# Comparing genome-based estimates of relatedness for use in pedigree-based conservation management

Samantha Hauser[1,*], Stephanie J. Galla[2,3], Andrea S. Putnam[4], Tammy E. Steeves[2], Emily K. Latch[1]

[1]Department of Biological Sciences, University of Wisconsin - Milwaukee, WI USA
[2]School of Biological Sciences, University of Canterbury, New Zealand
[3]Department of Biological Sciences, Boise State University, Boise, ID USA
[4]Department of Exhibit-Curators, San Diego Zoo Wildlife Alliance, San Diego, CA USA
[*]Embark Veterinary, Inc., Boston, Massachusetts, United States of America

Running head: Genomic estimates of relatedness.

ORCiD Hauser: 0000-0002-3404-5253
ORCiD Galla: 0000-0002-4650-8067
ORCiD Putnam: 0000-0003-1795-6589
ORCiD Steeves: 0000-0003-2112-5761
ORCiD Latch: 0000-0002-9892-1056

Keywords: relatedness, allele-sharing, captive breeding, *ex situ* management, conservation breeding program, wild pedigrees

**Abstract**

Researchers have long debated which estimator of relatedness best captures the degree of relationship between two individuals. In the genomics era, this debate continues, with relatedness estimates being sensitive to the method used to generate markers, marker quality, and levels of diversity in sampled individuals. Here, we compare six commonly used genome-based relatedness estimators (kinship genetic distance (KGD), Wang Maximum Likelihood (TrioML), Queller and Goodnight ($R_{xy}$), KING-robust, $R_{AB}$, allele-sharing co-ancestry) across five species bred in captivity–including three birds and two mammals–with varying degrees of reliable pedigree data, using reduced-representation and whole genome resequencing data. Relatedness estimates varied widely across estimators, sequencing method, and species, yet the most consistent results for known first order relationships were found using KING-robust and to a lesser extent KGD. Our combined results indicate there is not a single genome-based estimator that is ideal across different species and data types. To enable researchers to evaluate the most appropriate relatedness estimator for each new dataset, we provide a workflow that is broadly applicable to conservation breeding programs, particularly where genome-based estimates of relatedness can complement and complete poorly pedigreed populations. Given a growing interest in the application of wild pedigrees, our results and workflow are also applicable to *in situ* wildlife management.

## Introduction

Relatedness and kinship, concepts that quantify the relationship between two individuals (hereafter referred to relatedness for simplicity) are foundational in biology (Wright, 1922), with application in health sciences (Ott, 1974), agriculture (Cassell et al., 2003; Jannink et al., 2001) and species conservation (Fernández et al., 2005). Pedigrees track relatedness in a population by documenting the ancestry of individuals, and are used to manage small populations, including those in captivity (i.e., *ex situ* population management) or intensively managed wild or semi-wild populations (i.e., *in situ* or "*sorta situ"*; Wildt et al., 2019; Wolfe et al., 2012; Galla et al. 2021). Conservation practitioners can prioritize breeding individuals with low mean kinship when making pairing decisions in captivity, or population management decisions in wild or semi-wild populations (Ballou & Lacy, 1995; Giglio et al., 2016; Weeks et al., 2011). A management strategy that minimizes mean kinship in a population is effective at mitigating drift, inbreeding, and adaptation to captivity, while preserving genetic diversity and evolutionary potential to curtail extinction risk (Fernandez & Toro, 1999; Montgomery et al., 1997; Sonesson & Meuwissen, 2001; Spielman, 2004, Templeton & Read, 1994). Simulation studies have shown that pedigree-based management of small populations minimizes inbreeding while maximizing founder representation (e.g., Ballou & Lacy, 1995; Rudnick & Lacy, 2008). Empirical studies have demonstrated the importance of pedigree management for threatened species conservation (e.g., Tasmanian devil, *Sarcophilus harrisii*, Gooley et al, 2017; Ālala, *Corvus hawaiiensis*, Flanagan et al., 2021; Bison, *Bison bison*, Giglio et al., 2018). Because the collection of pedigree data is feasible for many captive and intensively managed wild populations and computer programs are readily available for studbook management (e.g., PopLink, ZIMS; Faust et al., 2019, Species360), pedigree analysis (e.g., PMx; Lacy et al., 2012), and population modeling

(e.g., VORTEX; Lacy & Pollak, 2014), pedigree-based genetic management of populations is an accessible and appealing tool for conservation management.

Pedigrees continue to provide high precision estimates of inbreeding and relatedness when pedigrees are robust (i.e., many generations deep with relatively low missing data; Putnam & Ivy, 2014; Robinson et al., 2013). However, there are several limitations common in pedigrees for small populations that can hinder their utility for genetic management (Galla et al. 2021). One is that the individuals forming the initial population (hereafter, founders) are often of unknown relationships and assumed to be equally unrelated (Ballou 1983; Hogg et al., 2019). In small populations that have experienced sustained bottlenecks, it is unlikely that remaining individuals are unrelated. For example, a genetic study of critically-endangered kākāpō (*Strigops habroptilus*) revealed high relatedness coefficients — including some first-order (i.e., parent-offspring or sibling) relationships — among founders that were previously assumed to be unrelated (Bergner et al., 2014). Assuming founders are unrelated can lead to underestimated kinship and inbreeding coefficients, which is exacerbated when pedigrees are shallow (< 5 generations deep; Balloux et al., 2004 but see Rudnick and Lacy 2008). Beyond the initial founding event, additional founders can be incorporated into pedigrees when populations are augmented by wild individuals with unknown relationships to each other or to individuals already in the pedigreed population (Galla et al., 2020; Spielman & Frankham, 1992).

Pedigree gaps indicating unknown relationships can also be a source of ambiguity constraining the utility of pedigrees for genetic management. Missing data can arise from uncertainties in parentage, for example through undetected extra-pair parentage or herd or colonial breeding systems (Overbeek et al., 2020; Ferrie, 2017; Mucha & Windig, 2009). In the wild, pedigree gaps may be introduced, or relationships incorrectly inferred due to individual

identification errors (e.g., dropped leg bands; Milligan et al., 2003). Parentage assignment errors may be rare (Henkel et al. 2012, Ferrie et al. 2013, Hammerly et al. 2016) but their effects can compound over generations, increasing the probability that related individuals will be unintentionally paired with negative impacts on population fitness. For example, in a captive population of Attwater's Prairie-chicken (*Tympanuchus cupido attwateri*), nearly 40% of the population were descendants of four individuals that had incorrect parentage assigned four generations prior (Hammerly et al., 2016). To alleviate pedigree shortcomings, including unknown founder relationships, missing pedigree data, and potential error, researchers can use genomic estimates of relatedness to complement and complete pedigrees.

Relatedness estimates generated using genomic markers have helped to resolve founder relationships (Bergner et al., 2014), reconstruct parentage (Flanagan & Jones, 2019), and vet pedigrees for potential errors (Hammerly et al., 2016). Factors that can affect the accuracy of genome-based relatedness estimates include marker type, number of loci, missingness, sequencing depth, and the specific estimator used. Single nucleotide polymorphisms (SNPs) - derived from SNP-assays (Santure et al., 2010), reduced representation sequencing (RRS) approaches like RAD-sequencing or GBS (Galla et al., 2019; Lemopoulos et al., 2019), or whole genome resequencing (WGS; Galla et al., 2020) - yield accurate genomic estimates of relatedness (Jones & Wang, 2010; Santure et al., 2010b; Skare et al., 2009). However, the number of markers necessary for a given system depends upon the genome size of the species, population size, and overall level of inbreeding in the population; insufficient marker power can lead to inaccurate relationship inference (Goudet et al. 2018; Morin et al., 2009; Smouse, 2010; Sun et al., 2016). Similarly, the specific estimator used to capture inheritance patterns and translate them into a relatedness estimate can impact the accuracy of relationship inference.

