ORIGINAL RESEARCH

A refined panel of 42 microsatellite loci to universally genotype catarrhine primates

Franziska Trede[1,2] | Niels Kil[2,3,4,5] | James Stranks[2,3,4,5] |
Andrew Jesse Connell[6] | Julia Fischer[1,4,7] | Julia Ostner[3,4,5] |
Oliver Schülke[3,4,5] | Dietmar Zinner[1,4,7] | Christian Roos[2,8]

[1]Cognitive Ethology Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[3]Department of Behavioral Ecology, University of Göttingen, Göttingen, Germany

[4]Leibniz ScienceCampus Primate Cognition, Göttingen, Germany

[5]Research Group Primate Social Evolution, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[6]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[7]Department of Primate Cognition, Georg-August-University, Göttingen, Germany

[8]Gene Bank of Primates, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

Correspondence
Franziska Trede, Cognitive Ethology Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany.
Email id: ftrede@dpz.eu

Funding information
Deutsche Forschungsgemeinschaft

Abstract

1. Microsatellite genotyping is an important genetic method for a number of research questions in biology. Given that the traditional fragment length analysis using polyacrylamide gel or capillary electrophoresis has several drawbacks, microsatellite genotyping-by-sequencing (GBS) has arisen as a promising alternative. Although GBS mitigates many of the problems of fragment length analysis, issues with allelic dropout and null alleles often remain due to mismatches in primer binding sites and unnecessarily long PCR products. This is also true for GBS in catarrhine primates where cross-species amplification of loci (often human derived) is common.

2. We therefore redesigned primers for 45 microsatellite loci based on 17 available catarrhine reference genomes. Next, we tested them in singleplex and different multiplex settings in a panel of species representing all major lineages of Catarrhini and further validated them in wild Guinea baboons (*Papio papio*) using fecal samples.

3. The final panel of 42 microsatellite loci can efficiently be amplified with primers distributed into three amplification pools.

4. With our microsatellite panel, we provide a tool to universally genotype catarrhine primates via GBS from different sample sources in a cost- and time-efficient way, with higher resolution, and comparability among laboratories and species.

KEYWORDS

apes, genotyping-by-sequencing, high-throughput sequencing, Old World monkeys, simple tandem repeats

1 | INTRODUCTION

Microsatellites have been and are still widely applied in various biological sciences including population genetics, kinship/pedigree analysis, human and wildlife forensics, linkage analysis, or disease association studies (e.g., Cunningham et al., 2001; Goodwin et al., 2011; Gulcher, 2012; Wasser et al., 2004). Population genetic information obtained by microsatellite genotyping is also

important for monitoring wild populations in conservation contexts, for reintroduction programs or to refine captive breeding management (Arandjelovic & Vigilant, 2018; Norman et al., 2019). Microsatellites are also often the markers of choice to genetically characterize (wild) populations in order to determine degrees of population fragmentation and hybridization, dispersal patterns, mating systems, and reproductive success (e.g., Charpentier et al., 2012; de Moor et al., 2020; Ferreira da Silva et al., 2018; Kheng et al., 2017; McCarthy et al., 2020). The ongoing popularity of microsatellites is largely based on their high abundancy in animal genomes (Hamada et al., 1982; Tautz & Renz, 1984), the high levels of allelic diversity (Ellegren, 2000), and the possibility to amplify them across related species. Accordingly, microsatellites are preferred, for example over SNPs, because of their higher statistical power per locus and their cross-species amplifiability, particularly when applied to small sample size datasets as typically found in forensic and kinship studies (Barbian et al., 2018; Guichoux et al., 2011).

However, traditional microsatellite genotyping via fragment length analysis (FLA) using polyacrylamide gel or capillary electrophoresis has several disadvantages, such as fragment size homoplasy, allele calling difficulties (stutter and split peaks, off-target PCR products), laborious work and relatively high laboratory costs, as well as poor comparability of results among laboratories (De Barba et al., 2017; Guichoux et al., 2011; Pasqualotto et al., 2007). Even with attempts to improve PCR amplification and more accurate/reliable genotyping procedures (Arandjelovic et al., 2009; Buchan et al., 2005; Navidi et al., 1992; Sefc et al., 2003; Taberlet et al., 1996), many of the problems remained.

With microsatellite genotyping-by-sequencing (GBS) using high-throughput sequencing technologies most of the difficulties can be mitigated (Barbian et al., 2018; Johannesen et al., 2017; Pimentel et al., 2018; Vartia et al., 2016). For instance, with GBS the exact length of the microsatellite alleles can be determined, which is a typical problem of FLA genotyping, particularly when alleles differ by only one basepair (bp) (Barbian et al., 2018; Vartia et al., 2016). Moreover, the nucleotide sequence is revealed so that cryptic alleles (alleles with the same length but containing a nucleotide variant) can be detected, resulting in an increased number of alleles and consequently greater statistical power per locus.

