DR. SARAH PERRY FLANAGAN (Orcid ID : 0000-0002-2226-4213)

Article type : Invited Reviews and Syntheses

The future of parentage analysis: from microsatellites to SNPs and beyond

Sarah P. Flanagan[1] and Adam G. Jones[2]

[1]*School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand 8140, spflanagan.phd@gmail.com (corresponding author)*

[2]*Department of Biological Sciences, 875 Perimeter Drive MS 3051, University of Idaho, Moscow ID 83844, adamjones@uidaho.edu*

Running Head: The future of parentage analysis

Abstract

Parentage analysis is a cornerstone of molecular ecology that has delivered fundamental insights into behavior, ecology, and evolution. Microsatellite markers have long been the king of parentage, their hypervariable nature conferring sufficient power to correctly assign offspring to parents. However, microsatellite markers have seen a sharp decline in use with the rise of next-generation sequencing technologies, especially in the study of population genetics and local adaptation. The time is ripe to review the current state of parentage analysis and see how it stands to be affected by the emergence of next-generation sequencing approaches. We find that single-nucleotide polymorphisms (SNPs), the typical next-generation sequencing marker, remain underutilized in parentage analysis but are gaining momentum, with 58 SNP-based parentage analyses published thus far. Many of these papers, particularly the earlier ones, compare the power of SNPs

and microsatellites in a parentage context. In virtually every case, SNPs are at least as powerful as microsatellite markers. As few as 100-500 SNPs are sufficient to resolve parentage completely in most cases. We also provide an overview of the analytical programs that are commonly used and compatible with SNP data. As the next-generation parentage enterprise grows, a reliance on likelihood and Bayesian approaches, as opposed to strict exclusion, will become increasingly important. We discuss some of the caveats surrounding the use of next-generation sequencing data for parentage analysis and conclude that the future is bright for this important realm of molecular ecology.

Key words: sibship reconstruction, paternity analysis, parentage assignment, Bayesian parentage, Illumina sequencing, next-generation sequencing, RAD-seq

Introduction

The analysis of parentage is a key facet of molecular ecology. Since the realization in the 1970s and 1980s that genetic data could potentially diagnose parent-offspring relationships in nature (Thompson 1976a, 1976b; Ellstrand 1984; Gowaty and Karlin 1984), parentage analysis has blossomed into an active enterprise spanning numerous fields of inquiry. In the realm of molecular ecology, an understanding of parentage patterns can provide indispensable information for the study of sexual selection (Coltman et al. 2002; Jones 2009), conservation biology (Haig 1998; Planes et al. 2009), effective population size (Araki et al. 2007), and even speciation and natural selection (Conner et al. 1996; Muhlfield et al. 2009). From its humble beginnings and slow start during the rise of allozyme markers, parentage analysis has matured into a sophisticated discipline that continues to evolve as new technology becomes available.

The fundamental idea underlying parentage analysis is actually quite simple. Given Mendel's laws, we know that each individual inherits genetic material from its parents. For diploids, each progeny receives precisely one allele from its mother and one from its father at each locus. Thus, a

suite of Mendelian loci, genotyped in a sample of individuals, can distinguish parent-offspring relationships from other sorts of relationships, including unrelated pairs (Thompson 1975; Thompson and Meagher 1987). This analysis requires that the marker loci be polymorphic (i.e., have two or more alleles per locus) and that the nature of inheritance at the loci is understood. Even though the application of parentage analysis often includes complex equations and arcane implementations, the simple fact of Mendelian inheritance always lies at its core.

The most important recent development in parentage analysis is the introduction of next-generation sequencing approaches. The concomitant potential to genotype thousands or even millions of loci in almost any species has the potential to dramatically transform parentage analysis (Glaubitz et al. 2003). The availability of such large genetic datasets makes the choice of genotyping marker more difficult than when a single marker type was clearly best, as many factors must now be weighed before embarking on a parentage study. Here, we review the current state of parentage analysis, with the goal of providing guidelines and important considerations for studies that require parentage analysis. We start with a short history of parentage analysis and a review of the current state of the field. We then discuss the molecular and statistical approaches currently in use. We follow those sections with a discussion of the potential pitfalls in parentage analysis and how to avoid them. We conclude with some recommendations regarding how researchers can easily transition into the next-generation sequencing era of parentage analysis.

A Brief History of Parentage Analysis

One of the remarkable features of evolutionary biology in the 20th Century is that much of the theory needed to analyze population-level patterns with genetic data had been developed well before appropriate markers became available. This statement is especially germane for parentage analysis. The most important ideas underlying this area of inquiry were developed in the 1970s and 1980s (Thompson 1975, 1976a, 1976b; Meagher and Thompson 1986). At the time, the only readily available molecular markers were allozymes (also called isozymes), an ingenious method that

launched evolutionary biology into the molecular marker era (Hubby and Lewontin 1966; Lewontin and Hubby 1966). The allozyme approach involved electrophoresis of proteins through a matrix, followed by visualization of the protein using stains that took advantage of the enzymatic properties of specific proteins to produce a visible smudge on a gel. Different alleles at a protein could differ with respect to size or net charge, both of which would affect the mobility of the protein, resulting in distinguishable allelic variation. In principle, allozymes, as the first easy-to-assay, codominant, Mendelian marker, were perfect for parentage analysis. In practice, however, they almost never harbored enough variation to diagnose parent-offspring relationships reliably.

Nevertheless, the advent of allozymes catalyzed the development of a robust body of theory related to parentage analysis. The most important realization was that the genotypes of pairs or triads of individuals could be used to develop a rigorous hypothesis-testing approach based on maximum-likelihood equations (Meagher and Thompson 1986). Around the same time, a method of fractional parentage allocation was developed, which retained the inevitable uncertainty in parentage assignments during downstream analysis (Devlin et al. 1988). Thus, by the end of the 1980s, just as the first wave of hypervariable Mendelian markers began to spread through the field (Tautz 1989), researchers already had access to a fairly robust analytical toolkit, which included strict exclusion, categorical allocation, and fractional allocation, with the latter two methods taking advantage of formal maximum likelihood approaches. Despite the growing toolkit, parentage analysis in the allozyme days was nearly impossible given the low information content of the markers, and the few successful examples could be counted using the fingers on one hand (Hanken and Sherman 1981; Ellstrand 1984; Gowaty and Karlin 1984; Meagher 1986).

The Modern Era of Parentage Analysis

We entered the modern era of parentage analysis in the 1990s, when this area of inquiry, along with everything else in molecular ecology, was swept forward on the wave of the microsatellite revolution (Tautz 1989; Queller et al. 1993; Jarne and Lagoda 1996; Powell et al. 1996). Before

microsatellites, multilocus DNA fingerprinting was explored as a way to diagnose extra-pair paternity, especially in birds (Burke et al. 1989; Birkhead et al. 1990), but DNA fingerprinting was crude in the sense that it relied on patterns of band sharing, and this feature prevented it from interfacing with theory developed for Mendelian markers. The fact that microsatellite markers were Mendelian (unlike multilocus DNA fingerprinting) and had multitudes of alleles per locus (unlike allozymes) allowed empiricism to finally catch up with the theory, and parentage studies in natural populations began to appear by the dozens (Brockmann et al. 1994; Morin et al. 1994; Kellogg et al. 1995; Dow and Ashley 1996; Jones and Avise 1997a, 1997b). Researchers took a second look at the theory and developed methods that accommodated the specific strengths and weaknesses of microsatellites (Sancristobal and Chevalet 1997; Marshall et al. 1998; Nielsen et al. 2001). As a result, we entered the new millennium with even better tools, coupled with the molecular techniques we needed to apply them (reviewed in Jones and Ardren 2003; Jones et al. 2010). Those who worked on parentage analysis in the 1990s could be forgiven if they looked back upon those times as the golden age of parentage.

Even though parentage analysis seemed to have been largely solved by microsatellites, some problems persisted. For instance, microsatellites were extremely successful in species in which they were abundant in the genome and highly polymorphic, such as most fishes (DeWoody and Avise 2005). However, many species harbor little polymorphism even at microsatellite loci, making robust parentage analysis difficult. In addition, microsatellite markers still require quite a large initial investment in terms of identifying loci, designing locus-specific primers, and optimizing PCR conditions. Moreover, the scoring of microsatellite markers is often an art form of its own, calling for often poorly documented criteria regarding the separation of true alleles from artifactual bands on sequencing gels. Thus, a successful microsatellite-based study of parentage still represents a significant investment in terms of labor and financial resources (see Hodel et al. 2016 for a recent review). Given these constraints, microsatellites still have their uses but there does seem to be room for newer technologies to supplement, or even supplant, microsatellites as the marker of choice for parentage.

Parentage Analysis Using Traditional SNPs

With the availability of genomic resources and multiplexed methods to assay many single-nucleotide polymorphisms (SNPs) simultaneously, researchers have moved toward using these types of SNP approaches to conduct parentage analysis (Table 1). Some researchers have turned to SNPs because of low polymorphism in microsatellites in their species (e.g., Cramer et al. 2011), but SNPs also provide other practical improvements such as easier automation and scoring (Anderson and Garza 2006), plus lower mutation rates (Amorim and Pereira 2005; Fisher et al 2009). Despite these favorable features of SNPs, we found a total of only 38 papers that used these traditional SNP approaches for parentage analysis (Table 1). A substantial number of these studies compared the power SNPs to microsatellites for parentage analysis (Table 1), and they universally concluded that SNPs were entirely appropriate for this endeavor.

Parentage Analysis in the Next-Generation Sequencing Era

The use of next-generation sequencing in parentage analysis is rapidly gaining momentum. The first application of next-generation sequencing in this arena was to use the easily obtainable sequence data to develop microsatellite markers (e.g., Santana et al. 2009; Castoe et al. 2010; Guichoux et al. 2011) or SNP markers that can be amplified via PCR and assayed using a highly multiplexed approach, such as SNPlex (Cramer et al. 2011), iPLEX/MassARRAY (Sellars et al. 2014), Fluidigm Integrated Fluidic Circuits (Lew et al. 2015) or any of a number of related approaches. Table 2 lists the studies that have taken a next-generation sequencing approach to the identification of SNPs in the genome, followed by one of these more traditional assays to score SNPs in the context of parentage analysis. With the exception of two studies that used next-generation sequencing to develop SNP arrays, these studies generally used less than 200 SNPs to assign parentage. The 10 studies listed in Table 2 reinforce the conclusion from traditional SNP studies (Table 1) that SNPs perform well in parentage.

A natural next step, facilitated by high throughput sequencing, is to genotype SNPs directly by using any one of a number of genotyping-by-sequencing approaches (Table 3). These next-generation approaches to parentage analysis are conceptually identical to the previous-generation approaches, and the new datasets can be analyzed by the current generation of parentage analysis software. Despite these similarities, some additional concerns begin to arise with true next-generation parentage analysis, and we will discuss these concerns as we describe each method. Due to the distinctions between these markers and SNPs derived through more traditional means, we will refer to these markers as 'next-generation markers' throughout the manuscript.

A notable feature of the studies of parentage using SNPs (Tables 1, 2, and 3) is that virtually all studies have concluded that a relatively small number of SNP markers, from 60-200, usually provides resolving power equal to or better than that provided by the available microsatellite markers for the species under consideration. The exact number of SNPs required will depend on a number of factors, including the minor allele frequencies of the SNPs, linkage disequilibrium among SNPs, the frequency of null alleles and genotyping errors, the number of parental pairs, the distribution of offspring numbers per family, and the mating design (Jones and Ardren 2003; Anderson and Garza 2006; Kalinowski et al. 2007). Despite these factors, the empirical results show clearly that as few as several hundred SNPs are sufficient for most parentage analyses. The power of SNPs for parentage analysis was appreciated over a decade ago by theoreticians (Anderson and Garza 2006), but empiricists have been patiently waiting for cheap and effective SNP approaches to catch up with theory, especially in non-model systems.

Given these new developments on the marker front, as well as continued progress in analytical approaches, the parentage analysis landscape is perhaps more confusing than it has ever been in the past. In some systems with well-established microsatellite markers, a more traditional approach may be best, whereas other systems may call for a next-generation approach. Most of the next-generation approaches are extremely cheap on a per-marker basis but expensive on a per-individual basis. Some of these approaches require well-developed molecular and bioinformatics skills, far beyond what is necessary for microsatellite genotyping. The decision of whether or not to

adopt a next-generation approach thus involves a careful weighing of the costs and benefits. In addition, not all parentage analysis software can accommodate the huge numbers of markers typically produced in a next-generation genotyping study, so the decision to use these approaches also narrows the scope of possible parentage techniques to some degree.

