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

Samantha Hauser[1], Stephanie J Galla[2-3], Tammy E Steeves[2], Emily K Latch[1]
[1]Department of Biological Sciences, University of Wisconsin - Milwaukee, USA
[2]School of Biological Sciences, University of Canterbury, New Zealand
[3]Department of Biological Sciences, Boise State University, USA

Running head: genomic estimates of relatedness

ORCiD Hauser: 0000-0002-3404-5253
ORCiD Galla: 0000-0002-4650-8067
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**

1. Researchers have long debated which genomic 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 genomic markers (e.g., reduced-representation sequencing, whole genome resequencing), marker quality, and levels of diversity in sampled individuals.
2. Here, we compare six commonly used 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.
3. Relatedness estimates varied widely across estimators, sequencing method, and species, yet the most consistent results with known pedigree data were found using KING-robust and to a lesser extent KGD. The allele-sharing estimator was sensitive to missing data and inbreeding, attributes that make this estimator ill-suited for use in captive breeding programs. Our combined results indicate there is not a single genomic based estimator that is ideal across different species and data types.
4. To enable researchers to evaluate the most appropriate relatedness estimator for each new data set, we provide a structured workflow that is broadly applicable to conservation breeding programs, particularly where genomic estimates of relatedness can complement and complete poorly pedigreed populations. Given a growing interest in 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 in the general sense), 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 have been an important tool for conservation and management of populations. In a conservation context, pedigrees have been fundamental to the management of 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). Practitioners can prioritize 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 in an effort to curtail extinction risk (Fernandez & Toro, 1999; Montgomery et al., 1997; Sonesson & Meuwissen, 2001; Spielman, 2004). Simulation studies have shown that pedigree-based management of small populations minimizes short-term inbreeding while maximizing founder diversity (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 genetic 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. Preprint). 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 (Spielman & Frankham (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 and negatively impacting population fitness. For example, in a captive population of Attwater's Prairie-chicken (*Tympanuchus cupido attwateri*), nearly 40% of the population were direct descendants of only 4 errors 4 years 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). Though, factors that can affect the accuracy of genetically-estimated relatedness include marker type, number of loci, and the specific estimator used. Single nucleotide polymorphisms (SNPs), which can be derived from SNP-assays and arrays (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) approaches (Galla et al., 2020), yield accurate genomic estimates of relatedness (Jones & Wang, 2010; Santure et al., 2010b; Skare et al., 2009). But, the quantity 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 (Morin et al., 2009; Smouse, 2010; Sun et al., 2016); insufficient markers can provide inaccurate relationship inference. Similarly, the

specific estimator used to capture the inheritance patterns and translate them into a relatedness estimate can impact the accuracy of relationship inference.

There are several commonly used methods to estimate genomic-based relatedness and reconstruct pedigrees, each with their own limitations. 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: the KGD estimator (Dodds et al., 2015) uses only pairs of individuals and therefore quantifies identity-by-state rather than IBD; $R_{AB}$ (Korneliussen & Moltke, 2015) and KING-robust (Waples et al., 2019) account for unrealistic assumptions such as a large panmictic population. A meaningful reference population is difficult to define 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 (Galla et al., 2020). 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 and has been found to be strongly correlated with mean kinships derived from pedigrees, thus making it a potentially useful relatedness estimator for populations where the assumptions of frequency-based estimators are violated, such as captive populations (Ivy et al., 2016).

In this study, we compared the accuracy of six genomic-based estimators relative to observed pedigree data, including allele-sharing and five different frequency-based relatedness estimators across 11 genomic datasets. Best practice guidance to estimate accurate genomic-based relatedness using high throughput sequencing data is needed to maximize conservation success for poorly-pedigreed and threatened populations. 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 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*). 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 conseration breeding programs will improve pairing recommendations, maximize genetic diversity and in turn increase species persistence; all critical goals of *ex-situ* conservation. Beyond broad applicability to conservation breeding programmes, use of genetically-derived wild pedigrees (Pemberton, 2008) will benefit from explicit evaluation of genomic estimators.

**Methods**

We used five species with conservation breeding programs: Addax, Inca Tern, Koala, Kakī and Kākāriki Karaka in the present study. The Addax, Inca Tern and Koala study populations are currently managed by the Association of Zoos and Aquariums (AZA) Population Management Center (USA), and Kakī and Kārāriki Karaka are managed by the New Zealand

Department of Conservation. These species represent a variety of realistic scenarios and challenges practitioners face when generating breeding recommendations (Table 1; e.g., unknown founder relationships, gaps in pedigrees, inbreeding). The Addax pedigree was historically not well-maintained; record keeping of parentage is challenging in herd managed species, and these uncertainties compound over the pedigree. In addition, a large portion of the AZA population originated from the private sector, adding to the uncertainty in the pedigree. Only 13% of the AZA Addax pedigree is known. The contemporary pedigree for the Koala is well known, but unknown relationships among founders and between contemporary imported individuals and the captive population add gaps in the pedigree. Currently, 39% of the Koala pedigree is known. The Inca Tern pedigree is 30% known, with unknown ancestry coming from contemporary individuals that are managed in colonies. The Kakī pedigree has been maintained since the early 1980's 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). 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 SNP data for each of the five species using WGS and two common RRS apparoches (double-digest restriction-site associated DNA (ddRAD) and genotype-by-sequencing (GBS): ddRAD sequencing data for the Addax, Inca Tern, and Koala, GBS data for Kakī (Galla et al., 2019), and WGS from Kakī and Kākāriki Karaka (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, a minimum SNP-depth 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 or a minimum average depth of 10, depending on the species (Galla et al. 2020).

The ddRAD library preparation, quality control, and sequencing were done as per Peterson et al. (2012) for the three AZA species at the Texas A&M AgriLife Genomics core facility using restriction enzymes SpeI and MboI (Koala and Addax) or SphI and MluCI (IncaTern) for paired-end 150 bp reads and sequenced on a portion of an Illumina NovaSeq 6000 lane. For all three AZA species, 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, 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 to quantify missing data impacts: low (10%) and high (staggered at 40% in Addax, 60% in Koala, and 80%

in Inca Tern). Maximum values for the high missing datasets were chosen based on the distribution of missing genomic data for each species.

These processing steps yielded 11 SNP datasets for comparison, including two datasets in which GBS and WGS Kakī data were subset to the same consensus 25 individuals to allow direct comparisons between SNP data generation approaches (Kakī GBS subset and Kakī WGS subset, respectively; Table 1).