Nevertheless, problems with null alleles due to relatively large PCR products and allelic dropout as a result of primers binding in unconserved regions remain with GBS (Pompanon et al., 2005). As many microsatellites can be cross-amplified in phylogenetically related species, primers designed for one species are often tested in related species and then applied if successfully amplified and informative (i.e., polymorph) (Barbara et al., 2007; De Barba et al., 2017). For example, various microsatellite loci characterized for humans can be successfully amplified in nonhuman catarrhine primates (Old World monkeys, apes) (Coote & Bruford, 1996; Ely et al., 1998; Kayser et al., 1996; Morin et al., 1998; Newman et al., 2002; Roeder et al., 2009; Smith et al., 2000) and have been used since then in numerous studies (e.g., Arandjelovic et al., 2014; Kopp et al., 2015;

Minkner et al., 2018; Städele et al., 2019). Yet, attempts to reduce PCR product size or to adapt primers specifically to the study species have been rare (but see Bradley et al., 2000; Engelhardt et al., 2017; Inoue et al., 2016). Furthermore, various research groups use different panels of microsatellites preventing a direct comparison of results, particularly of measures such as genetic diversity and heterozygosity, which are important in a conservation context (Kolleck et al., 2013).

In our study, we aimed to establish a microsatellite panel to universally genotype catarrhine primates via GBS from different sample sources in a cost- and time-efficient way, with higher resolution, and comparability among laboratories and species. Therefore, we screened a total of 269 microsatellite loci, widely targeted in catarrhine primates, and designed conserved primers for 45 loci based on available catarrhine genomes. We then tested the new microsatellite panel in ten primate species representing all major lineages of Catarrhini and further validated their applicability to low-quality DNA samples using fecal samples of wild Guinea baboons (*Papio papio*).

2 | MATERIAL AND METHODS

2.1 | In silico selection of microsatellite loci

We screened 269 human microsatellite loci widely used in catarrhine population genetic studies. We extracted the human (GRCh38/hg38) sequence of each locus with 500 bp flanking regions from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and performed BLAT searches against the 16 available (status: 5 December 2018) nonhuman Catarrhini reference genomes (Table S1) using the UCSC (http://genome.ucsc.edu) or Ensembl (www.ensembl.org) genome browsers with standard settings. In addition, we checked the human sequence for repetitive elements (SINEs, LINES, etc.) in flanking regions using the RepeatMasker Web Server (http://www.repeatmasker.org/) with standard settings. We generated alignments for each locus containing the 16 nonhuman catarrhine species, the human, and the human repeat-masked sequences with Muscle 3.8.31 (Edgar, 2004) in SeaView 4 (Gouy et al., 2010) and added published primer sequences to the alignments.

Loci were selected for further analysis if they fulfilled the following criteria: (a) primer binding sites are not in repetitive elements thus increasing locus-specific amplifiability and reducing the risk of off-target PCR products particularly in multiplex PCR reactions; (b) primer binding sites are conserved among catarrhines so that loci can be universally amplified in this taxonomic group with >180 species (Mittermeier et al., 2013); (c) the microsatellite motif is relatively short (max. 150 bp) to allow small amplicon size (max. 250 bp) and increase locus amplification success from degraded DNA samples, such as fecal samples; and (d) loci are evenly (1–3 loci per chromosome) distributed throughout the genome (using the genomes of *Homo sapiens*, *Nomascus leucogenys*, *Macaca mulatta*, and *Chlorocebus sabaeus* as reference) to avoid potential linkage problems.

For loci which passed the selection criteria, we designed new primers using Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). To allow for multiplexing, primers were designed to have similar annealing temperatures. Locus specificity of primers was checked by BLAT search against the 17 available catarrhine genomes. As primer binding sites were not always fully conserved among the 17 catarrhines, primers of 21 loci were designed with wobble positions. To simplify library preparation for GBS, we added adapter nucleotide sequences to the 5′ end of the locus-specific primers (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′ to forward primers, 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′ to reverse primers; locus-specific primers are provided in Table S2).

2.2 | Laboratory work

First, we tested in singleplex reactions for the locus specificity of selected primers and their universal applicability to catarrhine species in a panel representing all major lineages of catarrhines (Table 1). High-quality DNA from a male of each of the ten species was obtained from the Gene Bank of Primates at the German Primate Center. PCRs were performed in total volumes of 25 μl containing 1× Qiagen Multiplex PCR Master Mix (Qiagen), 0.4 μM of each primer, and 50 ng genomic DNA. Cycling conditions comprised of 15 min at 95°C, 30 cycles each with denaturation at 94°C for 30 s, annealing at 57°C for 90 s and extension at 72°C for 90 s, and a final extension step of 10 min at 72°C. All reactions were run together with no-template controls (NTCs) to check for contamination. PCR performance was checked on 2% agarose gels stained with ethidium bromide (Carl Roth GmbH). Sequencing of singleplex PCR reactions was omitted.