Approaches to Parentage Analysis

Parentage analysis using SNPs and next-generation sequencing relies on the same theoretical underpinnings as parentage studies based on traditional markers, which are still much more popular than SNP or next-generation methods. While, at the time of this writing, we found 58 studies that have employed SNPs for parentage analysis in the last decade (Tables 1, 2, and 3), dozens of parentage studies based on microsatellite markers are published annually, and microsatellites remain the most popular marker for all types of kinship and relatedness studies (e.g., Städele and Vigilant, 2016). Even the less popular classes of markers, such as amplified fragment length polymorphisms (AFLPs) and allozymes, have been used in more studies of parentage than SNPs. This large number of published studies produces a long list of best practices and potential pitfalls, many of which also apply to next-generation approaches.

Methods of Analysis. Theory regarding parentage analysis has not changed substantially in the last decade, and next-generation markers, which are essentially more of the same (but much more in some cases) as far as Mendelian markers are concerned, are not game-changers with respect to the analytical techniques needed for successful parentage analysis. Thus, reviews of parentage analysis from the last two decades (Jones and Ardren 2003; Jones et al. 2010) effectively summarize the underlying logic of the workhorses of the discipline: exclusion, parentage assignment, parental reconstruction, and Bayesian parentage analysis. Here, we summarize each approach briefly and discuss considerations for the current generation of markers and study designs. More detailed descriptions of each approach are given in Appendix 1.

Exclusion. The most intuitive approach to parentage analysis is to identify incompatibilities between a pair of individuals in a way that conclusively demonstrates that one could not be the parent of the other. Given Mendel's laws, we know that each diploid offspring receives exactly one allele per locus from its mother and exactly one from its father. Thus, if a putative offspring shares no alleles at even one locus with a putative parent, then the putative parent can be excluded from the pool of potential parents (Chakraborty et al. 1974). This exercise assumes that no mutations or scoring errors occurred during meiosis or the genotyping technique, an assumption whose validity varies depending on the type of locus under consideration.

While exclusion is conceptually appealing and easy to implement, it should be used with caution and its use should dwindle, hopefully to zero, over time. While mutations are relatively rare, even at most hypervariable microsatellite loci (Jones et al. 1999; Ellegren 2000), scoring errors are the real downfall for strict exclusion approaches. As studies grow in size, in terms of both numbers of samples and numbers of markers, an error-free dataset is virtually an impossibility. Consequently, most modern implementations of exclusion allow an arbitrary number of mismatches. While this approach seems logical and prevents complete exclusion of all candidate parents, including the true ones, it suffers from being arbitrary and ignoring the relevant literature regarding likelihoods and posterior probabilities of parentage.

An odd pattern in the literature is that some fields have tenaciously clung to the idea that exclusion approaches are the appropriate solution to parentage analysis. This pattern is especially evident among breeders of domestic cattle and sheep (Table 1). Not coincidentally, the cattle breeding field has arrived at a preliminary consensus that 500 or more SNPs are required to resolve parentage confidently (e.g., McClure et al. 2015; McClure et al. 2018), whereas fields using modern analytical methods tend to conclude that 100-200 SNPs are usually more than adequate (Abadía-Cardoso et al. 2013; Steele et al. 2013; Dussault and Boulding 2018). The reason that exclusion performs less well than formal maximum likelihood approaches is that it requires implicit assumptions that are ill defined and it discards much of the data. For instance, by choosing an arbitrary threshold of mismatches to constitute a true exclusion, researchers are implicitly imparting a level of confidence in

the reliability of the markers. Although the probability of a given number of Mendelian incompatibilities can be estimated from genotyping error rates (Chakraborty and Schull 1976), this extra step is rarely taken by users. Therefore the number of allowable mismatches arises from a vague feeling of confidence not estimated from genetic data. Furthermore, the exclusion approach for biallelic SNPs only uses data from homozygous loci at which a putative parent and offspring have different homozygous genotypes. Loci that are heterozygous in parents or offspring also carry information regarding the likelihood of parentage (Meagher and Thompson 1986; Marshall et al. 1998; Kalinowski et al. 2007), but this information is being discarded in an exclusion analysis. In general, the use of exclusion for parentage analysis should be phased out, particularly for parentage analysis using SNPs or next-generation data.

Parentage Assignment. The most commonly used approach to parentage analysis is assignment. Historically, parentage assignment could be divided into two categories: fractional and categorical (Jones and Ardren 2003; Jones et al. 2010). Fractional allocation assigns partial offspring to parents as a function of posterior probabilities (Devlin et al. 1988; Nielsen et al. 2001), and has now evolved into an approach known as "Bayesian parentage analysis", which we discuss in a separate section below. Categorical allocation, on the other hand, assigns each offspring entirely to the parent with the highest likelihood and treats the parentage analysis separately from subsequent estimates of population-level variables of interest.

Parentage assignment rests upon the calculation and comparison of the relative likelihoods of different hypotheses regarding the relationships among putative parent-offspring dyads or mother-father-offspring triads. The likelihood refers to the probability of observing the data given the hypothesis. In this case, the data are the genotypes, and the hypothesis is the proposed relationship among individuals. The likelihood can then be calculated easily by using the rules of Mendelian inheritance (see Marshall et al. 1998; Jones and Ardren 2003; Kalinowski et al. 2007). While absolute likelihoods are seldom of interest, they can be used to compare alternative hypotheses by constructing

a likelihood ratio of one hypothesis versus a second (often null) hypothesis. In parentage assignment, the ratio involves the hypothesis that the dyad or triad represents a true set of parents and offspring versus the hypothesis that the individuals are unrelated. Usually, we deal with the logarithms of likelihoods, and the likelihood ratio becomes a LOD score (Meagher 1986; Marshall et al. 1998). A positive LOD score indicates the parental hypothesis is more likely, whereas a negative LOD score indicates the unrelated hypothesis is more likely, given the genetic data. These LOD scores, while useful in obtaining a maximum likelihood solution, cannot be interpreted at face value in a statistical sense. A major breakthrough in parentage assignment occurred when Marshall et al. (1998) recognized that critical values for LOD scores could be determined by simulation. The approach used by Marshall et al. (1998) actually uses Δ (delta), the difference in LOD score between the most likely and second most likely parent (or the raw LOD score if only one candidate has a positive value), and simulates populations of parents and offspring to determine a critical value of Δ that results in a desired level of confidence in parentage assignment. This approach, which is still widely used 20 years after its introduction, was the first to control experiment-wise error in parentage analysis.

Bayesian Parentage Analysis. Bayesian parentage analysis originated as a technique to fractionally allocate offspring to parents as a function of posterior probabilities (Devlin et al. 1988). That is, the putative parent with the highest posterior probability would be assigned the largest fraction of the offspring, but the offspring would also be partially allocated to any parent with a non-zero probability of parentage. This fractional approach was intuitively unappealing in the early days of parentage because it must be strictly false from a biological standpoint, as fractional parentage has no biological analog. An adult cannot be 4% the parent of an offspring, for instance, even though this outcome is possible in fractional assignment. This intuitive distastefulness led to widespread adoption of categorical assignment, even though fractional assignment has better statistical properties for the estimation of many values of interest (Neff et al. 2001; Nielsen et al. 2001). With key developments in the mid-2000s, the technique of fractional allocation has matured into full-fledged Bayesian

parentage analysis, also called full probability parentage analysis (Hadfield et al. 2006; Jones et al. 2010).

The advantage of Bayesian parentage analysis is that various quantities of interest can be estimated simultaneously along with patterns of parentage. Some of these quantities could be variables whose values matter for the assignment of parentage, such as the proportion of candidate parents sampled (Nielsen et al. 2001). Other quantities could be population variables of interest, such as variance in mating success or the rate at which parentage decreases as a function of distance between prospective mates (Hadfield et al. 2006). The advantage to this approach for estimating variables is that any uncertainty in the parentage analysis is included as uncertainty in the ultimate estimates. In the case of categorical assignment, uncertainty in parentage is normally discarded at the next phase of analysis, as assignments are treated as the truth during the estimation of population-level parameters.

Another advantage to Bayesian parentage analysis is that prior information can be readily incorporated by modifying the priors. Thus, sources of information that imply even subtle differences in the probability of parentage for certain individuals can be incorporated naturally into the analysis. In categorical allocation, such information can be included, but in a very blunt fashion by either including or removing individuals from the list of candidate parents. Clearly, the Bayesian approach is more flexible and more statistically defensible. The biggest downside for Bayesian parentage analysis is that each analysis requires careful crafting of the posterior probability equation, and the analysis may be extremely sensitive to the decisions made during this step.

Parental and Sibship Reconstruction. Knowledge of the nature of Mendelian inheritance naturally leads to the conclusion that the full multilocus genotype of one parent can be reconstructed if the genotypes of the other parent and many offspring are known (Jones and Avise 1997b). Given that a full reconstruction requires many offspring from a family, coupled with *a priori* knowledge that these offspring are either full- or half-siblings, this technique is generally useful only in species in which

broods of reasonably large size can be collected together. If such a progeny array is available, then the full set of possible parents for the progeny array can be enumerated using a combinatorial or maximum likelihood algorithm (Jones 2001, 2005; Wang 2004). In general, maximum likelihood approaches are preferable, as exhaustive combinatorial algorithms are prohibitively slow for more than a handful of loci and require an error-free dataset of hypervariable loci.

Sibship reconstruction possesses many similarities to parental reconstruction, and some sibship reconstruction algorithms can be used to perform a very similar analysis (Wang 2004). Sibship reconstruction has the advantage that it requires no *a priori* knowledge of relationships among individuals in the genotyped sample. For a sample containing putative full-siblings and half-siblings, sibship reconstruction approaches use a clustering algorithm to arrange the genotyped individuals into families (Thomas and Hill 2000; Smith et al. 2001; Almudevar 2003). This review is concerned with parentage analysis, so we restrict our attention to sibship reconstruction methods that also allow the user to identify putative parents in the sample of genotyped individuals (Wang 2004; Jones and Wang 2010; Huisman 2017). Thus, many additional sibship reconstruction programs exist, beyond the ones we mention in the present review.

All approaches to sibship reconstruction use similar conceptual ideas for their algorithms. The methods use a maximum likelihood approach to evaluate and compare among proposed pedigrees, an approach that would, in principle, always yield the best solution given the underlying assumptions. However, the constellation of potential pedigrees is so large that an exhaustive search of all pedigrees is unfeasible in most cases. Consequently, the algorithms use an optimization procedure to restrict the search to a subset of pedigree space. Most of the algorithms use something akin to simulated annealing (Kirkpatrick et al. 1983; Almudevar 2003). For the latest developments in sibship and pedigree reconstruction (including distant relatives), outside the realm of parentage analysis *per se*, we direct the reader to recent work by Staples et al. (2014), Staples et al. (2016), and Ko and Nielsen (2017).

Methods of SNP Genotyping

Given the existing approaches and the history of parentage analysis, we now consider how this research endeavor should evolve with the advent of next-generation genotyping approaches. The decision of whether or not to embrace next-generation approaches boils down to whether or not the current project calls for the use of next-generation markers in place of the microsatellite markers that dominated the field of molecular ecology throughout the 1990s and 2000s. Many considerations come into play at the inception of a study regarding the choice of molecular markers. For instance, cost, ease of use, and the goals of a study should all be weighed before embarking on a research path that represents a significant investment of time and energy by the investigator.

The great advantage to next-generation markers is that they are extremely numerous, but their abundance comes with the cost that each individual marker, normally assayed as a SNP, harbors very little genetic variation. Thus, each individual SNP accomplishes very little but their collective strength can resolve almost any problem in parentage or kinship analysis. Next-generation SNP genotyping involves several categories of approaches, each of which has its own set of strengths and weaknesses.

