




Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias

Catherine R. M. Attard  | Luciano B. Beheregaray  | Luciana M. Möller 

Molecular Ecology Lab, College of Science and Engineering, Flinders University, Adelaide, SA, Australia

Correspondence

Catherine R. M. Attard, Molecular Ecology Lab, College of Science and Engineering, Flinders University, Adelaide, SA, Australia.
Email: catherine.attard@flinders.edu.au

Funding information

Sea World Research and Rescue Foundation Incorporated (SWRRFI); Winifred Violet Scott Charitable Trust (WVSCT); Australian Marine Mammal Centre; Flinders University; Macquarie University; Australian Research Council, Grant/Award Number: FT130101068

Abstract

There has been remarkably little attention to using the high resolution provided by genotyping-by-sequencing (i.e., RADseq and similar methods) for assessing relatedness in wildlife populations. A major hurdle is the genotyping error, especially allelic dropout, often found in this type of data that could lead to downward-biased, yet precise, estimates of relatedness. Here, we assess the applicability of genotyping-by-sequencing for relatedness inferences given its relatively high genotyping error rate. Individuals of known relatedness were simulated under genotyping error, allelic dropout and missing data scenarios based on an empirical ddRAD data set, and their true relatedness was compared to that estimated by seven relatedness estimators. We found that an estimator chosen through such analyses can circumvent the influence of genotyping error, with the estimator of Ritland (*Genetics Research*, 67, 175) shown to be unaffected by allelic dropout and to be the most accurate when there is genotyping error. We also found that the choice of estimator should not rely solely on the strength of correlation between estimated and true relatedness as a strong correlation does not necessarily mean estimates are close to true relatedness. We also demonstrated how even a large SNP data set with genotyping error (allelic dropout or otherwise) or missing data still performs better than a perfectly genotyped microsatellite data set of tens of markers. The simulation-based approach used here can be easily implemented by others on their own genotyping-by-sequencing data sets to confirm the most appropriate and powerful estimator for their data.

KEYWORDS

double-digest restriction site-associated DNA, low coverage, next-generation sequencing, pedigree, population genomics, relationships

1 | INTRODUCTION

Next-generation sequencing has allowed us to answer questions about ecology and evolution that were once confined to our imagination. The orders of magnitude increase in marker numbers in genomic (i.e., genome-wide) data results in greater power for use in traditional analyses that are based on neutral loci and the ability to identify the small proportion of adaptive loci in the genome (Allendorf, Hohenlohe, & Luikart, 2010; Narum, Buerkle, Davey, Miller, &

Hohenlohe, 2013; Stapley et al., 2010). Genotyping-by-sequencing (i.e., RADseq and similar methods) in particular allows markers to be simultaneously identified and genotyped, avoiding the need to first develop a panel of markers that then have inherent bias when new populations are surveyed (Davey et al., 2011). However, there has been remarkably little attention to using genotyping-by-sequencing for research that especially requires high resolution, such as questions that require identifying related individuals in the absence of a known pedigree. Ecologically relevant research that would benefit

from identifying related individuals includes investigations of breeding behaviour, social structure, inbreeding and inbreeding depression, and population demographics and connectivity (Coleman & Jones, 2011; Iacchei et al., 2013; Kjeldsen et al., 2016; Möller, 2012; Norman et al., 2017; Palsbøll, Peery, & Bérubé, 2010; Peery et al., 2008; Ross, 2001). Related individuals can be identified by estimating relatedness or inferring kinship categories: relatedness is a continuous measure of the proportion of alleles in a dyad (i.e., in a pair of individuals, regardless of whether the pair was observed together) that are identical by descent (IBD) relative to a reference population; kinship categories are discrete categories of dyad relationship such as parent–offspring or full-sibling.

Estimating relatedness or inferring kinship categories with molecular markers has traditionally been conducted using microsatellites (Jones, Small, Paczolt, & Ratterman, 2010; Pemberton, 2008). Microarray technology and, more recently, next-generation sequencing have opened up the possibility of using SNPs (Davey et al., 2011; Glaubitz, Rhodes, & Dewoody, 2003). SNPs are less informative per locus than microsatellites for identifying related individuals because they are bi-allelic and often more skewed in allele frequency, but they can make up for this by their greater abundance in the genome and therefore the potential for much larger marker sets. For model organisms and organisms of agricultural importance, there are now commercial SNP panels that contain thousands to hundreds of thousands of SNPs used for marker–trait association studies and genomic selection for breeding (Goddard & Hayes, 2007, 2009; Pe'er et al., 2006). In nonmodel species, time and expense have been placed into developing SNP panels of typically several tens to hundreds of loci to infer relationships (Kaiser et al., 2017; Labuschagne, Nupen, Kotzé, Grobler, & Dalton, 2015; Liu, Palti, Gao, & Rexroad Iii, 2016; Weinman, Solomon, & Rubenstein, 2015; Wright et al., 2015), or effort has been placed into assessing commercially available SNP panels for cross-amplification in these species (Haynes & Latch, 2012; Ivy, Putnam, Navarro, Gurr, & Ryder, 2016). SNP panels have been shown to be as powerful as (Glaubitz et al., 2003; Kaiser et al., 2017; Weinman et al., 2015) or more powerful (Labuschagne et al., 2015; Santure et al., 2010) than microsatellite data sets.

SNP panels have little genotyping error, have little missing data, and are highly repeatable compared with genotyping-by-sequencing data (Hoffman et al., 2012; Nielsen, Paul, Albrechtsen, & Song, 2011), making them a safe option for relatedness analyses. However, genotyping-by-sequencing is arguably the most popular approach currently used in population genomic studies of nonmodel organisms (Catchen et al., 2017; Narum et al., 2013). Studies that have compared genotyping-by-sequencing data of thousands of SNPs with traditional microsatellite data for relatedness analyses have suggested that the former shows better performance (e.g., Hellmann et al., 2016; Hoffmann et al., 2014; Rašić, Filipović, Weeks, & Hoffmann, 2014), potentially making genotyping-by-sequencing a viable option for assessing relatedness in wildlife populations. However, we have found no study that has assessed how the relatively high genotyping error and missing data in genotyping-by-sequencing data sets

may influence relatedness estimates (see literature search in Methods). Genotyping error in genotyping-by-sequencing data also includes allelic dropout, which occurs because of the high chance for only one of the two alleles to be sampled at relatively low depths of coverage. When only one of two alleles is sampled, a homozygote will be called correctly as a homozygote, but a heterozygote may be called incorrectly as a homozygote of the sampled allele. Note that this meaning of allelic dropout is different to that often used in the RAD literature to refer to null alleles from a mutation in the restriction enzyme cut site (e.g., Gautier et al., 2013), which is more predominant when comparing highly divergent populations or species, and therefore not examined further here. While the genotyping error and allelic dropout of typical genotyping-by-sequencing data sets are not thought to influence population-level analyses (Buerkle & Gompert, 2013), it could influence higher-resolution analyses such as when estimating relatedness and inferring kinship categories. Large data sets that are analysed poorly can produce precisely biased results (e.g., Peery et al., 2013). This is when there is little variance around the result (i.e., high precision), which may give the appearance of confidence in the result, but actually, the result is not close to the true value (i.e., low accuracy). For relatedness analyses using genotyping-by-sequencing, there is the expectation of high precision due to the large number of markers, but decreased accuracy through downward-biased estimates due to genotyping error.