Next, we tested for the possibility of running multiplex PCR reactions to reduce overall laboratory work and costs. Therefore, we pooled either all 45 primer pairs in a single PCR reaction (1-pool approach) or divided them into five PCR reactions each containing nine primer pairs (5-pool approach) or three PCR reactions containing 18 and 2 × 12 primer pairs (3-pool approach; for rationale of pooling and locus exclusion see Results section). Amplifications were conducted as described for the singleplex PCRs (same PCR set-up, DNA samples, cycling conditions, NTCs), but with different primer concentrations

(see Tables S3–S5 for pooling schemes and concentrations of single primers within pools). To minimize PCR errors, we ran PCR reactions in two independent replicates. PCR performance was again checked on 2% agarose gels. Replicate PCR products (including the NTCs) were pooled and then cleaned with the MinElute PCR Purification Kit (QIAGEN). DNA concentrations were measured with a Qubit 3.0 (Thermo Fisher Scientific) and 100 ng were subjected to indexing PCR. Indexing PCR was performed in total volumes of 25 μl containing 1× KAPA HiFi HotStart ReadyMix (Roche), 0.4 μM of each indexing primer and 100 ng purified PCR product. Cycling conditions comprised of 45 s at 98°C, 4 cycles each with denaturation at 98°C for 15 s, annealing at 62°C for 30 s and extension at 72°C for 30 s, and a final extension step of 1 min at 72°C. Subsequently, indexed PCR products were purified with the MinElute PCR Purification Kit (Qiagen) and ran on a Bioanalyzer 2100 (Agilent) to check for PCR performance and molarity. Libraries were diluted to a final concentration of 10 nM and then pooled and sequenced with 51 cycles forward and 251 cycles reverse on Illumina's MiSeq desktop sequencer.

To check for Mendelian inheritance and whether our new microsatellite panel is also applicable to low-quality and low-quantity DNA as typically extracted from fecal samples (Monteiro et al., 1997; Perry et al., 2010), we tested our panel in 12 fecal samples of wild Guinea baboons. The samples comprised of six males and two "families" each composed of a male, a female, and their known offspring. DNA from these 12 specimens was previously genotyped via FLA at 24 microsatellite loci (Dal Pesco, 2019). The amplification procedure and follow-up steps for the applied 3-pool approach were the same as described above, but the number of cycles in the initial amplification was increased to 40, the total DNA amount was increased to 200 ng, and each PCR was performed in triplicates (Barbian et al., 2018).

2.3 | Bioinformatic analysis

The data analysis was performed using the software package CHIIMP v.3.0.0 (Barbian et al., 2018). The raw data (FASTQ files) as well as all input files (config-file, sample-file, locus-attributes-file) are available in the online supplement resources. As our microsatellite

TABLE 1 Catarrhine species used to test the new microsatellite panel

Family	Subfamily	Tribe	Species
Hominidae	Homininae		_Pan troglodytes_
			Gorilla gorilla
	Ponginae		_Pongo abelii_
Hylobatidae			_Hylobates lar_
Cercopithecidae	Colobinae	Presbytini	_Trachypithecus obscurus_
			Pygathrix nemaeus
		Colobini	_Colobus guereza_
	Cercopithecinae	Cercopithecini	_Cercopithecus diana_
		Papionini	_Papio papio_
			Macaca mulatta

TABLE 2 Number of amplified loci and alleles, as well as the level of heterozygosity per species generated in three approaches with high quality DNA (blood) and degraded DNA from fecal samples

	High-quality DNA			Degraded DNA
	5-pool approach (45 loci)	1-pool approach (45 loci)	3-pool approach (42 loci)	3-pool approach (42 loci)
Mean number of loci amplified per sample/species (range)	40.2 (37–43)	25.9 (21–32)	37.8 (33–41)	38.8 (35–41)
Mean number of alleles amplified per sample/species (range)	60.1 (53–69)	36.9 (31–45)	55.5 (46–68)	52.2 (47–56)
Mean level of heterozygosity per sample/species (range)	49.8% (23.3%–69.4%)	43.5% (22.6%–71.4%)	47.1% (19.5%–66.7%)	34.3% (23.1%–47.4%)[a] 46.3% (31.0%–64.3%)[b]

[a]Including all 42 loci.

[b]Including only the 32 loci that were polymorphic in the study species.

panel included several di-repeat loci, which stutter more frequently than tetra-repeats, we increased the stutter count ratio to 0.70 (stutter.count.ratio_max: 0.7). We further implemented a broad range of possible allele lengths in the locus attributes by setting the length buffer to 100 bp. This ensured the inclusion of all tested species even if the allele sizes at a given locus varied between species according to the available reference genomes. The minimum number of reads per locus was set to 100 (counts.min: 100). All other parameters were set to default.