Genotyping-by-Sequencing. One set of approaches to next-generation genotyping can be categorized as genotyping-by-sequencing methods. These methods include restriction-site-associated DNA sequencing (RAD-seq; Baird et al. 2008), multiplexed shotgun genotyping (MSG; Andolfatto et al. 2011), exome sequencing (Ng et al. 2009), and related approaches where the genotypes are determined by short-read sequencing of a vast library of DNA fragments. The advantage of these approaches is that they produce genotypes at an extremely large number of loci, often numbering in the tens of thousands. However, this huge genotyping throughput carries a number of disadvantages. For instance, the reliability of genotypes depends on sequencing coverage, and even with relatively high sequencing depth, allelic dropout is guaranteed to occur at a substantial fraction of loci (due to simple rules of binomial sampling). Unfortunately, allelic dropout may be the worst type of sequencing error for parentage analysis because it is likely to result in apparent Mendelian

incompatibilities between parents and offspring (Chakraborty et al. 1994; Pemberton et al. 1995; Wang 2010). Other disadvantages include that library preparation requires substantial molecular skills, analyzing the data requires some degree of bioinformatics expertise (although pipelines have recently been developed to aid in the analysis of RAD-seq data in parentage; Andrews et al. 2018; Thrasher et al. 2018), the cost per individual is high (Fig. 1), and most of the methods include sources of bias that are not yet well understood. Thus far, only a few parentage studies have used genotyping-by-sequencing data directly to assign parentage (Table 3).

Targeted Capture. Targeted capture approaches use primers, oligos, or other probes developed from genotyping-by-sequencing, whole genomic, or transcriptomic data to sample a chosen subset of the genome using high-throughput sequencing technologies such as Illumina Hi-Seq (e.g. GT-Seq, Campbell et al. 2014; RADcap, Hoffberg et al. 2016; RAPTURE, Ali et al. 2016). In an attempt to reduce error rates, these methods minimize the number of PCR cycles used. Targeted capture approaches can produce hundreds or thousands of markers, allowing researchers to increase read depth per locus. Careful choice of markers can minimize the sources of bias and error inherent to genotyping-by-sequencing and maximize the quality of the generated genotypes. We view these methods as promising approaches for parentage analysis in the next-generation sequencing era, although they have yet to be widely applied in a parentage context.

SNP-PCR Approaches. Many SNP-based parentage approaches have opted to use some version of high-throughput genotyping of SNPs by designing PCR primers and amplifying specific regions containing SNPs (Tables 1 and 2). Many such approaches are commercially available, and they typically involve a highly multiplexed set of PCR primers, permitting dozens of SNPs to be amplified simultaneously. In the modern manifestations of these approaches, the genotypes from PCR products are usually obtained through MALDI-TOF mass spectrometry (iPLEX/MassARRAY) or real-time PCR (Fluidigm), although other approaches are also possible (e.g., capillary electrophoreses:

SNPlex). The only initial barrier to the SNP-PCR approach is the identification of genomic sequences containing SNPs. Thus, SNP-PCR is best used in a system in which some population genomic information is already available or can be generated. Studies have used data from RAD-seq and RNA-seq to identify loci suitable for SNP-PCR (Holman et al. 2017; Kaiser et al. 2017; Table 2), indicating that a relatively minor investment in next-generation sequencing can be enough to identify a suite of loci suitable for parentage analysis. Once the SNP-PCR loci are characterized, their assay generally requires specialized equipment, which can often be accessed through core facilities or commercial genotyping companies. In comparison to genotyping-by-sequencing approaches, SNP-PCR produces more reliable genotypes but involves orders of magnitudes fewer loci. This method also has the advantage that it is perhaps the least expensive method to genotype large numbers of individuals at a large enough number of loci to completely resolve parentage in most systems (Fig. 1).

The SNP-Chip or SNP Array. An alternative to SNP-PCR is the SNP-Chip or SNP array, which is a microarray that can be used to determine genotypes at a large number of SNPs, often tens of thousands, simultaneously. Spots on the SNP-Chip correspond to different alleles, and genotypes are determined by hybridizing labeled genomic DNA to the array. Because the development of a SNP-Chip requires detailed knowledge of the genome and is expensive, SNP-Chips are generally available only for heavily studied model organisms, such as humans, pigs, cattle, and so forth. The advantages to SNP-Chips are that they can be assayed easily and produce a large number of reliable genotypes at a much smaller cost than genotyping-by-sequencing. The disadvantages are that the cost per individual is high compared to SNP-PCR (Fig. 1) and that the investment in initial development is prohibitively large for all but the most heavily studied of organisms.

Choosing a Marker System

Given the proliferation of next-generation approaches, the choice of a marker system for parentage analysis may seem difficult. Gone are the days where we can casually recommend microsatellites as the resolution to all problems (Glaubitz et al. 2003; Jones and Ardren 2003; Jones et al. 2010). However, the classic markers should not be summarily discarded either as the field moves forward.

The choice of a marker system depends upon the parentage question, the natural history of the study system and the state of genomic resources for the target organism, because these factors determine the cost of the project and the resolution provided by the chosen marker system. Microsatellite markers may be a viable choice if they are already available and they have the power to provide the desired level of resolution. If microsatellite markers have not been characterized, however, a SNP-PCR or targeted capture approach will generally be easier and more cost effective (Fig. 1). The major limitation of any approach based on SNPs is that each marker is usually biallelic, a feature that limits its flexibility, especially in species where putative parents are difficult to sample. Moreover, if only low-quality DNA is available, SNPs generated by next-generation sequencing methods may be less reliable than microsatellites (Andrews et al. 2018) or traditional SNPs (Carroll et al. 2018). Nevertheless, for most parentage analysis problems, in which the sample includes a number of offspring and a pool of candidate parents, SNPs are entirely appropriate. Empirical work thus far indicates that a suite of 100 to 200 SNPs will generally provide resolving power exceeding that of a typical panel of microsatellite markers (Table 1). However, choosing the most informative SNPs is imperative whether pursuing a SNP-PCR, SNP-chip, or targeted capture method. To maximize power, the minor allele frequency of SNPs should be high and the likelihood of allelic dropout should be low. If developing SNPs from a genotype-by-sequencing dataset, the program GBSTOOLS (Cooke et al. 2016) can identify loci that have a likely low allelic dropout rate. Choosing informative SNPs will provide higher power for parentage analysis and maximize cost efficiency.

The choice of markers becomes more difficult when offspring can be collected in groups that are known to contain half- or full-siblings. This situation may be common when eggs are laid in egg masses (e.g., Liebgold et al. 2006; Croshaw et al. 2009) or one parent cares for a large group of related offspring (e.g., Mobley et al. 2009; Paczolt et al. 2016). In this scenario, the knowledge that progeny occur in family groups can provide additional power for parentage analysis. For instance, an approach that can reconstruct parental genotypes can be quite powerful, in some cases completely resolving parentage with only three or four hypervariable microsatellites (e.g., Jones et al. 1999, 2002). As per-locus allelic diversity drops, many more loci become necessary for successful parentage reconstruction and a maximum likelihood approach is required. For tightly linked SNPs, however, allelic diversity can be recovered by reconstructing haplotypes, as exemplified by a recent study of gilthead sea bream (Table 1; García-Fernández et al. 2018). Another viable approach when the sample includes groups of full- or half-siblings is to use a technique that reconstructs sibships, while also assigning parentage, as implemented in the programs COLONY (Wang 2004; Jones and Wang 2010) or SEQUIOA (Huisman 2017). These techniques appear to work equally well with all types of markers, so for most systems, SNPs will be the marker of choice, unless hypervariable microsatellites have already been developed.

The considerations above lead to some simple rules of thumb, given the current state of genotyping technology. For a new system, with no developed markers, the best approach is usually going to be to develop a set of SNP-PCR or targeted capture markers (Fig. 1). A single RNA-seq or RAD-seq analysis of a handful of individuals should be enough to permit the identification of promising SNP loci (e.g., Andrews et al. 2018; Thrasher et al. 2018). In the case of an organism for which microsatellite markers are available, the power of the markers can be assessed by calculating exclusion probabilities (Chakraborty et al. 1988) or simulating data in a program like CERVUS (Marshall et al. 1998), COLONY (Jones and Wang 2010; Wang 2012, 2013) or GERUD (Jones 2001, 2005). If the markers are sufficiently polymorphic to answer the question of interest, then the existing microsatellites will be a good choice. Other approaches, such as RAD-seq, exome sequencing, RNA-seq, whole-genome sequencing, and even SNP-Chips will usually be too expensive on a per-

individual basis to justify their use for most parentage applications (Fig. 1). However, this constraint is quickly changing, as evidenced by the recent papers summarized in Table 3. Nevertheless, these techniques typically produce genotypes at thousands or tens of thousands of SNPs, when only a few hundred SNPs are usually required for complete resolution of parentage (Tables 1, 2, 3). In addition, the types of errors typically encountered by microsatellites and traditional SNPs are already well accommodated by existing software packages, whereas the errors associated with genotyping-by-sequencing, especially PCR duplications and allelic dropouts, have yet to be integrated into most software packages. We expect this situation to change dramatically in the next few years, however, and these developments will make genotyping-by-sequencing methods more appealing for parentage analysis.

Choosing Software

The last few years have witnessed a proliferation of new software packages for parentage analysis. Here we focus on older software packages that have become the workhorses of parentage analysis, as well as newer promising programs, some of which may still need further testing. Many of the older, less popular software packages are still available, and the interested reader can track them down by consulting previous reviews (Jones and Ardren 2003; Jones et al. 2010). Here, we organize software packages by analysis approach. These programs are summarized in Table 4, and a bit more detail about each package is provided in Appendix 2.

Some important issues when choosing software for parentage analysis for next-generation markers, beyond whether they can analyze the type of parentage data collected (e.g., parent-offspring pairs, groups of putative siblings and putative parents, or parent-parent-offspring triads), are: (1) whether the program can handle the number of markers used in the study; and (2) if the method can accept genotype likelihoods that reflect the genotype uncertainties characteristic of next-generation sequencing or if additional consideration of errors will be required. All of the methods worth mentioning incorporate error rates, but most of those error rates are based on expectations for

microsatellites and will likely not properly incorporate error arising from sequencing errors, allelic dropout, and PCR bias, all of which can dramatically impact genotypes in next-generation sequencing datasets such as RAD-seq data (Flanagan and Jones 2017b).

Parentage Assignment Software. The granddaddy of parentage assignment software packages is CERVUS (Marshall et al. 1998; Kalinowski et al. 2007), which was the first mainstream program and remains by far the most popular, even for SNPs (Tables 1, 2) and next-generation data (Table 3). This program was groundbreaking in that it introduced a simulation approach, based on log-likelihoods, to control experiment-wise error. One advantage to CERVUS for most users is that it is user-friendly, with an intuitive graphical user interface. However, this apparent advantage is a drawback for some users, as it cannot be incorporated easily into command-line bioinformatics pipelines. The latest version also seamlessly incorporates next-generation SNP data. The most recent next-generation alternative to CERVUS is called SNPPIT (Anderson 2010; Anderson 2012). This program only allows the analysis of data for pairs of parents and their offspring. SNPPIT also only analyzes biallelic SNP genotypes, a design decision that significantly improves the speed of the analysis and allows a larger number of SNPs to be used.

Sibship Reconstruction and Parentage Assignment. Probably the second most popular parentage analysis program is COLONY, which was the first program to combine sibship reconstruction and parentage assignment in a single analytical framework (Wang 2004; Wang and Santure 2009; Jones and Wang 2010). On Windows operating systems, COLONY has an intuitive graphical user interface, which likely adds to its popularity. A more recent competitor program is the R package SEQUOIA (Huisman 2017), which has the advantage of running in R, a platform that supports many other analyses of next-generation datasets. SEQUOIA is optimized to run very quickly, relative to previous-generation parentage approaches. This optimization is achieved partially through an initial filtering step, based on loci where the individuals in question show alternatively homozygous genotypes,

which removes putative linkages between individuals that are unlikely to be relatives. SEQUIOA also differs from COLONY in considering a wider range of possible relationships, making grandparent-grandoffspring assignments possible, for instance. SEQUIOA performs well with a large number of independent SNPs; for datasets with less than about 200 SNPs, COLONY, with its more exhaustive search of pedigree space, performs better (Huisman 2017).

Bayesian Parentage Analysis. Bayesian parentage analysis, also known as full probability parentage analysis, was first implemented in MASTERBAYES (Hadfield et al. 2006), the R package that still provides the only reasonable framework for the implementation of this analysis technique. MASTERBAYES can be used to perform a parentage analysis per se, based on posterior probabilities (Nielsen et al. 2001), or to implement a full probability parentage analysis that simultaneously estimates population-level parameters during the parentage analysis. Of course, the Bayesian framework used by MASTERBAYES also allows the incorporation of any prior information that sheds light on parentage patterns. For instance, Walling et al. (2010) show that parentage analysis in red deer can be improved by using MASTERBAYES to include phenotypic data. MASTERBAYES has also been used, in a full-probability framework, to estimate a number of interesting ecological parameters, such as the fecundities of dispersing banner-tail kangaroo rats (Waser et al. 2013) and the relationships between mating order, sperm package size, and siring success in bushcrickets (Parker et al. 2017). Other full probability models, similar to those implemented in MASTERBAYES, have been developed independently using the tools available in R, for example to estimate distances of seed and pollen movement in the red oak (Moran and Clark 2011; see also Robledo 2012 and Chybicki 2017 for non-R implementations). Despite a number of successes of the Bayesian parentage approach, full probability models remain underutilized in the study of parentage.