For each of the 11 datasets, we calculated six pairwise relatedness estimators to compare estimates directly and to investigate their impact on downstream breeding recommendations: Allele-Sharing (AS; Gutiérrez et al., 2005)), Kinship Genetic Distance (KGD; Dodds et al., 2015)), Wang Maximum Likelihood (TrioML; Wang, 2002, 2011)), Queller and Goodnight ($R_{xy}$; Goodnight et al., 1999)), Kinship INference for Genome-wide association studies (KING-robust; Waples et al., 2019), and $R_{AB}$ (Korneliussen & Moltke, 2015). AS estimates (Gutiérrez et al., 2005) were calculated in the program CASC (Ivy & Putnam, 2019), subsampling 1,000 loci and using 1,000 iterations. AS values are scaled differently than genomic-based kinships, so we transformed AS values using a linear piecewise regression (hereafter Molecular Kinship, MolKin). KGD estimates were calculated in R using Dodds et al. (2015). TrioML and Rxy estimates were calculated in COANCESTRY (Wang, 2011), specifying unknown allele frequencies and inbreeding. KING-robust and $R_{AB}$ estimators were estimated using NGSRelate (Korneliussen & Moltke, 2015).

In comparisons to known pedigree values, we only included individuals with known pedigree information (i.e., those with kinship values > 0) from each species. Each pairwise relatedness estimator was directly compared to known pedigree kinships (PedKin) and pedigree relatedness (kinship/2; PedRel) and compared to all other genomic-based relatedness estimators

using Pearson's correlation coefficients in R. Distributions of each relatedness estimator and the pedigree kinship values were plotted using the density plots in R package ggplot2 (Wickham, 2011). Each distribution was compared to the pedigree kinship and pedigree relatedness distributions using the Kolmogorov-Smirnov test in R (R function ks.test; Marsaglia et al., 2003).

All genomic pairwise relatedness estimates, regardless if individuals had known pedigree information, were imported into the genetics tab in PMx (Lacy et al., 2012). This was in part to evaluate how pedigrees could be improved when incorporating genomic-based relatedness values. We generated a rank of all the mean kinships in the population to compare breeding recommendations across relatedness estimators. Low mean kinship ranks indicate individuals with the fewest relatives in the population that are most valuable for breeding. All of the relatedness estimators (MolKin, KGD, $R_{xy}$, TrioML, KING-robust, and $R_{AB}$) were imported directly into PMx as relatedness values. $R_{xy}$, TrioML, KING-robust, and $R_{AB}$ do not generate pairwise self-to-self estimates; we manually added values of 0.5 (the theoretical expected self-self-relatedness value) where absent to allow PMx import. We compared mean kinship rank lists (top 20 males and females) to each other and to the known pedigree kinships using Kendall's tau correlations for ranked data (R function kendall.tau).

**Results**

We generated six relatedness estimates for each of 11 genomic datasets (Table 1). All of the relatedness estimator distributions overlap pedigree-based kinships (PedKin) except for AS, which is on an inflated scale to the rest of the estimators (Figure 2). Transforming the AS values (MolKin) brought them within the range of pedigree kinships. Often, the restricted variance we observed in AS values was also observed in the transformed MolKin values; however, there are

notable instances where the MolKin distribution almost entirely overlapped the distribution of pedigree kinships (Koala 10 and Koala 60). Distributions for genomic relatedness estimators TrioML and RAB are zero inflated because relatedness values truncate at 0. This zero-inflated skew was substantially reduced when using WGS compared to reduced representation methods. WGS was able to provide relatedness values for relatively unrelated individuals whereas RRS truncated those values at 0, thereby lumping less related individuals as unrelated. The $R_{xy}$ distribution was skewed downward, lower than the pedigree kinship distribution. Though several distributions are visually similar to the known pedigree kinship values, only one of the estimated relatedness distributions was statistically similar: KING-robust in Koala 10.

The highest correlation values between genomic-based relatedness and known pedigree-based kinship values were found in datasets with low missing data or high power, but we found overall similar correlations values regardless of the level of missing data (Figure 3). Correlations for a species were often higher for the low (10%) missing dataset counterpart for Koala (Koala 10: r = 0.8 – 0.88, Koala 60: r = -0.038 - 0.82) and Inca Tern (Inca Tern 10: r = 0.63 - 0.84, Inca Tern 80: r = 0.61 - 0.87). This is particularly prevalent when comparing the low correlation values of KING-robust and $R_{AB}$ compared to the pedigree kinship values (0.025 -0.038, respectively). Though, low correlation values in certain estimators (i.e., KING-robust) were not consistently found in high missing genomic dataset, which could signal to another factor not represented in our comparisons. Similarly, higher correlations were also found in WGS data (r = 0.69 - 0.82) than in GBS (r = 0.31 - 0.72), which has considerably fewer loci and thus statistical power (Kaki 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 Kaki (WGS

subset mean r = 0.78; GBS subset mean r = 0.64) but also when comparing ddRAD datasets across species. For example, we observed a higher correlation between relatedness estimators and known pedigree kinship for koala datasets with >50,000 SNPs (r = 0.68) than for addax datasets with < 20,000 SNPs (r = 0.56) holding missing data low (Figure 3). Relatedness estimates produced from datasets with less missing data were more correlated across estimators (e.g., Koala 60 and Koala 10, Figure 3). That being said, there are a number of confounding variables (e.g., pedigree accuracy, number of individuals, species genome size) across our datasets that could affect the resulting relatedness values and be the underlying reason why there are inconsistencies across datasets.

Kendall's tau correlations on mean-kinship rank data were low overall and varied widely across species and datasets (Figure 4). There was a weak pattern in which species with more complete pedigrees had more high correlation values than those with more incomplete pedigrees. For instance, the Addax 10 dataset, with 13% known pedigree, showed a much lower correlation (average 0.06) between pedigree-based and genomic-based mean-kinship rank than determined based on the pedigree or on empirical data than Kakī WGS (average correlation 0.19 with 56.6% of the pedigree known). However, this pattern did not hold across all species; Kākāriki Karaka in which 99.5% of the pedigree was known prior, the average Kendall's tau correlation value was 0.10. In datasets with higher power, i.e., the WGS datasets, there are consistently higher Kendall's tau correlation values among all estimators compared to those by reduced representation methods. Datasets with less missing data echo this pattern of higher and more consistent correlation values.