The issue of genotyping error, including allelic dropout, could perhaps be circumvented by choosing a relatedness estimator that is most appropriate for the data set at hand. It is well known from microsatellites that the accuracy and precision of a particular relatedness estimator depends on the data set, altering with the number of loci, number of alleles, allele frequency distribution and relationship structure in the population (Van de Castele, Galbusera, & Matthysen, 2001). Ideally, the power and reliability of different estimators should be assessed to choose the most appropriate estimator for the data. This requires knowing the true relatedness or kinship categories of a number of individuals, and then assessing how close are different relatedness estimates to the true relatedness (e.g., Santure et al., 2010). Unfortunately, it is rare for pedigrees to be well known when studying wild populations. A solution to this is to use known relationships simulated from empirical data (Taylor, 2015; Wang, 2011). A user-friendly simulation-based approach has already been developed in the program COANCESTRY with microsatellites in mind: pairs of known relatedness are simulated *in silico* based typically on the allele frequencies of the empirical population, the relatedness of these simulated pairs is then estimated using the candidate relatedness estimators, and the estimator with the strongest correlation to true relatedness is usually chosen to estimate the relatedness of empirical individuals (Wang, 2011). There has been poor uptake of simulation-based approaches, even with their user-friendly availability for microsatellite studies. A review by Taylor (2015) of literature citing COANCESTRY highlighted that only 9% of studies used simulations to, in some way, select and report the performance of the best estimator. Negligible understanding of an estimator's performance can

make subsequent inferences debatable, especially when they are about specific dyads rather than average relatedness of groups.

The behaviour of different relatedness estimators when using thousands of bi-allelic markers with relatively high rates of genotyping error is, to our knowledge, unknown. Here, we assess the applicability of genotyping-by-sequencing for relatedness inferences given its relatively high genotyping error rate—especially allelic dropout—and missing data. We also demonstrate, using empirical ddRAD data, how a user-friendly simulation-based approach can be implemented to make such an assessment on any genotyping-by-sequencing data set. To the best of our knowledge, this is the first time that the influence of genotyping error, allelic dropout or missing data on relatedness estimates has been assessed in genotyping-by-sequencing data sets (see literature search in Methods). We found remarkable precision of relatedness estimates, despite genotyping errors and missing data, but the choice of estimator was imperative for circumventing the downward-biased inaccuracies caused by genotyping error and allelic dropout. Typically used correlation assessments were found to be insufficient for determining the most appropriate estimator; even relatedness estimators that had downward-biased estimates showed > 0.99 correlation to true relatedness. We also compared the empirical ddRAD data to microsatellite data from the same population to add to the growing evidence of the greater power of large SNP data sets for estimating relatedness, even with high proportions of genotyping error and allelic dropout.

2 | METHODS

2.1 | Literature search

To ascertain how SNPs have been used to date for relatedness inferences in ecologically relevant studies, we searched for articles published up to and including September 2017 in Web of Science. The term combination used was as follows: Topic (TS) = (SNP* AND relatedness) AND Web of Science Category (WC) = (Environmental Sciences OR Evolutionary Biology OR Zoology OR Ecology). WC search terms were used to avoid papers that were about agricultural or human marker–trait association studies, or genomic selection for breeding, which often use commercially available SNP panels. These studies also often use realized genomic similarity (i.e., identical by state [IBS]) rather than pedigree-based relatedness (i.e., IBD given a reference population) (Goddard, Hayes, & Meuwissen, 2011; Speed & Balding, 2015). Of the remaining studies, we recorded the study species, the number of SNPs in the final data set, and whether the SNP data was obtained through genotyping-by-sequencing or a SNP panel (the former involving simultaneously discovering and genotyping SNPs; the latter involving genotyping a set of predefined SNPs). We recorded whether the study assessed the robustness of the relatedness estimates by comparing different relatedness estimators, determining the influence of genotyping errors (including allelic dropout) or missing data, comparing the data set to a microsatellite data set, or comparing the estimates to simulated or empirical individuals of known relatedness.

2.2 | Relatedness analyses

We chose to conduct all relatedness analyses using `COANCESTRY` 1.0.1.6 (Wang, 2011) because it implements multiple pairwise relatedness estimators, has a built-in module for assessing the reliability of different estimators by simulation, is amenable to simulating genotyping error, allelic dropout and missing data, and has a user-friendly interface. The relatedness estimators available in the program are the moment estimators of Queller and Goodnight (1989), Li, Weeks, and Chakravarti (1993), Ritland (1996), Lynch and Ritland (1999), and Wang (2002) (which reduces to Li et al. (1993) for bi-allelic loci (Wang, 2016)), and the dyadic maximum-likelihood estimator of Milligan (2003) and triadic maximum-likelihood estimator of Wang (2007) (we used 100 reference individuals for the triadic estimator). These have already been used to estimate relatedness in genotyping-by-sequencing data sets, but without assessing the influence of genotyping error (Escoda, González-Esteban, Gómez, & Castresana, 2017; de Fraga, Lima, Magnusson, Ferrão, & Stow, 2017; Hellmann et al., 2016). Only the maximum-likelihood estimators have the capability to incorporate genotyping error into their estimations, but this requires the genotyping error rate to be known for each locus (which is rare for genotyping-by-sequencing data), and these estimators do not specifically consider allelic dropout, so this capability was not considered here. As `COANCESTRY` and most other relatedness or kinship category programs were designed with tens of loci in mind, we encountered issues when using thousands of loci. We consulted with the author of `COANCESTRY`, Jinliang Wang, who consequently updated `COANCESTRY` (to 1.0.1.6 and 1.0.1.7) to make it amenable to large data sets and capable of simulating allelic dropout. We note that the R package `RELATED` 1.0 (Pew, Muir, Wang, & Frasier, 2015)—which is based on `COANCESTRY`—also has a simulation module, but could not handle large data sets when conducting simulations at the time of this study (T. R. Frasier pers. comm.).

