RESOURCE ARTICLE

Pedigree reconstruction and distant pairwise relatedness estimation from genome sequence data: A demonstration in a population of rhesus macaques (*Macaca mulatta*)

Lauren E. Petty[1] | Kathrine Phillippi-Falkenstein[2] | H. Michael Kubisch[2] |
Muthuswamy Raveendran[3] | R. Alan Harris[3] | Eric J. Vallender[2,4] | Chad D. Huff[5] |
Rudolf P. Bohm[2] | Jeffrey Rogers[3] | Jennifer E. Below[1]

[1]Vanderbilt Genetics Institute and Department of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[2]Division of Veterinary Medicine, Tulane National Primate Research Center, Covington, LA, USA

[3]Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

[4]Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA

[5]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Correspondence
Jeffrey Rogers, Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.
Email: jr13@bcm.edu

Jennifer Below, Vanderbilt Genetics Institute and Department of Genetic Medicine, Vanderbilt University Medical Center, Light Hall #519B, 2215 Garland Avenue, Nashville, TN 37232, USA.
Email: jennifer.e.below@vanderbilt.edu

Abstract

A primary challenge in the analysis of free-ranging animal populations is the accurate estimation of relatedness among individuals. Many aspects of population analysis rely on knowledge of relatedness patterns, including socioecology, demography, heritability and gene mapping analyses, wildlife conservation and the management of breeding colonies. Methods for determining relatedness using genome-wide data have improved our ability to determine kinship and reconstruct pedigrees in humans. However, methods for reconstructing complex pedigree structures and estimating distant relatedness (beyond third-degree) have not been widely applied to other species. We sequenced the genomes of 150 male rhesus macaques from the Tulane National Primate Research Center colony to estimate pairwise relatedness, reconstruct closely related pedigrees, estimate more distant relationships and augment colony records. Methods for determining relatedness developed for human genetic data were applied and evaluated in the analysis of nonhuman primates, including identity-by-descent-based methods for pedigree reconstruction and shared segment-based inference of more distant relatedness. We compared the genotype-based pedigrees and estimated relationships to available colony pedigree records and found high concordance (95.5% agreement) between expected and identified relationships for close relatives. In addition, we detected distant relationships not captured in colony records, including some as distant as twelfth-degree. Furthermore, while deep sequence coverage is preferable, we show that this approach can also provide valuable information when only low-coverage (5×) sequence data is available. Our findings demonstrate the value of these methods for determination of relatedness in various animal populations, with diverse applications to conservation biology, evolutionary and ecological research and biomedical studies.

KEYWORDS
bioinfomatics/phyloinfomatics, kinship, mammals, quantitative genetics

1 | INTRODUCTION

One fundamental element of mammalian social organization and population biology is the pattern of kinship relationships among individuals in a local population (de Waal & Tyack, 2003). The social groups of many diverse mammalian species include matrilines or other clusters of genealogical relatives who develop, invest in and benefit from long-term relationships (bottlenose dolphins [*Tursiops* sp.: Krutzen et al., 2003]; lions [*Panthera leo*: Spong et al., 2002]; elephants [*Loxodonta cyclotis*: Schuttler et al., 2014]; baboons [*Papio* sp.: Altmann, 1980; Silk et al., 2003; Stadele et al., 2016]). Investigating and understanding behavioural interactions among individuals, patterns of mate selection (Vigilant et al., 2001), dispersal and intergroup migration (Cheney & Seyfarth, 1983; Langergraber et al., 2014; Van Cise et al., 2017), group fission and critical aspects of natural selection such as variation in interbirth intervals or infant survivorship (Silk et al., 2003) requires information about the relationships among the actors in the continual drama of mammalian social behaviour and life history. Unfortunately, documenting kinship among individuals in natural populations of animals, especially among long-lived mammals such as primates, cetaceans, elephants and other large-bodied species, has traditionally required many years (often decades) of continuous observation of recognizable individuals (Alberts & Altmann, 2012; Ishizuka et al., 2020; McComb et al., 2001). Even then, paternity is often uncertain without some form of genetic testing. In addition, investigations of smaller animals with shorter generation times but that are migratory or are otherwise difficult to repeatedly observe or identify can also present challenges for the assessment of kinship among study individuals. Research in various aspects of the ethology, social organization, ecology, demography and population genetics of natural populations benefits directly from methods to identify or confirm relationships among individuals that do not require long-term observation of habituated populations (Foroughirad et al., 2019; Ishizuka et al., 2020; Langergraber et al., 2014; Snyder-Mackler et al., 2016; Stadele et al., 2016; Van Cise et al., 2017). In addition, biomedical research using captive populations of nonhuman primates or other species can benefit from the capacity to determine parentage or other relationships among individuals whenever reliable pedigree information is not available (Kanthaswamy et al., 2006; Rogers et al., 2006; Vinson et al., 2013).

In previous work spanning more than twenty years, researchers have used highly polymorphic genetic markers (usually microsatellites) to infer relationships in nonhuman species (Morin et al., 1994; Stadele & Vigilant, 2016). These inferences have relied on genotyping microsatellites, mitochondrial DNA or other markers and have focused on identifying parent-offspring pairs, full- and half-siblings ([Kanthaswamy et al., 2006; Morin et al., 1994; Stadele et al., 2016; Vigilant et al., 2001; Walling et al., 2010] and see Flanagan and Jones (2019) for a recently published excellent review). More recently, researchers have begun to use single nucleotide polymorphisms (SNPs) for parentage and kinship analyses in long-lived mammals such as yellow baboons (*Papio cynocephalus*: Snyder-Mackler et al., 2016), dolphins (*Tursiops aduncus*: Foroughirad et al., 2019), and pilot whales (*Globicephala macrorhynchus*: Van Cise et al., 2017). In some cases these studies have used thousands of SNPs from across the genome (Andrews et al., 2018; Premachandra et al., 2019; Snyder-Mackler et al., 2016; Strucken et al., 2016; Zhang et al., 2018). The methods employed in these studies can accurately infer very close pairwise relationships and some can cluster closely related individuals together into a family group with assigned parentage (as in COLONY: Wang, 2013) and grandparentage (as in SEQUOIA: Huisman, 2017).

Many studies of mammalian populations, both natural and domesticated, have used genetic markers to identify parent-offspring relationships (see references cited above) but few have attempted to identify more distant relationships (Foroughirad et al., 2019; Goudet et al., 2018; Snyder-Mackler et al., 2016; Van Cise et al., 2017). One major hurdle has been the lack of adequate methods for generating sufficiently informative genetic data to detect relatedness of third-degree or greater. Microsatellite loci can be powerful for detecting close relationships, but lack the capacity to accurately determine more distant relationships, and require specific knowledge of high heterozygosity loci in the particular population under study. Given recent advances in DNA sequencing (both whole genome sequencing and RAD-seq methods) and genotyping technology, it is now feasible to generate whole genome genotype data for significant numbers of study subjects from any vertebrate or invertebrate population (Andrews et al., 2018; Flanagan & Jones, 2019; Snyder-Mackler et al., 2016; Xue et al., 2016). This has dramatically improved prospects for kinship analysis in various species.

In parallel with advances in molecular genomics, there has been considerable progress in the development of analytical methods to reconstruct closely related pedigrees and infer distant pairwise relationships using genetic data in humans (Conomos et al., 2016; Li et al., 2014; Staples et al., 2014, 2015, 2016). In addition, as methods to detect cryptic or unrecognized relatedness have been developed, this has been increasingly recognized as a source of bias in genetic studies when not adequately addressed (Conomos et al., 2015). Our ability to investigate, leverage or control for relatedness is of course limited by our ability to accurately assess it.

Related individuals often share identical segments of the genome that have been inherited from a recent common ancestor. Due to recombination events in homologous chromosomes during meiosis, we expect fewer and smaller identical segments to be shared between pairs of individuals that are more distantly related compared to more and larger segments shared between more closely related pairs. Therefore, the length and distribution of these genomic segments shared identical-by-descent (IBD) can be used to estimate the degree of relatedness between a pair of densely genotyped individuals (Browning & Browning, 2011; Huff et al., 2011; Li et al., 2014). If IBD sharing is widespread across the genome, the genome-wide average proportion of IBD sharing is also a useful metric because the expected mean proportions

of genetic sharing between relatives can be used to distinguish between degrees of relatedness (Huff et al., 2011; Purcell et al., 2007).

Effectiveness of these different methods depends on the degree of relatedness between the individuals (Staples et al., 2014). Genome-wide average IBD proportions are better suited to predict close relationships (e.g., third-degree or closer), while IBD segment-based approaches are preferable for more distant relationships, generally fourth- to eighth-degree. Beyond eighth-degree, the likelihood that a pair of individuals shares any genomic segment at all from a recent common ancestor drops below 50% and therefore pairwise relationships beyond the eighth-degree are often not detectable using genetic data (Huff et al., 2011). For closely related kin (third-degree relatives or closer) genome-wide average sharing of common variant genotypes that are IBD provides robust estimates (Staples et al., 2014). Software tools such as Pedigree Reconstruction and Identification of a Maximum Unrelated Set (PRIMUS; Staples et al., 2014, 2015), can be used to reconstruct all possible pedigrees consistent with the pattern of pairwise IBD sharing for any given set of study samples linked by third-degree or closer relationships. Estimation of recent shared ancestry (ERSA; Huff et al., 2011; Li et al., 2014) is a tool for estimating pairwise more distant (first- through eighth-degree) relatedness from dense genotype data.