Niche Programs. A number of niche programs with potential utility for various specific parentage scenarios are listed in Table 5. Some of these programs deal with specific issues that apply to certain types of biological systems or datasets. For example, FAMOZ (Gerber et al. 2003) provides a CERVUS-like analysis framework that also works with dominant markers, such as amplified fragment length polymorphisms (AFLPs). Another program, ORCHARD (Spielmann et al. 2015), tackles the thorny problem of parentage analysis in autotetraploids. Other niche programs implement similar algorithms to mainstream programs, but with modifications that increase computational speed or improve them in other ways. For instance, ACCURASSIGN (Boichard et al. 2014) uses a CERVUS-like algorithm, restricted to sire-dam-offspring trios, which is optimized to run more quickly than a full CERVUS analysis. Similarly, FAPS (Ellis et al. 2018) improves upon the speed of COLONY by implementing a hierarchical clustering approach, at the expense of being less general, as FAPS requires arrays of half-siblings whose mothers are known. Finally, some of the niche programs perform functions that are unique. For instance, GERUD2.0 (Jones 2005) exhaustively searches for the genotypes of the minimum number of parents necessary to explain a progeny array, but requires a small number of highly polymorphic markers. GRANDPARENT FINDER (Christie et al. 2011) skips the parentage analysis step altogether and matches offspring to grandparents through an exclusion-based approach. See Table 5 for additional programs of interest beyond those mentioned here.

Other Methods of Note. The most widely used techniques for parentage analysis rely upon the Mendelian likelihoods derived by Meagher and Thompson (1986) and extended by Marshall et al. (1998), but these approaches assume unlinked markers and require knowledge of details like the proportion of parents sampled and genotyping error rates. As datasets get larger, some of these assumptions will be hard to meet, and researchers are beginning to explore other methods with less exacting requirements. For instance, Grashei et al. (2018) develop a method based on the genomic relationship likelihood, which uses metrics based on pairwise relatedness estimates among individuals that are insensitive to linkage disequilibrium (VanRaden 2008). Their method outperforms COLONY when applied to datasets consisting of very large numbers of marker loci (approx. 54,000 SNPs).

Other recently developed approaches include constrained genomic regression (Boerner 2017) and counting of the number of opposing homozygote loci (Wiggans et al. 2009; Hayes 2011). Before any of these new approaches can be considered mainstream alternatives to the tried-and-true approaches discussed above, they will have to be subjected to additional rigorous testing.

Avoiding Parentage Analysis Pitfalls

A number of additional questions should be considered to avoid critical issues during parentage analysis. Linkage disequilibrium is moving to the forefront as a major issue as datasets grow to a genomic scale. The calculation of the likelihoods underlying parentage assignment is greatly simplified by the assumption that loci are in linkage equilibrium (Thompson and Meagher 1998), and virtually all mainstream techniques embrace this simplification (but see the previous section). For now, the most common solution to this problem is to filter loci so that only a subset of statistically independent loci remains. For example, Andrews et al. (2018) illustrate one possible approach in their pipeline for RAD-seq data by using the program PLINK (Purcell et al. 2007) to filter out tightly linked loci.

Historically, one of the most troubling issues in parentage analysis was the occurrence of null (i.e., non-amplifying) alleles at microsatellite loci (Callen et al. 1993; Jones et al. 1998; Dakin and Avise 2004), and this problem persists in next-generation approaches as allelic dropout (Andrews et al. 2016). Null alleles and allelic dropout can lead to false exclusions, and the models of error in most assignment programs are not designed to accommodate this source of error. Rather, the advice is to remove loci suffering from null alleles or allelic dropout from the analysis, a solution that is relatively easy to apply to small microsatellite or SNP datasets but perhaps difficult to apply to the extremely large datasets produced by genotyping-by-sequencing approaches. In addition, the rate of allelic dropout may vary based on the type of next-generation method used (Flanagan and Jones 2017b). Possible approaches are to use a program like GBSTOOLS to estimate which SNPs in the dataset are most likely suffering from allelic dropout (Cooke et al. 2016) or to strictly filter loci for adherence to

Hardy-Weinberg Equilibrium. The effects of allelic dropout still need to be investigated in more detail, particularly in situations in which some parents are missing from the sample, as many parentage programs have the potential to assign incorrect parents with surprisingly high confidence when the true parents have not been genotyped.

Parentage assignment requires some level of understanding of the population from which the samples originated, usually including estimates allele frequencies and at least a rough idea of the proportion of candidate parents sampled (see Appendix 2). Allele frequencies can be calculated from the sample of potential parents, but it is often advisable to have a separate sample of individuals from the population. In addition, the progeny usually should not be included in allele frequency estimates, because the existence of groups of close relatives can lead to bias and spurious departures from Hardy-Weinberg Equilibrium. Most parentage assignment methods also assume that marker loci are neutral, but as these loci come to represent a greater fraction of the genome, some may be targets of selection (Flanagan and Jones 2017a), possibly affecting the outcome of parentage analysis. In terms of the proportion of parents sampled, ignoring this parameter has a direct influence on the confidence of assignments (Nielsen et al. 2001). Thus, the study should be designed to deliver population-level allele frequencies as well as an estimate of the proportion of candidate parents sampled.

Another challenge facing parentage assignment using any type of marker is the accidental inclusion of family members other than the parents in the pool of candidate parents. Most parentage assignment programs assume that the parents are unrelated to each other and that no relatives of the offspring other than parents are included in the sample of adults. However, generations often overlap, resulting in half- or full-siblings of some of the progeny being included in the putative parent pool. This situation can cause problems, because full sibs can have higher assignment likelihoods than the true parent (Thompson 1976a; Thompson 1976b; Thompson and Meagher 1987; Marshall et al. 1998). Sometimes related males will be clustered together, leading to set of related putative fathers, which will also bias the results of parentage assignment (Double et al. 1997). If many relatives are likely to be present in the sample, one solution might be to use an approach that estimates a broader pedigree than just parent-offspring relationships. Programs such as SEQUOIA (Huisman 2017) and a

growing list of pedigree reconstruction programs (Ko and Nielsen 2017) can be used to perform these sorts of analyses. As genomic resources become available for more species, it may even be possible to infer pedigrees using the length and distribution of genome segments that are identical by descent (Hill and White 2013), but this sort of approach still requires some development.

Polyploids pose additional challenges, and consequently fewer approaches exist for assigning parentage in polyploids. The primary challenge facing parentage in polyploids is allele dosage. Allele dosage occurs when an individual has multiple copies of the same allele – for example, if an individual is genotyped with *a*, *b*, and *c* alleles, it could have the genotypes *aabc*, *abbc*, or *abcc* – and its exact genotype can only be diagnosed if it is homozygous for one allele or heterozygous for all four alleles. Some researchers have ignored the problem of allele dosage by (1) treating alleles as loci and transforming codominant microsatellite data into a binary dataset that can be treated like AFLPs (e.g., Gerber et al. 2003; Rodzen et al. 2004) or (2) uncertain alleles are recorded as missing data (Riday et al. 2013). The program ORCHARD (Spielmann et al. 2015) implements a method that tests all possible genotypes in tetraploids, thereby incorporating allele dosage. ORCHARD uses a combination of exclusion and likelihood to assign parentage. While ORCHARD is limited to tetrapolyploids, POLYPATEX accommodates autopolyploids with 4*n*, 6*n*, or 8*n* duplications (Zwart et al. 2016). However, unlike ORCHARD, POLYPATEX conducts only exclusion-based parentage analysis, which is the least desirable of the parentage analysis approaches (see above). In short, much work remains to be done with respect to parentage assignment in polyploids.

Conclusions and Recommendations

Parentage analysis continues to play an important role in molecular ecology, and recent technological advances have made generating data used in parentage analysis even more accessible. We encourage researchers to carefully consider their questions and budgets before they embrace next-generation genotyping approaches. Although SNPs may now be the best marker type for parentage analysis, the number required is far below what is typically included in genotyping-by-sequencing

approaches. Additionally, error arising from allelic dropout can be especially problematic in parentage analysis, and more development of analytical techniques will be required to accommodate these types of errors. Many of these problems can be circumvented with SNP-PCR approaches, but these methods require some up-front development, in the same vein as microsatellite markers. Regardless of how SNPs are generated, researchers should ensure that they carefully select markers that will adhere to the assumptions of parentage analysis programs and should preferentially choose loci with high minor allele frequencies. Researchers will have to weigh the relative strengths and weaknesses of next-generation approaches against traditional molecular markers, such as microsatellites, which have repeatedly proven their worth.

In terms of analytical approach, parentage assignment and sibship reconstruction are the leading methods. Both methods perform well with a wide range of markers and existing mainstream software packages can use next-generation data. Some can even combine data from microsatellites and SNPs into a single analysis. Bayesian methods are also a viable alternative, and they are particularly appropriate when prior information, such as age or location, can be included in the analysis. They also permit a full-probability parentage analysis, which simultaneously estimates patterns of parentage and population-level parameters of interest. This latter method is the most defensible from a statistical standpoint, but it also requires careful consideration in the construction of probability equations. For certain sampling schemes that allow groups of related offspring to be collected together, parental reconstruction remains a viable technique and often substantially reduces the number of markers required. The one method that should be phased out is strict exclusion. Since an error-free dataset is a near impossibility, especially as the number of markers increases, exclusion requires arbitrary decisions regarding the number of mismatches required for a true exclusion. This number cannot be determined from first principles, and exclusion is divorced from statistical theory. Consequently, exclusion-based approaches make poor use of the data and also provide no method to assess confidence of assignments.

The final consideration, which we did not discuss extensively throughout this review, is the sampling design of the study. The success of parentage analysis depends strongly on the nature of samples that can be obtained from the system of interest. Thus, a major consideration is whether or not the organism can be sampled in a way that allows parents to be present in the dataset. Now that abundant genetic markers can be obtained more easily than at any other time in the history of molecular ecology, the sampling plan is perhaps the single most important factor in parentage analysis. Thus, a substantial investment of time and effort in the planning and execution of fieldwork will certainly pay dividends in the study of parentage. Regardless, the future of parentage analysis is bright, and next-generation sequencing promises to deliver answers in systems that were difficult to study with traditional markers. With careful thought and experimental design, parentage analysis can now be conducted with sufficient power to completely resolve virtually any question in this area of inquiry.

Acknowledgments

This work was supported by and conducted in part while SPF was a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville.

Data Availability

All data supporting this review are contained in the tables and appendices.

Author Contributions

SPF and AGF contributed equally to the design, implementation, and writing of this review.

References

Abadía-Cardoso, A., E. C. Anderson, D. E. Pearse, and J. C. Garza. 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). Molecular Ecology 22:4733-4746.

Alam, M., J. Neal, K. O'Connor, A. Kilian, and B. Topp. 2018. Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. PLOS One 13:e0203465.

Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics 202:389-400.

Almudevar, A. 2003. A simulated annealing algorithm for maximum likelihood pedigree reconstruction. Theoretical Population Biology 63:63-75.

Amorim, A., and Pereira, L. 2005. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Science International 150:17-21.

Anderson, E. C. 2010. Computational algorithms and user-friendly software for parentage based tagging of Pacific salmonids. Final report submitted to the Pacific Salmon Commission's Chinook Technical Committee (US Section).

Anderson, E.C. 2012. Large-scale parentage inference with SNPs: an efficient algorithm for statistical confidence of parent pair allocations. Statistical Applications in Genetics and Molecular Biology 11:1544-6115.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567-2582.

Anderson, E. C. and T. C. Ng. 2014. Comment on 'Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching'. Bioinformatics 30:743-745.

Andolfatto, P., D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, and D. L. Stern. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research 21:610-617.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81-92.

Andrews, K. R., J. R. Adams, E. F. Cassirer, R. K. Plowright, C. Gardner, M. Dwire, P. A. Hohenlohe, and L. P. Waits. 2018. A bioinformatic pipeline for identifying informative single nucleotide polymorphism panels for parentage assignment from RADseq data. Molecular Ecology Resources, in press.