There are several commonly used relatedness estimators appropriate for genomic data, each with its own limitations (Box 1). Briefly, frequency-based estimates of relatedness quantify the probability that shared alleles are identical-by-descent (IBD) relative to a reference population using probabilities (i.e., moment-methods) or correlations (likelihood-based; Wang, 2014). Frequency-based estimators assume that populations are large, randomly mating, and outbred (but see Hedrick & Lacy, 2015; Wang, 2007), effectively sampled (Wang 2017), and that allele frequencies at each marker are reliably estimated (Csillery et al. 2006, Galla et al. 2020). However, there are notable exceptions (see Box 1). A meaningful reference population for frequency-based relatedness estimation may be difficult to define, especially for captive populations where individuals are managed through non-random mating strategies and often divided across numerous institutions. Further, the precision of frequency-based approaches will rely on marker completeness and richness, with more markers leading to more precise estimates of relatedness (Baruch & Weller, 2008; Morin et al., 2009). As an alternative, pairwise allele sharing (i.e., molecular co-ancestry or similarity index; Gutiérrez et al., 2005) does not rely upon reference population allele frequencies. Further, Ivy et al. (2016) found allele-sharing to be strongly correlated with mean kinships derived from pedigrees using genomic data, thus making it a potentially useful estimator for populations where the assumptions of frequency-based estimators are violated, such as captive populations.

Examining the relative performance of genome-based estimators using empirical data can expose estimator pitfalls that cannot be found solely through simulation studies. Best practice guidance to estimate accurate genome-based relatedness using high throughput sequencing data is needed to maximize conservation success for poorly-pedigreed threatened populations. In this empirical study, we extend beyond previous study of relatedness estimators (Attard et al., 2017;

Csillery et al., 2006; Goudet et al. 2018, Santure et al., 2010) to a comparison with conservation systems across mammal and bird taxa that provide non-model and realistic challenges to genome-based relatedness estimation. We compared the accuracy of six genome-based relatedness estimators relative to observed pedigree data, including allele-sharing and five different frequency-based estimators across 9 genomic datasets. Relatedness estimates are derived from SNP data generated using three different approaches (WGS, and RRS with high and low levels of missing data), collected from five different species with *ex-situ* conservation breeding programs, including Addax (*Addax nasomaculatus*), Inca Tern (*Larosterna inca*), Koala (*Phascolarctos cinereus*), Kakī (Black Stilt; *Himantopus novaezelandiae*), and Kākāriki Karaka (Orange-fronted Parakeet; *Cyanoramphus malherbi*). In addition to several interactive factors related to the data itself (i.e., genome size, inbreeding, marker type, number of loci, missingness, sequencing depth), these five species offer a variety of realistic challenges in breeding programs, including the size of the managed population, depth of the pedigree, and the amount and structure of pedigree gaps (Figure 1). Imperfect pedigrees, as featured in the present study and common in small populations, expose limitations of existing relatedness estimators. Evaluation and identification of optimal relatedness estimators for use in conservation breeding programs that minimizes a population's mean kinship is effective at mitigating drift, inbreeding, and adaptation to captivity, while preserving genetic diversity and evolutionary potential to curtail extinction risk, all critical goals of *ex-situ* conservation. Beyond broad applicability to conservation breeding programs, pedigree-informed management of wild populations (Pemberton, 2008) will benefit from explicit evaluation of genome-based relatedness estimators.

**Methods**

Of the five species with breeding programs in the present study, Addax, Inca Tern and Koala study populations are cooperatively managed through the Association of Zoos and Aquariums (AZA) Species Survival Plans®, and Kakī and Kārāriki Karaka are managed by the New Zealand Department of Conservation.

Addax: The current population of addax in AZA descends from a small number of individuals (~15) imported from Chad and the Khartoum Zoo, Sudan in the 1960s and 1970s (Enright, 2019). Until the early 2000s, facilities with large herds did not reliably record parentage and frequent exchanges with the private sector occurred. At the time of analysis, only 13% of the AZA addax population could be traced to specific founders.

Inca Tern: Wild Inca terns were imported by AZA facilities in large numbers during the 1960s through the mid-1990s. Parentage historically and currently has not been reliably recorded at a specific facility that holds and propagates the largest number of terns, resulting in a steady increase in the proportion of the pedigree that cannot be traced to founders (30%; Nelson, 2017, Nelson and Lynch, 2021).

Koala: The contemporary pedigree for the AZA Queensland koala population is well recorded. However, unresolved relationships in the historical pedigree of captive-bred Australian importations, mostly from a single facility in the 1970s – 1980s, generates an overall pedigree for the living AZA population that is only 39% is known (Hamlin, 2019).

Kakī: Kakī are a critically endangered wading bird endemic to the South Island of Aotearoa New Zealand. While Kakī were once more widespread in Aotearoa, they experienced decline due to introduced mammalian predators and habitat loss (Sanders & Maloney, 2002). By the early 1980s, only 23 birds remained (Galla, 2020). Since then, intensive management strategies—including predator control efforts and a conservation breeding program—were initiated to

prevent extinction and enhance recovery. The Kakī pedigree has been maintained for this population and includes intensively managed wild and captive individuals. About half of the Kakī pedigree is known (56.6%). Pedigree gaps come from unbanded birds (gaps) and extra-pair copulation (errors) in the wild (Overbeek et al., 2020).

Kākāriki Karaka: Kākāriki Karaka were once more widespread on the South Island of Aotearoa, and experienced decline due to invasive mammalian predators and habitat loss. During the 20th century, this species was considered extinct twice, likely due to its cryptic nature. In 2003, a conservation breeding program was initiated with 12 individuals collected from the wild, with sparse supplementation in recent years (Galla, 2020). The Kākāriki Karaka pedigree is a nearly complete pedigree, with 99.5% of the pedigree known. While complete, this pedigree is shallow (0-3 generations deep), and as a result may be more impacted by incorrect assumptions about founder relationships.

We generated SNPs for each of the five species using two types of data: RRS (double-digest restriction-site associated DNA (ddRAD) or genotype-by-sequencing (GBS) or WGS: ddRAD sequencing data for the Addax, Inca Tern, and Koala (herein), GBS data for Kakī (Galla et al., 2019), and WGS for Kakī and Kākāriki Karaka (Galla et al., 2020). A direct comparison of Kakī samples was possible with the WGS and GBS, while the remaining comparisons across species and sequencing type were holistic. The bioinformatics methods and parameters for each dataset vary as they were chosen based on best practice optimization protocols for each sequencing approach. As such, these datasets represent realistic datasets found across studies and systems wherein methods and parameters are tailored to the technique and observed data.

We isolated high quality DNA from blood, feather, or tissue samples for Kakī and Kākāriki Karaka using a lithium chloride chloroform extraction method (Galla et al. 2019) and

quantified using a Qubit 2.0 Fluorometer. Kakī GBS paired-end libraries were prepared using the Illumina TruSeq DNA PCR-free protocol (350 bp DNA fragment) and sequenced on 5 lanes of an Illumina HiSeq 2500 (Galla et al 2019). Kakī WGS libraries were prepared using the TruSeq Nano DNA Library Prep Kit and sequenced on 34 lanes of Illumina HiSeq 400 (Galla et al 2020) Kākāriki Karaka WGS libraries were prepared using the Nextera DNA Flex Library Prep Kit and sequenced on 1 lane of an Illumina Novaseq 6000 (Galla et al. 2020). SNP data from WGS and GBS data for Kakī and Kākāriki Karaka were processed, and genotypes were called as per Galla et al (2019, 2020). Briefly, Kakī GBS data was generated from two batches of individuals, Illumina Hi-Seq reads underwent reference-guided SNP discovery using Tassel 5.0 (Glaubitz et al., 2014), and biallelic SNPs were filtered for a minimum MAF of 0.05, an average minimum SNP-depth per locus of five, maximum of 10% missing data, and linkage disequilibrium ($r^2$ = 0.6 over 1000 sites; Galla et al. 2019). For Kakī and Kākāriki Karaka WGS, reads were generated from Illumina Hi-Seq or NovaSeq platforms, SNPs were discovered using BCFtools v. 1.9 (Li et al., 2009), and biallelic SNPs were filtered for a minimum MAF of 0.05, a quality score greater than 20, a maximum of 10% missing data, linkage disequilibrium ($r^2$ = 0.6 over 1000 sites), and either a minimum depth of 5 (Kākāriki Karaka) or a minimum average depth of 10 (Kakī; for more details, see Galla et al. 2020). GBS and WGS Kakī data were the same consensus 25 individuals to allow direct comparisons between SNP data generation approaches (Table 1).