With the current version of CHIIMP, wobble positions in primer sequences cannot be accounted for. Hence, for loci with a wobble position in a primer sequence, alternative nucleotides of the wobble are erroneously recognized as different alleles. Moreover, the repeat motif needs to be specified in CHIIMP, but as repeat motifs can vary in the investigated species, correct (orthologous) reads remain unrecognized for some species if CHIIMP is fed with a wrong repeat motif. Due to these reasons, the output for all loci was checked manually and corrected if needed. Additionally, we screened the processed reads for the general level of amplification per locus and the occurrence of PCR artifacts (off-target amplification, primer dimer, false primer pairings, etc.).

3 | RESULTS

3.1 | In silico selection of microsatellite loci

In total, 217 of the 269 investigated loci were not optimal for microsatellite genotyping of catarrhines. For 147 of them, one or both primer binding sites or the complete locus were located in repetitive elements. This increases the likelihood to amplify various off-target PCR products, particularly in multiplex settings when many primers that can bind multiple times in the respective genome are combined in a single PCR reaction. For an additional 32 loci, we could not find conserved primer binding sites near the microsatellite and a further 15 loci contained relatively long microsatellite repeat regions for one or more species, resulting in long PCR products (>250 bp). Longer

PCR products are often difficult to generate if only degraded DNA material is available and can result in null alleles. Further problems included, for instance, the location of loci directly next to each other on the same chromosome and thus increasing the risk of linkage. Additionally, double entries of loci under different names or gaps in some of the reference genomes (especially for Y-chromosomal loci) impeded the screening process. A full list of screened loci including the respective reasons for their exclusion is provided in Table S6. Of the 52 loci which fulfilled our criteria, we selected 45 (1–3 loci per human chromosome including gonosomes) for downstream analyses. The chromosomal locations of the chosen loci in the genomes of *H. sapiens*, *N. leucogenys*, *M. mulatta*, and *C. sabaeus* are provided in the supplement (Table S2). We found no indication for the presence of linkage between any of the loci in any of the four investigated species (minimal distance between two loci 5.35 million bp).

The newly designed primers for the 45 loci (consisting of di-, tri-, and tetra-repeats) amplify PCR products between 56–215 bp (according to available genome data; Table S2). Compared to the original published primers, we were able to reduce PCR product sizes by 2–225 bp (mean 75.9 bp) in 37 loci whereas for five loci, the new primers amplify a moderately longer fragment (elongation by 2 – 15 bp; mean 7.6 bp). PCR product size for the remaining three loci did not change. As primer binding sites were not always perfectly conserved among the 17 investigated catarrhine reference genomes, primers for 21 loci contain wobble positions. Mismatches in primer binding sites, found only in a few (1–2) of the investigated species, were neglected in primer design and probably result in less efficient or no amplification of the respective locus in the given species (0–12 loci with mismatches per species, mean 3.4; Table S2).

3.2 | Singleplex PCR test

Singleplex PCR reactions of the 45 loci in ten species representing all major lineages of catarrhines were run on agarose gels and resulted, for all loci and species, in PCR products within the expected size range with no signs of amplifying any off-target PCR products (data

not shown). Thus, locus specificity and universal applicability of our primer set to catarrhine primates was indicated.

3.3 | Multiplexing approaches

Sequenced alleles ranged in size from 71 bp (D3s1768) to 211 bp (D12s372) and nine loci contained cryptic alleles in at least one species (Tables S7–S10). The level of amplification and obtained sequence reads varied across samples/species and loci in all three approaches. The amplification of all loci in one pool (1-pool approach) was least effective, resulting in the lowest number of amplified loci (mean 25.9) and alleles (mean 36.9; Table 2). In some cases, the reason for allelic dropouts could be attributed to wrong primer pairing/primer mismatches (primer dimer or off-target amplification of short products). Most loci amplified less efficiently than in the other two approaches and some (N = 11) failed to amplify at all. Only nine loci recovered the same number of alleles as in the 5-pool approach. Interestingly, even though the number of amplified alleles was reduced from a mean of 60.1 to 36.9 compared to the 5-pool approach, the level of heterozygosity was not affected to the same extent with a reduction from 49.8% to 43.5% (Table 2).

The best results, that is, the highest amplification levels for loci (mean 40.2) and alleles (mean 60.1), were generated applying the 5-pool approach (Table 2). Nevertheless, we observed again primer dimers and short off-target PCR products potentially as a result of interacting primers from different loci. Moreover, three loci (D11s1366, D12s67.2, and D15s1007) neither amplified in the 1-pool nor in the 5-pool approach and were excluded from further testing. To further improve amplification success and to reduce primer interactions among primers of different loci (based on the knowledge obtained from the 1-pool and 5-pool approaches), we distributed the 42 remaining loci into three amplification pools containing 18, 12 and 12 loci, respectively (Table S5). Using the 3-pool approach, we were able to largely minimize primer interactions, but amplification success for loci (mean 37.8) and alleles (mean 55.5) per species was slightly reduced compared to the 5-pool approach, but higher than in the 1-pool approach (Table 2). The reduced amplification success was due to allelic dropouts of single alleles or whole loci in some species (see Table S9).