We conducted our assessment using empirical data from a long-lived animal, the blue whale (*Balaenoptera musculus*), and specifically from the population of pygmy blue whales (*B. m. breviceauda*) inhabiting Australian waters. To our knowledge, this is the first study assessing relatedness or inferring kinship categories in blue whales, emphasizing how this approach can be used even in nonmodel systems with nonexistent pedigree information. We chose the population inhabiting Australian waters as we have both a genotyping-by-sequencing data set of 8,294 filtered SNPs ($n = 68$) and a microsatellite data set of 20 loci ($n = 110$) from this population (Attard et al., 2018) [see Attard et al. (2010, 2012, 2015); Attard, Beheregaray, and Möller (2016) for associated studies], allowing a comparison of these data types. The SNPs were developed by preparing libraries following the ddRAD protocol of Peterson, Weber, Kay, Fisher, and Hoekstra (2012) modified as detailed in Brauer, Hammer, and Beheregaray (2016). Data are based on 100-bp paired-end sequences from an Illumina HiSeq 2000, and resulting reads were processed using the de novo pipeline of `STACKS` (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). The final SNP data set

consisted of SNPs that were in at least 70% of samples, had a minor allele frequency of at least 0.05 and were the first SNP from the associated ddRAD locus. Individuals had no more than 40% missing data.

The most appropriate relatedness estimator for the SNP and microsatellite data sets was assessed independently by simulating one thousand pairs of dyads for each of four kinship categories—unrelated, half-sibling, full-sibling and parent-offspring—based on the allele frequencies of the empirical individuals. There is essentially an infinite number of possible kinship categories that could be simulated, and so we assessed the most common categories considered in ecological studies. Relatedness was then estimated for each simulated pair based on the allele frequencies of the empirical individuals using independently the seven relatedness estimators. We conducted simulations with either no missing data or 40% missing data, and either with no genotyping error, 0.04 genotyping error, or 0.2 allelic dropout, resulting in six different simulated data sets for each marker type. While the actual genotyping and allelic dropout rates are unknown, and can vary considerably between individuals regardless of sample quality (Fountain, Pauli, Reid, Palsbøll, & Peery, 2016), the same pattern in the relative performance of the estimators is expected regardless of the level of error. The amount of simulated missing data was chosen based on our maximum allowed 40% missing data for individuals in the empirical data. We simulated allelic dropout using a custom script (see Data Accessibility) as the *COANCESTRY* version available at the time of our analysis, 1.0.1.6, could not simulate allelic dropout (*COANCESTRY* 1.0.1.7 can now simulate allelic dropout, which we recommend others to use as it is more user-friendly than the script). The script considers each simulated heterozygous genotype, randomly samples a decimal number between 0 and 1, and changes the heterozygote to a homozygote if this number is less than the user-supplied value for allelic dropout. The final estimator for each data set was chosen based on the accuracy (closeness to the true value) and precision (variation in estimated values) of each estimator across kinship categories, which was assessed by calculating means and standard deviations for each kinship category, visualization of percentiles using box plots and the correlation to true relatedness as determined by Pearson's correlation coefficient calculated in *COANCESTRY*. The chosen estimator was then used for empirical relatedness estimates.

3 | RESULTS

3.1 | Literature search

We found 63 peer-reviewed papers, 19 of which were manually identified as ecologically relevant studies that involved estimating relatedness using SNPs (Table S1). Of these, none assessed the influence of genotyping errors (including allelic dropout) or missing data. Only four used genotyping-by-sequencing data; all of these were published within the last 3 years. Two of these assessed different relatedness estimators to choose an estimator for their study,

specifically by simulating known individuals using the R package *RELATED*, but assuming a perfectly genotyped data set with no missing data (Escoda et al., 2017; de Fraga et al., 2017). The SNP panel studies had 22–771 SNPs, except for two that were based on a commercial 65-k equine and 500-k human SNP chip but were conservatively kept in the literature review due to their potential ecological relevance for relatedness inferences (see Table S1). Conversely, the genotyping-by-sequencing data sets typically had markers in the thousands, ranging from 720 to 7,805 SNPs. The studies that compared SNPs with microsatellites tended to conclude that SNPs performed better (Hellmann et al., 2016; Santure et al., 2010), except when there were a few hundred or less SNPs (Glaubitz et al., 2003; Ross et al., 2014; Santure et al., 2010; Seddon, Parker, Ostrander, & Ellegren, 2005), which is rare in genotyping-by-sequencing studies.

3.2 | Relatedness analyses

In the simulated SNP data, relatedness estimators showed similar, high accuracy when there was no genotyping error or allelic dropout, and showed a decrease in their precision when there was missing data (Figure 1; see Table S2 for means and standard deviations). However, the estimators differed in accuracy when there was genotyping error or allelic dropout. Only the estimator of Ritland (1996) was unaffected by allelic dropout for the simulated kinship categories. This estimator was also the most accurate when there was genotyping error. Despite this, all relatedness estimators had >0.99 correlation to true relatedness for the SNPs, regardless of the simulated properties of the data set (i.e., genotype error, allelic dropout or missing data; Table 1). While Ritland (1996) tended to not be as precise as other estimators, with it showing comparatively high variance in relatedness estimates of simulated relationships (except for unrelated individuals; Figure 1, Table S2), the difference was considered negligible given the high precision of the SNPs and the estimator's robustness to genotyping error and allelic dropout. This led us to choose the Ritland (1996) estimator for analysing the empirical SNP data. In contrast, the simulated microsatellite data showed a much lower correlation of relatedness estimates to true relatedness (0.395–0.875; Table 1) and a lower precision compared with the SNP data, even when SNPs were simulated under genotyping error, allelic dropout or missing data scenarios (Figure 1; Table S2). The dyadic maximum-likelihood estimator had the greatest correlation in the microsatellites across all simulated scenarios, with a minimum correlation of 0.704, which led us to choose this as the best estimator for the microsatellite data.