We extracted high quality DNA from blood samples for Inca Tern, Koala, and Addax using the Qiagen DNeasy Blood and Tissue Kit. All DNA isolates were quantified using a Qubit 2.0 Fluorometer. For the three AZA species, ddRAD library preparation (per Peterson et al., 2012), quality control, and sequencing were done at the Texas A&M AgriLife Genomics core

facility. ddRAD libraries were prepared using restriction enzymes *Spe*I and *Mbo*I (Koala and Addax) or *Sph*I and *Mlu*CI (IncaTern) for paired-end 150 bp reads and sequenced on a portion of an Illumina NovaSeq 6000 lane. Raw sequencing data were demultiplexed, filtered, and genotypes were called using the bioinformatics pipelines STACKS v.2.0 and VCFTOOLS (Catchen et al., 2013; Danecek et al., 2011). SNP genotypes for Addax and Inca Tern were generated using the de novo pipeline in STACKS with the following parameters chosen through optimization (Paris et al., 2017), respectively: m = 3, M = 3, n = 0, min_maf = 0.02, r = 0.7 and m = 3, M = 3, n = 0, min_maf = 0.02, r = 0.6. Genotypes were called for Koala using the reference pipeline using a reference genome (GenBank Accession: GCA_002099425.1) with min_maf = 0.02 and r = 0.80 parameters. Individuals were omitted using VCFTOOLS based on maximum missingness thresholds, set to two settings for each dataset to quantify missing data impacts: low (10% for all datasets) and high (staggered at 40% in Addax, 60% in Koala, and 80% in Inca Tern). Missing data is a prevalent and sometimes unavoidable problem in conservation breeding systems where the combination of suboptimal samples and low species-wide genetic diversity leads to high missing data (Graham et al., 2015). As such, we tested a range of missing data filters to encapsulate the high missing datasets likely to burden many genomic studies in conservation breeding programs. Maximum values for the high missing datasets were chosen based on the distribution of missing genomic data found for each species.

These processing steps yielded 9 SNP datasets for comparison, for which, we calculated six pairwise relatedness estimators to compare estimates directly and to investigate their impact on downstream breeding recommendations: Allele-Sharing (AS), Kinship Genetic Distance (KGD; Dodds et al., 2015), Wang Maximum Likelihood (TrioML; Wang, 2007, 2011), Queller and Goodnight ($R_{xy}$; Queller & Goodnight, 1989), Kinship INference for Genome-wide

association studies (KING-robust; Waples et al., 2019), and $R_{AB}$ (Korneliussen & Moltke, 2015). Estimators were directly compared for the effects of missing data in Addax, Koala, and Inca Tern and sequencing approach (RRS vs WGS) in Kakī. We holistically compared relative estimator performance across species. The AS estimates were calculated in the program CASC (Ivy & Putnam, 2019), subsampling 1,000 loci and using 1,000 iterations. AS values are scaled differently than genome-based kinships, so we transformed AS values using a linear piecewise regression on known pedigree-based kinship estimates (>0). Pedigree-based kinship estimates were calculated in PMx (Lacy 2012) based on PopLink studbook data and were scaled to relatedness estimates (2*kinship) for comparative analyses. KGD estimates were calculated in R (R Core Team, 2018) using Dodds et al. (2015). TrioML and Rxy estimates were calculated in COANCESTRY (Wang, 2011), specifying for the settings 'unknown allele frequencies' (i.e., allele frequencies are calculated from the imported data) and 'inbreeding'. KING-robust and $R_{AB}$ estimators were estimated using NGSRelate (Hanghøj et al, 2019). All KING-robust, a kinship estimator values were multiplied by 2 to approximately scale them to the rest of the relatedness estimators and pedigree relatedness (0 - 1.0). It should be noted that if inbreeding is present, relatedness may be less than 2*kinship as the derivation of their calculations differ slightly.

For comparisons to known pedigree-based relatedness values, we only included individuals with known pedigree information (i.e., those with relatedness values > 0) from each species. Each pairwise relatedness estimator was directly compared to known pedigree relatedness (PedRel) and compared to all other genome-based relatedness estimators using Pearson's correlation coefficients in R. Relatedness estimators and pedigree-based relatedness are both most robust with low expected variance for inferring first-order relationships, making subsets of known first-order relationships ideal to explicitly test the relative performance of

genome-based estimators to the known pedigree values. Relatedness estimator variance is expected to be larger for more distant relationships Thus, first order relationships, i.e. those with pedigree relatedness > 0.50 were subset from each dataset, and boxplots of each relatedness estimator and pedigree relatedness values were plotted in R package ggplot2 (Wickham, 2011). The Addax datasets contained too few pairwise first order relationships (<10) for comparison and thus were omitted from the boxplots.

**Results**

We generated six relatedness estimates for each of 9 genomic datasets (Table 1). The most consistently high correlation values between genome-based and known pedigree-based relatedness values were found in datasets with low missing data and a large number of loci (Figure 2). This is particularly evident when looking within a species; stronger correlations were observed for the datasets with low missing data (10%) (Koala 10: r = 0.8 – 0.88, Koala 60: r = -0.04 - 0.82) (Inca Tern 10: r = 0.63 - 0.84, Inca Tern 80: r = 0.61 - 0.87). Lower correlation values between genome-based and pedigree-based relatedness values in datasets that retained individuals with more missing data were simultaneously reflected in poor concordance between genome-based relatedness metrics (e.g., Koala 60, Figure S1), pointing to poor data quality. Low correlation values in certain estimators (i.e., KING-robust and $R_{AB}$) were not consistently found in the high missing genomic datasets, which could signal to another factor influencing estimator performance not represented in our comparisons. Similarly, higher correlations were also found in WGS data (r = 0.71 - 0.82) than in GBS data (r = 0.31 - 0.72), which has considerably fewer loci and thus lower statistical power (Kakī WGS = 68,144 and GBS = 19,395). In general, datasets with more SNPs produced estimates of relatedness with higher precision than datasets with smaller numbers of SNPs (Figure 3). This pattern was observed when comparing WGS and

GBS datasets in Kakī (WGS mean r = 0.79; GBS mean r = 0.64; Figure 3). Relatedness estimates produced from datasets with less missing data were more correlated across estimators (e.g., Koala 60 and Koala 10, Figure 2).

Genome-based relatedness estimators underestimated pedigree relatedness values across species, missing data, and sequencing type (Figure 3). Similar to patterns found in the metric correlation results, high missing data elicited poor estimator performance and more frequent inconsistencies with pedigree relatedness. The KING-robust estimator appeared sensitive to high missing data in RRS datasets, indicated by broader distributions relative to the other estimators and deviation from pedigree relatedness (see Kakī GBS and Koala 60 datasets; Figure 3). However, across the higher quality datasets (i.e., low missing data, high number of SNP loci), the KING-robust distribution was the most consistent and similar to that of the known pedigree relatedness (Koala 10, Inca 10, Kakī WGS, Kakariki; Figure 3).