## Discussion

Theoretical expectations assert that allele-sharing would be the most appropriate genomic-based relatedness estimator for intensively managed conservation breeding populations (Ivy et al., 2016); however, our results indicate that allele-sharing is be the least appropriate estimator to improve conservation pedigrees. Allele-sharing values are on a vastly different scale than pedigree kinships and frequency-based relatedness estimators. 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 is 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 such, any potential use of allele-sharing would require rescaling using a complex transformation for incorporation into pedigree management, for example, the linear piecewise transformation used here. However, transformations can introduce errors and reduce accuracy and precision of relatedness estimates; in addition, a estimator that requires a complex transformation for use suggests that it is ill fit for purpose. An essential quality of genomic relatedness measures for conservation management is that there is sufficient variance to distinguish relationships. Our results show that allele-sharing estimators and their transformed molecular kinships provide little variance in either allele-sharing or the transformed molecular kinship values to provide the necessary statistical power to confidently distinguish relationships. This inconsistency may be influenced by the structure of the underlying pedigrees and each of their idiosyncrasies. Regardless, for small populations of conservation concern, we would not recommend using either allele-sharing or its derived molecular kinship due to the high inconsistency across datasets.

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). $R_{xy}$, TrioML and $R_{AB}$ estimates were biased downward and consistently underestimated relatedness across all datasets. The range of KGD and KING-robust estimates were the most consistent with pedigree-based relatedness values, yet there was still some inconsistency relative to the known pedigree values across the datasets. There was no clear pattern of bias in over-estimating (or under-estimating) relatedness compared to the known pedigree values. While there is no single optimal relatedness estimator as each has its advantages and caveats, we recommend that researchers formally test candidate estimators, including KING-robust and KGD, for their systems.

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 the opportunistic sampling is that sequencing is often done in batches as samples become available through routine veterinary visits. However, batch effects

are common and can be problematic, reducing the common loci 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 than 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 AZA Frozen Zoo (e.g., Chemnick et al., 2009), Cryo-Intiative (Comizzoli, 2014), Frozen Ark (Clarke, 2009).

There was a wide range of results in correlations among mean kinship rank data derived from pedigree kinship and the genomic-based relatedness estimates. For pedigrees with little to no information, the addition of genomic-based relatedness estimates replace unknown relationship information (pedigree kinship = 0) and subsequently change the mean kinship ranks more than for known relationships. This is well illustrated with the Addax results for which the pedigree was 13% known prior to relatedness estimates and 22.1% known after their implementation. Though correlations between genomic-based relatedness estimators and pedigree kinship were generally high (> 0.82; omitting the AS-derived molecular kinship), differences in relatedness estimator distribution shape and range act as factor of discordance among mean kinship rank. For instance, the variance for pedigree metrics was much wider for the Kākāriki Karaka than for the other metrics, and the distributions were shaped quite differently for genomic estimates (normal distribution) compared to those for the pedigree (trimodal distribution). While the Kākāriki Karaka pedigree is almost complete, it is shallow and therefore subject to outbreeding among the recent founders due to in-situ population structure (Andrews, 2013; Galla et al., 2019, 2020), which could be the underlying cause behind deviations with mean kinship rank. Nevertheless, across datasets, we optimistically found

relatively high correspondence among estimators for mean kinship rank in species with little of the pedigree known. We acknowledge that mean kinship is just one of many factors that inform conservation breeding pairing recommendations, including other genetic metrics (e.g., inbreeding), demography at institutions, health, and behavior of individuals (Lacy et al., 2012). Though resolving more relationships in a species pedigree using genomic-based estimates will change mean kinship rankings, the new rankings will be utilizing more evidence and provide more insight into the captive population. Subsequent breeding recommendations will improve and captive population will benefit in the long-term.

There was variable performance of relatedness estimators across species, sequencing method, degree of missing data, and pedigree knowness and depth. Thus, we advocate for explicit evaluation of relatedness estimators for each new dataset and system. Researchers can use our methods as a structured workflow to test the relative performance of relatedness estimators against a set of known pedigree relationships for their specific system. 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. Our results also highlight potential pitfalls in relatedness estimation. Firstly, allele-sharing is not recommended due to its low variance, sensitivity to missing data, and large bias requiring a statistical transformation which performs inconsistently and adds a potential source of error. Further, several estimators ($R_{xy}$, TrioML, and $R_{AB}$) showed systematic downward biases due to slightly differing ranges compared to kinship or truncation of values to 0, features that were designed specifically for inbred populations (Korneliussen & Moltke, 2015; Wang, 2011). As such, we cannot recommend

AS, $R_{xy}$, TrioML, and $R_{AB}$ as candidate estimators to be used for small populations. 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.

**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 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, A. Putnam, and Association of Zoos and Aquarium's 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 it. 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 Aquarium, we will make the genomic data available 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.

**Table 1.** Summary information on each of the 11 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 | 518 | 13.0% | 6.3 ±3.2 | ddRAD | 10% | 53 | 17184 | 22.1% |
| Addax | 518 | 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% |
| Kakī | 533 | 56.6% | 2.6 ±1.4 | WGS | <1% | 36 | 68144 | 66.6% |
| Kakī | 533 | 56.6% | 2.6 ±1.4 | GBS | <1% | 82 | 19395 | 68.0% |
| Kākāriki Karaka | 184 | 99.5% | 2.5 ±0.7 | WGS | 0% | 36 | 90949 | 99.5% |
| Kakī (subset) | 533 | 56.6% | 2.6 ±1.4 | WGS | <1% | 25 | 68144 | 65.0% |
| Kakī (subset) | 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.** Distribution density plots of each relatedness estimator. Please note that the x-axes vary across panels.

**Figure 3.** Heatmaps depicting the pairwise Pearson's correlation coefficient between relatedness estimators with darker gradient colors denoting higher correlation values.

**Figure 4.** Heatmaps depicting the pairwise Kendall's tau correlation coefficient for ranked data between sets of mean kinship rank derived from relatedness estimators. Above the gray diagonal are the values for females and below the gray diagonal for males. Darker gradient colors denoting higher correlation values.

**References:**


Andrews, B. J. (2013). *Conservation Genetics of the orange-fronted kakariki (Cyanoramphus malherbi)*. University of Canterbury.

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.

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.

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.

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

Fernández, J., & Toro, M. A. (1999). The use of mathematical programming to control inbreeding in selection schemes. *Journal of Animal Breeding and Genetics*, *116*(6), 447–466.