Combining the closely related pedigree structures generated in PRIMUS with the distant relatedness estimations from ERSA, it is possible to aggregate evidence over all distantly related pairs of individuals in two pedigrees to more accurately identify even very distant (up to thirteenth-degree) relationships using Pedigree-Aware Distant Relatedness Estimation (PADRE; Staples et al., 2016). For example, in Figure 1, if individual D is demonstrated to be related at sixth-degree to individual F, and individuals D and G are estimated to be related at seventh-degree, we can maximize likelihood across these multiple distant relationship measures to increase our confidence in our inference of the distant relationship between founders D and M.

PRIMUS and PADRE overcome many of the limitations of previous approaches by leveraging both pairwise genome-wide IBD proportions and the length and distribution of shared segments. Additionally, unlike other approaches, PRIMUS reconstructs all non-consanguineous pedigrees of arbitrary size and structure that are consistent with the genetic data, spanning any number of generations, allowing for data missingness from missing samples within pedigrees, establishing directionality of relationships, and generating pedigree images and corresponding FAM files (Chang et al., 2015). PRIMUS can even reconstruct some consanguineous pedigrees with offspring of parents who are third-degree relatives. PADRE leverages multiple relationships across pedigrees to provide more accurate estimates of distant relationships between pedigree founders. Together, these tools represent major advances in pedigree reconstruction that, to our knowledge, have not been applied for pedigree reconstruction and distant relatedness estimation outside of the context of human data analysis.

Access to software tools that can estimate both close pedigree relationships and distant pairwise relatedness, combined with whole genome genotype data, creates new opportunities to determine kinship relationships across a set of study subjects using genetic information alone and opens remarkable new avenues of research into social behaviour, social organization, various aspects of demography (dispersal, group histories) and other elements of population biology and socioecology. We sought to test these methods using a population in which partial pedigree information was available, thus providing an opportunity to explore the power of these pedigree reconstruction methods for nonhuman mammals and to assess their accuracy.

Rhesus macaques (*Macaca mulatta*) are the most commonly used nonhuman primates in biomedical research. A whole genome reference sequence assembly for this species was first published in 2007 (Gibbs et al., 2007), and has been subsequently improved (e.g., GenBank accession GCA_000772875.3; also called Mmul_8.0.1). Accurate determination of relatedness in nonhuman primates is essential to facilitate gene mapping studies, the development and analysis of primate models of human genetic diseases, and other aspects of biomedical investigation (Ackermann et al., 2014; Bimber et al., 2017; Johnson et al., 2015; Rogers et al., 2013; Vinson et al., 2013).

Despite the fact that the human and macaque genomes are ~93% identical at the basepair level, various factors such as the higher levels of genetic variation found in the macaques (greater than two-fold higher in macaques of Indian origin relative to humans, Xue et al., 2016), complex mating patterns, often including large paternal half-sibships and the potential for inbreeding in captive populations, make the generalizability of algorithms and software tools developed for use in humans to the nonhuman primates uncertain (Bercovitch, 1997; Gibbs et al., 2007; Smith, 1982; Xue et al., 2016). Therefore, we evaluated the performance of existing methods for estimating relatedness from genetic data in a colony of 148 presumed Indian-origin and two presumed Chinese-origin rhesus macaques from the Tulane National Primate Research Center and compared our findings to colony breeding group records.

2 | MATERIALS AND METHODS

2.1 | Subjects

The 150 animals analysed were selected from a set of active breeding males and their male offspring at the Tulane National Primate Research Center (TNPRC) colony. In total, the Tulane NPRC rhesus macaque colony consists of more than 4000 animals that are managed as several distinct multimale, multifemale breeding groups. Males are occasionally exchanged between breeding groups to maintain genetic health and natural social behaviour. The full history of the TNPRC rhesus macaque population includes many more animals going back multiple generations to the 1960s, but complete pedigree records are not available. Males were selected
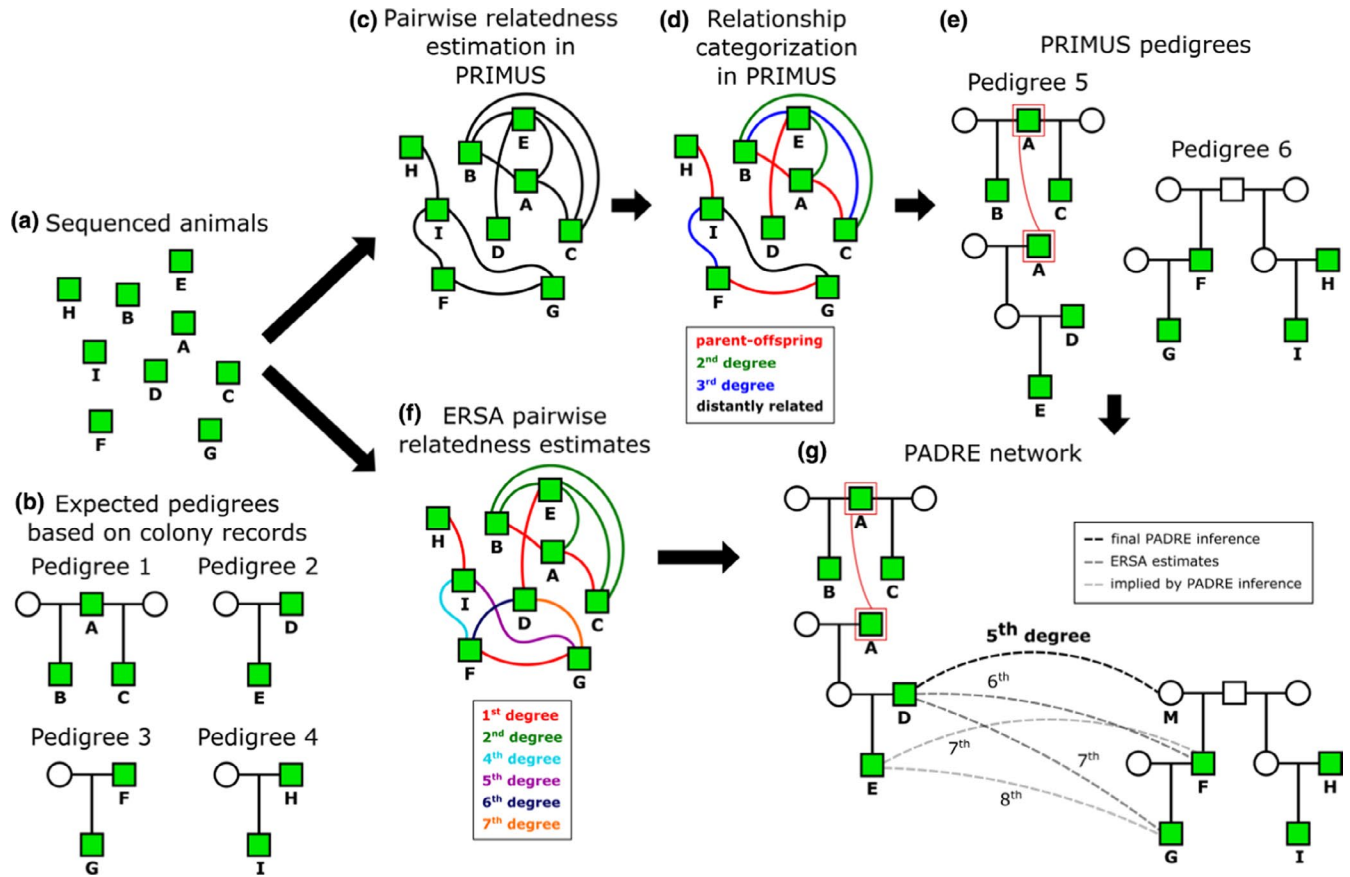
FIGURE 1 Example of pedigree reconstruction and distant relatedness estimation. This figure illustrates, for a single PADRE network, the process of going from individual sequenced and QCed samples (a), to PADRE network (g). Expected pedigrees based on colony records are given in (b). Step (c) depicts estimation of pairwise relatedness based on genome-wide identical-by-descent (IBD) proportions. In PRIMUS, this step uses estimates from PLINK, however other tools could be used, including those that use genotype-likelihood data such as NgsRelate (Korneliussen & Moltke, 2015). Step (d) shows the categorization of each pairwise relationship in PRIMUS, using a trained kernel density estimation for each of relationship category. Then, PRIMUS reconstructs all possible pedigree structures that match the relationships present, shown in step (e). At this stage, we compared these pedigrees to colony records, including animal date of birth, to select the pedigree that matched this data most closely. Step (f) shows, in a separate pipeline, the estimation of pairwise relatedness in estimation of recent shared ancestry (ERSA), using IBD segments identified in the sequencing data. Here, we used GERMLINE to identify these segments (used as input to ERSA), but other tools, including hap-ibd (Zhou et al., 2020), RAPID (Naseri et al., 2019), or TRUFFLE (Dimitromanolakis et al., 2019), could also be used. Finally, the PRIMUS pedigree structures and the ERSA pairwise relatedness estimates are combined to improve estimates of distant relatedness between the pedigrees in PADRE in step (g). In this step, all distant pairwise relationships between genotyped individuals detected by ERSA (depicted by medium-grey dotted lines) are leveraged to maximize the likelihood of the relationship estimate between founders in each pedigree, D and M. The light-grey dotted lines depict additional distant relationships between the pedigrees implied by the PADRE inference, not detected in ERSA. Additional details for steps c–g are available in the original methods manuscripts (PRIMUS, ERSA, and PADRE). The red boxes with connecting lines indicate that the same individual is depicted twice in the same pedigree. Sequenced samples are highlighted in green