In the empirical SNP data, we identified three first-degree relatives (parent-offspring or full-sibling) (Figure 2). This was based on these dyads having Ritland (1996) estimates from 0.492 to 0.436, with biased downward estimates in accordance with our simulations of genotyping error, and the next closest relatedness estimate being 0.211 and a likely second-degree relationship (half-sibling, avuncular or grandparent-grandchild). Only under a situation of inbreeding would one expect true relatedness values between these estimated

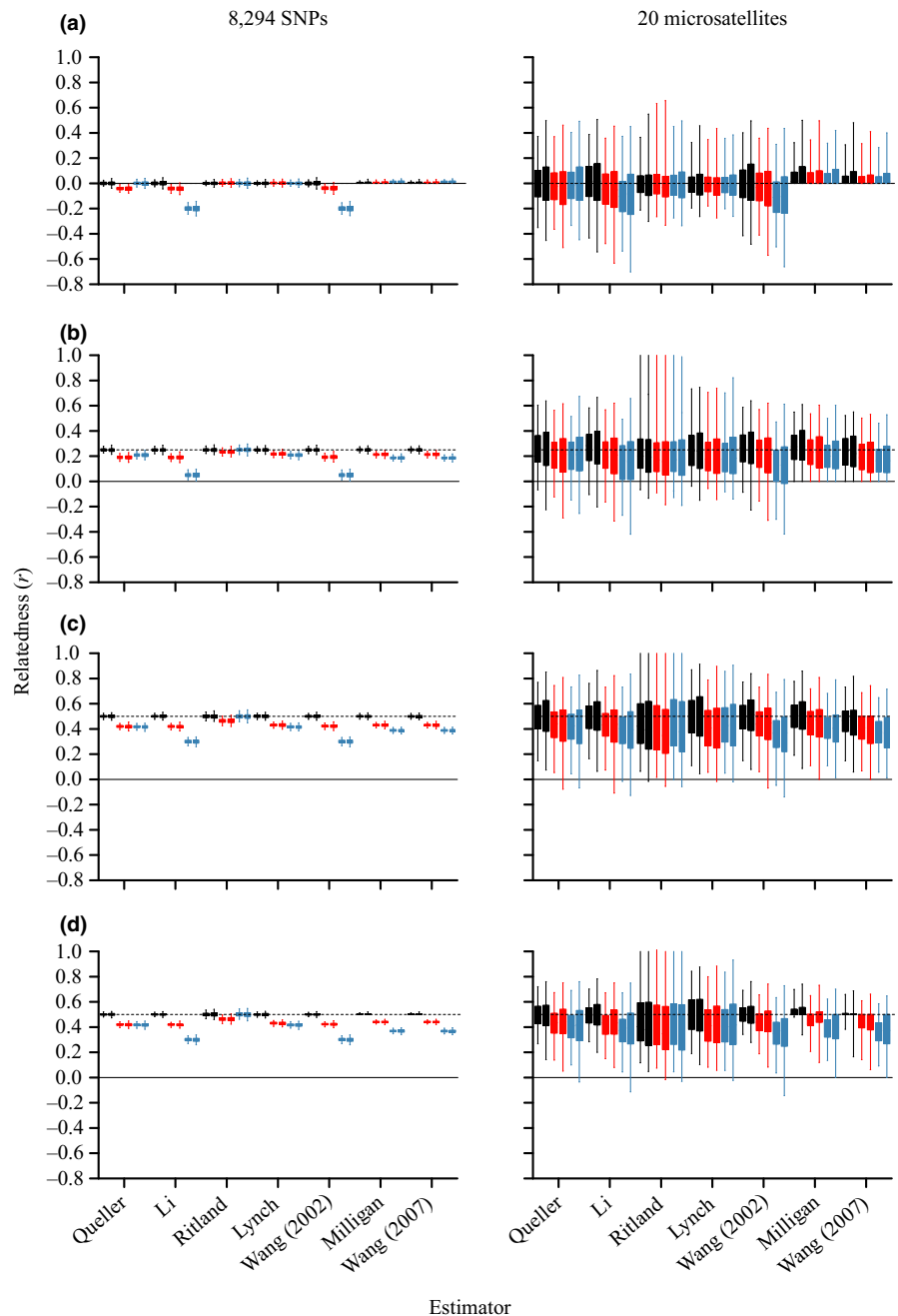


FIGURE 1 Box plots of relatedness estimates from simulated (a) unrelated, (b) half-sibling, (c) full-sibling and (d) parent-offspring dyads for SNP and microsatellite data sets of pygmy blue whales from Australia. Simulations consisted of 1,000 dyads per kinship category, either without genotyping error (black), with 0.04 genotyping error (red) or with 0.2 allelic dropout (blue). Each of these was also simulated without missing data (first box plot) or with 0.4 missing data (second box plot). The boxes represent the 25–75 percentiles, and the whiskers represent the 1–99 percentiles (the 99 percentile has been truncated to 1 in the Ritland estimator for related dyads in microsatellites to allow easier visualization). Zero relatedness (solid horizontal line) and the expected true (average or absolute) relatedness for related categories (dashed horizontal line) are marked

values. We conservatively classified three additional dyads as second-degree relatives based on their relatively high relatedness estimates compared with the remainder of the data set, with other relatives of lower relatedness values also detected in the data but conservatively not classified into a kinship category. A similar categorization based on the microsatellite data using the dyadic maximum-likelihood estimator resulted in a large amount of false positives, with 254 dyads (when only counting dyads of individuals present in the SNP data) having an equal or greater relatedness than the seven first- and second-degree relatives detected in the SNP data (Figure 2). This illustrates the inability to reliably classify the degree of relationship in even highly related individuals when using the microsatellite data.

4 | DISCUSSION

The greater abundance of SNPs in the genome is able to counteract their low information content per locus, leading to genome-wide SNP data often being recognized as more powerful than traditional microsatellite data. However, this fails to consider the relatively high genotyping error expected from genotyping-by-sequencing, which may influence high-resolution analyses such as relatedness estimation. We showcased the use of a simulation-based approach for assessing the reliability of relatedness estimators for genotyping-by-sequencing data, taking into account the relatively high genotyping error, allelic dropout and missing data rates typical of this data type. We found that the choice of estimator can circumvent the influence

TABLE 1 Pearson's correlation coefficient between relatedness estimates and true relatedness for simulations conducted in COANCESTRY using different proportions of missing data, genotype error, and allelic dropout, and using empirical data from pygmy blue whales in Australia

Genotype error	0	0	0.04	0.04	0	0
Allelic dropout	0	0	0	0	0.2	0.2
Missing data	0	0.4	0	0.4	0	0.4

8,294 SNPs						
Queller	0.998	0.998	0.998	0.997	0.997	0.996
Li	0.998	0.997	0.998	0.996	0.997	0.995
Ritland	0.998	0.997	0.997	0.996	0.997	0.995
Lynch	0.999	0.998	0.998	0.997	0.998	0.996
Wang (2002)	0.998	0.997	0.998	0.996	0.997	0.995
Milligan	0.999	0.999	0.998	0.998	0.997	0.996
Wang (2007)	0.999	0.999	0.998	0.998	0.997	0.996

20 microsatellites						
Queller	0.827	0.753	0.792	0.724	0.762	0.661
Li	0.823	0.749	0.787	0.717	0.763	0.665
Ritland	0.578	0.464	0.525	0.435	0.506	0.395
Lynch	0.799	0.735	0.732	0.681	0.733	0.665
Wang (2002)	0.833	0.761	0.800	0.734	0.765	0.672
Milligan	0.875	0.803	0.832	0.764	0.800	0.704
Wang (2007)	0.871	0.796	0.821	0.751	0.790	0.686

of genotyping error and allelic dropout and that this choice should not only rely on the degree of correlation between estimated and true relatedness. We also showed that genotyping-by-sequencing data with genotyping error (allelic dropout or otherwise) or missing data can perform better than traditional microsatellite data that has been perfectly genotyped.