Fernández, Jesús, 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., 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. (Preprint). The relevance of pedigrees in the conservation genomics era. Authorea. doi:
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.
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.
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 Deigo 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, Robert 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.
Marsaglia, G., Tsang, W. W., & Wang, J. (2003). Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, *8*(1), 1–4.
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.
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).
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.
Robinson, S. P., Simmons, L. W., & Kennington, W. J. (2013). Estimating relatedness and inbreeding using molecular markers and pedigrees: the effect of demographic history. *Moleculary 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.

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.

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.

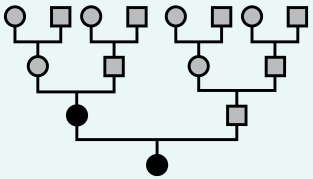
Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

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*.

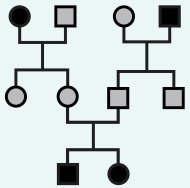
Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, *56*(645), 330–338.

## Addax
*Addax nasomaculatus*



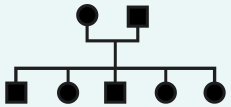
Poor records, high uncertainty

## Inca Tern
*Larosterna inca*



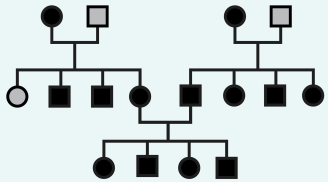
Contemporary missing data

## Kākāriki Karaka
*Cyanoramphus malherbi*



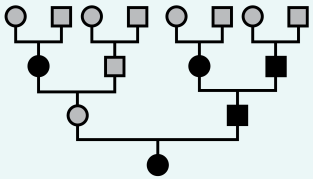
High completeness, but with low depth.

## Kakī
*Himantopus novaezelandiae*

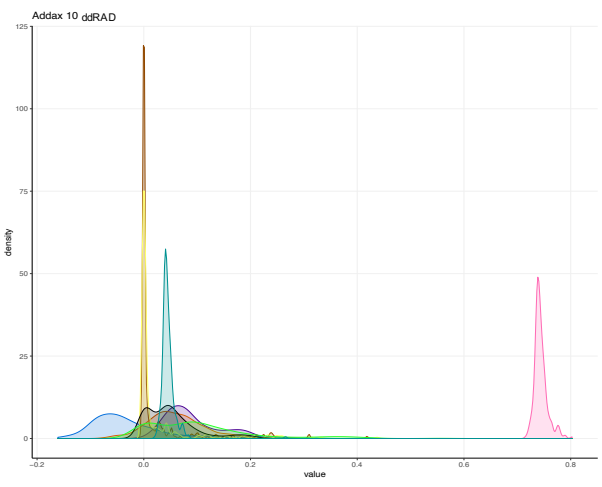


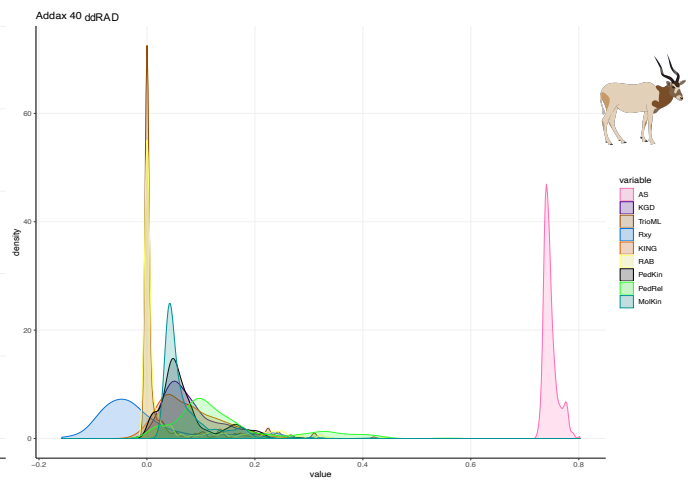
Semi-wild with missing data and uncertainty.