for this study because higher ranking males sire a disproportionately large proportion of all births, and their male offspring were included because enriching the sample population for expected first-degree relationships helps to anchor distant relationship detection by providing more pairwise relationship comparisons (Figure 1 gives an illustration of how close pedigree relationships improve distant relatedness estimation). Available colony records included the known dam for each animal and either the known sire or likely sires based on breeding group composition. These records were used to construct a set of expected relationships for each pair of animals.

2.2 | Sequencing and quality control

DNA sequence data for 150 TNPRC rhesus macaques were generated via whole genome sequencing using 2 × 150 bp reads generated on the Illumina HiSeq X platform (workflow shown in Figure 2). The average genome-wide read coverage across samples exceeded 35×. Reads were mapped to the rhesus reference genome (NCBI Mmul_8.0.1) using BWA-MEM and SNPs were called using GATK with standard quality filtering and workflow processes (DePristo et al., 2011; Li, 2013; McKenna et al., 2010; Van der Auwera et al., 2013). We compared observed and expected X-chromosome heterozygosity to check sample
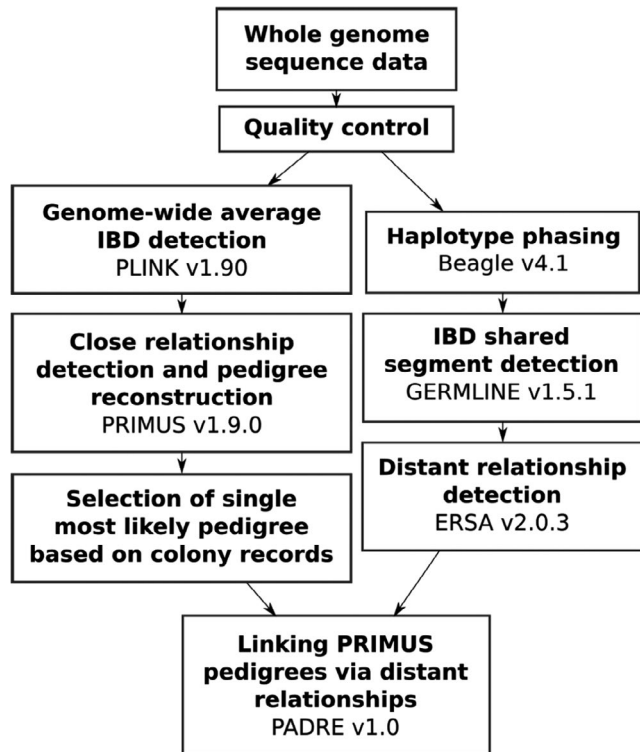
FIGURE 2 Analysis workflow for genetic determination of relatedness within a colony of rhesus macaques from the Tulane National Primate Research Center. This flow chart outlines our analytic approach, including software that was used. The left branch gives the steps used for close relationship detection based on genome-wide IBD proportions between each pair of animals; the right branch gives the steps used for distant relationship detection via pairwise identical-by-descent (IBD) segmental sharing. The final step, PADRE, combines these results and reclassifies distant relationships based on maximizing likelihood across pedigree structures identified in PRIMUS

sex, as well as comparing genome-wide heterozygosity F coefficients for all samples (Figure S1). Finally, to verify sample ancestry, we merged colony data with a reference data set of 133 unrelated rhesus macaques, including nine known Chinese-origin samples, and performed principal components analysis using PLINK version 1.9 (Figure 3; Chang et al., 2015; Xue et al., 2016). Due to allele frequency differences between Indian and Chinese rhesus macaques that can bias close relationship detection, all Chinese-origin samples were removed from further analysis. After excluding four Chinese-origin and two extreme heterozygosity Indian-origin animals, we retained a total of 144 samples for our final analysis. Variant filtering included removal of low frequency (MAF < 1%) or high missingness (>5%) variants. Of 39,193,704 autosomal variants genotyped, 19,237,207 remained after all filtering steps and were included in our PRIMUS analysis. For our ERSA analysis, the IBD segment detection method used is sensitive to marker density (Li et al., 2014), and our sequence data is substantially more dense than typical human genotype array data (we observe a mean pair difference of ~26,900 per cM, compared to ~454 per cM for Illumina Expanded Multi-Ethnic Genotyping Array [MEGA] human data). Thus, we performed additional variant filtering to more closely match the marker density observed in humans. To do so, we used GATK, excluding any variants with quality by depth <10.0, strand bias using Fisher's exact test >2.0, root mean square of mapping quality <59.0 or >61.0, strand bias using the symmetric odds ratio test >1.5, or MAF <45%, retaining a total of 989,979 variants.

2.3 | Data analysis

The cleaned and filtered data were used to generate genome-wide average IBD proportions for each pair of individuals using PLINK version 1.9 within the pre-PRIMUS pipeline (Chang et al., 2015). Other tools can also be applied for IBD estimation, including methods that allow for use of genotype likelihood data. Close relationship detection and pedigree reconstruction was performed by supplying IBD proportion estimates to PRIMUS version 1.9.0 (Staples et al., 2014). We first reconstructed all pedigrees of first-degree relatives ($\hat{\pi} > 0.375$), then combined the most closely related first-degree pedigrees into groups of two or three pedigrees and reconstructed these combined samples to third-degree ($\hat{\pi} > 0.09375$). Using the data filtered for ERSA, haplotype phasing was completed in Beagle version 4.1 using default settings (Browning & Browning, 2007). GERMLINE version 1.5.1 was used to detect IBD shared segments for all animals; we allowed a maximum of one mismatching heterozygous marker and two mismatching homozygous markers in each identified IBD segment, and required a minimum segment length of 2.5 cM (Gusev et al., 2009). Conversion to cM location assumed the average genome-wide recombination rate observed in a prior study of rhesus macaques, 0.433 cM/Mb (Xue et al., 2016). ERSA version 2.0.33 was used to detect distant relationships, with the number of chromosomes set to 20 and recombination events per meiosis set to 13.6. Recombination events per meiosis were calculated by taking the observed recombination rate in Xue et al. (2016) (0.433 cM/Mb) and multiplying by the total length of the sequencing data in Mb. To characterize the sensitivity of our ERSA estimates to an accurate estimate of the recombination rate, we performed additional ERSA analyses, varying the number of recombination events per meiosis from 6.0 to 30.0. Using PADRE version 1.02, we identified second- through ninth-degree relationships between pairs of individuals without genotyped parents ("founders") in different PRIMUS pedigrees (Staples et al., 2016). This approach jointly considers the likelihood of all distant relationships implied between each pedigree by a certain degree of relatedness between a founder from each of the pedigrees to more accurately determine the founder relationship. To characterize the accuracy of our genetically determined results, we compared the expected relationship status to the genetically determined relationship for each pair.