**Discussion**

Theoretical expectations assert that allele-sharing would be a suitable genome-based relatedness estimator for intensively managed conservation breeding populations, given that it does not require a reference population (Ivy et al., 2016). However, our results indicate that untransformed allele-sharing may be the least appropriate estimator to improve conservation pedigrees. Because allele-sharing provides a direct measure of identity-by-state, as opposed to a frequency-based estimate of identity-by-descent, this estimator is expected to overestimate pairwise relatedness when genetic diversity or more specifically minor allele frequencies are low. This is salient for estimating relatedness in genetically-depauperate conservation breeding programs, where the probability of identity-by-state is inflated (Henkel et al., 2012). As a result, allele-sharing values are on a vastly different scale than pedigree relatedness and frequency-

based relatedness estimators and any potential use of allele-sharing requires rescaling using a complex transformation for incorporation into pedigree management, for example, the linear piecewise transformation used here. Our results show while allele-sharing estimators without a transformation are inappropriate for use, transformed allele-sharing values perform comparably to the other relatedness estimators. Though, it is important to note that these transformations were based upon existing pedigree data, which may not be possible across systems, and may inherently bias correlations upward. Further, transformations can introduce errors and reduce accuracy and precision of relatedness estimates; future use should use caution in reducing potential errors and including the transformation in their interpretations.

We anticipated that the inability to accurately estimate population allele frequencies in captive breeding programs would lead to biased relatedness estimates when using frequency-based estimators. However, we generally found high correspondence between relatedness estimators and known pedigree values, especially in cases of low missing data, more fully resolved pedigrees, and high-resolution genomic data (WGS and large numbers of SNPs). All genome-based estimators consistently underestimated relatedness across all datasets. The range of KING-robust estimates were the most consistent with first order relatedness values, yet there was still some inconsistency relative to the known pedigree values across the datasets. Whereas KING-robust was an optimal estimator for high quality datasets, it performed poorly with higher missing data and RRS compared to WGS, indicating it may not be ideal for all systems. It is important to point out that there are several confounding variables (e.g., pedigree accuracy, degree of inbreeding, number of individuals, species genome size) across our datasets that could affect the resulting relatedness values and contribute to inconsistencies across datasets. There is no single optimal relatedness estimator as each has its advantages and caveats, thus we

recommend that researchers formally test candidate estimators for their systems. Poor correlations between known pedigree-based and genome-based relatedness likely indicate an unsuitable estimator for the data. If poor correlations are found across a variety of estimators that use different approaches, the data itself may be insufficient for relatedness estimation because of sampling and sequencing issues mentioned throughout this paper.

Missing data can occur in genomic datasets generated for captive populations for several reasons. While proactive collection and banking of high-quality DNA samples for genomic analyses is ideal to mitigate many of the limitations due to missing data, logistical and ethical constraints in handling captive animals (i.e., minimizing handling and the associated stress and risks) makes collection of high-quality and representative samples a challenge. These limitations often compel researchers to rely on low-quality samples opportunistically taken from veterinary samples including mammalian blood or plasma samples. Missing data can be the direct result of suboptimal quality samples as genetic material is low yield and/or degraded (Graham et al., 2015). A secondary effect of opportunistic sampling is that sequencing is often done in batches as samples become available through routine veterinary visits. Though we ran all samples for each species in a single batch and thus had no batch effects in our datasets, batch effects can be common and potentially problematic, reducing the number of loci shared among sequence batches and subsequently increasing the amount of missing data in a genomic dataset (Leigh et al., 2018). Thus, proactive sampling and storage of high-quality samples will be important in reducing the downstream complications that can arise from opportunistic sampling (i.e., missing data). Conservation breeding managers can learn from ongoing efforts to establish proactive and standardized sample collection by groups including San Diego Zoo Wildlife Alliance's Frozen

Zoo (e.g., Chemnick et al., 2009), Cryo-Initiative (Comizzoli, 2014), Frozen Ark (Clarke, 2009), and EAZA's Biobank (see: https://www.eaza.net/conservation/research/eaza-biobank/).

There was variable performance of relatedness estimators across species, sequencing method, degree of missing data, and pedigree knownness and depth. Thus, we advocate for explicit evaluation of relatedness estimators for each new dataset and system. Researchers can use our methods as a generalized workflow (Figure 4) to test the relative performance of relatedness estimators against a set of known pedigree relationships for their specific system. While this approach leverages known first order relationships to evaluate the appropriateness of relatedness evaluators, we appreciate this may not be feasible for all systems, especially wild systems without pedigree data. We therefore advocate for the need of a simulation software, similar to COANCESTRY (Wang 2011), that can compare newer genomic relatedness estimators to evaluate estimator effectiveness. In the meantime, these results from in this manuscript provide a few general recommendations that can be used in systems with and without pedigree data. The most accurate estimates of relatedness are predicted to come from genomic datasets generated using sequencing approaches that maximize statistical power (e.g., WGS) with nominal missing data (e.g., high-quality samples, single batch sequencing). We recommend using WGS data over RRS, when it is not cost- or computationally prohibitive, and further advocate for high quality DNA sampling, to provide the best possible sequencing reads. When pedigree information is unavailable for comparison, particular care must be used to ensure sufficient data quality. While we recommend researchers carefully test several relatedness estimators to find one that is appropriate for their study system and molecular data set using known relationships, our results also indicate that some estimators (e.g., KING) provided accurate estimates of first order relationships in most of our systems and may be broadly

applicable across genomic data sets. Lastly, even a robust relatedness estimator cannot overcome sampling deficiencies, so we urge proactive sample collection across the breadth and depth of pedigrees to facilitate accurate reconstruction of relationships among individuals to improve pedigree-based conservation management. While resolving more relationships in a species pedigree using genome-based estimates will change pairwise kinship, the new kinships will be using more evidence to provide deeper insight into relationships within the captive population. If a new baseline of kinships utilizing genomic data can be established, breeding recommendations can be improved which will assist in maximizing long-term gene diversity retention.

**Acknowledgments**

We are grateful for the continued support of Te Rūnanga o Ngāi Tahu, Te Ngāi Tūāhuriri Rūnanga, Te Rūnanga o Arowhenua, Te Rūnanga o Waihao and Te Rūnanga o Moeraki. We thank all members of the Kakī and Kākāriki Karaka Recovery Programmes for their ongoing support. The Kakī and Kākāriki Karaka research was funded by the Ministry of Business, Innovation and Employment Endeavour Fund (UOCX1602 awarded to TES), the Brian Mason Scientific and Technical Trust (awarded to SJG and TES), and the Mohua Charitable Trust (awarded to TES).We would like to thank the Institute of Museum and Library Services National Leadership, Grant MG-30-15-0102-15 awarded to EKL. Support for SSH and SJG came from University of Wisconsin-Milwaukee College of Letters and Sciences, and a NSF EPSCoR RII Track-2 award (OIA-1826801), respectively. Special thanks to the AZA institutions that supplied samples: Buffalo Zoo, Brookfield Zoo, Dallas Zoo, Fossil Rim Wildlife Center, Frozen Zoo at the San Diego Zoo Wildlife Alliance, Kansas City Zoo, Louisville Zoo, Cleveland Metroparks Zoo, Omaha's Henry Doorly Zoo and Aquarium, The Living Desert Zoo and Gardens, San Diego Zoo, and Saint Louis Zoo. Lastly, thanks to J. Ivy and Association of Zoos and

Aquariums' Molecular Data for Population Management Scientific Advisory Group who facilitated this collaboration.

**Conflict of Interest Statement**

There are no conflicts of interest to report.