3.4 | Degraded DNA samples

For the degraded DNA samples, we applied the 3-pool approach as this represented the best compromise between amplification efficiency and laboratory effort and costs (see Results Multiplex approaches). The amplification from fecal samples was successful except for four (out of 42) loci (two autosomal and two gonosomal loci; Table S10). The number of loci and alleles amplified per sample was comparable to the results obtained from high-quality DNA samples (Table 2). However, 10 of the 42 amplified loci were

monomorphic in our *P. papio* population, that is, all twelve individuals showed the same allele. The remaining 32 loci showed a level of 46.3% heterozygosity (Table 2).

All autosomal loci were in accordance with Mendelian inheritance besides D7s503 and D13s1291. For D7s503, the two alleles with the highest read counts for male MRX (99/109) did not match one of the two alleles of his offspring THL (111/113; mother MMI: 103/113). A closer look at the data revealed that MRX also had many reads for allele 111 (only 23 reads less than for allele 109), indicating that the allele 109 of MRX was likely an overamplified stutter sequence. A genotype with the allele combination 99/111 also corresponds to the genotype derived from FLA for this locus (152/164; 12bp distance between alleles, based on >20 amplifications; Dal Pesco, 2019). For D13s1291, the two alleles of male MLK (130b/132b) did not match with his offspring PTC (130a/132a; mother LCY: 128/130a). MLK was the only individual with these two potential cryptic alleles (each with a G→A point mutation) and further showed reads with 130 and 132 bp length without this mutation. At this point, we can neither exclude the occurrence of a PCR artifact in this particular case, nor that this locus does not follow the rules of Mendelian inheritance.

4 | DISCUSSION

From a set of 269 microsatellite loci widely applied in catarrhine primates, we selected a total of 45 loci that can be universally applied to genotype catarrhine primates. Due to the relatively small amplicon sizes, even low-quality DNA could be genotyped and since the selected loci were evenly distributed throughout the genome (at least according to the human genome), the risk of linkage was significantly reduced. Moreover, our panel could be multiplexed to a great extent. The testing of different multiplex settings revealed that a 5-pool approach produced the best result, but that a 3-pool approach containing one pool of 18 and two of 12 loci is the best compromise between locus amplification efficiency and laboratory effort and costs.

We tested the panel with high-quality DNA samples from all major lineages of catarrhines in multiplex settings and revealed successful amplification rates of 33 to 41 (average 38) loci per species (Table 2). We additionally showed the applicability of the 3-pool approach to degraded DNA samples such as fecal samples, which is a common material in many noninvasive wildlife studies (Carroll et al., 2018; Waits & Paetkau, 2005). The results for fecal samples were similar to the results of the high-quality samples with respect to the mean number of loci and alleles amplified per sample (Table 2). All loci, besides D13s1291, were in accordance with Mendelian inheritance, demonstrating the suitability of the new microsatellite panel for kinship and relatedness analyses. To ensure high-quality genotypes from fecal samples, and depending on DNA quality, further adaptions to the protocol might be necessary. Quantifying the endogenous DNA content via quantitative PCR prior to genotyping will help to select only those samples with sufficient endogenous DNA content (e.g., >25 pg endogenous

DNA as suggested by Barbian et al., 2018). Additionally, multiple samples per individual can be analyzed or the number of PCR replicates per sample can be increased.

Through multiplexed GBS, cryptic alleles can be detected (Barbian et al., 2018; Sarhanova et al., 2018; Vartia et al., 2016), and even in our test panel of only ten catarrhine species with one individual each, we found cryptic alleles at nine loci (Tables S7–S10). Although our results are based on only two or three replicates per approach (depending on the sample type) and hence should be interpreted with caution, we are confident that these alleles are indeed cryptic alleles and not PCR artifacts. In case of PCR artifacts, we would expect mixed sequence reads showing more than two alleles or highly imbalanced sequence read counts for the "true allele" and the "artifact allele," as it is highly unlikely that the same PCR artifact occurs in all replicates. As more individuals per species get tested, the number of cryptic alleles will most likely increase and provide further accuracy and a higher statistical power of our panel.

Another advantage of GBS is that the resulting genetic data, in form of allele sequences, are independent of the used sequencing platform. Thus, data produced by different laboratories can be easily shared and compared. By applying validated bioinformatics pipelines, such as the CHIIMP pipeline (Barbian et al., 2018), one can further ensure that the resulting data are reproducible and less prone to arbitrary allele calling by different researchers while still allowing the customization of, for example, filtering parameters to fit different datasets.