Specifically, while all seven relatedness estimators assessed here were strongly correlated with true relatedness when using the SNP data (>0.99 for all simulations; Table 1), inference of accuracy and precision through box plots (Figure 1) and means and standard

deviations (Table S2) revealed substantial differences in the accuracy of the estimators under simulations of genotyping error or allelic dropout. Therefore, unlike previous recommendations which were based on traditional data sets (Wang, 2011), an estimator should not be chosen based on only the strongest correlation between true and estimated relatedness; a strong correlation inherently does not mean the estimates are close to true relatedness. Of the seven examined estimators, we found that only the relatedness estimator of Ritland (1996) was not influenced by allelic dropout and that it was also the least influenced by other genotyping errors. While Ritland (1996) tended to not be as precise as other estimators, with it showing comparatively high variance in relatedness estimates of simulated relationships (except for unrelated individuals; Figure 1, Table S2), we considered the difference in precision negligible given the high power of the SNP data and the estimator's robustness to genotyping error and allelic dropout.

To the best of our knowledge, no previous study has assessed how genotyping error and allelic dropout can influence relatedness estimates from genotyping-by-sequencing data sets, despite the current popularity of such data in population genomics. Most SNP-based relatedness estimates in nonmodel organisms still rely on SNP panels (Table S1), presumably because of their lower genotyping error and missing data rates. The closest study found to ours was by Escoda et al. (2017), who assessed a genotyping-by-sequencing data set for estimating relatedness based on perfectly genotyped simulations. Similar to our study, they found all estimators had >0.97 correlation to true relatedness. Extremely high correlations may be expected from genotyping-by-sequencing data due to the higher number of markers compared with a typical SNP panel for a non-model species or a microsatellite data set. Escoda et al. (2017) also considered the precision of the estimators (their standard deviations were 0.01–0.06; ours for a perfectly genotyped data set were 0.004–0.016, Table S2). Taking together their assessments of accuracy and precision, they named the dyadic and triadic maximum-likelihood estimators as the best estimators for their data. For the

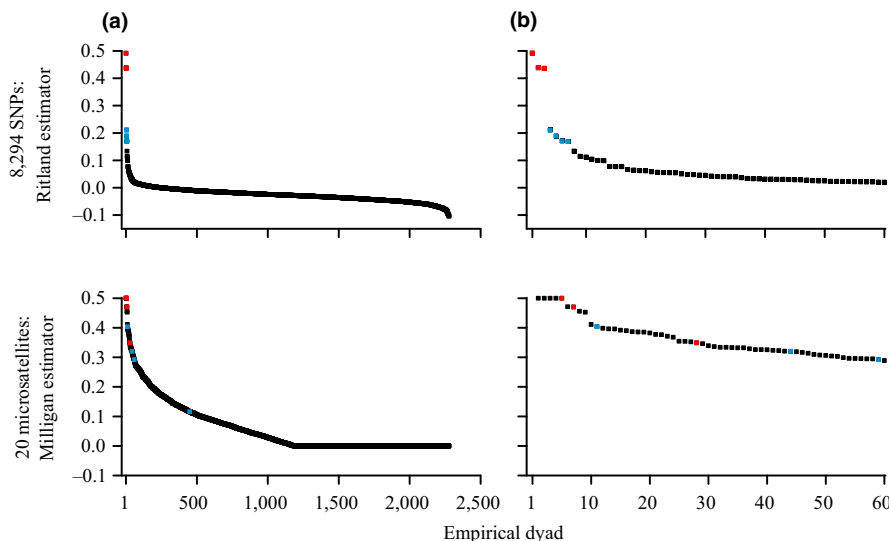


FIGURE 2 Relatedness estimates of empirical dyads of pygmy blue whales from Australia in descending order based on SNP and microsatellite data sets and the best chosen estimator for each data set. These are for (a) all potential dyads for which SNP data are available (those with only microsatellite data are not shown to allow easier visual comparison), with a zoom in on (b) the 60 dyads with the highest estimated relatedness. The first-order relatives (red) and likely second-order relatives (blue) detected in the SNP data set are highlighted in both data sets

Ritland (1996) estimator, they found it had the lowest precision except for unrelated individuals (Escoda et al. (2017) Figure S4), as also shown here. What they did not assess was the potential influence of genotyping error, instead choosing an estimator based on simulations and then further filtering their SNP data so that four duplicated samples had estimated relatedness close to one. The only other study that made a similar assessment to Escoda et al. (2017) and ourselves for genotyping-by-sequencing data selected Ritland (1996) as the best estimator (de Fraga et al., 2017); this was based on Pearson's correlation coefficients, but we are unable to fully compare the findings of their study to our own as they did not report the coefficients. Simulations by others on their own genotyping-by-sequencing data sets are required to ascertain whether Ritland (1996) remains the most appropriate estimator across data sets for mitigating the influence of genotyping error. It is now easy for others to conduct simulations on their own data because the current study has sparked improvements in the user-friendly program *COANCESTRY*: the program is now amenable to simulating large data sets and allelic dropout. In addition to relatedness, simulations can also be run in *COANCESTRY* to assess estimates of individual inbreeding coefficients for those interested in estimating inbreeding or investigating inbreeding depression.