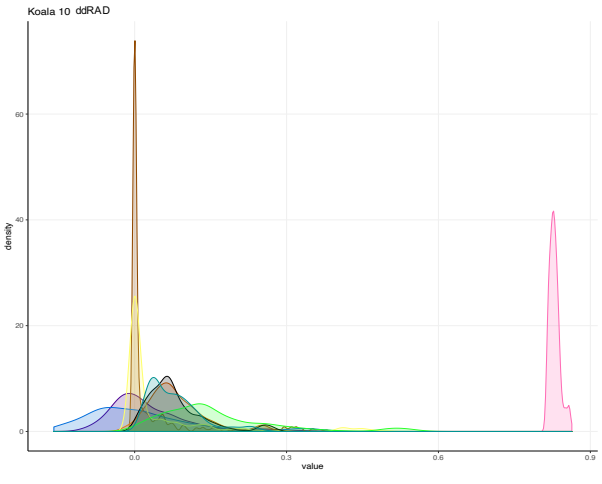
## Koala
*Phascolarctos cinereus*

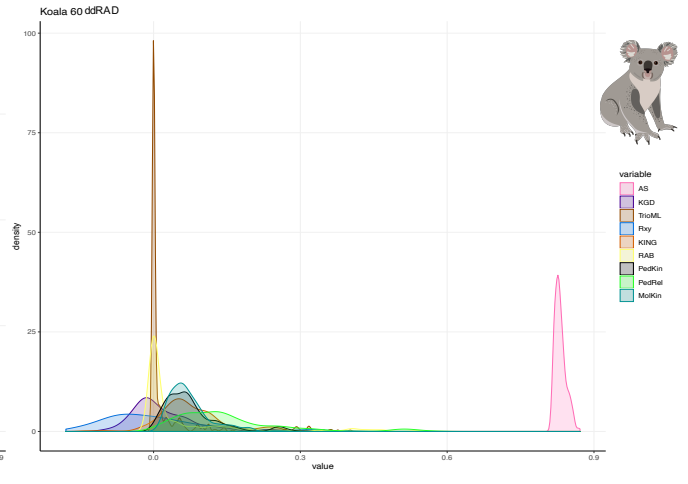


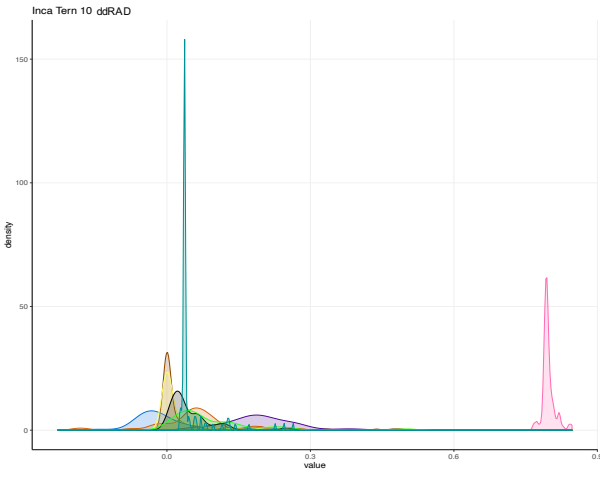
High historic missing data

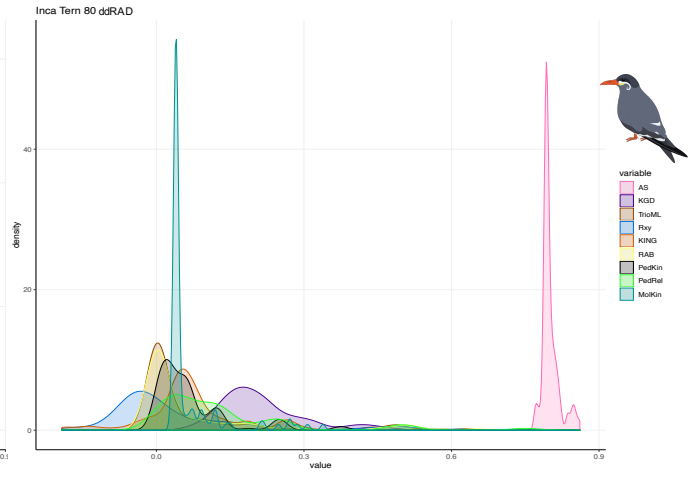

Addax 10 ddRAD
density
value

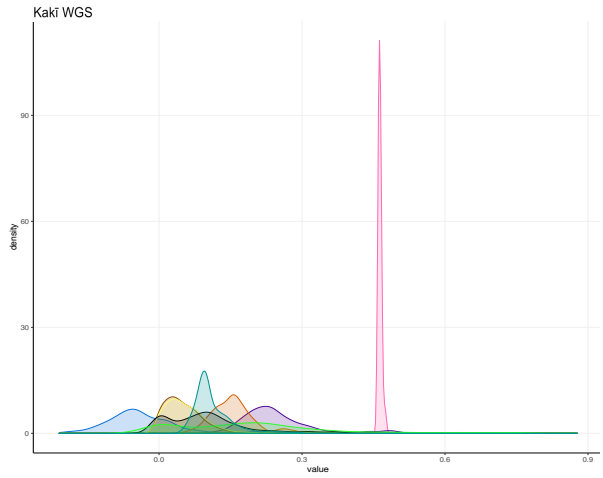

Addax 40 ddRAD
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin

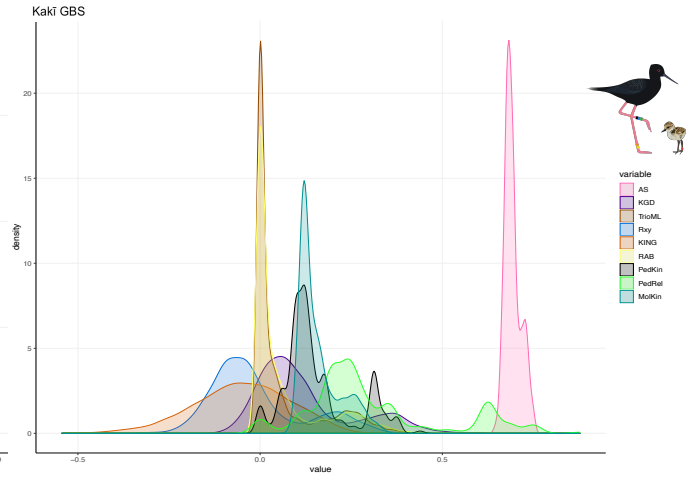

Koala 10 ddRAD
density
value

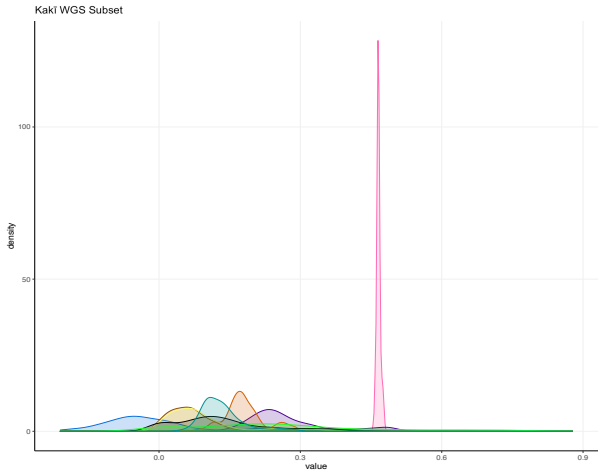

Koala 60 ddRAD
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin

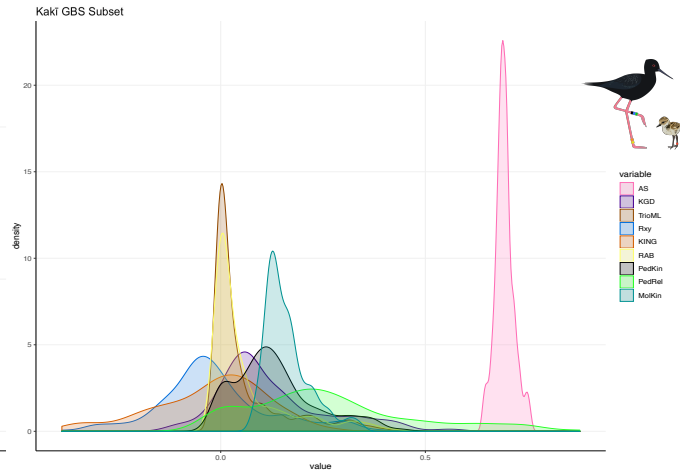

Inca Tern 10 ddRAD
density
value

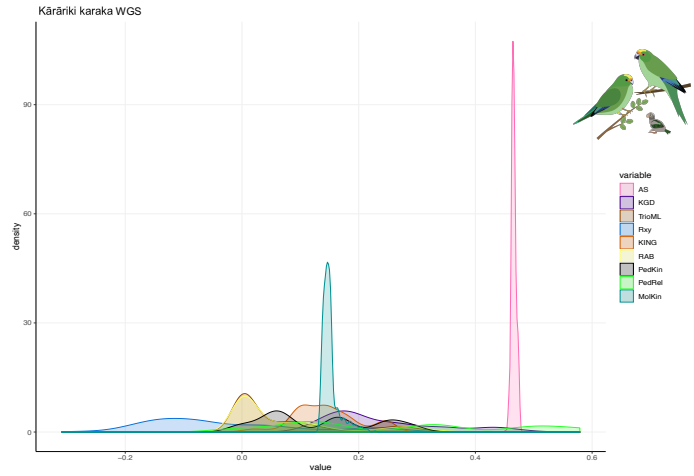

Inca Tern 80 ddRAD
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin


Kakī WGS
density
value

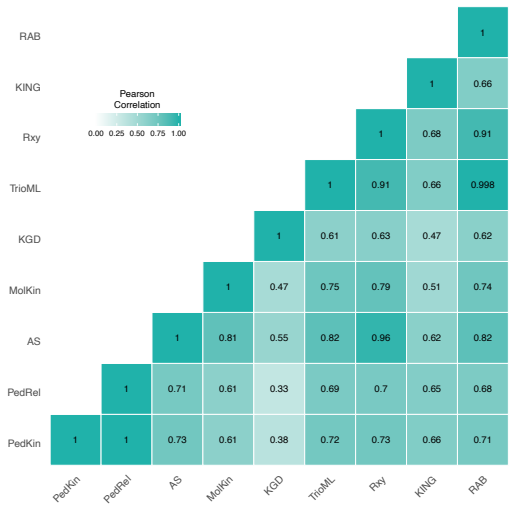

Kakī GBS
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin


Kakī WGS Subset
density
value

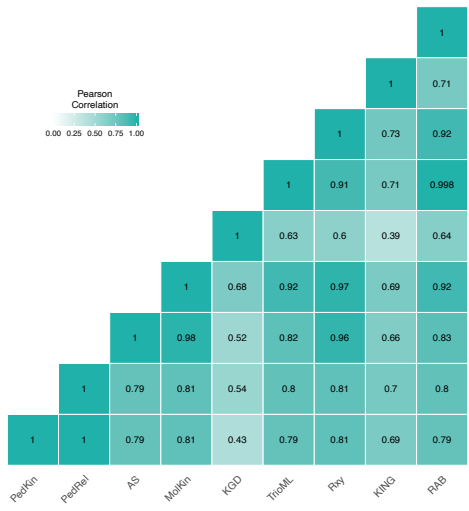

Kakī GBS Subset
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin


Kārāriki karaka WGS
density
value
variable
AS
KGD
TrioML
Rxy
KING
RAB
PedKin
PedRel
MolKin