**Data availability**

A subset of the genomic data used in this project were previously derived from two culturally significant species, namely Kakī/Black Stilt and Kākāriki Karaka/Orange-Fronted Parakeet (Galla et al. 2019; Galla et al. 2020) and are currently stored on a password protected server (http://www.ucconsert.org/data/). Kakī and kākāriki karaka are taonga (treasured) species. For Māori (the Indigenous Peoples of Aotearoa New Zealand), all genomic data obtained from taonga species have whakapapa (genealogy that includes people, plants and animals, mountains, rivers and winds) and are therefore taonga in their own right (Collier-Robinson et al. 2019). Thus, these data are tapu (sacred), and tikanga (customary practices, protocols, and ethics) determine how people interact with them. To this end, the passwords for the genomic data in this manuscript will be made available to researchers on the recommendation of the kaitiaki (guardians) for the iwi (tribes) and hapū (subtribes) that affiliate with them. Having said this, if one or more of the editors or reviewers would like access to these data during the review process, we are able to provide time-sensitive passwords upon request.

For Addax, Koala, and Inca Tern, species managed by the Association of Zoos and Aquariums, we will make the genomic data available in VCF format upon acceptance on Dryad.

**Authors' Contributions**

SSH, SJG, EKL, and TES conceived the research ideas and designed the methodology. SSH and SJG generated the genomic datasets. SSH analyzed the data. SSH and SJG led the writing of the

manuscript. EKL and TES supervised the analysis and interpretation of the research. All authors contributed to the manuscript preparation and gave final approval for submission.

434

435

**Table 1.** Summary information on each of the 9 genomic datasets used including: Number of individuals currently managed in the captive population (# Inds Managed), percentage of the pedigree known prior to genomic data incorporation (% Pedigree Known), average generation per individual with standard deviation (Pedigree Depth), type of genomic technique used (Type), the maximum missingness threshold used (Max Missing), number of individuals with SNP data in the dataset (# Inds), number of SNP loci in the dataset ( #Loci), percentage of the pedigree known after genomic data incorporation (% analytic known)

| Species | # Living Inds Managed | % Pedigree Known | Pedigree Depth | Type | Max Missing | # Inds | # Loci | % Analytic Known |
|---|---|---|---|---|---|---|---|---|
| Addax | 182 | 13.0% | 6.3 ±3.2 | ddRAD | 10% | 53 | 17184 | 22.1% |
| Addax | 182 | 13.0% | 6.3 ±3.2 | ddRAD | 40% | 58 | 17184 | 22.3% |
| Inca Tern | 305 | 55.5% | 1.2 ±1.2 | ddRAD | 10% | 72 | 38356 | 59.2% |
| Inca Tern | 305 | 55.5% | 1.2 ±1.2 | ddRAD | 80% | 76 | 38356 | 59.3% |
| Koala | 50 | 36.0% | 3.7 ±1.4 | ddRAD | 10% | 44 | 50875 | 48.0% |
| Koala | 50 | 36.0% | 3.7 ±1.4 | ddRAD | 60% | 49 | 50875 | 49.0% |
| Kākāriki Karaka | 184 | 99.5% | 2.5 ±0.7 | WGS | 0% | 36 | 90949 | 99.5% |
| Kakī | 533 | 56.6% | 2.6 ±1.4 | WGS | <1% | 25 | 68144 | 65.0% |
| Kakī | 533 | 56.6% | 2.6 ±1.4 | GBS | <1% | 25 | 19395 | 65.0% |

## Figure Captions

**Figure 1.** Study species and key pedigree challenges faced by each conservation breeding program. Schematic pedigrees denote missing data with white circles/squares and existing data with black circles/squares.

**Figure 2.** Heatmaps depicting the pairwise Pearson's correlation coefficient between pedigree relatedness and each relatedness estimators with darker gradient colors denoting higher correlation values.

**Figure 3.** Boxplots of the relatedness values of known first order relationships calculated from each genome-based estimator and pedigree relatedness. Relatedness values from 0 - 1 are on the y-axis with the associated estimator names on the x-axis.

**Figure 4.** A general workflow for evaluating candidate genome-based relatedness estimators for a genomic dataset.

## Supplemental Figure Captions

**Figure S1.** Heatmaps depicting the pairwise Pearson's correlation coefficient between all relatedness estimators with darker gradient colors denoting higher correlation values.

**References:**


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

Ballou, J. D., & Lacy, R. C. (1995). Identifying genetically important individuals for management of genetic variation in pedigreed populations. *Population management for survival and recovery*, 76–111.

Balloux, F., Amos, W., & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, *13*(10), 3021–3031.

Baruch, E., & Weller, J. I. (2008). Estimation of the number of SNP genetic markers required for parentage verification. Animal Genetics, 39(5), 474-479.

Bergner, L. M., Jamieson, I. G., & Robertson, B. C. (2014). Combining genetic data to identify relatedness among founders in a genetically depauperate parrot, the Kakapo (Strigops habroptilus). *Conservation Genetics*, *15*(5), 1013–1020.

Cassell, B. G., Adamec, V., & Pearson, R. E. (2003). Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *Journal of Dairy Science*, *86*(9), 2967–2976.

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

Csilléry, K., Johnson, T., Beraldi, D., Clutton-Brock, T., Coltman, D., Hansson, B., ... & Pemberton, J. M. (2006). Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, *173*(4), 2091-2101.

Chemnick, L. G., Houck, M. L., & Ryder, O. A. (2009). Banking of Genetic Resources. *Conservation genetics in the age of genomics* (pp. 124–130).

Clarke, A. G. (2009). The Frozen Ark Project: The role of zoos and aquariums in preserving the genetic material of threatened animals. *International Zoo Yearbook*, *43*(1), 222–230. Retrieved June 9, 2021, from http://www.frozenark.org

Comizzoli, P. (2014). C-29: The Pan-Smithsonian cryo-initiative-freezing for the future. *Cryobiology*, *69*(3), 509.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

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

Enright W. (2019) International Addax Studbook. Gland Switzerland: World. Association of Zoos & Aquariums.

Faust, L., Bergstrom, Y., Thompson, S., & Bier, L. (2019). PopLink Version 2.5. Lincoln Park Zoo. Chicago, IL.

Fernández, J., Villanueva, B., Pong-Wong, R., & Toro, M. Á. (2005). Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics*, *170*(3), 1313–1321.

Ferrie, G. (2017). *Using Molecular Genetic and Demographic Tools to Improve Management of Ex Situ Avian Populations.*

Flanagan, A. M., Masuda, B., Grueber, C. E., & Sutton, J. T. (2021). Moving from trends to benchmarks by using regression tree analysis to find inbreeding thresholds in a critically endangered bird. *Conservation Biology*.

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

Galla, S. J. (2019) Conservation genomic management of two critically-endangered New Zealand birds. Doctoral Thesis, University of Canterbury. http://dx.doi.org/10.26021/8081

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

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., Richardson, A., et al. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications*, *13*(5), 991–1008.

Galla, S. J., Brown, L., Couch-Lewis, Y… Steeves, T. E. (2021). The relevance of pedigrees in the conservation genomics era. *Molecular Ecology*, doi: https://doi.org/10.1111/mec.16192,

Giglio, R. M., Ivy, J. A., Jones, L. C., & Latch, E. K. (2016). Evaluation of alternative management strategies for maintenance of genetic variation in wildlife populations. *Animal Conservation*, *19*(4), 380–390.

Giglio, Rachael M., Ivy, J. A., Jones, L. C., & Latch, E. K. (2018). Pedigree-based genetic management improves bison conservation. *Journal of Wildlife Management*, *82*(4), 766–774.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, *9*(2), e90346.

Goodnight, K. F., Queller, D. C., & Goodnight, K. F. (1999). Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*, *8*(7), 1231–1234.

Gooley, R., Hogg, C. J., Belov, K., & Grueber, C. E. (2017). No evidence of inbreeding depression in a Tasmanian devil insurance population despite significant variation in inbreeding. *Scientific Reports*, *7*(1).