Araki, H., R. S. Waples, W. R. Ardren, B. Cooper, and M. S. Blouin. 2007. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. Molecular Ecology 16:953-966.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLOS One 10:e3376.

Beacham, T.D., C. Wallace, C. MacConnachie, K. Jonsen, B. McIntosh, J.R. Candy, R.H. Devlin, and R.E. Withler. 2017. Population and individual identification of coho salmon in British Columbia through parentage-based tagging and genetic stock identification: an alternative to coded-wire tags. Canadian Journal of Fisheries and Aquatic Sciences 74: 1391-1410.

Beacham, T.D., C. Wallace, K. Jonsen, B. McIntosh, J.R. Candy, D. Willis, C. Lynch, J.-S. Moore, L. Bernatchez, and R.E. Withler. 2018a. Comparison of coded-wire tagging with parentage-based tagging and genetic stock identification in a large-scale coho salmon fisheries application in British Columbia, Canada. Evolutionary Applications 2018. https://doi.org/10.1111/eva.12711.

Beacham, T.D., C. Wallace, C. MacConnachie, K. Jonsen, B. McIntosh, J.R. Candy, and R.E. Withler. 2018a. Population and individual identification of Chinook salmon in British Columbia through parentage-based tagging and genetic stock identification with single

nucleotide polymorphisms. Canadian Journal of Fisheries and Aquatic Sciences 75: 1096-1105.
Bell, A. M., J. M. Henshall, S. Gill, K. Gore, and J. W. Kijas. 2013. Success rates of commercial SNP based parentage assignment in sheep. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20:278-281.
Birkhead, T. R., T. Burke, R. Zann, F. M. Hunter, and A. P. Krupa. 1990. Extra-pair paternity and intraspecific brood parasitism in wild zebra finches *Taeniopygia guttata*, revealed by DNA fingerprinting. Behavioral Ecology and Sociobiology 27:315-324.
Boerner, V. 2017. On marker-based parentage verification via non-linear optimization. Genetics Selection Evolution 49:50.
Boichard, D. 2014. AccurAssign, software for accurate maximum-likelihood parentage assignment. 10th World Congress of Genetics Applied to Livestock Production.
Borrell, Y. J., V. Gallego, C. García-Fernández, I. Mazzeo, L. Pérez, J. F. Asturiano, C. E. Carlos, E. Vázquez, J. A. Sánchez, and G. Blanco. 2011. Assessment of parental contributions to fast- and slow-growing progenies in the sea bream *Sparus aurata* L. using a new multiplex PCR. Aquaculture 314:58-65.
Boyle, J. H., D. J. Martins, J. Pelaez, P. M. Musili, S. Kibet, S. K. Ndung'u, D. Kenfack, and N. E. Pierce. 2018. Polygyny does not explain the superior competitive ability of dominant ant associates in the African ant-plant, *Acacia* (*Vachellia*) *drepanolobium*. Ecology and Evolution 8:1441-1450.
Broccanello, C., C. Chiodi, A. Funk, J. M. McGrath, L. Panella, and P. Stevanato. 2018. Comparison of three PCR-based assays for SNP genotyping in plants. Plant Methods 14:28.
Brockmann, H. J., T. Colson, and W. Potts. 1994. Sperm competition in horseshoe crabs (*Limulus polyphemus*). Behavioral Ecology and Sociobiology 35:153-160.
Buchanan, J. W., G. N. Woronuk, F. L. Marquess, K. Lang, S. T. James, H. Deobald, B. T. Welly, and A. L. Van Eenennaam. 2017. Analysis of validated and population-specific single nucleotide polymorphism parentage panels in pedigreed and commercial beef cattle populations. Canadian Journal of Animal Science 97:231-240.
Burke, T., N. B. Davies, M. W. Bruford, and B. J. Hatchwell. 1989. Parental care and mating behaviour of polyandrous dunnocks *Prunella modularis* related to paternity by DNA fingerprinting. Nature 338:249-251.
Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, J. C. Mulley, and G. R. Sutherland. 1993. Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite markers. American Journal of Human Genetics 52:922-927.
Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resrources 15:855-867.
Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., Wang, J. 2018. Genetic and genomic monitoring with minimally invasive sampling methods. Evolutionary Applications. 11:1094-1119.
Castoe, T. A., A. W. Poole, W. J. Gu, A. P. de Koning, J. M. Daza, E. N. Smith, and D. D. Pollock. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. Molecular Ecology Resources 10:341-347.
Chakraborty, R., M. Shaw, M. J. Schull. 1974. Exclusion of paternity: the current state of the art. American Journal of Human Genetics 26:477-488.
Chakraborty, R., Schull, W. 1976. A note on the distribution of the number of exclusions to be expected in paternity testing. American Journal of Human Genetics 28:615–618.
Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527-536.
Chakraborty, R., J. Li, and Y. X. Zhong. 1994. Paternity evaluation in cases lacking a mother and nondetectable alleles. International Journal of Legal Medicine 107:127-131.

Christie, M. R., M. Marine, L., and M. S. Blouin. 2011. Who are the missing parents? Grandparentage analysis identifies multiple sources of gene flow into a wild population. Molecular Ecology 20:1263-1276.
Christie, M. R., J. A. Tennessen, and M. S. Blouin. 2013. Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching. Bioinformatics 29:725-732.
Chybicki, I. J. 2017. NMπ – improved re-implementation of NM+, a software for estimating gene dispersal and mating systems. Molecular Ecology Resources 18:159-168.
Clarke, S. M., H. M. Henry, K. G. Dodds, T. W. D. Jowett, T. R. Manley, R. M. Anderson, and J. C. McEwan. 2014. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PLoS One 9:e93392.
Coltman, D. W., M. Festa-Bianchet, J. T. Jorgenson, and C. Strobeck. 2002. Age-dependent sexual selection in bighorn rams. Proceedings of the Royal Society of London B 269:165-172.
Conner, J. K., S. Rush, S. Kercher, and P. Jennetten. 1996. Measurements of natural selection on floral traits in wild radish (*Raphanus raphanistrum*) II. Selection through lifetime male and total fitness. Evolution 50:1137-1146.
Cooke, T. F., M.-C. Yee, M. Muzzio, A. Sockell, R. Bell, O. E. Cornejo, J. L. Kelley, G. Bailliet, C. M. Bravi, C. D. Bustamante, and E. E. Kenny. 2016. GBStools: A statistical method for estimating allelic dropout in reduced representation sequencing data. PLoS Genetics 12:e1005631.
Cramer, E. R. A., M. L. Hall, S. R. de Kort, I. J. Lovette, and S. L. Vehrencamp. 2011. Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine. Condor 113:637-645.
Croshaw, D. A., M. B. Peters, and T. C. Glenn. 2009. Comparing the performance of analytical techniques for genetic parentage of half-sib progeny arrays. Genetics Research 91:313-325.
Dakin, E. E. and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity 93:504.
Devlin, B., K. Roeder, and N. C. Ellstrand. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. Theoretical and Applied Genetics 76:369-380.
DeWoody, J. A. and J. C. Avise. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. Journal of Fish Biology 56:461-473.
Double, M. C., A. Cockburn, S. C. Barry, and P. E. Smouse. 1997. Exclusion probabilities for single- locus paternity analysis when related males compete for matings. Molecular Ecology 6:1155-1166.
Dow, B. D. and M. V. Ashley. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Molecular Ecology 5:615-627.
Dussault, F. M. and E. G. Boulding. 2018. Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: A methodology illustrated using Atlantic salmon. Aquaculture Research 49:1368-1372.
Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics 12:551-558.
Ellis, T. J., D. L. Field, and N. H. Barton. 2018. Efficient inference of paternity and sibship inference given known maternity via hierarchical clustering. Molecular Ecology Resources 18:988-999.
Ellstrand, N. C. 1984. Multiple paternity within the fruits of the wild radish, *Raphanus sativus*. The American Naturalist 123:819-828.
Ferdosi, M. H., B. P. Kinghorn, J. H. J. van der Werf, S. H. Lee, and C. Gondro. 2014. hsphase: An R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. BMC Bioinformatics 15:172.
Fisher, P. J., B. Malthus, M. C. Walker, G. Corbett, and R. J. Spelman. 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. Journal of Dairy Science 92:369-374.
Flanagan, S. P. and A. G. Jones. 2017a. Genome-wide selection components analysis in a fish with male pregnancy. Evolution 71:1096-1105.
Flanagan, S. P. and A. G. Jones. 2017b. Substantial differences in bias between single-digest and double-digest RAD-seq libraries: A case study. Molecular Ecology Resources 18:264-280.

García-Fernández, C., J. A. Sánchez, and G. Blanco. 2018. SNP-haplotypes: An accurate approach for parentage and relatedness inference in gilthead sea bream (*Sparus aurata*). Aquaculture 495:582-591.

Gerber, S., P. Chabrier, and A. Kremer. 2003. Famoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Molecular Ecology Notes 3:479-481.

Glaubitz, J.C., O. E. Rhodes Jr., J. A. DeWoody. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology 12:1039-1047.

Gowaty, P. A. and A. A. Karlin. 1984. Multiple maternity and paternity in single broods of apparently monogamous eastern bluebirds (*Sialia sialis*). Behavioral Ecology and Sociobiology 15:91-95.

Grashei, K. E., J. Ødegård, T. H. E. Meuwissen. 2018. Using genomic relationship likelihood for parentage assessment. Genetics Selection Evolution 50:26.

Gudex, B., M. Walker, P. Fisher, and R. Spelman. 2014. Validation of a single-nucleotide polymorphism panel for parentage testing of farmed red deer. Animal Genetics 45:142-143.

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Léger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin, and R. J. Petit. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11:591-611.

Gutierrez, A. P., F. Turner, K. Gharbi, R. Talbot, N. R. Lowe, C. Peñaloza, M. McCullough, P. A. Prodöhl, T. P. Bean, and R. D. Houston. 2017. Development of a medium density combined-species SNP array for Pacific and European oysters (*Crassostrea gigas* and *Ostrea edulis*). G3: Genes, Genomes, Genetics 7:2209-2218.

Hadfield, J. D., D. S. Richardson, and T. Burke. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. Molecular Ecology 15:3715-3730.

Haig, S. 1998. Molecular contributions to conservation. Ecology 79:413-425.

Hanken, J. and P. W. Sherman. 1981. Multiple paternity in Belding's ground squirrel litters. Science 212:351-353.

Harlizius, B., M. S. Lopes, N. Duijvesteijn, L. van de Goor, W. A. van Haeringen, H. Panneman, S. E. F. Guimaraes, J. W. M. Merks, and E. F. Knol. 2011. A single nucleotide polymorphism set for paternal identification to reduce the costs of trait recording in commercial pig breeding. Journal of Animal Science 89:1661-1668.

Harney, E., B. Dubief, P. Boudry, O. Basuyaux, M.B. Schihabel, S. Huchette, C. Paillard, and F.L.D. Nunes. 2016. *De novo* assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome. Marine Genomics 28: 11-16.

Harney, E., S. Lachambre, S. Roussel, S. Huchette, F. Enez, R. Morvezen, P. Haffray, and P. Boudry. 2018. Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*. Aquaculture 491:105-113.

Hauser, L., M. Baird, R. A. Y. Hilborn, L. W. Seeb, and J. E. Seeb. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Molecular Ecology Resources 11:150-161.

Hayes, B. 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science 94:2114-2117.

Head, M. L., A. T. Kahn, J. M. Henshaw, J. S. Keogh, and M. D. Jennions. 2017. Sexual selection on male body size, genital length and heterozygosity: Consistency across habitats and social settings. Journal of Animal Ecology 86:1458-1468.

Heaton, M. P., K. A. Leymaster, T. S. Kalbfleisch, J. W. Kijas, S. M. Clarke, J. McEwan, J. F. Maddox, V. Basnayake, D. T. Petrik, B. Simpson, T. P. L. Smith, C. G. Chitko-McKown, and C. Int Sheep Genomics. 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS One 9:e94851.

Hess, J. E., M. W. Ackerman, J. K. Fryer, D. J. Hasselman, C. A. Steele, J. J. Stephenson, J. M. Wheteaker and S. R. Narum. 2016. Differential adult migration-timing and stock-specific abundance of steelhead in mixed stock assemblages. ICES Journal of Marine Science 73, 2606-2615. doi:10.1093/icesjms/fsw138

Hess, M. A., J. E. Hess, A. P. Matala, R. A. French, C. A. Steele, J. C. Lovtang, and S. R. Narum. 2016. Migrating adult steelhead utilize a thermal refuge during summer periods with high water temperatures. ICES Journal of Marine Science 73:2616-2624.