2.4 | Downsampled analyses

To broaden the utility of this report to applications with lower coverage sequencing data, we downsampled the sequencing data to
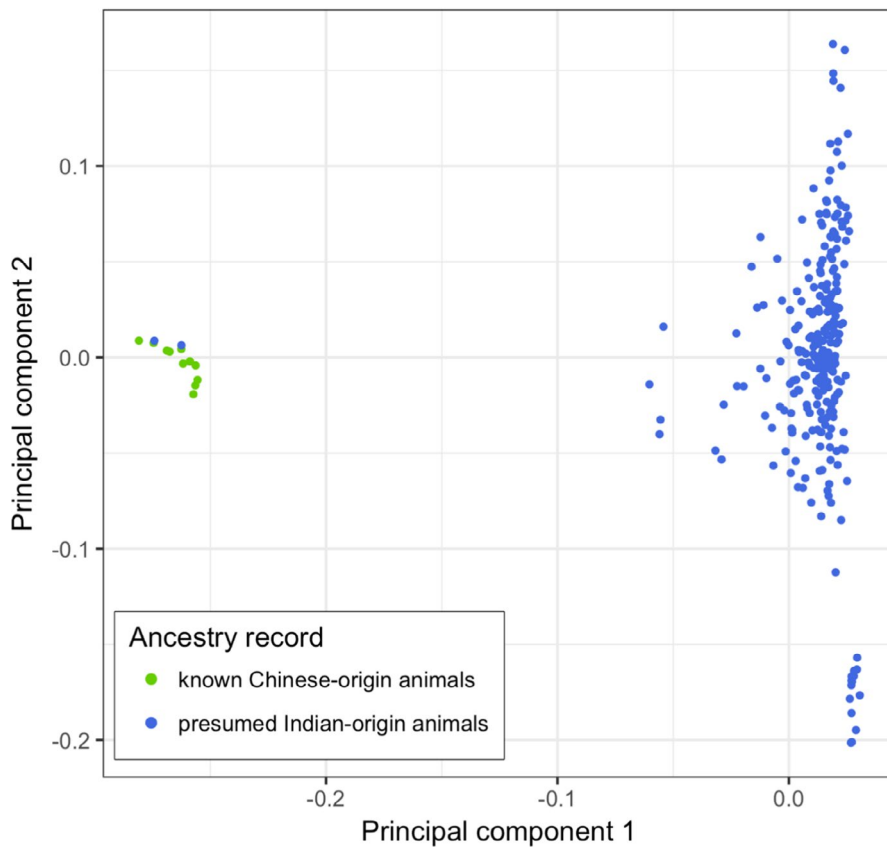
FIGURE 3 Principal component plot illustrating sample ancestry for a colony of rhesus macaques from the Tulane National Primate Research Center merged with known ancestry reference animals. Principal component analysis (PCA) was performed for the colony data set merged with a reference data set of animals of known ancestry, resulting in the identification of two additional Chinese-origin samples within the Tulane colony

average coverage of 5× and 3× using samtools and performed the same analyses as in the full coverage data. Quality control and analysis of the downsampled data followed the procedures outlined for the primary analysis, with the exception of the variant filtering steps. For the 5× analysis, variants with read depth >10 or genotype quality >14 were included in the PRIMUS analysis, retaining a total of 33,436,831 variants. For the 3× analysis, variants with read depth >7 or genotype quality >14 were included in the PRIMUS analysis, retaining a total of 29,059,190 variants. PRIMUS typically downsamples densely genotyped data at random to reduce computational load; however, for these analyses, because of the reduced the number of variants that passed filtering steps, we removed this filter. ERSA analyses were performed for the 5× downsampled data, using IBD segments generated from GERMLINE allowing for 19 mismatching heterozygous variants and five mismatching homozygous variants.

3 | RESULTS

3.1 | Genetically determined relationships

To illustrate the methods used, we will first walk through an example of pedigree reconstruction for close relationships and distant relationship inference anchored by the genetically determined pedigrees for a single PADRE network. For a group of nine animals (Figure 1a), colony records gave the expected relationship structures shown in Figure 1b, constituting four sire-offspring pedigrees. Using PRIMUS, we determined that the same animals comprise two closely related networks (Figure 1c–e), identifying close cryptic relatedness among the four expected pedigrees. Expected pedigrees 1 and 2 are connected to form PRIMUS pedigree 5 (Figure 1e), with a grandparental relationship between individuals A and E and half-avuncular relationships between pairs B–E and C–E. Expected pedigrees 3 and 4 are connected to form PRIMUS pedigree 6, with the discovery of a cryptic half-sibship between individuals F and H. These networks and distant relationship predictions from ERSA (Figure 1f) were supplied to PADRE for inference of distant relationships between founders of each pedigree, and PADRE identified a fifth-degree relationship between founder D from pedigree 5 and missing founder M from pedigree 6 ($p = 4.06 \times 10^{-4}$, Figure 1g). This relationship can be more confidently asserted as fifth-degree due to evaluation in the context of the closely related pedigrees for each founder, with a fifth-degree relationship between founder D and missing founder M implying additional relationships not identified by ERSA of seventh-degree between individuals E and F and eighth-degree between individuals E and G.

We applied PRIMUS, ERSA, and PADRE across the entire genotyped sample, and in total, using PADRE, we identified 2749 relationships first to twelfth-degree among the included animals (Table 1, Figure S2). We identified 85 parent-offspring (first-degree), 36 second-degree, 14 third-degree, and four fourth-degree relationships using PRIMUS. These close relationships comprised 54 sire

Genetically determined degree of relatedness	PRIMUS results[a]	ERSA results	Combined PADRE results
Parent-offspring	85	85	85
Full sibling	0	0	0
2	36	197	91
3	14	178	215
4	4	368	369
5	—	536	570
6	—	387	674
7	—	143	386
8	—	9	140
9	—	—	58
10	—	—	102
11	—	—	49
12	—	—	10

[a]PRIMUS only detects relationships that can be connected by non-missing relationships up to third-degree; ERSA only detects relationships up to eighth-degree. Dashes in all columns indicate that this degree of relationship was not assessed.

groups, i.e., pedigrees for each founder sire including only first-degree relationships. The largest sire group included six animals, and the smallest included only two animals in a single parent-offspring relationship. Combining the most closely related sire groups to create larger anchoring pedigrees for PADRE resulted in 47 pedigrees (https://github.com/belowlab/tnprc-pedigrees), containing a mean of 2.96 animals. We identified 1809 relationships from first- to eighth-degree using ERSA, based on pairwise IBD segmental sharing. We note that for the PADRE results, ninth-degree relationships are beyond the limit of genetic detection using ERSA, so we can only observe ninth and greater degree relationships when they are linked through PRIMUS close relationships. Therefore, we expect to observe fewer ninth- and greater degree relationships relative to other distant relationships.

We also explored the sensitivity of ERSA estimates to potential error in the recombination rate. We varied the estimated number of recombination events from what has been previously observed (13.6), to 30.0. We observed that small variation in the estimated number of recombination events per meiosis does not dramatically impact the inferred degree of relatedness (Figure 4). The results indicate that if the number of recombinations per meiosis is overestimated by as much as 65%, the proportion of pairwise estimates
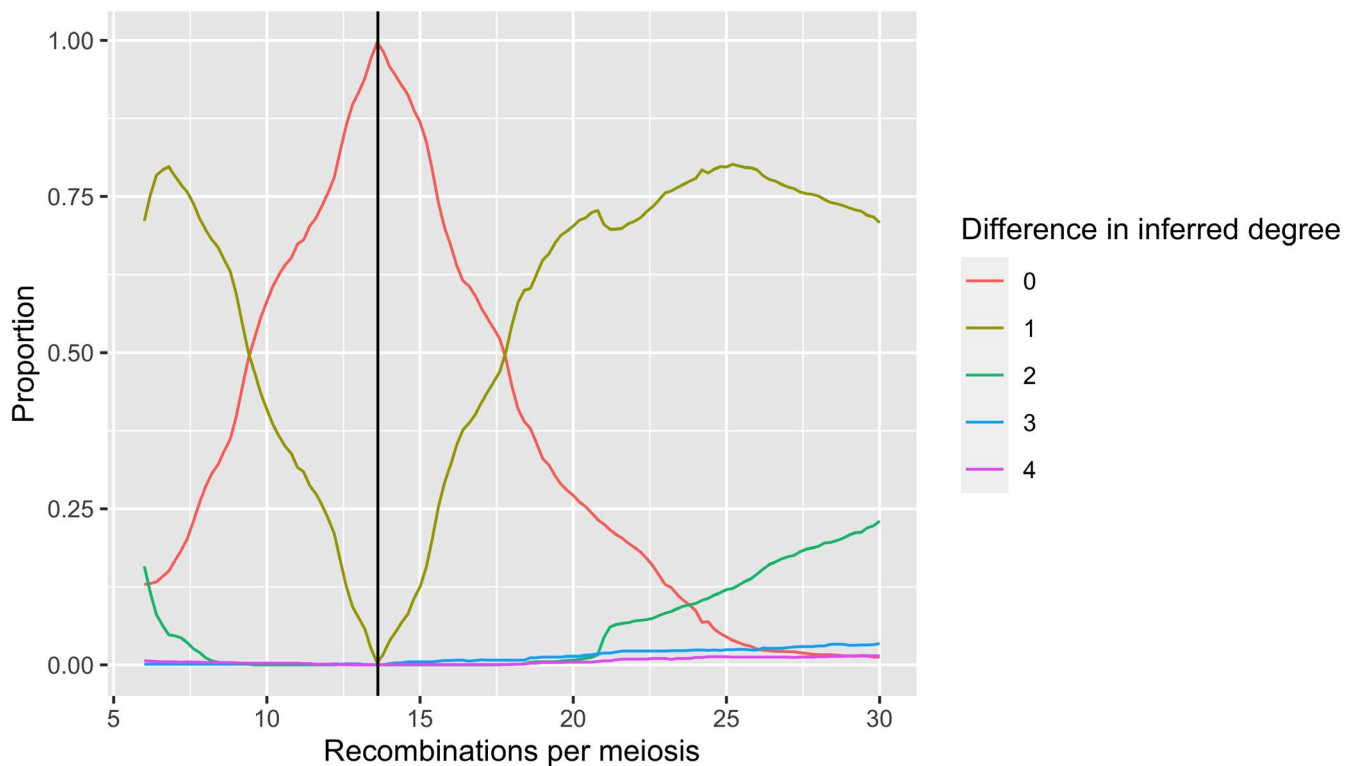
FIGURE 4 Comparison of inferred degree of relatedness by the number of recombination events per meiosis supplied to estimation of recent shared ancestry (ERSA). For each pair of animals predicted to be related by ERSA in the primary analysis, the proportion of pairs with each difference in the inferred degree of relatedness is plotted for different numbers of recombinations per meiosis supplied to ERSA, compared to the inferred degree of relatedness based on the previously observed recombination rate in rhesus macaques (13.6, denoted by the black vertical line here)

of relatedness that are correct or within one degree of relatedness remains 89.6%. However, if the number of recombination events is off by approximately two times the true recombination rate, 97.8% of inferred degrees of relatedness will be inaccurate, though 95.8% of these predictions were within two degrees of the gold standard estimate. If the number of recombinations per meiosis is under-estimated by as much as one half of the true recombination rate, 85.9% of inferred degess of relatedness will be inaccurate, though 94.8% of these predictions will be within one degree of relatedness (Figure 4 shows results for all pairs of animals inferred to be related in the original ERSA estimates, Figure S3 shows results for all pairs of animals, including those inferred to be unrelated.)