Although we recommend the 3-pool approach, the amplification success of individual loci can be improved, for example, by amplifying all loci in individual reactions and then pooling before or after the indexing PCR. However, this would largely increase workload in the laboratory and costs. It is also important to check which loci are polymorphic in the species of interest, so that monomorphic loci can be excluded from large-scale population genetic investigations. Likewise, as several species exhibit mismatches in primer binding sites (0–12 loci with mismatches per species), primer design for a given species can be adjusted and optimized, which becomes easier to do with an increasing number of sequenced catarrhine genomes.

In summary, with our microsatellite panel, we provide a tool to universally genotype catarrhine primates via GBS from samples of varying DNA quality in a cost- and time-efficient way, with higher resolution, better comparability among laboratories, and largely mitigated problems of traditional FLA.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Franziska Trede: Investigation (equal); Writing-original draft (equal); Writing-review & editing (equal). **Niels Kil:** Investigation (equal); Writing-review & editing (equal). **James Stranks:** Investigation (equal); Writing-review & editing (equal). **Andrew Jesse Connell:** Software (lead); Writing-review & editing (equal). **Julia Fischer:** Funding acquisition (equal); Writing-review & editing (equal). **Julia Ostner:** Funding acquisition (equal); Writing-review & editing (equal). **Oliver Schülke:** Funding acquisition (equal); Writing-review & editing (equal). **Dietmar Zinner:** Writing-original draft (equal); Writing-review & editing (equal). **Christian Roos:** Conceptualization (equal); Investigation (supporting); Writing-original draft (equal); Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

The sequencing data have been submitted to the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA672243 (http://www.ncbi.nlm.nih.gov/bioproject/672243). Input files for the bioinformatics analysis (config-, sample-, and locus-attributes-files) are available online in the supporting information (Appendix S1).

ORCID

Franziska Trede https://orcid.org/0000-0003-3690-1006
Niels Kil https://orcid.org/0000-0002-8823-2599
James Stranks https://orcid.org/0000-0002-5557-9051
Andrew Jesse Connell https://orcid.org/0000-0002-9393-3074
Julia Fischer https://orcid.org/0000-0002-5807-0074
Julia Ostner https://orcid.org/0000-0001-6871-9976
Oliver Schülke https://orcid.org/0000-0003-0028-9425
Dietmar Zinner https://orcid.org/0000-0003-3967-8014
Christian Roos https://orcid.org/0000-0003-0190-4266

REFERENCES

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T. R., Thalmann, O., Siedel, H., & Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources*, *9*(1), 28–36. https://doi.org/10.1111/j.1755-0998.2008.02387.x

Arandjelovic, M., Head, J., Boesch, C., Robbins, M. M., & Vigilant, L. (2014). Genetic inference of group dynamics and female kin structure in a western lowland gorilla population (*Gorilla gorilla gorilla*). *Primate Biology*, *1*(1), 29–38. https://doi.org/10.5194/pb-1-29-2014

Arandjelovic, M., & Vigilant, L. (2018). Non-invasive genetic censusing and monitoring of primate populations. *American Journal of Primatology*, *80*(3), e22743. https://doi.org/10.1002/ajp.22743

Barbara, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology*, *16*(18), 3759–3767. https://doi.org/10.1111/j.1365-294X.2007.03439.x

Barbian, H. J., Connell, A. J., Avitto, A. N., Russell, R. M., Smith, A. G., Gundlapally, M. S., & Hahn, B. H. (2018). CHIIMP: An automated high-throughput microsatellite genotyping platform reveals greater allelic diversity in wild chimpanzees. *Ecology and Evolution*, *8*(16), 7946–7963. https://doi.org/10.1002/ece3.4302

Bradley, B. J., Boesch, C., & Vigilant, L. (2000). Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces.

Conservation Genetics, 1, 289–292. https://doi.org/10.1023/A:10115 14007741

Buchan, J. C., Archie, E. A., Van Horn, R. C., Moss, C. J., & Alberts, S. C. (2005). Locus effects and sources of error in noninvasive genotyping. *Molecular Ecology Notes*, 5(3), 680–683. https://doi.org/10.1111/j.1471-8286.2005.01002.x

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, 11(7), 1094–1119. https://doi.org/10.1111/eva.12600

Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., Barthès, N., Alberts, S. C., & Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. *Molecular Ecology*, 21(3), 715–731. https://doi.org/10.1111/j.1365-294X.2011.05302.x

Coote, T., & Bruford, M. W. (1996). Human microsatellites applicable for analysis of genetic variation in apes and old world monkeys. *Journal of Heredity*, 87, 406–410. https://doi.org/10.1093/oxfordjournals.jhered.a023026

Cunningham, E. P., Dooley, J. J., Splan, R. K., & Bradley, D. G. (2001). Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics*, 32(6), 360–364. https://doi.org/10.1046/j.1365-2052.2001.00785.x