Hill, W. G. and I. M. S. White. 2013. Identification of pedigree relationship from genome sharing. G3: Genes|Genomes|Genetics 3:1553-1571.

Hodel, R. G. J., M. C. Segovia-Salcedo, J. B. Landis, A. A. Crowl, M. Sun, X. Liu, M. A. Gitzendanner, N. A. Douglas, C. C. Germain-Aubrey, S. Chen, D. E. Soltis, and P. S. Soltis. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st Century. Applications in Plant Sciences 4:1600025.

Hoffberg, S. L., T. J. Kieran, J. M. Catchen, A. Devault, B. C. Faircloth, R. Mauricio, T. C. Glenn. 2016. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources 16:1264-1278.

Holl, H. M., J. Vanhnasy, R. E. Everts, K. Hoefs-Martin, D. Cook, S. A. Brooks, M. L. Carpenter, C. D. Bustamante, and C. Lafayette. 2017. Single nucleotide polymorphisms for DNA typing in the domestic horse. Animal Genetics 48:669-676.

Holman, L. E., D. Garcia de la serrana, A. Onoufriou, B. Hillestad, and I. A. Johnson. 2017. A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. Aquaculture 476:59-64.

Hu, H., X. Liu, W. Jin, H. H. Ropers, and T. F. Wienker. 2015. Evaluating information content of SNPs for sample-tagging in resequencing projects. Scientific Reports 5:10247.

Hubby, J. L. and R. C. Lewontin. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54:577-594.

Huisman, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular Ecology Resources 17:1009-1024.

Jarne, P. and P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11:424-429.

Ji, K., D. P. Zhang, L. A. Motilal, M. Boccara, P. Lachenaud, and L. W. Meinhardt. 2013. Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genetic Resources and Crop Evolution 60:441-453.

Jin, Y. L., L. F. Kong, H. Yu, and Q. Li. 2014. Development, inheritance and evaluation of 55 novel single nucleotide polymorphism markers for parentage assignment in the Pacific oyster (*Crassostrea gigas*). Genes & Genomics 36:129-141.

Johnson, J. S., K. D. Gaddis, D. M. Cairns, and K. V. Krutovsky. 2017. Seed dispersal at alpine treeline: an assessment of seed movement within the alpine treeline ecotone. Ecosphere 8:e01649.

Jones, A. G. 2001. Gerud1.0: a computer program for the reconstruction of parental genotypes from progeny arrays using multilocus DNA data. Molecular Ecology Notes 1:215-218.

Jones, A. G. 2005. Gerud 2.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. Molecular Ecology Notes 5:708-711.

Jones, A. G. 2009. On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. Evolution 63:1673-1684.

Jones, A. G. and W. R. Ardren. 2003. Methods of parentage analysis in natural populations. Molecular Ecology 12:2511-2523.

Jones, A. G. and J. C. Avise. 1997a. Microsatellite analysis of maternity and the mating system in the Gulf pipefish *Syngnathus scovelli*, a species with male pregnancy and sex-role reversal. Molecular Ecology 6:203-213.

Jones, A. G. and J. C. Avise. 1997b. Polygynandry in the dusky pipefish *Syngnathus floridae* revealed by microsatellite DNA markers. Evolution 51:1611-1622.

Jones, A. G., C. A. Stockwell, D. Walker, and J. C. Avise. 1998. The molecular basis of a microsatellite null allele from the White Sands pupfish. Journal of Heredity 89:339-342.

Jones, A. G., G. Rosenqvist, A. Berglund, and J. C. Avise. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152:1057-1063.

Jones, A. G., J. R. Arguello, and S. J. Arnold. 2002. Validation of Bateman's principles: a genetic study of sexual selection and mating patterns in the rough-skinned newt. Proceedings of the Royal Society of London B 269:2533-2539.
Jones, A. G., C. M. Small, K. A. Paczolt, and N. L. Ratterman. 2010. A practical guide to methods of parentage analysis. Molecular Ecology Resources 10:6-30.
Jones, O. R. and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551-555.
Kaiser, S. A., S. A. Taylor, N. Chen, T. S. Sillett, E. R. Bondra, and M. S. Webster. 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Molecular Ecology Resources 17:183-193.
Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106.
Kellogg, K. A., J. A. Markert, J. R. Stauffer, and T. D. Kocher. 1995. Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. Proceedings of the Royal Society of London B 260:79-84.
Kess, T., J. Gross, F. Harper, and E.G. Boulding. 2016. Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. Journal of Molluscan Studies 82: 104-109.
Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. Science 220:671-680.
Ko, A. and R. Nielsen. 2017. Composite likelihood method for inferring local pedigrees. PloS Genetics 13:e1006963.
Labuschagne, C., L. Nupen, A. Kotze, P. J. Grobler, and D. L. Dalton. 2015. Assessment of microsatellite and SNP markers for parentage assignment in ex situ African Penguin (*Spheniscus demersus*) populations. Ecology and Evolution 5:4389-4399.
Landegren, U., R. Kaiser, J. Sanders, and L. Hood. 1988. A ligase-mediated gene detection technique. Science 241:1077.
Laucou, V., A. Launay, R. Bacilieri, T. Lacombe, A.-F. Adam-Blondon, A. Bérard, A. Chauveau, M. T. de Andrés, L. Hausman, J. Ibáñez, M.-C. Le Paslier, D. Maghradze, J. M. Martinez-Zapater, E. Maul, M. Ponnaiah, R. Töpfer, J.-P. Péros, and J.-M. Boursiquot. 2018. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. PLoS One 13:e0192540.
Lew, R. M., A. J. Finger, M. R. Baerwald, A. Goodbla, B. May, and M. H. Meek. 2015. Using next-generation sequencing to assist a conservation hatchery: a single-nucleotide polymorphism panel for the genetic management of endangered delta smelt. Transactions of the American Fisheries Society 144:767-779.
Lewontin, R. C. and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595-609.
Liebgold, E. B., P. R. Cabe, R. G. Jaeger, and P. L. Leberg. 2006. Multiple paternity in a salamander with socially monogamous behaviour. Molecular Ecology 15:4153-4160.
Liu, S. X., Y. Palti, G. T. Gao, and C. E. Rexroad. 2016. Development and validation of a SNP panel for parentage assignment in rainbow trout. Aquaculture 452:178-182.
Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.
McClure, M. C., J. McCarthy, P. Flynn, J. C. McClure, E. Dair, D. K. O'Connell, and J. F. Kearney. 2018. SNP data qality control in a national beef and dairy cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics 9:84.
McClure, M. C., J. McCarthy, P. Flynn, R. Weld, M. Keane, K. O'Connell, M. P. Mullen, S. Waters, and J. F. Kearney. 2015. SNP selection for nationwide parentage verification and identification in beef and dairy cattle. Pp. 175-181. International Committee for Animal Recording (ICAR), Rome.
Meagher, T. R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely male parents. The American Naturalist 128:199-215.

Meagher, T. R. and E. A. Thompson. 1986. The relationship between single and parent pair genetic likelihoods in genealogy reconstruction. Theoretical Population Biology 29:87-106.
Mobley, K. B., T. Amundsen, E. Forsgren, P. A. Svensson, and A. G. Jones. 2009. Multiple mating and a low incidence of cuckoldry for nest-holding males in the two-spotted goby, *Gobiusculus flavescens*. BMC Evolutionary Biology 9:6.
Moran, E. V. and J. S. Clark. 2011. Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data. Molecular Ecology 20:1248-1262.
Morin, P. A., J. Wallis, J. J. Moore, and D. S. Woodruff. 1994. Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. Molecular Ecology 3:469-478.
Muhlfield, C. C., S. T. Kalinowski, T. E. McMahon, M. L. Taper, S. Painter, R. F. Leary, and F. W. Allendorf. 2009. Hybridization rapidly reduces fitness of a native trout in the wild. Biology Letters 5:328-331.
Neff, B. D., J. Repka, and M. R. Gross. 2001. A Bayesian framework for parentage analysis: The value of genetic and other biological data. Theoretical Population Biology 59:315-331.
Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D. A. Nickerson, and J. Shendure. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461:272-276.
Nguyen, T. T. T., B. J. Hayes, K. Guthridge, E. S. Ab Rahim, and B. A. Ingram. 2011. Use of a microsatellite-based pedigree in estimation of heritabilities for economic traits in Australian blue mussel, *Mytilus galloprovincialis*. Journal of Animal Breeding and Genetics 128:482-490.
Nguyen, T. T. T., B. J. Hayes, and B. A. Ingram. 2014. Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. Aquaculture 420-421:295-301.
Nielsen, R., D. K. Mattila, P. J. Clapham, and P. J. Palsbøll. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. Genetics 157:1673-1682.
O'Brien, D. M., J. S. Keogh, A. J. Silla, and P. G. Byrne. 2018. The unexpected genetic mating system of the red-backed toadlet (*Pseudophryne coriacea*): A species with prolonged terrestrial breeding and cryptic reproductive behaviour. Molecular Ecology 27:3001-3015.
Paczolt, K. A., W. E. Martin, N. L. Ratterman, and A. G. Jones. 2016. A low rate of multiple maternity for pregnant male northern pipefish *Syngnathus fuscus*. Journal of Fish Biology 88:1614-1619.
Palaiokostas, C., M. Kocour, M. Prchal, and R. D. Houston. 2018. Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. Frontiers in Genetics 9:82.
Panetto, J. C. C., M. A. Machado, M. V. G. B. da Silva, R. S. Barbosa, G. G. dos Santos, R. M. H. Leite, and M. G. C. D. Peixoto. 2017. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. Livestock Science 204:33-38.
Parker, D. J., J. Zaborowska, M. G. Ritchie, and K. Vahed. 2017. Paternity analysis of wild-caught females shows that sperm package size and placement influence fertilization success in the bushcricket *Pholidoptera griseoaptera*. Molecular Ecology 26:3050-3061.
Pemberton, J. M., J. Slate, D. R. Bancroft, and J. A. Barrett. 1995. Non-amplifying alleles at microsatellite loci: a caution for parentage and population studies. Molecular Ecology 4:249-252.
Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.
Planes, S., G. P. Jones, and S. R. Thorrold. 2009. Larval dispersal connects fish populations in a network of marine protected areas. Proceedings of the National Academy of Sciences 106:5693-5697.

Powell, W., G. C. Machray, and J. Provan. 1996. Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215-222.

Premachandra, H. K. A., N. H. Nguyen, and W. Knibb. 2019. Effectiveness of SNPs for parentage analysis and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. Aquaculture 499:24-31.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, J. Maller, P. Sklar, P. de Bakker, M. Daly, and P. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81:559-575.

Queller, D. C., J. E. Strassman, and C. R. Hughes. 1993. Microsatellites and kinship. Trends in Ecology and Evolution 8:285-288.

Riday, H., D. W. Johnson, K. Heyduk, J. A. Raasch, M. E. Darling, and J. M. Sandman. 2013. Paternity testing in an autotetraploid alfalfa breeding polycross. Euphytica 194:335-349.

Riester, M., P. F. Stadler, and K. Klemm. 2009. FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics 25:2134-2139.

Robledo, J. J. 2012. Joint estimation of contemporary seed and pollen dispersal rates among plant populations. Molecular Ecology Resources 12:299-311.

Rodzen, J. A., T. R. Famula, and B. May. 2004. Estimation of parentage and relatedness in the polyploid white sturgeon (*Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci. Aquaculture 232:165-182.

Sancristobal, M. and C. Chevalet. 1997. Error tolerant parent indentification from a finite set of individuals. Genetical Research 70:53-62.

Santana, Q. C., M. P. A. Coetzee, E. T. Steenkamp, O. X. Mlonyeni, G. N. A. Hammond, M. J. Wingfield, and B. D. Wingfield. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. Biotechniques 46:217-223.

Schunter, C., M. Pascual, J.C. Garza, N. Raventos, and E. Macpherson. 2014. Kinship analyses identify fish dispersal events on a temperate coastline. Proceedings of the Royal Socity B 281: 20140556.

Sekino, M. and S. Kakehi. 2012. PARFEX v1.0: an EXCEL™-based software package for parentage allocation. Conservation Genetics Resources 4:275-278.

Sellars, M. J., L. Dierens, S. McWilliam, B. Little, B. Murphy, G. J. Coman, W. Barendse, and J. Henshall. 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. Aquaculture Research 45:417-426.