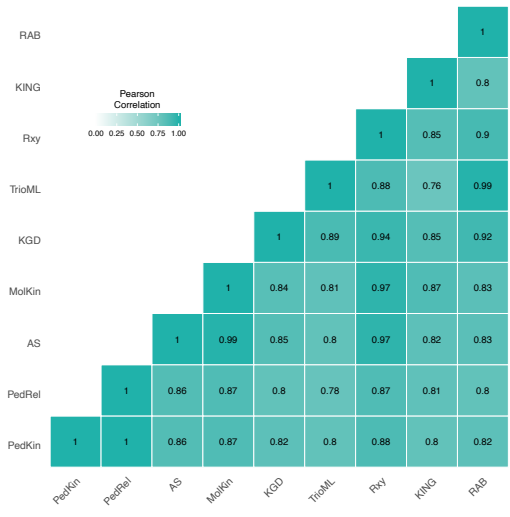
Addax 10 ddRAD



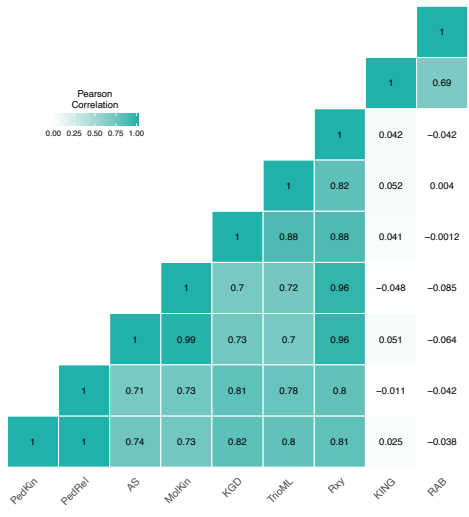

Addax 40 ddRAD



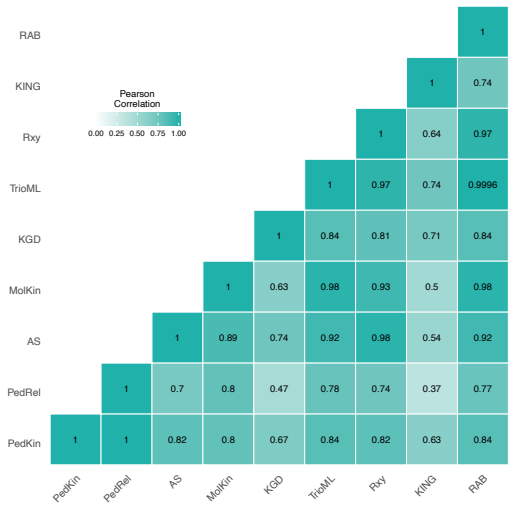
Koala 10 ddRAD



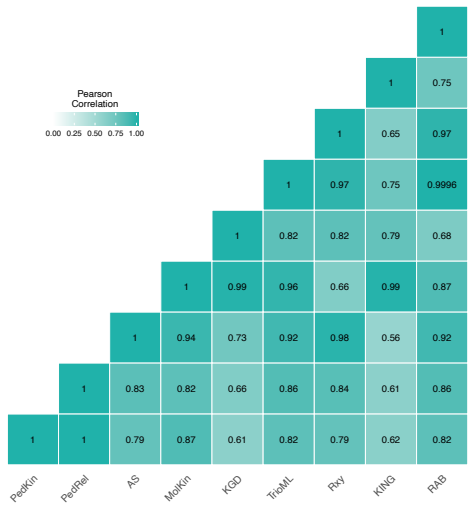

Koala 60 ddRAD



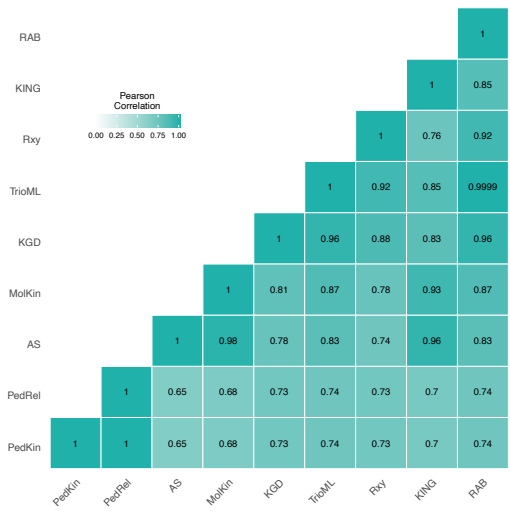
Inca Tern 10 ddRAD



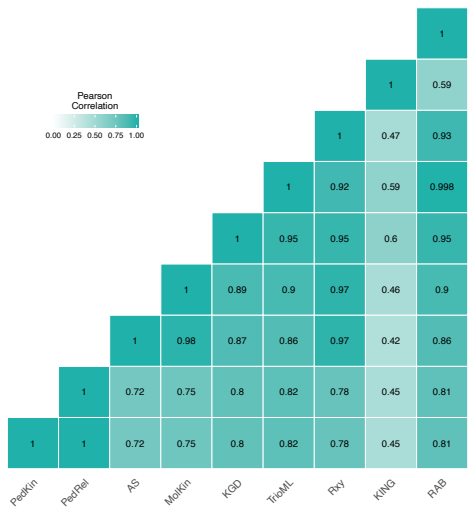

Inca Tern 80 ddRAD



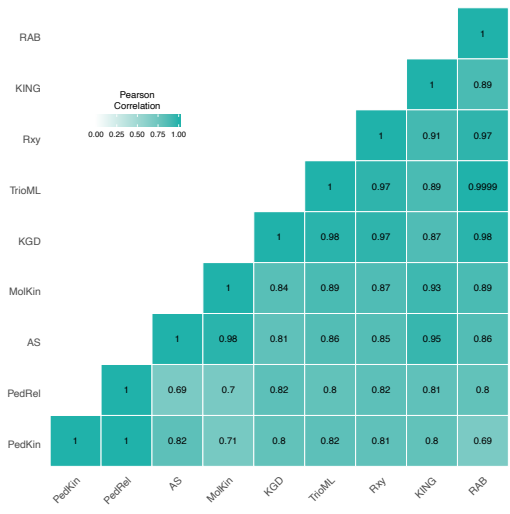
Kakī WGS



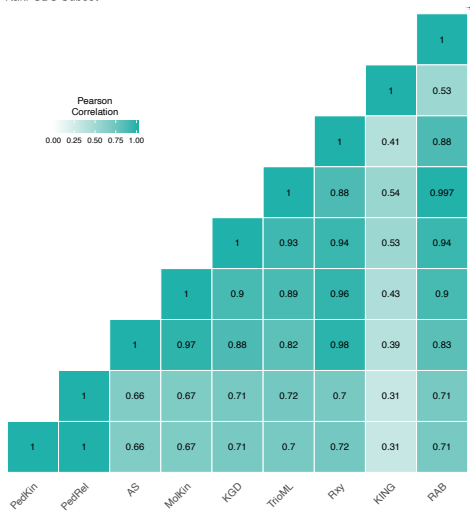

Kakī GBS



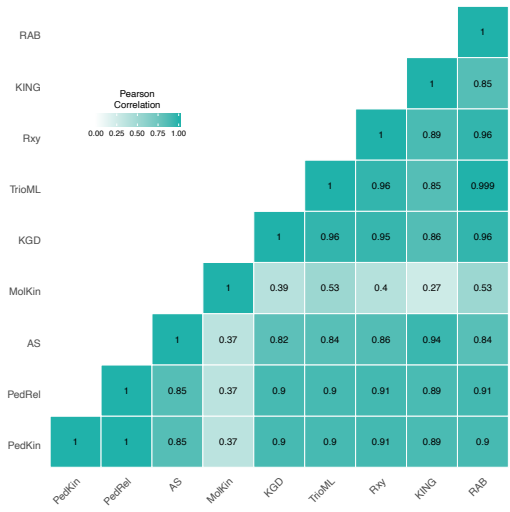
Kakī WGS Subset




Kakī GBS Subset



Kārāriki karaka WGS



Addax 10 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.2 | 0.08 | 0.03 | 0.67 | 0.09 | −0.07 | |
| KING | 0.07 | 0.06 | 0.09 | −0.11 | 0.33 | | −0.32 |
| Rxy | 0.05 | −0.08 | 0.31 | 0.06 | | 0 | −0.11 |
| TrioML | 0.17 | 0.05 | 0.13 | | −0.23 | −0.26 | 0.35 |
| KGD | 0.2 | −0.04 | | 0.16 | −0.17 | 0.18 | 0.14 |
| AS | 0.13 | | −0.01 | −0.12 | 0 | 0.09 | 0.07 |
| Pedigree | | 0.13 | −0.04 | 0.27 | 0.14 | −0.3 | 0.17 |

Addax 40 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | −0.05 | 0.2 | 0.03 | 0.26 | −0.21 | 0.2 | |
| KING | 0.07 | 1 | −0.03 | −0.07 | 0.13 | | −0.06 |
| Rxy | 0 | 0.13 | 0.17 | 0.15 | | −0.04 | 0.31 |
| TrioML | 0.09 | −0.07 | 0.24 | | −0.36 | −0.18 | 0.17 |
| KGD | 0.09 | −0.03 | | −0.21 | 0.09 | 0.17 | −0.24 |
| AS | 0.07 | | 0.17 | −0.18 | −0.04 | 1 | −0.06 |
| Pedigree | | −0.3 | −0.03 | 0.13 | 0.14 | −0.3 | 0.12 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0

Koala 10 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0 | 0 | 0.08 | 0.58 | −0.01 | 0.11 | |
| KING | 0.18 | 0.09 | −0.12 | 0.4 | 0.27 | | 0.06 |
| Rxy | 0.21 | 0.02 | −0.04 | 0.14 | | 0.22 | 0.02 |
| TrioML | 0.15 | 0.15 | 0.04 | | −0.13 | −0.08 | 0.62 |
| KGD | −0.47 | 0.22 | | 0.29 | −0.33 | −0.05 | 0.08 |
| AS | −0.03 | | −0.05 | −0.02 | 0.12 | 0.01 | 0.02 |
| Pedigree | | −0.19 | −0.08 | 0.03 | −0.06 | 0.02 | 0.05 |

