 7: 955–961. <https://doi.org/10.1111/j.1471-8286.2007.01829.x>
46. Murphy MA, Kendall KC, Robinson A, Waits LP. The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics*. 2007; 8: 1219–1224. <https://doi.org/10.1007/s10592-006-9264-0>
47. Brinkman TJ, Schwartz MK, Person DK, Pilgrim KL, Hundertmark KJ. Effects of time and rainfall on PCR success using DNA extracted from deer fecal pellets. *Conservation Genetics*. 2010; 11: 1547–1552. <https://doi.org/10.1007/s10592-009-9928-7>
48. Forcina G, Leonard JA. Tools for Monitoring Genetic Diversity in Mammals: Past, Present, and Future. In Ortega J, Maldonado JE, editors. *Conservation Genetics in Mammals: Integrative Research Using Novel Approaches*. Springer; 2020. pp. 13–27. [https://doi.org/10.1007/978-3-030-33334-8\\_2](https://doi.org/10.1007/978-3-030-33334-8_2)