3.2 | Comparison to colony records

We compared our genetically determined sex, ancestry and relation-ships to colony records. All animals were genetically male, consist-ent with colony records. We identified two additional Chinese-origin animals and verified primarily Indian-origin ancestry in the remaining animals (Figure 3). Twenty-one animals were expected to be unre-lated to the rest of the colony based on records, and all of these animals were inferred to be unrelated based on genotypes. Table 2 presents, for each pair of animals, the degree of relatedness that was expected from colony records and the degree of relatedness that was genetically determined from PADRE. Concordant relationships lie on the diagonal, with off-diagonal cells representing relationships that were not validated (either due to record misspecification or sample swap), or relationships that were not present in the colony records. It is also conceivable that discordant relationships could

be due to the genetically determined relationship being misspeci-fied, though this is extremely unlikely for close relationships such as those examined here (Staples et al., 2014). Genetically determined parent-offspring relationships were overwhelmingly concordant (95.5%) with expected relationships, although one expected parent-offspring relationship was determined to be unrelated and one was determined to be a third-degree relationship. No full-sibling rela-tionships were expected or observed. We expected 28 half-sibling relationships based on colony records and identified 25 concordant genetically determined half-sibships. Two expected half-sibships were found to be related at a greater degree, and one expected half-sibship was genetically unrelated. We also identified a total of 66 additional second-degree relationships.

3.3 | Downsampled analyses

We next tested the power and accuracy of this approach using low-coverage sequence data. Using genotype calls based on the same Illiumina sequencing reads downsampled to an average 5× genome wide coverage, we found that although IBD proportion estimates were impacted by the downsampling (Figure 5), the PRIMUS analysis of first-degree relationships were identical to those using the full deep coverage sequencing. When the sequence reads were downsampled to genome wide 3× coverage, PRIMUS produced different (incorrect) pedigrees in 16 of the original 52 pedigrees, compared to the pri-mary results.
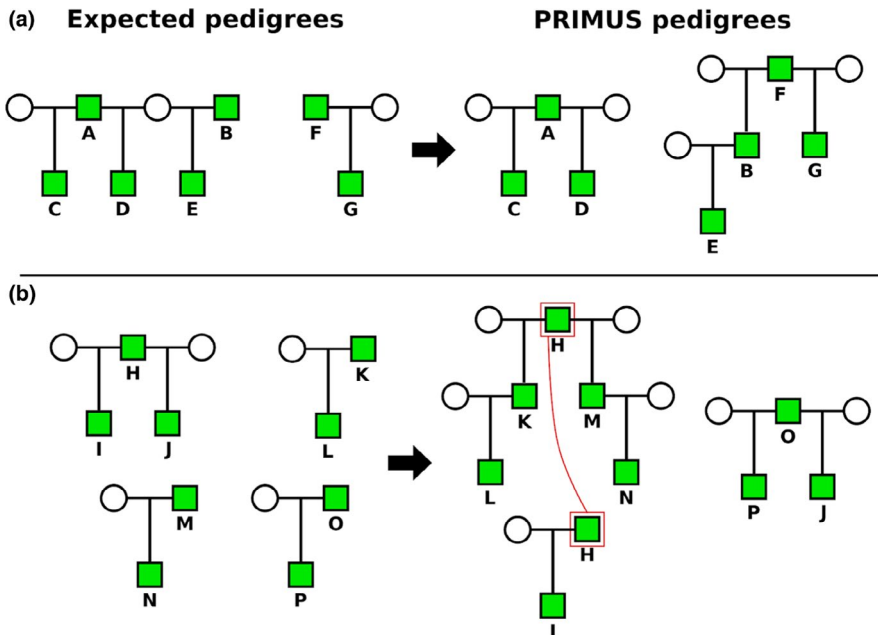
Given the discrepancies in even close relationships for the 3× downsampled coverage data, ERSA analyses were only undertaken in the 5× downsampled data. We compared the degree of the rela-tionship from PRIMUS and inferred degree of relatedness from ERSA based on the 5× downsampled data to analogous estimates from the original full-coverage data (Tables 3 and 4). Sixteen relationships from PRIMUS pedigrees that could be reconstructed with the full-cov-erage data could not be reconstructed with the 5× downsampled data. We observed some deflation of ERSA-inferred relatedness in the downsampled analysis, i.e., pairs of animals observed to be re-lated at a higher degree in the original analyses were predicted to be somewhat more distantly related in the downsampled analysis, especially among pairs identified as more that fourth-degree rela-tives in the full-coverage data. However, we do note that most errors consisted of overestimating pairwise relatedness by just one degree.

4 | DISCUSSION

As access to whole genome, RAD-Seq or SNP array methods has improved, an increasing number of investigators have begun to use SNPs rather than microsatellites for the analysis of parentage and pairwise relatedness. The cost per sample of generating large-scale whole genome or RAD-seq SNP data is still higher than that of a mod-est number of microsatellite genotypes, but the difference in cost is declining and the additional value of SNPs, e.g., information about

TABLE 2 Comparison of genetically determined relationships with presumed relationships based on colony records among rhesus macaques from the Tulane National Primate Research Center colony

Genetically determined degree of relatedness from PADRE	Presumed degree of relatedness from colony records		
	1	2	Unknown
1	81	0	4
2	0	25	66
3	1	1	213
4	0	1	368
5	0	0	570
6	0	0	674
7	0	0	386
8	0	0	140
9	0	0	58
10	0	0	102
11	0	0	49
12	0	0	10
Unrelated	1	1	7545

FIGURE 5 Illustration of nonconcordance of colony records and genetically determined relationships. For two half-sibships that were nonconcordant between colony records and final PADRE results, we show, on the left, expected pedigrees based on colony records. On the right, we show the genetically determined pedigrees from PRIMUS. Sequenced samples are highlighted in green

Furthermore, methods related to low quality DNA, such as DNA from faecal samples, are improving. This will provide an opportunity to use noninvasive sample collection methods and therefore significantly expand the range of populations that can be analysed (Chiou & Bergey, 2018; Snyder-Mackler et al., 2016; White et al., 2019).

Using genetic data from 150 rhesus macaques, we applied IBD-based methods of pedigree reconstruction and distant relationship inference and compared the identified relationships against colony records. We observed nearly complete concordance between the genetically determined relationships and those expected based on colony records (95.5% of expected relationships were verified by IBD), with only five conflicting relationships. For first-degree relationships, this echos previous studies that have shown that parentage testing is reliable using just a few thousands SNPs (Flanagan & Jones, 2019; Foroughirad et al., 2019; Snyder-Mackler et al., 2016). We also inferred degrees of distant relationship between founders of each of the closely related pedigrees, demonstrating our ability to predict relationships as distant as twelfth-degree in rhesus

TABLE 3 Comparison of downsampled 5× PRIMUS results to full-coverage 35× PRIMUS results

Degree of relatedness in PRIMUS 5× results	Degree of relatedness in full-coverage PRIMUS results			
	1	2	3	4
1	85	0	0	0
2	0	32	0	0
3	0	0	5	0
4	0	0	0	1
NA	0	13	3	0

Abbreviation: NA, not assessed

predicted functional variation in protein-coding sequences as well as the potential for inferences about more distant relationships has been recognized (Foroughirad et al., 2019; Stadele & Vigilant, 2016). In addition, some populations such as genetically depauperate endangered or captive populations may not present adequate microsatellite variation.

TABLE 4 Comparison of downsampled 5× estimation of recent shared ancestry (ERSA) results to full-coverage 35× ERSA results

| Degree of relatedness in ERSA 5× results	Degree of relatedness in full-coverage ERSA results								
	1	2	3	4	5	6	7	8	Unrelated
1	85	0	0	0	0	0	0	0	0
2	0	52	0	0	0	0	0	0	0
3	0	110	6	0	0	0	0	0	0
4	0	34	156	61	0	0	0	0	0
5	0	0	14	269	110	1	0	0	0
6	0	0	0	32	319	147	10	0	20
7	0	0	0	0	77	185	98	4	150
8	0	0	0	0	4	12	24	5	198
Unrelated	0	0	0	3	25	42	11	0	7860

macaques and to identify 2640 relationships not present in colony records. The initial expected relationships comprised only 4.0% of the total identified relationships.