Koala 60 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.16 | 0.06 | 0.27 | 0.31 | 0.19 | −0.35 | |
| KING | 0.01 | 0 | −0.19 | −0.16 | 0.02 | | 0.32 |
| Rxy | 0.13 | 0.09 | 0.09 | 0.38 | | 0.02 | −0.41 |
| TrioML | 0.12 | 0.13 | 0.17 | | −0.02 | 0.37 | 0.36 |
| KGD | −0.27 | 0.14 | | 0.09 | 0.06 | 0.16 | 0.06 |
| AS | −0.13 | | −0.11 | 0.27 | 0.01 | −0.11 | 0.05 |
| Pedigree | | −0.12 | −0.08 | 0.02 | −0.26 | −0.04 | 0.01 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0

Inca Tern 10 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0 | 0 | 0.08 | 0.58 | −0.01 | 0.11 | |
| KING | 0.18 | 0.09 | −0.12 | 0.4 | 0.27 | | 0.06 |
| Rxy | 0.21 | 0.02 | −0.04 | 0.14 | | 0.22 | 0.02 |
| TrioML | 0.15 | 0.15 | 0.04 | | −0.13 | −0.08 | 0.62 |
| KGD | −0.47 | 0.22 | | 0.29 | −0.33 | −0.05 | 0.08 |
| AS | −0.03 | | −0.05 | −0.02 | 0.12 | 0.01 | 0.02 |
| Pedigree | | −0.19 | −0.08 | 0.03 | −0.06 | 0.02 | 0.05 |

Inca Tern 80 ddRAD

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | −0.16 | −0.07 | −0.07 | 0.53 | 0.18 | 0.12 | |
| KING | 0.2 | −0.24 | −0.24 | 0.19 | −0.07 | | 0.36 |
| Rxy | −0.14 | 0.09 | 0.09 | −0.02 | | −0.24 | −0.15 |
| TrioML | −0.06 | −0.19 | −0.19 | | −0.14 | 0.37 | 0.74 |
| KGD | 0.14 | 1 | | 0.02 | 0 | −0.04 | 0.01 |
| AS | 0.14 | | 1 | 0.02 | 0 | −0.04 | 0.01 |
| Pedigree | | 0.12 | 0.12 | −0.15 | −0.17 | 0.06 | −0.07 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0

Kakī WGS

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.16 | 0.31 | 0.35 | 0.48 | −0.34 | 0.22 | |
| KING | 0.33 | 0.47 | 0.54 | 0.21 | −0.36 | | 0.25 |
| Rxy | −0.25 | −0.48 | −0.38 | −0.24 | | −0.07 | −0.11 |
| TrioML | 0.29 | 0.29 | 0.23 | | −0.11 | 0.19 | 0.89 |
| KGD | 0.18 | 0.75 | | 0.31 | −0.19 | 0.4 | 0.28 |
| AS | 0.24 | | 0.71 | 0.18 | −0.27 | 0.53 | 0.24 |
| Pedigree | | 0.53 | 0.53 | 0.21 | −0.18 | 0.37 | 0.19 |

Kakī GBS

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.38 | 0.47 | 0.41 | 0.53 | −0.24 | 0.27 | |
| KING | 0.28 | 0.13 | −0.04 | 0.12 | −0.11 | | 0.19 |
| Rxy | −0.32 | 0.12 | 0.03 | −0.04 | | −0.02 | 0.16 |
| TrioML | 0.2 | 0.42 | 0.53 | | 0.12 | 0.02 | 0.37 |
| KGD | 0.17 | 0.52 | | 0.33 | −0.03 | −0.02 | 0.43 |
| AS | 0.08 | | 0.51 | 0.51 | 0.04 | 0.07 | 0.34 |
| Pedigree | | 0.2 | 0.27 | 0.32 | 0.02 | −0.14 | 0.25 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0

Kakī WGS Subset

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.29 | 0.54 | 0.61 | 0.8 | 0.03 | 0.62 | |
| KING | 0.4 | 0.71 | 0.67 | 0.57 | 0.03 | | 0.28 |
| Rxy | 0.13 | 0.05 | 0.06 | 0.06 | | 0.09 | 0.12 |
| TrioML | 0.28 | 0.48 | 0.56 | | 0.15 | 0.44 | 0.61 |
| KGD | 0.26 | 0.78 | | 0.33 | −0.04 | 0.63 | 0.27 |
| AS | 0.38 | | 0.8 | 0.34 | −0.09 | 0.73 | 0.24 |
| Pedigree | | 0.24 | 0.25 | 0.17 | −0.16 | 0.22 | 0.39 |

Kakī GBS Subset

| Female \ Male | Pedigree | AS | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | 0.21 | 0.34 | 0.45 | 0.83 | −0.06 | −0.02 | |
| KING | 0.31 | −0.05 | 0.15 | 0.02 | −0.28 | | −0.04 |
| Rxy | −0.01 | −0.09 | −0.13 | −0.13 | | −0.14 | −0.02 |
| TrioML | 0.27 | 0.36 | 0.49 | | 0.29 | 0 | 0.26 |
| KGD | 0.4 | 0.63 | | 0.27 | 0.14 | −0.03 | 0.21 |
| AS | 0.26 | | 0.56 | 0.34 | 0.22 | 0.14 | 0.15 |
| Pedigree | | 0.25 | 0.19 | 0.2 | 0 | 0.29 | 0.22 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0

Kārāriki karaka WGS

| Female \ Male | Pedigree | MolKin | KGD | TrioML | Rxy | KING | RAB |
|---|---|---|---|---|---|---|---|
| RAB | −0.04 | −0.23 | −0.06 | 0.49 | −0.12 | 0.21 | |
| KING | −0.14 | 0.07 | −0.22 | 0.17 | 0.08 | | −0.04 |
| Rxy | 0.51 | 0.38 | 0.38 | −0.14 | | −0.11 | 0.03 |
| TrioML | −0.13 | −0.21 | −0.13 | | 0.13 | −0.14 | 0.46 |
| KGD | 0.52 | 0.16 | | 0.38 | 0.03 | −0.06 | 0.35 |
| MolKin | 0.26 | | −0.03 | 0.21 | 0.16 | −0.08 | 0.12 |
| Pedigree | | 0.49 | 0.05 | 0.21 | 0.01 | −0.08 | 0.33 |

Kendall−Tau Correlation: 1.0, 0.5, 0.0, −0.5, −1.0