We also showed here, for the first time, that a genotyping-by-sequencing data set with a relatively high error rate (i.e., 0.04 genotyping error or 0.2 allelic dropout) or missing data (i.e., 0.4) can still remain more powerful than a microsatellite data set without any error and without missing data (Figure 1). The choice of relatedness estimator overcame issues with genotyping error or allelic dropout, and individuals simulated with missing data still had thousands of SNPs to provide the high resolution needed for relatedness analyses. The greater power of genome-wide SNP data, as already attested to in the literature (Hellmann et al., 2016; Santure et al., 2010; Sun et al., 2016), was also confirmed by the contrast in the empirical relatedness distribution of the data sets (Figure 2). Most dyads in the empirical SNP data were unrelated, with instead hundreds of false positives in the microsatellite data. The SNP data identified three first-degree relatives, four likely second-degree relatives, and other second-degree or higher degree relatives with lower relatedness values. None of the seven highly related dyads consisted of individuals sampled in the same pod (data not shown), suggesting that kinship may not be of importance for associations in groups of blue whales. This is in agreement with other, limited knowledge that indicates associations in baleen whales tend to be short term and unstable, with no to very little evidence of an influence of kinship (e.g., Valsecchi, Hale, Corkeron, & Amos, 2002; Weinrich, Rosenbaum, Baker, Blackmer, & Whitehead, 2006).

We have only touched here on the potential of genome-wide SNPs for estimating relatedness in nonmodel organisms. Relatedness estimates and associated simulations could also take into account the likelihood of genotype calls with the aim of minimizing the downward bias in estimates caused by genotyping errors. The program *LCMLKIN* (Lipatov, Sanjeev, Patro, & Veeramah, 2015) can already take likelihood information into account. This is currently

limited as, to our knowledge, *LCMLKIN*, and no other SNP-specific program that can estimate relatedness (e.g., Danecek et al., 2011; Zheng et al., 2012) has yet a simulation component to assess reliability. In addition, the popular bioinformatics pipeline *STACKS* does not output likelihoods or a similar index for all three possible genotypes at each locus in each individual (the latest version we checked was 1.44), which is required by *LCMLKIN*. Another promising potential of genome-wide SNP data is for identifying individuals who are equally related to each other according to a pedigree relationship, but share by descent a different proportion of the genome due to the probabilities associated with Mendelian inheritance (Hill & Weir, 2011). If the marker set is powerful enough, as seen here (Figure 1), estimates of relatedness in full-siblings should have higher variance than parent-offspring relationships because the latter always has an IBD of one allele per locus. As the number of loci and therefore linkage continues to increase, and if a linkage map is known, information about SNPs in linkage disequilibrium could also be used to provide further power for estimating relatedness (Albrechtsen et al., 2009). These advances could be especially important in refining captive breeding programmes for conservation when they already consist of highly related individuals (Attard, Brauer, et al., 2016; Attard, Möller, et al., 2016). Pedigree-based ideas of relatedness in conservation breeding could also be replaced with measures of genome similarity (Ivy et al., 2016), such as used in agricultural breeding (Speed & Balding, 2015). This is because the allele frequencies of the original wild population are often poorly known, and so it is difficult to accurately estimate pedigree-based relatedness (Ivy et al. (2016), but see Svengren, Prettejohn, Bunge, Fundi, and Björklund (2017)).

We showed here how simulations can be used to improve relatedness inferences from genotyping-by-sequencing data. The relatedness estimator of Ritland (1996) was found to perform best when considering the relatively high error rate of such data, with similar simulation assessments required for other genotyping-by-sequencing data sets to confirm whether this is the case across data sets. A rigorous, technical comparison of estimators would be required to ascertain the reason for any differences in performance. Relatedness estimates from genotyping-by-sequencing data will allow higher-resolution assessments of breeding behaviour, social structure, inbreeding and inbreeding depression, and population demographics and connectivity than previously possible, and will only improve as relatedness estimators are developed that are specific to this type of data set.

ACKNOWLEDGEMENTS

Collection of SNP data was funded by the Sea World Research and Rescue Foundation Incorporated (SWRRFI) and the Winifred Violet Scott Charitable Trust (WVSCT). Collection of microsatellite data was funded by the Australian Marine Mammal Centre within the Australian Antarctic Division and Macquarie University in Australia. We thank Jinliang Wang for support in running *COANCESTRY* with large data sets, Timothy Frasier for information about *RELATED*, Jonathan Sandoval-Castillo for bioinformatics support, Julian Catchen for

information about STACKS, and associate editor Andrew DeWoody and three anonymous reviewers for their comments on the manuscript. We also thank Curt Jenner, Peter Gill, Micheline-Nicole Jenner, Margaret Morrice and their funding sources for blue whale samples. These were collected under the ethics and research permit requirements of the country in which the work was conducted, as detailed elsewhere (Attard et al., 2012; Attard et al., 2018). We are also grateful for a Future Fellowship from the Australian Research Council to L.B.B. (FT130101068) and the associated Flinders University component for providing the salary for C.R.M.A.

DATA ACCESSIBILITY

SNP genotypes are available in Dryad entry <https://doi.org/10.5061/dryad.t8ph5>, and microsatellite genotypes are available under Attard et al. (2012) in Dryad entry <https://doi.org/10.5061/dryad.8m0t6>, with the exception of one microsatellite-genotyped sample that is in Dryad entry doi: <https://doi.org/10.5061/dryad.t8ph5>. The Perl script used to simulate allelic dropout is available from <https://github.com/CatherineAttard>.

AUTHOR CONTRIBUTIONS

C.R.M.A. conceived and designed the study, acquired, analysed and interpreted the data, and drafted the article. L.M.M. and L.B.B. contributed to the acquisition of the data and critically revised the article.

ORCID

Catherine R. M. Attard  <http://orcid.org/0000-0003-1157-570X>
Luciano B. Beheregaray  <http://orcid.org/0000-0003-0944-3003>
Luciana M. Möller  <http://orcid.org/0000-0002-7293-5847>