Dal Pesco, F. (2019). *Dynamics and fitness benefits of male-male sociality in wild Guinea baboons* (*Papio papio*). PhD Thesis, Georg-August-University Göttingen. Retrieved from http://hdl.handle.net/21.11130/00-1735-0000-0005-134F-E

De Barba, M., Miquel, C., Lobreaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17(3), 492–507. https://doi.org/10.1111/1755-0998.12594

De Moor, D., Roos, C., Ostner, J., & Schülke, O. (2020). Bonds of bros and brothers: Kinship and social bonding in post-dispersal male macaques. *Molecular Ecology*, 29(17), 3346–3360. https://doi.org/10.1111/mec.15560

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ellegren, H. (2000). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, 24(4), 400–402. https://doi.org/10.1038/74249

Ely, J. J., Gonzalez, D. L., Reeves-Daniel, A., & Stone, W. H. (1998). Individual identification and paternity determination in chimpanzees (*Pan troglodytes*) using human short tandem repeat (STR) markers. *International Journal of Primatology*, 19(2), 255–271. https://doi.org/10.1023/A:1020379531874

Engelhardt, A., Muniz, L., Perwitasari-Farajallah, D., & Widdig, A. (2017). Highly polymorphic microsatellite markers for the assessment of male reproductive skew and genetic variation in critically endangered crested macaques (*Macaca nigra*). *International Journal of Primatology*, 38(4), 672–691. https://doi.org/10.1007/s10764-017-9973-x

Ferreira da Silva, M. J., Kopp, G. H., Casanova, C., Godinho, R., Minhós, T., Sá, R., Zinner, D., & Bruford, M. W. (2018). Disrupted dispersal and its genetic consequences: Comparing protected and threatened baboon populations (*Papio papio*) in West Africa. *PLoS One*, 13(4), e0194189. https://doi.org/10.1371/journal.pone.0194189

Goodwin, W., Linacre, A., & Hadi, S. (2011). *An introduction to forensic genetics* (vol. 2). John Wiley & Sons.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multi-platform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2), 221–224. https://doi.org/10.1093/molbev/msp259

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., & Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11(4), 591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

Gulcher, J. (2012). Microsatellite markers for linkage and association studies. *Cold Spring Harbor Protocols*, 2012(4), 425–432. https://doi.org/10.1101/pdb.top068510

Hamada, H., Petrino, M. G., & Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 79(21), 6465–6469. https://doi.org/10.1073/pnas.79.21.6465

Inoue, E., Ogata, M., Seino, S., & Matsuda, I. (2016). Sex identification and efficient microsatellite genotyping using fecal DNA in proboscis monkeys (*Nasalis larvatus*). *Mammal Study*, 41(3), 141–148. https://doi.org/10.3106/041.041.0304

Johannesen, J., Fabritzek, A. G., Ebner, B., & Bikar, S. E. (2017). Characterisation of microsatellite and SNP markers from Miseq and genotyping-by-sequencing data among parapatric *Urophora cardui* (Tephritidae) populations. *PeerJ*, 5, e3582. https://doi.org/10.7717/peerj.3582

Kayser, M., Ritter, H., Bercovitch, F., Mrug, M., Roewer, L., & Nürnberg, P. (1996). Identification of highly polymorphic microsatellites in the rhesus macaque *Macaca mulatta* by cross-species amplification. *Molecular Ecology*, 5(1), 157–159. https://doi.org/10.1111/j.1365-294X.1996.tb00302.x

Kheng, V., Zichello, J. M., Lumbantobing, D. N., Lawalata, S. Z. S., Andayani, N., & Melnick, D. J. (2017). Phylogeography, population structure, and conservation of the Javan gibbon (*Hylobates moloch*). *International Journal of Primatology*, 39(1), 5–26. https://doi.org/10.1007/s10764-017-0005-7

Kolleck, J., Yang, M., Zinner, D., & Roos, C. (2013). Genetic diversity in endangered Guizhou snub-nosed monkeys (*Rhinopithecus brelichi*): Contrasting results from microsatellite and mitochondrial DNA data. *PLoS One*, 8(8), e73647. https://doi.org/10.1371/journal.pone.0073647

Kopp, G. H., Fischer, J., Patzelt, A., Roos, C., & Zinner, D. (2015). Population genetic insights into the social organization of Guinea baboons (*Papio papio*): Evidence for female-biased dispersal. *American Journal of Primatology*, 77(8), 878–889. https://doi.org/10.1002/ajp.22415

McCarthy, M. S., Lester, J. D., Cibot, M., Vigilant, L., & McLennan, M. R. (2020). Atypically high reproductive skew in a small wild chimpanzee community in a human-dominated landscape. *Folia Primatologica*, 91, 1–9. https://doi.org/10.1159/000508609

Minkner, M. M. I., Young, C., Amici, F., McFarland, R., Barrett, L., Grobler, J. P., Henzi, S. P., & Widdig, A. (2018). Assessment of male reproductive skew via highly polymorphic STR markers in wild vervet monkeys, *Chlorocebus pygerythrus*. *Journal of Heredity*, 109(7), 780–790. https://doi.org/10.1093/jhered/esy048

Mittermeier, R. A., Wilson, D. E., & Rylands, A. B. (2013). *Handbook of the mammals of the world: Primates* (vol. 3). Lynx Edicions.