The observed nonconcordance of some expected and observed relationships can be explained by pedigree misspecification or sample swaps (two or more samples that were misidentified as another sample due to an error in sample labelling or handling). We observed two expected half-sibships that were actually distantly related. As illustrated in Figure 5a, for one of these sibships, we observe that the presumed shared dam of D and E was misspecified for one, and observe a cryptic sire-offspring relationship between F and B. For the other sibship (Figure 5b), based on genetically determined relationships, we observe that H was misspecified as the sire of J and also observe cryptic sire-offspring relationships between pairs H–K and H–M. Because genetically determined relationships are remarkably accurate given sufficient data (>90% accuracy even with pedigree missingness up to 40% (Staples et al., 2014), whereas missingness in our closely related pedigrees is far below this) for first- and second-degree relationships, we believe that they represent the true pedigree structure in this case, providing an example of how our methods can augment known relationship records.

Changes required to apply our methods to rhesus macaques were minimal. For inferences of close relationships, we did not observe any systematic bias in the inferred degree of relationship due to differing patterns of genetic variation in rhesus macaques relative to humans. Therefore, no adjustment to PRIMUS procedures used for human data was required. For distant relationship inference using PADRE, we only had to adjust model input parameters for recombination rate and

expected segmental sharing in unrelated population samples in ERSA to match segmental sharing patterns in rhesus macaques. We did observe that ERSA identified more second-degree relationships than PADRE did. This may be due to the fact that PADRE does not allow for multiple common ancestors when connecting closely related pedigrees. Alternatively, since the IBD proportion estimates used by PRIMUS to infer degree of relatedness are more accurate than the shared segments used by ERSA for closely related pairs, ERSA may be inaccurately characterizing pairs as closely related, and incorporating a lack of evidence of close relatedness from PRIMUS in PADRE may indicate that the pair is actually related at a greater degree.

We note that a limiting factor for some applications of these techniques may be the quantity of genetic data available per sample, which was not a major concern in our high-coverage data. To explore the applicability of our methods to lower coverage data, we undertook additional analyses with data downsampled to 5× and 3× coverage. We then performed systematic depth and genotype quality filtering in our 5× downsampled analysis and plotted IBD proportion estimates against the "gold standard" IBD proportion estimates from the primary analysis (Figure 6). Our final filtering scheme for our 5× downsampled data did result in some bias in the IBD proportion estimates, however this did not ultimately impact the first-degree pedigrees reconstructed. It is clear from these analyses that lower-coverage data is more sensitive to quality control procedures. Therefore, careful filtering to retain only high quality variants is important when applying these methods in low coverage data to obtain accurate IBD proportion estimates. Approximately 6,000 high quality variants are needed in common for each pair of individuals
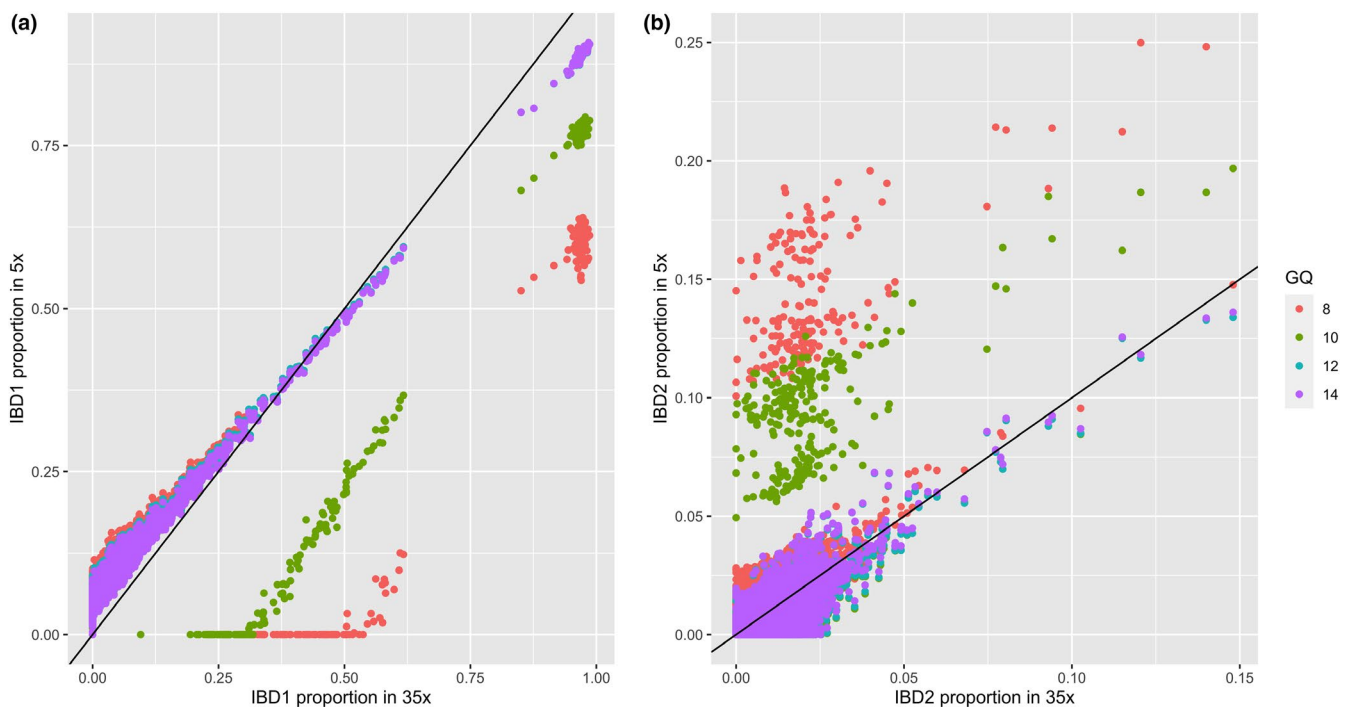
FIGURE 6 Comparison of identical-by-descent (IBD) proportion estimates in original and downsampled sequencing data. For our 5× downsampled analysis, we compare here the (a) IBD1 and (b) IBD2 proportion estimates from PLINK for all pairs of animals to the original IBD proportion estimates from PLINK in the primary, 35× coverage, analysis

in humans to get accurate IBD proportion estimates in PLINK (Staples et al., 2015); this number may vary depending on species total genome size and the IBD estimation approach chosen.

Using ERSA to identify relatedness from IBD segment sharing is also applicable in lower coverage data, and depending on the IBD segment identification tool used, can be more tolerant to genotyping error due to low coverage. GERMLINE, which we applied here, allows the user to specify the number of mismatching variants to tolerate. Compared to our "gold standard" full-coverage results, accuracy is reduced; among pairs predicted to be related in either the full-coverage or downsampled analyses, 74.0% of pairs were inferred to be related within one degree of the full-coverage estimate. For many applications of distant relatedness, this may be sufficient, as the difference between, for example, fifth- and sixth-degree relatedness, may be less important than the simple observation of some degree of distant relatedness.

There are of course some limitations to our present study. We sampled only two generations of animals, meaning that we were missing common ancestors in the inferred pedigrees. Greater inclusion of missing common ancestors will decrease the level of uncertainty in our distant relationship inference (Staples et al., 2016; Taylor, 2015). Additionally, expected relationship status is only available for close relationships, meaning we are unable to definitively verify the accuracy of the distant relationships we identified based on colony records. However, simulations in human data have demonstrated the utility of the PADRE method to identify true distant relatedness through 13th-degree (Staples et al., 2016). Furthermore, any ability to ascertain distant relationships represents an improvement over the current state of the field, and accuracy will improve as better characterization of population genetic parameters specific to rhesus macaques (or any subject species) occurs. Finally, since the inference of distant relationships relies on having genetic reference data, including both a reference genome similar enough to the species under study that reliable SNP genotypes can be produced, and accurate estimates of recombination rate in the species, our results inferring very distant relationships may not be representative of what we would observe for species lacking both. However, reference genome assemblies for more species are being generated at a rapid pace, and recombination rates generally do not differ significantly among closely related species (Stevison et al., 2016). If the number of animals studied is large enough (previous estimates in chimpanzees have used as few as 10 individuals; Auton et al., 2012), it is possible to estimate recombination rate within the study sample as well. In this case, we recommend performing the calculation of recombination rate in the maximally unrelated subset from PRIMUS to remove any close relatedness in the data. Galla et al. (2019) have examined the impact of using congeneric, confamilial, and conordinal reference genomes for SNP discovery on pairwise estimates of close relatedness, and find that estimates of relatedness are highly correlated with estimates using a conspecific reference genome. Our comparisons of ERSA-inferred relatedness with varied recombination rates demonstrated that inferred relatedness will be accurate within two degrees of relatedness even when the number of recombinations per meiosis is incorrect by a factor of two (Figure 4). Additionally, close relationship determination would be applicable and accurate even when estimates of recombination rates are not available. Using PRIMUS alone, distant relationships cannot be estimated for all pairs of individuals, but any distant relationships interlinked by pairs of closer relationships would still be identified.