Smith, B. R., C. M. Herbinger and H. R. Merry. 2001. Accurate partition of individuals into fullsib families from genetic data without parental information. Genetics 158:1329-1338.

Spielmann, A., S. A. Harris, D. H. Boshier, and C. C. Vinson. 2015. Orchard: Paternity program for autotetraploid species. Molecular Ecology Resources 15:915-920.

Städele, V. and L. Vigilant. 2016. Strategies for determining kinship in wild populations using genetic data. Ecology and Evolution 6:6107-6120.

Staples, J., D. Qiao, M. H. Cho, E. K. Silverman, University of Washington Center for Mendelian Genomics, Deborah A. Nickerson, and Jennifer E. Below. 2014. PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. The American Journal of Human Genetics 95:553-564.

Staples, J., D. J. Witherspoon, L. B. Jorde, D. A. Nickerson, the University of Washington Center for Mendelian Genomics, J. E. Below, and C. D. Huff. 2016. PADRE: Pedigree-aware distant-relationship estimation. The American Journal of Human Genetics 99:154-162.

Steele, C. A., E. C. Anderson, M. W. Ackerman, M. A. Hess, N. R. Campbell, S. R. Narum, and M. R. Campbell. 2013. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Canadian Journal of Fisheries and Aquatic Sciences 70:1046-1054.

Strucken, E. M., B. Gudex, M. H. Ferdosi, H. K. Lee, K. D. Song, J. P. Gibson, M. Kelly, E. K. Piper, L. R. Porto-Neto, S. H. Lee, and C. Gondro. 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics 45:572-575.

Strucken, E. M., S. H. Lee, H. K. Lee, K. D. Song, J. P. Gibson, and C. Gondro. 2015. How many markers are enough? Factors influencing parentage testing in different livestock populations. Journal of Animal Breeding and Genetics 133:13-23.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17:6463-6471.
Telfer, E. J., G. T. Stovold, Y. J. Li, O. B. Silva, D. G. Grattapaglia, and H. S. Dungey. 2015. Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: A comparative analysis of marker data power and robustness. PLoS One 10: e0130601.
Thomas, S. C. and W. G. Hill. 2000. Estimating quantitative genetic parameters using sibships reconstructed from marker data. Genetics 155:1961-1972.
Thompson, E. A. 1975. The estimation of pairwise relationships. Annals of Human Genetics 39:173-188.
Thompson, E. A. 1976a. A paradox of genealogical inference. Advances in Applied Probability 8:648-650.
Thompson, E. A. 1976b. Inference of genealogical structure. Social Science Information 15:477-526.
Thompson, E. A. and T. R. Meagher. 1987. Parental and sib likelihoods in genealogy reconstruction. Biometrics 43:585-600.
Thompson, E. A. and T. R. Meagher. 1998. Genetic linkage in the estimation of pairwise relationship. Theoretical and Applied Genetics 97:857-864.
Thongda, W., H. Zhao, D. Zhang, L. N. Jescovitch, M. Liu, X. Guo, M. Schrandt, S. P. Powers, and E. Peatman. 2018. Development of SNP panels as a new tool to assess the genetic diversity, population structure, and parentage analysis of the eastern oyster (*Crassostrea virginica*). Marine Biotechnology 20:385-395.
Thrasher, D. J., B. G. Butcher, L. Campagna, M. S. Webster, and I. J. Lovette. 2018. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. Molecular Ecology Resources, in press.
Tokarska, M., T. Marshall, R. Kowalczyk, J. M. Wojcik, C. Pertoldi, T. N. Kristensen, V. Loeschcke, V. R. Gregersen, and C. Bendixen. 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity 103:326-332.
Tortereau, F., C. R. Moreno, G. Tosser-Klopp, B. Servin, and J. Raoul. 2017. Development of a SNP panel dedicated to parentage assignment in French sheep populations. BMC Genetics 18:11.
Vandeputte, M., S. Mauger, and M. Dupont-Nivet. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes 6:265-267.
VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91:4414-4423.
Walling, C. A., J. M. Pemberton, J. D. Hadfield, and L. E. B. Kruuk. 2010. Comparing parentage inference software: reanalysis of a red deer pedigree. Molecular Ecology 19:1914-1928.
Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963-1979.
Wang, J. 2010. Effects of genotyping errors on parentage exclusion analysis. Molecular Ecology 19:5061-5078.
Wang, J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. Genetics 191:183-194.
Wang, J. 2013. A simulation module in the computer program COLONY for sibship and parentage analysis. Molecular Ecology Resources 13:734-739.
Wang, J. and A. W. Santure. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. Genetics 181:1579-1594.
Waser, P. M., K. M. Nichols, and J. D. Hadfield. 2013. Fitness consequences of dispersal: Is leaving hom the best of a bad lot? Ecology 94:1287-1295.
Weinman, L. R., J. W. Solomon, and D. R. Rubenstein. 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.
Weller, J. I., G. Glick, E. Ezra, Y. Zeron, E. Seroussi, and M. Ron. 2010. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. Animal Genetics 41:551-553.

Wiggans, G. R., T. S. Sonstegard, P. M. Vanraden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science 92:3431-3436.

Xu, J., J. Y. Feng, W. Z. Peng, X. Liu, J. X. Feng, and P. Xu. 2017. Development and evaluation of a high-throughput single nucleotide polymorphism multiplex assay for assigning pedigrees in common carp. Aquaculture Research 48:1866-1876.

Yu, G. C., Q. Z. Tang, K. R. Long, T. D. Che, M. Z. Li, and S. R. Shuai. 2015. Effectiveness of microsatellite and single nucleotide polymorphism markers for parentage analysis in European domestic pigs. Genetics and Molecular Research 14:1362-1370.

Zhang, T., L. Guo, M. Shi, L. Xu, Y. Chen, L. Zhang, H. Gao, J. Li, and X. Gao. 2018. Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. Gene 673:211-216.

Zhao, H. G., C. Li, J. S. Hargrove, B. R. Bowen, W. Thongda, D. D. Zhang, H. Mohammed, B. H. Beck, J. D. Austin, and E. Peatman. 2018. SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). Aquaculture 485:30-38.

Zwart, A. B., C. Elliott, T. Hopley, D. Lovell, and A. Young. 2016. Polypatex: an R package for paternity exclusion in autopolyploids. Molecular Ecology Resources 16:694-700.

Table 1. Studies of parentage using 'classical' SNPs – including Illumina SNP-Chips. Each method used an existing panel of reference SNPs. SNP-PCR refers to any SNP genotyping technique that prominently features a PCR amplification of a small number of specific loci. Examples include SNPlex (ABI; e.g., Cramer et al. 2011), Oligo Ligation Assay (OLA; e.g., Landegren et al. 1988), TaqMan (ABI; e.g., Hauser et al. 2011), iPLEX/MassARRAY (Agena Bioscience; e.g., Sellars et al. 2014, Weinman et al. 2015, Zhao et al. 2018), Fluidigm Integrated Fluidic Circuits (Fluidigm; e.g., Lew et al. 2015), and Ion AmpliSeq (Thermo Fisher Scientific: e.g., Beacham et al. 2017, 2018a,b).

Reference	Organism	Number of Micro-satellites	SNP Method, Number of SNPs	Parentage Analysis Software	Parentage Relevance and Conclusions
Abadía-Cardoso et al. (2013)	Steelhead (fish)	0	SNP-PCR, 95	SNPPIT	Reconstruction of pedigrees via parentage analysis provided insights into aspects of life history and heritability. A panel of 95 SNPs resolved parentage in this very large study (*N*=4,895 samples).
Bell et al. (2013)	Sheep (mammal)	0	SNP-PCR, 383	Custom Maximum Likelihood Approach	Tested SNP panels for parentage analysis in domestic sheep. As few as 127 SNPs resolved parentage well.
Beacham et al. (2017)	Coho Salmon (fish)	0	SNP-PCR, 304	SNPPIT, COLONY 2.0.6.2	Tested direct sequencing of SNPs located in amplicons for parentage-based tagging and genetic stock identification. With the 304 SNPs, 92% of individuals of known age and origin were assigned to parents with 100% accuracy using SNPPIT and 99.9% of individuals were assigned with 99.9% accuracy via COLONY.
Beacham et al. (2018a)	Coho Salmon (fish)	0	SNP-PCR, 304	COLONY 2.0.6.2	Compared parentage-based tagging and genetic stock identification to coded wire tags to identify coho salmon sampled in fisheries and escapements. Found 100% population assignment accuracy using parentage-based tagging, with a total of 94.8% of known-origin individuals assigned. Found that the parentage-based tagging system is less expensive than the coded wire tagging system and performs at least as well, if not better, than coded wire tags in assigning individuals to populations of origin.
Beacham et al.	Chinook	0	SNP-	SNPPIT,	Tested the ability to directly sequence SNPs in amplicons for

(2018b)	salmon (fish)		PCR, 321	COLONY 2.0.6.2	parentage-based tagging and genetic stock identification in Chinook salmon. 82% of 656 one-year-old individuals of known origin were assigned with 100% accuracy to their source population using SNPPIT. COLONY assigned 96.9% of known-origin individuals to the correct population with 99.8% accuracy.
Buchanan et al. (2017)	Beef Cattle (mammal)	14	SNP-PCR, 109; SNP-Chip, 1000	Custom Exclusion Approach	This comparison of microsatellites and SNPs found both to resolve parentage accurately.
Clarke et al. (2014)	Sheep (mammal)	11	SNP-PCR, 84	CERVUS 3.0.3, Custom Maximum Likelihood	Explored the utility of SNPs for parentage analysis in sheep. The SNPs resolved parentage almost perfectly, a better result than that obtained with microsatellites.
Cramer et al. (2011)	Banded Wren (bird)	7	SNP-PCR, 41	CERVUS 3.0	In a study of extra-pair paternity, 41 SNP loci had greater power than 7 microsatellites. The combined dataset assigned all offspring with >99% confidence.
Dussault and Boulding (2018)	Atlantic Salmon (fish)	0	SNP-Chip, 150	CERVUS 3.0.7	This study investigated how the SNP minor allele frequency impacts parentage assignment. Results show that loci with minor allele frequencies nearer to 0.5 are more powerful. A set of 50-150 SNPs was sufficient to resolve parentage completely.
Fisher et al. (2009)	Dairy Cattle (mammal)	14	SNP-PCR, 72	Custom Exclusion Software	In a comparison of marker types, 40 or more SNPs were as effective for parentage analysis as 14 microsatellites.
García-Fernández et al. (2018)	Gilthead Sea Bream (fish)	9	SNP-PCR, 58	CERVUS 3.0	The 58 SNPs were from 7 genes and were used to resolve haplotypes, resulting in 3-41 haplotypes per locus. The haplotypes (with 99.2 percent accuracy) performed better than either 9 microsatellites (95.7 percent) or the 58 original SNPs (88.7 percent). Microsatellite data are from Borrell et al. (2011).
Gudex et al.	Red Deer	12	SNP-	Custom	In a comparison of microsatellites and SNPs, 100 SNPs