REFERENCES

- Albrechtsen, A., Korneliusen, T. S., Moltke, I., Hansen, T. v. O., Nielsen, F. C., & Nielsen, R. (2009). Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genetic Epidemiology*, 33, 266–274. <https://doi.org/10.1002/gepi.20378>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11, 697–709. <https://doi.org/10.1038/nrg2844>
- Attard, C. R. M., Beheregaray, L. B., Jenner, C., Gill, P., Jenner, M., Morrice, M., ... Möller, L. (2010). Genetic diversity and structure of blue whales (*Balaenoptera musculus*) in Australian feeding aggregations. *Conservation Genetics*, 11, 2437–2441. <https://doi.org/10.1007/s10592-010-0121-9>
- Attard, C. R. M., Beheregaray, L. B., Jenner, K. C. S., Gill, P. C., Jenner, M.-N., Morrice, M. G., ... Möller, L. M. (2012). Hybridization of Southern Hemisphere blue whale subspecies and a sympatric area off Antarctica: impacts of whaling or climate change? *Molecular Ecology*, 21, 5715–5727. <https://doi.org/10.1111/mec.12025>
- Attard, C. R. M., Beheregaray, L. B., Jenner, K. C. S., Gill, P. C., Jenner, M.-N. M., Morrice, M. G., ... Möller, L. M. (2015). Low genetic diversity in pygmy blue whales is due to climate-induced diversification rather than anthropogenic impacts. *Biology Letters*, 11, 20141037. <https://doi.org/10.1098/rsbl.2014.1037>
- Attard, C. R. M., Beheregaray, L. B., & Möller, L. M. (2016). Towards population-level conservation in the critically endangered Antarctic blue whale: The number and distribution of their populations. *Scientific Reports*, 6, 22291. <https://doi.org/10.1038/srep22291>
- Attard, C. R. M., Beheregaray, L. B., Sandoval-Castillo, J., Jenner, K. C. S., Gill, P. C., Jenner, M.-N. M., ... Möller, L. M. (2018). From conservation genetics to conservation genomics: A genome-wide assessment of blue whales (*Balaenoptera musculus*) in Australian feeding aggregations. *Royal Society Open Science*, 5, 170925. <https://doi.org/10.1098/rsos.170925>
- Attard, C. R. M., Brauer, C. J., Van Zoelen, J. D., Sasaki, M., Hammer, M. P., Morrison, L., ... Beheregaray, L. B. (2016). Multi-generational evaluation of genetic diversity and parentage in captive southern pygmy perch (*Nannoperca australis*). *Conservation Genetics*, 17, 1469–1473. <https://doi.org/10.1007/s10592-016-0873-y>
- Attard, C. R. M., Möller, L. M., Sasaki, M., Hammer, M. P., Bice, C. M., Brauer, C. J., ... Beheregaray, L. B. (2016). A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conservation Biology*, 30, 1060–1069. <https://doi.org/10.1111/cobi.12699>
- Brauer, C. J., Hammer, M. P., & Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, 25, 5093–5113. <https://doi.org/10.1111/mec.13830>
- Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, 22, 3028–3035. <https://doi.org/10.1111/mec.12105>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 17, 362–365. <https://doi.org/10.1111/1755-0998.12669>
- Coleman, S. W., & Jones, A. G. (2011). Patterns of multiple paternity and maternity in fishes. *Biological Journal of the Linnean Society*, 103, 735–760. <https://doi.org/10.1111/j.1095-8312.2011.01673.x>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510. <https://doi.org/10.1038/nrg3012>
- Escoda, L., González-Esteban, J., Gómez, A., & Castresana, J. (2017). Using relatedness networks to infer contemporary dispersal: application to the endangered mammal *Galemys pyrenaicus*. *Molecular Ecology*, 26, 3343–3357. <https://doi.org/10.1111/mec.14133>
- Fountain, E. D., Pauli, J. N., Reid, B. N., Palsbøll, P. J., & Peery, M. Z. (2016). Finding the right coverage: The impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Molecular Ecology Resources*, 16, 966–978. <https://doi.org/10.1111/1755-0998.12519>
- de Fraga, R., Lima, A. P., Magnusson, W. E., Ferrão, M., & Stow, A. J. (2017). Contrasting patterns of gene flow for Amazonian snakes that actively forage and those that wait in ambush. *Journal of Heredity*, 108, 524–534. <https://doi.org/10.1093/jhered/esx051>

- Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., ... Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22, 3165–3178. <https://doi.org/10.1111/mec.12089>
- Glaubitz, J. C., Rhodes, O. E., & Dewoody, J. A. (2003). Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology*, 12, 1039–1047. <https://doi.org/10.1046/j.1365-294X.2003.01790.x>
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124, 323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews: Genetics*, 10, 381–391. <https://doi.org/10.1038/nrg2575>
- Goddard, M. E., Hayes, B. J., & Meuwissen, T. H. E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics*, 128, 409–421. <https://doi.org/10.1111/j.1439-0388.2011.00964.x>
- Haynes, G. D., & Latch, E. K. (2012). Identification of novel Single Nucleotide Polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *PLoS ONE*, 7, e36536. <https://doi.org/10.1371/journal.pone.0036536>
- Hellmann, J. K., Sovic, M. G., Gibbs, H. L., Reddon, A. R., O'Connor, C. M., Ligoeki, I. Y., ... Hamilton, I. M. (2016). Within-group relatedness is correlated with colony-level social structure and reproductive sharing in a social fish. *Molecular Ecology*, 25, 4001–4013. <https://doi.org/10.1111/mec.13728>
- Hill, W. G., & Weir, B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, 93, 47–64. <https://doi.org/10.1017/S0016672310000480>
- Hoffman, J. I., Tucker, R., Bridgett, S. J., Clark, M. S., Forcada, J., & Slate, J. (2012). Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: A case study in the Antarctic fur seal. *Molecular Ecology Resources*, 12, 861–872. <https://doi.org/10.1111/j.1755-0998.2012.03158.x>
- Hoffmann, J., Simpson, F., David, P., Rijks, J., Kuiken, T., & Thorne, M. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3775–3780. <https://doi.org/10.1073/pnas.1318945111>
- Iacchei, M., Ben-Horin, T., Selkoe, K. A., Bird, C. E., García-Rodríguez, F. J., & Toonen, R. J. (2013). Combined analyses of kinship and F_{ST} suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. *Molecular Ecology*, 22, 3476–3494. <https://doi.org/10.1111/mec.12341>
- 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, 403–412. <https://doi.org/10.1093/jhered/esw029>
- Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, 10, 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
- Kaiser, S. A., Taylor, S. A., Chen, N., Sillett, T. S., Bondra, E. R., & Webster, M. S. (2017). A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources*, 17, 183–193. <https://doi.org/10.1111/1755-0998.12589>
- Kjeldsen, S. R., Zenger, K. R., Leigh, K., Ellis, W., Tobey, J., Phalen, D., ... Raadsma, H. W. (2016). Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. *Conservation Genetics*, 17, 337–353. <https://doi.org/10.1007/s10592-015-0784-3>
- Labuschagne, C., Nupen, L., Kotzé, A., Grobler, P. J., & Dalton, D. L. (2015). Assessment of microsatellite and SNP markers for parentage assignment in ex situ African penguin (*Spheniscus demersus*) populations. *Ecology and Evolution*, 5, 4389–4399. <https://doi.org/10.1002/ece3.1600>
- Li, C. C., Weeks, D. E., & Chakravarti, A. (1993). Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity*, 43, 45–52. <https://doi.org/10.1159/000154113>
- Lipatov, M., Sanjeev, K., Patro, R., & Veeramah, K. (2015). Maximum likelihood estimation of biological relatedness from low coverage sequencing data. *bioRxiv*. <https://doi.org/10.1101/023374>
- Liu, S., Palti, Y., Gao, G., & Rexroad Iii, C. E. (2016). Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture*, 452, 178–182. <https://doi.org/10.1016/j.aquaculture.2015.11.001>
- Lynch, M., & Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Genetics*, 152, 1753–1766.
- Milligan, B. G. (2003). Maximum-likelihood estimation of relatedness. *Genetics*, 163, 1153–1167.
- Möller, L. M. (2012). Sociogenetic structure, kin associations and bonding in delphinids. *Molecular Ecology*, 21, 745–764. <https://doi.org/10.1111/j.1365-294X.2011.05405.x>
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, 2841–2847. <https://doi.org/10.1111/mec.12350>
- Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews: Genetics*, 12, 443–451. <https://doi.org/10.1038/nrg2986>
- Norman, A. J., Stronen, A. V., Fuglstad, G.-A., Ruiz-Gonzalez, A., Kindberg, J., Street, N. R., & Spong, G. (2017). Landscape relatedness: Detecting contemporary fine-scale spatial structure in wild populations. *Landscape Ecology*, 32, 181–194. <https://doi.org/10.1007/s10980-016-0434-2>
- Palsbøll, P. J., Peery, M. Z., & Bérubé, M. (2010). Detecting populations in the 'ambiguous' zone: Kinship-based estimation of population structure at low genetic divergence. *Molecular Ecology Resources*, 10, 797–805. <https://doi.org/10.1111/j.1755-0998.2010.02887.x>
- Pe'er, I., de Bakker, P. I. W., Maller, J., Yelensky, R., Altshuler, D., & Daly, M. J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nature Genetics*, 38, 663–667. <https://doi.org/10.1038/ng1816>
- Peery, M. Z., Beissinger, S. R., House, R. F., Bérubé, M., Hall, A., Sellas, A., & Palsbøll, P. J. (2008). Characterizing source-sink dynamics with genetic parentage assignments. *Ecology*, 89, 2746–2759. <https://doi.org/10.1890/07-2026.1>
- Peery, M. Z., Reid, B. N., Kirby, R., Stoelting, R., Doucet-Béer, E., Robinson, S., ... Palsbøll, P. J. (2013). More precisely biased: Increasing the number of markers is not a silver bullet in genetic bottleneck testing. *Molecular Ecology*, 22, 3451–3457. <https://doi.org/10.1111/mec.12394>
- Pemberton, J. M. (2008). Wild pedigrees: The way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275, 613–621. <https://doi.org/10.1098/rspb.2007.1531>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: An R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, 15, 557–561. <https://doi.org/10.1111/1755-0998.12323>
- Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, 43, 258–275. <https://doi.org/10.1111/j.1558-5646.1989.tb04226.x>
- Rasić, G., Filipović, I., Weeks, A. R., & Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*, 15, 275. <https://doi.org/10.1186/1471-2164-15-275>

- Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research*, 67, 175–185. <https://doi.org/10.1017/S0016672300033620>
- Ross, K. G. (2001). Molecular ecology of social behaviour: Analyses of breeding systems and genetic structure. *Molecular Ecology*, 10, 265–284. <https://doi.org/10.1046/j.1365-294x.2001.01191.x>
- Ross, C. T., Weise, J. A., Bonnar, S., Nolin, D., Trask, J. S., Smith, D. G., ... Kanthaswamy, S. (2014). An empirical comparison of Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs) for relatedness estimation in Chinese rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, 76, 313–324. <https://doi.org/10.1002/ajp.22235>
- 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, 1439–1451. <https://doi.org/10.1111/j.1365-294X.2010.04554.x>
- Seddon, J. M., Parker, H. G., Ostrander, E. A., & Ellegren, H. (2005). SNPs in ecological and conservation studies: A test in the Scandinavian wolf population. *Molecular Ecology*, 14, 503–511. <https://doi.org/10.1111/j.1365-294X.2005.02435.x>
- Speed, D., & Balding, D. J. (2015). Relatedness in the post-genomic era: Is it still useful? *Nature Reviews: Genetics*, 16, 33–44. <https://doi.org/10.1038/nrg3821>
- Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., ... Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, 25, 705–712. <https://doi.org/10.1016/j.tree.2010.09.002>
- 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. <https://doi.org/10.1016/j.tpb.2015.10.002>
- Svengren, H., Prettejohn, M., Bunge, D., Fundi, P., & Björklund, M. (2017). Relatedness and genetic variation in wild and captive populations of Mountain Bongo in Kenya obtained from genome-wide single-nucleotide polymorphism (SNP) data. *Global Ecology and Conservation*, 11, 196–206. <https://doi.org/10.1016/j.gecco.2017.07.001>
- Taylor, H. R. (2015). The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, 5, 3140–3150. <https://doi.org/10.1002/ece3.1541>
- Valsecchi, E., Hale, P., Corkeron, P., & Amos, W. (2002). Social structure in migrating humpback whales (*Megaptera novaeangliae*). *Molecular Ecology*, 11, 507–518. <https://doi.org/10.1046/j.0962-1083.2001.01459.x>
- Van de Castele, T., Galbusera, P., & Matthysen, E. (2001). A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, 10, 1539–1549. <https://doi.org/10.1046/j.1365-294X.2001.01288.x>
- Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, 160, 1203–1215.
- Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetics Research*, 89, 135–153. <https://doi.org/10.1017/s0016672307008798>
- Wang, J. (2011). COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 11, 141–145. <https://doi.org/10.1111/j.1755-0998.2010.02885.x>
- Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, 107, 4–13. <https://doi.org/10.1016/j.tpb.2015.08.006>
- Weinman, L. R., Solomon, J. W., & Rubenstein, D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, 15, 502–511. <https://doi.org/10.1111/1755-0998.12330>
- Weinrich, M. T., Rosenbaum, H., Baker, C. S., Blackmer, A. L., & Whitehead, H. (2006). The influence of maternal lineages on social affiliations among humpback whales (*Megaptera novaeangliae*) on their feeding grounds in the southern Gulf of Maine. *Journal of Heredity*, 97, 226–234. <https://doi.org/10.1093/jhered/esj018>
- Wright, B., Morris, K., Grueber, C. E., Hogg, C. J., O'mELLY, D., Hamede, R., ... Belov, K. (2015). Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. *BMC Genomics*, 16, 1–11. <https://doi.org/10.1186/s12864-015-2020-4>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Attard CRM, Beheregaray LB, Möller LM. Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias. *Mol Ecol Resour*. 2018;00:1–10. <https://doi.org/10.1111/1755-0998.12739>