Our results represent a step forward toward powerful and widely applicable strategies to estimate pairwise relatedness among sets of animals when pedigree relationships are not known. While we have demonstrated the utility of this approach in a population of captive macaques, the methods will be applicable to any diploid species (vertebrate or invertebrate) for which suitable genotype data can be generated. The approach presented here also allows for verification of relationship status when expected relationships are available and permits correction when any inconsistencies are found. This also enables inference of close pedigree structure among animals where relationship status is unknown but relatedness is presumed, e.g., in the case of the majority of our identified second- and third-degree relationships. Finally, we demonstrate that we are able to infer distant relationships between pedigree founders, who are often presumed unrelated due to lack of information; this assumption is rarely founded and can have profound impacts on colony management or inferences drawn from observational pedigrees in wild populations (Hogg et al., 2019). These results illustrate that use of our methods results in a more complete and accurate assessment of relatedness in nonhuman study subjects, enabling expanded studies of natural populations, improved management of captive colonies, more effective study of genetic risk for disease and increased power to determine pedigree relationships in any populations where that information is not available. Future advances in sequencing or genotyping technologies, reduction in SNP genotyping costs and improvements in protocols for the use of non-invasively collected samples will certainly increase the potential of this approach.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

L.E.P. performed quality control of genotype data and conducted all genetic relationship inference analyses. K.P-F., H.M.K., and E.J.V. identified animals to sample, verified demographic data including observed and/or genetic birth parents to determine possible relationships, collected and extracted DNA. The process of DNA sequencing was overseen by M.R. M.R. and R.A.H. analysed whole genome sequence data to generate genotype calls. R.P.B. is PI of

the grant supporting the colony and contributed to study design. J.R. and J.E.B. conceived and designed the study. J.E.B. and C.D.H. participated in the development of the software tools employed for pedigree reconstruction and distant relatedness inference and advised on their use. L.E.P., J.R. and J.E.B. wrote the paper. All authors reviewed and commented on the paper.

DATA AVAILABILITY STATEMENT

All the whole genome sequence data from this study has been deposited in the NCBI sequence archive, with Biosample accessions SAMN07168963–SAMN07169031 and SAMN11842252–SAMN11842332. The Variant Call File (VCF) that presents all the genotypes generated in used for pedigree and relationship analyses employing the deep sequencing coverage is available at https://zenodo.org/record/4066891. Code used for analysis is available at https://github.com/belowlab/tnprc-pedigrees.

ORCID

Lauren E. Petty https://orcid.org/0000-0003-1619-6303
R. Alan Harris https://orcid.org/0000-0002-7333-4752
Eric J. Vallender https://orcid.org/0000-0003-3506-0540

REFERENCES

Ackermann, R. R., Schroeder, L., Rogers, J., & Cheverud, J. M. (2014). Further evidence for phenotypic signatures of hybridization in descendant baboon populations. *Journal of Human Evolution*, *76*, 54–62. https://doi.org/10.1016/j.jhevol.2014.05.004

Alberts, S. C., & Altmann, J. (2012). The Amboseli Baboon Research Project: 40 years of continuity and change. In P. M. Kappeler & D. P. Watts (Eds.), *Long-term field studies of primates* (pp. 261–288). Springer-Verlag.

Altmann, J. (1980). *Baboon mothers and infants*. University of Chicago Press.

Andrews, K. R., Adams, J. R., Cassirer, E. F., Plowright, R. K., Gardner, C., Dwire, M., Hohenlohe, P. A., & Waits, L. P. (2018). A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. *Molecular Ecology Resources*, *18*(6), 1263–1281. https://doi.org/10.1111/1755-0998.12910

Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Segurel, L., Street, T., Leffler, E. M., Bowden, R., Aneas, I., Broxholme, J., Humburg, P., Iqbal, Z., Lunter, G., Maller, J., Hernandez, R. D., Melton, C., Venkat, A., Nobrega, M. A., Bontrop, R., ... McVean, G. (2012). A fine-scale chimpanzee genetic map from population sequencing. *Science*, *336*(6078), 193–198. https://doi.org/10.1126/science.1216872

Bercovitch, F. B. (1997). Reproductive strategies of rhesus macaques. *Primates*, *38*(3), 247–263.

Bimber, B. N., Ramakrishnan, R., Cervera-Juanes, R., Madhira, R., Peterson, S. M., Norgren, R. B. Jr, & Ferguson, B. (2017). Whole genome sequencing predicts novel human disease models in rhesus macaques. *Genomics*, *109*(3–4), 214–220. https://doi.org/10.1016/j.ygeno.2017.04.001

Browning, B. L., & Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *American Journal of Human Genetics*, *88*(2), 173–182. https://doi.org/10.1016/j.ajhg.2011.01.010

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Cheney, D. L., & Seyfarth, R. M. (1983). Nonrandom dispersal in free-ranging vervet monkeys: Social and genetic consequences. *American Naturalist*, *122*, 392–412.

Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, *8*(1), 1975. https://doi.org/10.1038/s41598-018-20427-9

Conomos, M. P., Miller, M. B., & Thornton, T. A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genetic Epidemiology*, *39*(4), 276–293. https://doi.org/10.1002/gepi.21896

Conomos, M. P., Reiner, A. P., Weir, B. S., & Thornton, T. A. (2016). Model-free estimation of recent genetic relatedness. *American Journal of Human Genetics*, *98*(1), 127–148. https://doi.org/10.1016/j.ajhg.2015.11.022

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

de Waal, F. B., & Tyack, P. L. (Eds.) (2003). *Animal social complexity*. Harvard University Press.

Dimitromanolakis, A., Paterson, A. D., & Sun, L. (2019). Fast and accurate shared segment detection and relatedness estimation in un-phased genetic data via TRUFFLE. *American Journal of Human Genetics*, *105*(1), 78–88. https://doi.org/10.1016/j.ajhg.2019.05.007

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, *28*(3), 544–567. https://doi.org/10.1111/mec.14988

Foroughirad, V., Levengood, A. L., Mann, J., & Frere, C. H. (2019). Quality and quantity of genetic relatedness data affect the analysis of social structure. *Molecular Ecology Resources*, *19*(5), 1181–1194. https://doi.org/10.1111/1755-0998.13028

Galla, S. J., Forsdick, N. J., Brown, L., Hoeppner, M. P., Knapp, M., Maloney, R. F., Moraga, R., Santure, A. W., & Steeves, T. E. (2019). Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes*, *10*(1), 9.

Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., Remington, K. A., Strausberg, R. L., Venter, J. C., Wilson, R. K., Batzer, M. A., Bustamante, C. D., Eichler, E. E., Hahn, M. W., Hardison, R. C., Makova, K. D., Miller, W., Milosavljevic, A., Palermo, R. E., ... Zwieg, A. S. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, *316*(5822), 222–234. https://doi.org/10.1126/science.1139247

Goudet, J., Kay, T., & Weir, B. S. (2018). How to estimate kinship. *Molecular Ecology*, *27*(20), 4121–4135. https://doi.org/10.1111/mec.14833

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M., & Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, *19*(2), 318–326. https://doi.org/10.1101/gr.081398.108

Hogg, C., Wright, B., Morris, K., Lee, A., Ivy, J., Grueber, C., & Belov, K. (2019). Founder relationships and conservation management: empirical kinships reveal the effect on breeding programmes when founders are assumed to be unrelated. *Animal Conservation*, *22*(4), 348–361.

Huff, C. D., Witherspoon, D. J., Simonson, T. S., Xing, J., Watkins, W. S., Zhang, Y., Tuohy, T. M., Neklason, D. W., Burt, R. W., Guthery, S. L., Woodward, S. R., & Jorde, L. B. (2011). Maximum-likelihood

estimation of recent shared ancestry (ERSA). *Genome Research*, *21*(5), 768–774. https://doi.org/10.1101/gr.115972.110

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, *17*(5), 1009–1024. https://doi.org/10.1111/1755-0998.12665

Ishizuka, S., Takemoto, H., Sakamaki, T., Tokuyama, N., Toda, K., Hashimoto, C., & Furuichi, T. (2020). Comparisons of between-group differentiation in male kinship between bonobos and chimpanzees. *Scientific Reports*, *10*(1), 251. https://doi.org/10.1038/s41598-019-57133-z

Johnson, Z., Brent, L., Alvarenga, J. C., Comuzzie, A. G., Shelledy, W., Ramirez, S., Cox, L., Mahaney, M. C., Huang, Y.-Y., Mann, J. J., Kaplan, J. R., & Rogers, J. (2015). Genetic influences on response to novel objects and dimensions of personality in Papio baboons. *Behavior Genetics*, *45*(2), 215–227. https://doi.org/10.1007/s10519-014-9702-6

Kanthaswamy, S., von Dollen, A., Kurushima, J. D., Alminas, O., Rogers, J., Ferguson, B., Lerche, N. W., Allen, P. C., & Smith, D. G. (2006). Microsatellite markers for standardized genetic management of captive colonies of rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, *68*(1), 73–95. https://doi.org/10.1002/ajp.20207

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics*, *31*(24), 4009–4011. https://doi.org/10.1093/bioinformatics/btv509

Krutzen, M., Sherwin, W. B., Connor, R. C., Barre, L. M., Van de Casteele, T., Mann, J., & Brooks, R. (2003). Contrasting relatedness patterns in bottlenose dolphins (*Tursiops* sp.) with different alliance strategies. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1514), 497–502. https://doi.org/10.1098/rspb.2002.2229

Langergraber, K. E., Rowney, C., Crockford, C., Wittig, R., Zuberbuhler, K., & Vigilant, L. (2014). Genetic analyses suggest no immigration of adult females and their offspring into the Sonso community of chimpanzees in the Budongo Forest Reserve, Uganda. *American Journal of Primatology*, *76*(7), 640–648. https://doi.org/10.1002/ajp.22258

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv preprint arXiv:1303.3997.