Graham, C. F., Glenn, T. C., Mcarthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., et al. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, *15*(6), 1304–1315.

Gutiérrez, J. P., Royo, L. J., Álvarez, I., & Goyache, F. (2005). MolKin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity*, *96*(6), 718–721.

Hamlin C. (2019) AZA Regional Queensland Koala (Phascolarctos cinereus) Studbook. Maryland: Association of Zoos & Aquariums.

Hammerly, S. C., de la Cerda, D. A., Bailey, H., & Johnson, J. A. (2016). A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. *Animal Conservation*, *19*(3), 296–303.

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. Gigascience, 8(5), giz034.

Hedrick, P. P. W., Lacy, R. C., & Baker, C. S. (2015). Measuring relatedness between inbred individuals. *Journal of Heredity*, *106*(1), 20–25.

Henkel, J. R., Jones, K. L., Hereford, S. G., Savoie, M. L., Leibo, S. P., & Howard, J. J. (2012). Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (Grus canadensis pulla). *Zoo Biology*, *31*(3), 322–335.

Hogg, C. J., Wright, B., Morris, K. M., Lee, A. V., Ivy, J. A., Grueber, C. E., & 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.

Ivy, J. A., & Putnam, A. S. (2019). Calculate Allele Sharing Coefficients (CASC) (version 1.0). San Diego, CA, USA: San Diego Zoo Global.

Ivy, J. A., Putnam, A. S., Navarro, A. Y., Gurr, J., & Ryder, O. A. (2016). Applying SNP-derived molecular coancestry estimates to captive breeding programs. *Journal of Heredity*, *107*(5), 403–412.

Jannink, J. L., Bink, M. C. A. M., & Jansen, R. C. (2001). Using complex plant pedigrees to map valuable genes. *Trends in Plant Science*, *6*(8), 337–342.

Jones, O. R., & Wang, J. (2010). Molecular marker-based pedigrees for animal conservation biologists. *Animal Conservation*, *13*(1), 26–34.

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics*, *31*(24), 4009–4011.

Lacy, R C, & Pollak, J. P. (2014). Vortex: A Stochastic Simulation of the Extinction Process. Manual. *Version 9.99c*.

Lacy, R.C., Ballou, J. D., & Pollak, J. P. (2012). PMx: Software package for demographic and genetic analysis and management of pedigreed populations. *Methods in Ecology and Evolution*, *3*(2), 433–437.

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*(4), 778–788.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M. L., et al. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Milligan, J. L., Davis, A. K., & Altizer, S. M. (2003). Errors associated with using colored leg bands to identify wild birds. *Journal of Field Ornithology*, *74*(2), 111–118.

Montgomery, M. E., Ballou, J. D., Nurthen, R. K., England, P. R., Briscoe, D. A., & Frankham, R. (1997). Minimizing kinship in captive breeding programs. *Zoo Biology*, *16*(5), 377–389.

Morin, P. A., Martien, K. K., & Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, *9*(1), 66–73.

Mucha, S., & Windig, J. J. (2009). Effects of incomplete pedigree on genetic management of the Dutch Landrace goat. *Journal of Animal Breeding and Genetics*, *126*(3), 250–256.

Nelson S. (2017) AZA Regional Inca Tern (Larosterna inca) Studbook. Maryland: Association of Zoos & Aquariums.

Nelson S., Lynch C. (2021) Population Analysis & Breeding and Transfer Plan, Inca Tern (Larosterna inca) Species Survival Plan. Maryland: Association of Zoos & Aquariums

Ott, J. (1974). Estimation of the recombination fraction in human pedigrees: efficient

computation of the likelihood for human linkage studies. *American Journal of Human Genetics*, *26*(5), 588–597.

Overbeek, A., Galla, S. J., Brown, L., Cleland, S., Thyne, C., Maloney, R., & Steeves, T. E. (2020). Pedigree validation using genetic markers in an intensively managed taonga species, the critically endangered kakī (Himantopus novaezelandiae). *Notornis*, *67*, 709–716.

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

Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 613–621.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). protocol to ddRAD. *PLoS ONE*.

Putnam, A. S., & Ivy, J. A. (2014). Kinship-based management strategies for captive breeding programs when pedigrees are unknown or uncertain. *Journal of Heredity*, *105*(3), 303–311.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*(2), 258-275.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson, S. P., Simmons, L. W., & Kennington, W. J. (2013). Estimating relatedness and inbreeding using molecular markers and pedigrees: the effect of demographic history. *Molecular Ecology*, *22*, 5779–5792.

Rudnick, J. A., & Lacy, R. C. (2008). The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies. Conservation Genetics, 9(6), 1439–1450.

Sanders, M. D., & Maloney, R. F. (2002). Causes of mortality at nests of ground-nesting birds in the Upper Waitaki Basin, South Island, New Zealand: a 5-year video study. Biological conservation, 106(2), 225-236.

Santure, A. W., Stapley, J., Ball, A. D., Birkhead, T. R., Burke, T., & Slate, J. (2010). On the use of large marker panels to estimate inbreeding and relatedness: Empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, *19*(7), 1439–1451.

Skare, Ø., Sheehan, N., & Egeland, T. (2009). Identification of distant family relationships. *Bioinformatics*, *25*(18), 2376–2382.

Smouse, P. E. (2010). How many SNPs are enough? *Molecular Ecology*.

Sonesson, A. K., & Meuwissen, T. H. E. (2001). Minimization of rate of inbreeding for small populations with overlapping generations. *Genetical Research*, *77*(3), 285–292.

Spielman, D., Brook, B. W., Briscoe, D. A., & Frankham, R. (2004). Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, *5*(4), 439–448.

Spielman, D., & Frankham, R. (1992). Modeling problems in conservation genetics using captive Drosophila populations: Improvement of reproductive fitness due to immigration of one individual into small partially inbred populations. *Zoo Biology*, *11*(5), 343–351.

Templeton, A. R., & Read, B. (1994). Inbreeding: one word, several meanings, much confusion. In *Conservation genetics* (pp. 91-105).

Sun, M., Jobling, M. A., Taliun, D., Pramstaller, P. P., Egeland, T., & Sheehan, N. A. (2016). On the use of dense SNP marker data for the identification of distant relative pairs. *Theoretical Population Biology*, *107*, 14–25.

Wang, J. (2007). Parentage and sibship exclusions: Higher statistical power with more family members. *Heredity*, *99*(2), 205–217.

Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: An assessment of current methods. *Journal of Evolutionary Biology*, *27*(3), 518–530.

Wang, Jinliang. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203–1215.

Wang, Jinliang. (2011). Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*(1), 141–145.

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*(1), 35–48.

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., et al. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, *4*(6), 709–725.

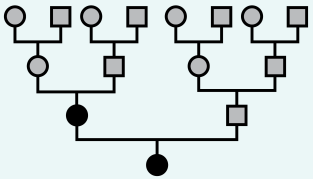
Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

Wildt, D., Miller, P., Koepfli, K. P., Pukazhenthi, B., Palfrey, K., Livingston, G., Beetem, D., et al. (2019). Breeding Centers, Private Ranches, and Genomics for Creating Sustainable Wildlife Populations. *BioScience*.

Wolfe, B. A., Aguilar, R. F., Aguirre, A. A., Olsen, G. H., & Blumer, E. S. (2012). Sorta situ: The new reality of management conditions for wildlife populations in the absence of "Wild" spaces. *New Directions in Conservation Medicine: Applied Cases of Ecological Health*.

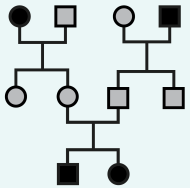
Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, *56*(645), 330–338.