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K. G., Bonnet, J., Vidal, R., Cabrita, J., & Mégraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. *Journal of Clinical Microbiology*, 35(4), 995–998. https://doi.org/10.1128/JCM.35.4.995-998.1997

Morin, P. A., Mahboubi, P., Wedel, S., & Rogers, J. (1998). Rapid screening and comparison of human microsatellite markers in baboons: Allele size is conserved, but allele number is not. *Genomics*, 53, 12–20. https://doi.org/10.1006/geno.1998.5460

Navidi, W., Arnheim, N., & Waterman, M. S. (1992). A multiple-tubes approach for accurate genotyping of very small DNA samples by using

PCR: Statistical considerations. *American Journal of Human Genetics*, *50*(2), 347–359.

Newman, T. K., Fairbanks, L. A., Pollack, D., & Rogers, J. (2002). Effectiveness of human microsatellite loci for assessing paternity in a captive colony of vervets (*Chlorocebus aethiops sabaeus*). *American Journal of Primatology*, *56*(4), 237–243. https://doi.org/10.1002/ajp.1078

Norman, A. J., Putnam, A. S., & Ivy, J. A. (2019). Use of molecular data in zoo and aquarium collection management: Benefits, challenges, and best practices. *Zoo Biology*, *38*(1), 106–118. https://doi.org/10.1002/zoo.21451

Pasqualotto, A. C., Denning, D. W., & Anderson, M. J. (2007). A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *Journal of Clinical Microbiology*, *45*(2), 522–528. https://doi.org/10.1128/JCM.02136-06

Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, *19*(24), 5332–5344. https://doi.org/10.1111/j.1365-294X.2010.04888.x

Pimentel, J. S. M., Carmo, A. O., Rosse, I. C., Martins, A. P. V., Ludwig, S., Facchin, S., Pereira, A. H., Brandão-Dias, P. F. P., Abreu, N. L., & Kalapothakis, E. (2018). High-throughput sequencing strategy for microsatellite genotyping using neotropical fish as a model. *Frontiers in Genetics*, *9*, 73. https://doi.org/10.3389/fgene.2018.00073

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847–859. https://doi.org/10.1038/nrg1707

Roeder, A. D., Bonhomme, M., Heijmans, C., Bruford, M. W., Crouau-Roy, B., Doxiadis, G., & Otting, N. (2009). A large panel of microsatellite markers for genetic studies in the infra-order catarrhini. *Folia Primatologica*, *80*(2), 63–69. https://doi.org/10.1159/000211121

Sarhanova, P., Pfanzelt, S., Brandt, R., Himmelbach, A., & Blattner, F. R. (2018). SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. *Ecology and Evolution*, *8*(22), 10817–10833. https://doi.org/10.1002/ece3.4533

Sefc, K. M., Payne, R. B., & Sorenson, M. D. (2003). Microsatellite amplification from museum feather samples: Effects of fragment size and template concentration on genotyping errors. *The Auk*, *120*(4), 982–989. https://doi.org/10.1093/auk/120.4.982

Smith, K. L., Alberts, S. C., Bayes, M. K., Bruford, M. W., Altmann, J., & Ober, C. (2000). Cross-species amplification, non-invasive genotyping, and non-Mendelian inheritance of human STRPs in savannah baboons. *American Journal of Primatology*, *51*, 219–227.

Städele, V., Roberts, E. R., Barrett, B. J., Strum, S. C., Vigilant, L., & Silk, J. B. (2019). Male-female relationships in olive baboons (*Papio anubis*): Parenting or mating effort? *Journal of Human Evolution*, *127*, 81–92. https://doi.org/10.1016/j.jhevol.2018.09.003

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194. https://doi.org/10.1093/nar/24.16.3189

Tautz, D., & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, *12*(10), 4127–4138. https://doi.org/10.1093/nar/12.10.4127

Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., Carlsson, J. E. L., Gauthier, D. T., McGinnity, P., Cross, T. F., FitzGerald, R. D., Mirimin, L., Crispie, F., Cotter, P. D., & Carlsson, J. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society Open Science*, *3*(1), 150565. https://doi.org/10.1098/rsos.150565

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, *69*(4), 1419–1433.

Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B., & Stephens, M. (2004). Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14847–14852. https://doi.org/10.1073/pnas.0403170101

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.