Li, H., Glusman, G., Hu, H., Shankaracharya, Caballero, J., Hubley, R., Witherspoon, D., Guthery, S. L., Mauldin, D. E., Jorde, L. B., Hood, L., Roach, J. C., & Huff, C. D. (2014). Relationship estimation from whole-genome sequence data. *PLoS Genetics*, *10*(1), e1004144. https://doi.org/10.1371/journal.pgen.1004144

McComb, K., Moss, C., Durant, S. M., Baker, L., & Sayialel, S. (2001). Matriarchs as repositories of social knowledge in African elephants. *Science*, *292*, 491–494.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J., & Woodruff, D. S. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science*, *265*(5176), 1193–1201. https://doi.org/10.1126/science.7915048

Naseri, A., Liu, X., Tang, K., Zhang, S., & Zhi, D. (2019). RaPID: Ultra-fast, powerful, and accurate detection of segments identical by descent (IBD) in biobank-scale cohorts. *Genome Biology*, *20*(1), 143. https://doi.org/10.1186/s13059-019-1754-8

Premachandra, H. K. A., Nguyen, N. H., & Knibb, W. (2019). Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. *Aquaculture*, *499*, 24–31. https://doi.org/10.1016/j.aquaculture.2018.09.022

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Skiar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Rogers, J., Garcia, R., Shelledy, W., Kaplan, J., Arya, A., Johnson, Z., Bergstrom, M., Novakowski, L., Nair, P., Vinson, A., Newman, D., Heckman, G., & Cameron, J. (2006). An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics*, *87*(1), 30–38. https://doi.org/10.1016/j.ygeno.2005.10.004

Rogers, J., Raveendran, M., Fawcett, G. L., Fox, A. S., Shelton, S. E., Oler, J. A., Cheverud, J., Muzny, D. M., Gibbs, R. A., Davidson, R. J., & Kalin, N. H. (2013). CRHR1 genotypes, neural circuits and the diathesis for anxiety and depression. *Molecular Psychiatry*, *18*(6), 700–707. https://doi.org/10.1038/mp.2012.152

Schuttler, S. G., Philbrick, J. A., Jeffery, K. J., & Eggert, L. S. (2014). Fine-scale genetic structure and cryptic associations reveal evidence of kin-based sociality in the African forest elephant. *PLoS One*, *9*(2), e88074. https://doi.org/10.1371/journal.pone.0088074

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science*, *302*(5648), 1231–1234. https://doi.org/10.1126/science.1088580

Smith, D. G. (1982). Inbreeding in three captive groups of rhesus monkeys. *American Journal of Physical Anthropology*, *58*(4), 447–451.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics*, *203*(2), 699–714. https://doi.org/10.1534/genetics.116.187492

Spong, G., Stone, J., Creel, S., & Bnjorklund, M. (2002). Genetic structure of lions (*Panthera leo*) in the Selous Game Reserve: implications for the evolution of sociality. *Journal of Evolutionary Biology*, *15*, 945–953.

Stadele, V., Pines, M., Swedell, L., & Vigilant, L. (2016). The ties that bind: Maternal kin bias in a multilevel primate society despite natal dispersal by both sexes. *American Journal of Primatology*, *78*(7), 731–744. https://doi.org/10.1002/ajp.22537

Stadele, V., & Vigilant, L. (2016). Strategies for determining kinship in wild populations using genetic data. *Ecology and Evolution*, *6*(17), 6107–6120. https://doi.org/10.1002/ece3.2346

Staples, J., Ekunwe, L., Lange, E., Wilson, J. G., Nickerson, D. A., & Below, J. E. (2015). PRIMUS: Improving pedigree reconstruction using mitochondrial and Y haplotypes. *Bioinformatics*, *32*(4), 596–598. https://doi.org/10.1093/bioinformatics/btv618

Staples, J., Qiao, D., Cho, M. H., Silverman, E. K., University of Washington Center for Mendelian Genomics, Nickerson, D. A., & Below, J. E. (2014). PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *American Journal of Human Genetics*, *95*(5), 553–564. https://doi.org/10.1016/j.ajhg.2014.10.005

Staples, J., Witherspoon, D. J., Jorde, L. B., Nickerson, D. A., University of Washington Center for Mendelian Genomics, Below, J. E., & Huff, C. D. (2016). PADRE: Pedigree-Aware Distant-Relationship Estimation. *American Journal of Human Genetics*, *99*(1), 154–162. https://doi.org/10.1016/j.ajhg.2016.05.020

Stevison, L. S., Woerner, A. E., Kidd, J. M., Kelley, J. L., Veeramah, K. R., McManus, K. F., Bustamante, C. D., Hammer, M. F., & Wall, J. D. (2016). The time scale of recombination rate evolution in great apes. *Molecular Biology and Evolution*, *33*(4), 928–945. https://doi.org/10.1093/molbev/msv331

Strucken, E. M., Lee, S. H., Lee, H. K., Song, K. D., Gibson, J. P., & Gondro, C. (2016). How many markers are enough? Factors influencing parentage testing in different livestock populations. *Journal of Animal*

Breeding and Genetics, *133*(1), 13–23. https://doi.org/10.1111/jbg.12179

Taylor, H. R. (2015). The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, *5*(15), 3140–3150.

Van Cise, A. M., Martien, K. K., Mahaffy, S. D., Baird, R. W., Webster, D. L., Fowler, J. H., Oleson, E. M., & Morin, P. A. (2017). Familial social structure and socially driven genetic differentiation in Hawaiian short-finned pilot whales. *Molecular Ecology*, *26*(23), 6730–6741. https://doi.org/10.1111/mec.14397

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vigilant, L., Hofreiter, M., Siedel, H., & Boesch, C. (2001). Paternity and relatedness in wild chimpanzee communities. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 12890–12895. https://doi.org/10.1073/pnas.231320498

Vinson, A., Prongay, K., & Ferguson, B. (2013). The value of extended pedigrees for next-generation analysis of complex disease in the rhesus macaque. *ILAR Journal*, *54*(2), 91–105. https://doi.org/10.1093/ilar/ilt041

Walling, C. A., Pemberton, J. M., Hadfield, J. D., & Kruuk, L. E. (2010). Comparing parentage inference software: Reanalysis of a red deer pedigree. *Molecular Ecology*, *19*(9), 1914–1928. https://doi.org/10.1111/j.1365-294X.2010.04604.x

Wang, J. (2013). An improvement on the maximum likelihood reconstruction of pedigrees from marker data. *Heredity (Edinb)*, *111*(2), 165–174. https://doi.org/10.1038/hdy.2013.34

White, L. C., Fontsere, C., Lizano, E., Hughes, D. A., Angedakin, S., Arandjelovic, M., Granjon, A. C., Hans, J. B., Lester, J. D., Rabanus-Wallace, M. T., Rowney, C., Städele, V., Marques-Bonet, T., Langergraber, K. E., & Vigilant, L. (2019). A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. *Molecular Ecology Resources*, *19*(3), 609–622. https://doi.org/10.1111/1755-0998.12993

Xue, C., Raveendran, M., Harris, R. A., Fawcett, G. L., Liu, X., White, S., Dahdouli, M., Deiros, D. R., Below, J. E., Salerno, W., Cox, L., Fan, G., Ferguson, B., Horvath, J., Johnson, Z., Kanthaswamy, S., Kubisch, H. M., Liu, D., Platt, M., & Rogers, J. (2016). The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome Research*, *26*(12), 1651–1662. https://doi.org/10.1101/gr.204255.116

Zhang, T., Guo, L., Shi, M., Xu, L., Chen, Y., Zhang, L., Gao, H., Li, J., & Gao, X. (2018). Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. *Gene*, *673*, 211–216. https://doi.org/10.1016/j.gene.2018.06.054

Zhou, Y., Browning, S. R., & Browning, B. L. (2020). A fast and simple method for detecting identity-by-descent segments in large-scale data. *American Journal of Human Genetics*, *106*(4), 426–437. https://doi.org/10.1016/j.ajhg.2020.02.010

SUPPORTING INFORMATION

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