## Addax
*Addax nasomaculatus*



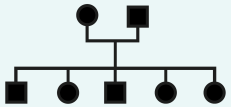
Poor records, high uncertainty

## Inca Tern
*Larosterna inca*



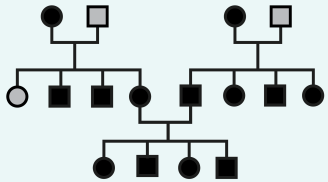
Contemporary missing data

## Kākāriki Karaka
*Cyanoramphus malherbi*



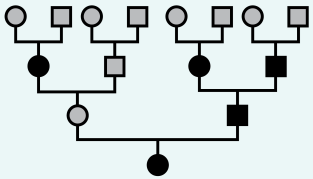
High completeness, but with low depth.

## Kakī
*Himantopus novaezelandiae*

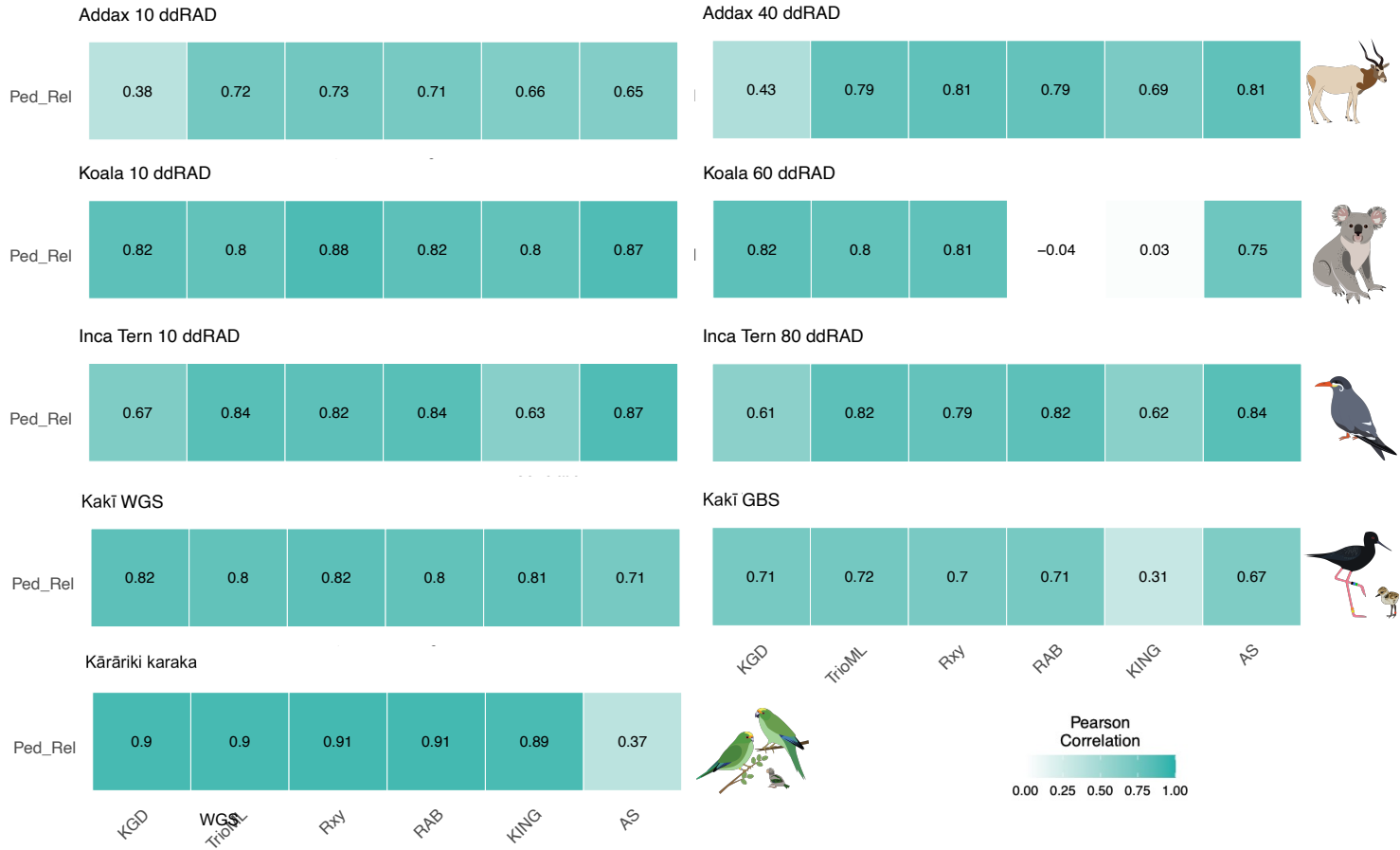


Semi-wild with missing data and uncertainty.

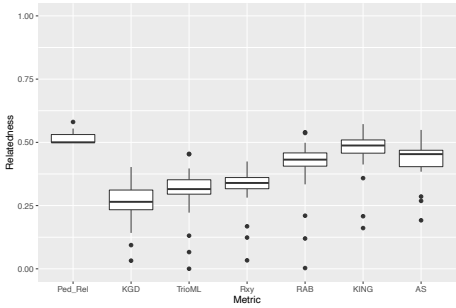
## Koala
*Phascolarctos cinereus*



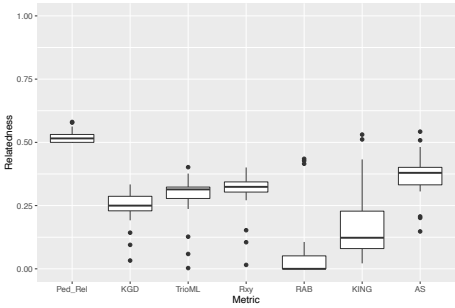
High historic missing data


Addax 10 ddRAD
Ped_Rel
0.38
0.72
0.73
0.71
0.66
0.65
Addax 40 ddRAD
0.43
0.79
0.81
0.79
0.69
0.81
Koala 10 ddRAD
Ped_Rel
0.82
0.8
0.88
0.82
0.8
0.87
Koala 60 ddRAD
0.82
0.8
0.81
−0.04
0.03
0.75
Inca Tern 10 ddRAD
Ped_Rel
0.67
0.84
0.82
0.84
0.63
0.87
Inca Tern 80 ddRAD
0.61
0.82
0.79
0.82
0.62
0.84
Kakī WGS
Ped_Rel
0.82
0.8
0.82
0.8
0.81
0.71
Kakī GBS
0.71
0.72
0.7
0.71
0.31
0.67
KGD
TrioML
Rxy
RAB
KING
AS
Kārāriki karaka
Ped_Rel
0.9
0.9
0.91
0.91
0.89
0.37
KGD
WGS
TrioML
Rxy
RAB
KING
AS
Pearson Correlation
0.00
0.25
0.50
0.75
1.00

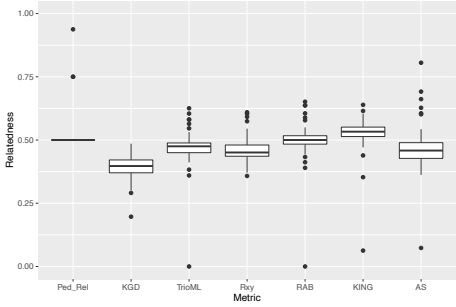
Koala 10 ddRAD



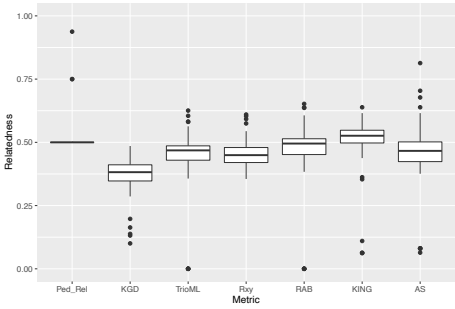

Koala 60 ddRAD



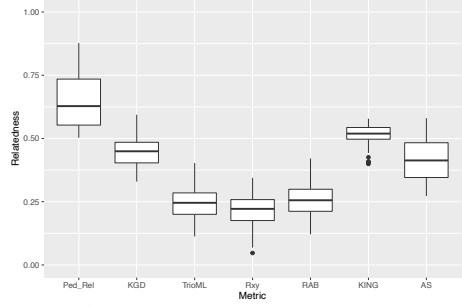
Inca Tern 10 ddRAD



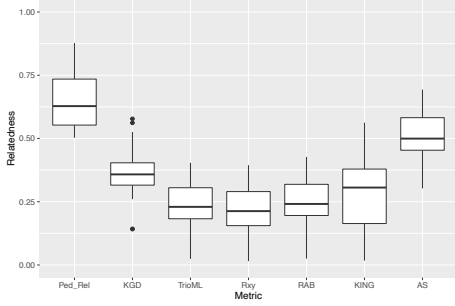

Inca Tern 80 ddRAD



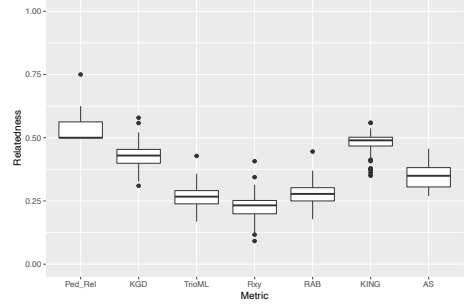
Kakī WGS :

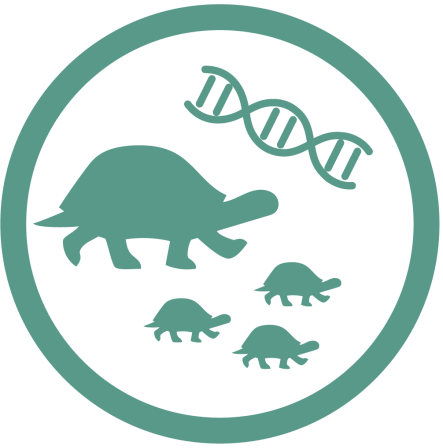



Kakī GBS



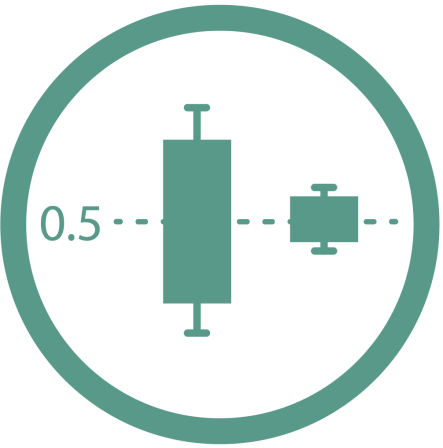
Kārāriki karaka WGS



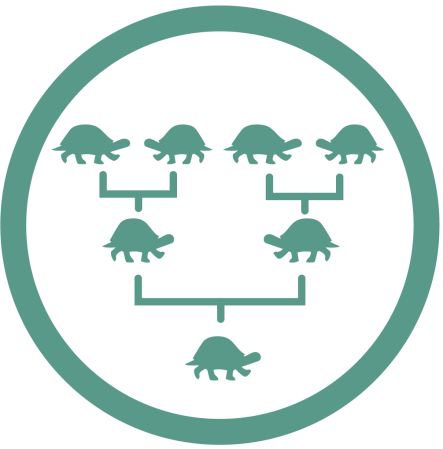


1) Sequence known relationships for downstream comparisons.

2) Optimize SNP filtering strategies for the data at hand.




3) Compare relatedness estimators to see which are most precise for parent-offspring relationships.



4) Use best genetic estimator to complement pedigree relationships.

## Addax 10 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.54 |
| RAB | | | | | 0.66 | 0.75 |
| Rxy | | | | 0.91 | 0.68 | 0.83 |
| TrioML | | | 0.91 | 1 | 0.66 | 0.75 |
| KGD | | 0.61 | 0.63 | 0.62 | 0.47 | 0.49 |
| Ped_Rel | 0.38 | 0.72 | 0.73 | 0.71 | 0.66 | 0.65 |

## Addax 40 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.68 |
| RAB | | | | | 0.71 | 0.9 |
| Rxy | | | | 0.92 | 0.73 | 0.97 |
| TrioML | | | 0.91 | 1 | 0.71 | 0.9 |
| KGD | | 0.63 | 0.6 | 0.64 | 0.39 | 0.55 |
| Ped_Rel | 0.43 | 0.79 | 0.81 | 0.79 | 0.69 | 0.81 |

## Koala 10 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.82 |
| RAB | | | | | 0.8 | 0.85 |
| Rxy | | | | 0.9 | 0.85 | 0.97 |
| TrioML | | | 0.88 | 0.99 | 0.76 | 0.83 |
| KGD | | 0.89 | 0.94 | 0.92 | 0.85 | 0.86 |
| Ped_Rel | 0.82 | 0.8 | 0.88 | 0.82 | 0.8 | 0.87 |

## Koala 60 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.05 |
| RAB | | | | | 0.69 | −0.06 |
| Rxy | | | | −0.04 | 0.04 | 0.96 |
| TrioML | | | 0.82 | 0 | 0.05 | 0.75 |
| KGD | | 0.88 | 0.88 | 0 | 0.04 | 0.75 |
| Ped_Rel | 0.82 | 0.8 | 0.81 | −0.04 | 0.03 | 0.75 |

## Inca Tern 10 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.67 |
| RAB | | | | | 0.74 | 0.97 |
| Rxy | | | | 0.97 | 0.64 | 0.94 |
| TrioML | | | 0.97 | 1 | 0.74 | 0.97 |
| KGD | | 0.84 | 0.81 | 0.84 | 0.71 | 0.78 |
| Ped_Rel | 0.67 | 0.84 | 0.82 | 0.84 | 0.63 | 0.87 |

## Inca Tern 80 ddRAD

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.69 |
| RAB | | | | | 0.75 | 0.97 |
| Rxy | | | | 0.97 | 0.65 | 0.95 |
| TrioML | | | 0.97 | 1 | 0.75 | 0.97 |
| KGD | | 0.82 | 0.79 | 0.82 | 0.68 | 0.78 |
| Ped_Rel | 0.61 | 0.82 | 0.79 | 0.82 | 0.62 | 0.84 |

## Kakī WGS

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.93 |
| RAB | | | | | 0.89 | 0.89 |
| Rxy | | | | 0.97 | 0.91 | 0.87 |
| TrioML | | | 0.97 | 1 | 0.89 | 0.89 |
| KGD | | 0.98 | 0.97 | 0.98 | 0.87 | 0.84 |
| Ped_Rel | 0.82 | 0.8 | 0.82 | 0.8 | 0.81 | 0.71 |

## Kakī GBS

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.43 |
| RAB | | | | | 0.53 | 0.9 |
| Rxy | | | | 0.88 | 0.41 | 0.96 |
| TrioML | | | 0.88 | 1 | 0.54 | 0.89 |
| KGD | | 0.93 | 0.94 | 0.94 | 0.53 | 0.9 |
| Ped_Rel | 0.71 | 0.72 | 0.7 | 0.71 | 0.31 | 0.67 |

## Kārāriki karaka WGS

Pearson Correlation: 0.00 0.25 0.50 0.75 1.00

| | KING | RAB | Rxy | TrioML | KGD | Ped_Rel |
|---|---|---|---|---|---|---|
| KING | | | | | | 0.27 |
| RAB | | | | | 0.85 | 0.53 |
| Rxy | | | | 0.96 | 0.89 | 0.4 |
| TrioML | | | 0.96 | 1 | 0.85 | 0.53 |
| KGD | | 0.96 | 0.95 | 0.96 | 0.86 | 0.39 |
| Ped_Rel | 0.9 | 0.9 | 0.91 | 0.91 | 0.89 | 0.37 |