(2014)	(mammal)		PCR, 100	Exclusion Software from Fisher et al. (2009)	outperformed 12 microsatellites for parentage exclusion.
Harlizius et al. (2011)	Pig (mammal)	0	SNP-PCR, 120	CERVUS	A known pig pedigree was used to develop a battery of SNPs for parentage testing. At least 60 SNPs were required for reliable assignment. A troubling pattern was that incorrect assignments increased with more loci when the true parent was not sampled.
Hauser et al. (2011)	Sockeye Salmon (fish)	11	SNP-PCR, 80	CERVUS 3.0, COLONY 2.0, SNPPIT	Compared assignment based on microsatellites and SNPs. The SNP panel generally outperformed the microsatellites.
Heaton et al. (2014)	Sheep (mammal)	0	SNP-PCR, 109	Manual Exclusion	An examination of 95 tetrad families (one mother, one father, two offspring) demonstrated that 109 SNPs can effectively resolve parentage.
Hess JE et al. (2016)	Steelhead (fish)	0	SNP-PCR, 95	SNPPIT	Used parentage-based tagging to assign individuals to stock locations, with the ultimate goal of estimating variation among stocks in abundance and migration timing.
Hess MA et al. (2016)	Steelhead (fish)	0	SNP-PCR, 95	SNPPIT	Parentage analysis was used to show that steelhead use thermal refuges to avoid prolonged exposure to high water temperatures.
Holl et al. (2017)	Domestic Horse (mammal)	0	SNP-PCR, 101	Custom Exclusion Approach	A panel of SNPs was used to assess parentage in known horse families. The 101 SNPs correctly resolved the parentage of 99.9% of offspring.
Ji et al. (2013)	Cacao (tree)	0	SNP-PCR, 70	CERVUS 3.0	Parentage analysis used to assign parents of farmer varieties of cacao. The SNPs assigned parentage for 28/53 varieties at 80% confidence. The reference "parents" (cacao clones) were not the direct parents of the farmer varieties, hence the low assignment rate.
Jin et al. (2014)	Pacific Oyster (mollusk)	0	SNP-PCR, 48	CERVUS 3.0	An analysis of six families showed that 40 SNPs proved sufficient to assign all offspring to parents.
Labuschagne	African	10	SNP-	CERVUS	For parentage analysis in a small captive population, SNPs

et al. (2015)	Penguin (bird)		PCR, 31	3.03, PARFEX	performed at least as well as microsatellites.
Liu et al. (2016)	Rainbow Trout (fish)	0	SNP-PCR, 95	CERVUS 3.0.7, SNPPIT	Parentage was assigned for fish with a known pedigree. The 95-SNP panel was sufficient to completely resolve parentage. CERVUS slightly outperformed SNPPIT.
McClure et al. (2015)	Cattle (mammal)	0	SNP-Chip, 800	Custom Exclusion Approach	Different numbers of SNPs were tested for parentage validation in cattle. The authors recommend using at least 500 SNPs.
McClure et al. (2018)	Cattle (mammal)	0	SNP-Chip, 800	Custom Exclusion Approach	The goal of this study was to develop better quality control practices for SNP-based pedigree validation. The authors recommend at least 500 SNPs be used for parentage validation.
Panetto et al. (2017)	Cattle (mammal)	0	SNP-Chip, 3894	Exclusion	An analysis of Red Sindhi cattle showed that as few as 71 SNPs were sufficient for parentage verification with extremely high confidence.
Schunter et al. (2014)	Black-Faced Blenny (fish)	0	SNP-PCR, 192	CERVUS, COLONY	Genotyped recruits to test whether larvae settle back in their natal location. Parentage analysis revealed limited successful dispersal and only short-term dispersal events. Sibship reconstruction revealed that recruitment can still occur quite distant from the natal site. This study shows that information provided by parentage analysis and kinship reconstruction can be complementary.
Steele et al. (2013)	Steelhead (fish)	17	SNP-PCR, 188	SNPPIT for SNPs, CERVUS 3.0.3 for microsatellites	Tested the feasibility of parentage-based tagging (PBT) based on SNP markers. As few as 72 SNPs can perform as well for parentage analysis as 17 microsatellite markers.
Strucken et al. (2014)	East Asian Cattle (mammal)	0	SNP-Chip, 200	Custom Exclusion Approach (Hayes 2011)	In a test of two commonly used SNP panels in cattle, results showed that 100 SNPs produced about 3-4% false positives, whereas 200 SNPs resolved parentage perfectly.
Strucken et al.	Cattle and	0	SNP-	Custom	Starting with a panel including tens of thousands of SNPs, multiple

(2015)	Sheep (mammals)		Chip, 33159 (cattle) or 48599 (sheep)	Exclusion Approach	smaller panels were tested. The authors conclude that at least 200 SNPs are necessary for reliable parentage testing.
Telfer et al. (2015)	Eucalyptus (tree)	13, 16	SNP-Chip, 106	Manual Exclusion	In an exclusion-based analysis, the 106 SNP markers outperformed a panel of 13 microsatellites and equaled the performance of 16 microsatellites.
Thongda et al. (2018)	Eastern Oyster (mollusk)	16	SNP-PCR, 58	CERVUS 3.0.7, SNPPIT 1.0	This study developed a panel of 58 SNPs and tested them in farmed populations of oysters. Parentage assignments from 58 SNPs largely agreed (98.74%) with assignment based on 16 microsatellite loci.
Tokarska et al. (2009)	European Bison (mammal)	17	SNP-Chip, 960	CERVUS 3.0.3	Low genetic diversity rendered the microsatellites unsuitable for parentage analysis. Simulations showed that as few as 60-100 SNPs completely resolved parentage.
Tortereau et al. (2017)	Sheep (mammal)	0	SNP-PCR, 249	Custom Maximum Likelihood (Boichard et al. 2014)	A SNP-Chip was used to identify SNPs suitable for a parentage testing panel. At least 175 of these SNPs were required for accurate parentage assignment.
Weller et al. (2010)	Holstein Cattle (mammal)	0	SNP-Chip, 38828	Custom Exclusion Approach	This study genotyped fathers and sons at nearly 40,000 SNPs. These genotypes allowed exclusion of bulls incorrectly assumed to be sires and an estimate of the SNP genotyping error rate (~0.05%).
Wiggans et al. (2009)	Dairy Cattle (mammal)	0	SNP-Chip, 40874	Custom Exclusion Approach	A comparison of known parent-offspring pairs showed that this panel of SNPs could easily distinguish between true parents and unrelated individuals.
Xu et al. (2017)	Common Carp (fish)	0	SNP-PCR, 48	CERVUS 3.0.7	Efficacy of SNP-based parentage analysis was tested on sire-dam-offspring trios of common carp. Assignment rates based on CERVUS were approximately 87.3%.
Yu et al. (2015)	Pig (mammal)	12	SNP-Chip, 960	CERVUS 3.0	In a sample of 24 pigs, as few as 30 SNPs provided better parentage resolving power than a dozen microsatellites.

Zhang et al. (2018)	Cattle (mammal)	0	SNP-Chip, 774,660	CERVUS 3.0	After genotyping 1074 calves with an Illumina BovineHD BeadChip, 303 SNPs were identified as highly informative. The top 50 of these were tested in paternity analysis and were found to resolve paternity in 99.89% of cases in Chinese Simmental cattle.

Table 2. Parentage analysis using some form of novel next-generation sequencing to identify SNPs to subsequently assay. All of these studies used the Illumina sequencing platform during the next-generation sequencing step. SNPs were subsequently assayed using techniques such as those listed in the Table 1 caption.

Reference	Organism	Number of Micro-satellites	Next-Gen Method	Number of SNPs	Parentage Analysis Software	Parentage Relevance and Conclusions
Gutierrez et al. (2017)	Pacific and European Oysters (mollusk)	0	Whole-genome sequencing, RAD-seq	~27,000, ~11,000	CERVUS 3.0.7, identity-by-state clustering	Whole genome sequencing and RAD-seq were used to identify SNPs, which provided the markers for a SNP array. The SNP array was tested in a parentage analysis involving three nuclear families (with 5 parents and 161 offspring). The SNP array performed well in assigning offspring to families.
Harney et al. (2018)	European Abalone (mollusk)	0	Transcriptome (Harney et al. 2016)	123	CERVUS 3.0, VITASSIGN 8.5	Analysis of known families of abalone revealed that CERVUS correctly assigned 99.99% of offspring to their parents. CERVUS slightly outperformed the exclusion method implemented in VITASSIGN.
Holman et al. (2017)	Atlantic salmon (fish)	0	sdRAD-seq	94	COLONY 2.0.6.2	Used RAD-seq to identify SNPs to be developed into a marker set. In populations involving known crosses, 94 SNPs resolved parentage with 100 percent accuracy.
Kaiser et al. (2017)	Blue Warbler (bird)	6	RNA-seq	97	CERVUS 3.0	Novel SNPs, developed using RNA-Seq but genotyped by a SNP-PCR technique, were compared to microsatellites. A panel of 97 SNPs had approximately the same parentage resolving power as 6 microsatellite loci.
Laucou et al.	Grapevine	20	Whole-	10,207	FAMOZ	Whole-genome resequencing, via Illumina,

(2018)	(plant crop)		genome sequencing			was used to develop a SNP-Chip. These SNPs were used to investigate parentage for 783 grapevine cultivars, and the results were verified using microsatellite data.
Lew et al. (2015)	Delta Smelt (fish)	0	sdRAD-seq	24	CERVUS 3.0	A SNP panel consisting of as few as 24 markers can effectively resolve parentage in captive populations.
Nguyen et al. (2014)	Blue Mussel (mollusk)	10	GBS	179	MasterBayes	In a hatchery setting, mussels were assigned to families to estimate heritabilities and selection on traits. Microsatellites assigned only 62.6%, whereas SNPs assigned 92.5%. Microsatellite data are from Nguyen et al. (2011).
Sellars et al. (2014)	Black Tiger Shrimp (crustacean)	13	RNA-seq	122	Custom CERVUS-like analysis	Compared assignment based on microsatellites and SNPs. The SNPs provided more power.
Weinman et al. (2015)	Superb Starlings (bird)	15	RNA-seq	102	CERVUS 3.0, COLONY 2.0	Compared markers for parentage assignment in a cooperative breeder. Microsatellites and SNPs performed similarly.
Zhao et al. (2018)	Florida Bass (fish)	10	GBS	58	CERVUS 3.0, SNPPIT	This study developed a novel SNP panel. A comparison of SNPs to microsatellites showed that 58 SNPs performed better than 10 microsatellites. The program SNPPIT slightly outperformed CERVUS.

Table 3. Parentage analysis using next-generation sequencing to genotype the SNPs. All of these studies used the Illumina sequencing platform during the next-generation sequencing step.

Reference	Organism	Number of Micro-satellites	Next-Gen Method	Number of SNPs	Parentage Analysis Software	Parentage Relevance and Conclusions
Alam et al. (2018)	Macadamia (tree)	0	DArT-seq	3,956	CERVUS 3.0.7	DArT-seq is a reduced representation sequencing approach, similar to RAD-seq. Parentage analysis based on DArT-seq SNPs was used to identify the parental cultivars for 18 seedlings.
Andrews et al. (2018)	Mexican gray wolf and bighorn sheep (mammal)	22 (wolf), 14 (sheep)	ddRAD-seq	139-363 (wolf), 142-523 (sheep)	CERVUS 3.0	This study developed a pipeline for assigning parentage using RAD-seq data. SNPs that were generated from *de novo* RAD-seq analysis and from reference-guided analysis were compared to microsatellites. The results demonstrated that SNPs had higher power than the microsatellite panels.
Boyle et al. (2018)	Acacia Ants (insect)	0	ddRAD-seq	309-764	COLONY	A ddRAD-seq approach was used to generate several hundred SNPs for four species of acacia-associated ants. Parentage analysis revealed the mating systems of these ants and confirmed that ddRAD-seq is a viable method for this type of study.
Head et al. (2017)	Mosquitofish (fish)	0	DArT-seq	3.171	Custom Exclusion	A parentage analysis in captive populations was used to examine the effects of sex ratio and habitat complexity on sexual selection in mosquitofish. Samples were genotyped using DArT-seq, and parentage was

						assigned by comparing Hamming Distances (Hu et al. 2015) among offspring and putative parents. All offspring were unambiguously assigned to parents (but only 10 candidate males were present per population).
Johnson et al. (2017)	Mountain Hemlock (tree)	0	ddRAD-seq	353	CERVUS 3.0.3	The goal was to assess mode of reproduction and seed sources at the arctic treeline. Only 18/161 plants were assigned a parent from within the 860m × 600m study transect, indicating substantial seed dispersal or serious technical artifacts.
Kess et al. (2017)	Marine Snail (mollusk)	0	ddRAD-seq	1131	COLONY 2.0.4.4	Used ddRAD-seq to determine if the sequenced male is the true father of putative full-siblings from a cross of a hybrid father and parental ecotype mother. All offspring were assigned to the sequenced father, providing evidence that the female had not been multiply mated during crosses.
O'Brien et al. (2018)	Red-Backed Toadlet (amphibian)	0	DArT-seq	15,746	Custom Exclusion	This study examined parentage in the red-backed toadlet, and assigned parentage using the Hamming Distance (Hu et al. 2015). Cutoffs were determined empirically. Results showed that this species is polygynous with intense male-male competition.
Palaiokostas et al. (2018)	Common Carp (fish)	0	RAD-seq	12,311	Sibship Analysis with R/hsphase	Offspring were produced from controlled crosses but raised together, requiring parentage analysis to determine family of origin. The parentage analysis was

						conducted with the pedigree analysis software R/hsphase (Ferdosi et al. 2014), and the data were used to estimate heritability and perform a GWAS for body size.
Premachandra et al. (2019)	Yellowtail Kingfish (fish)	8	DArT-seq	2,128	COLONY 2.0.6.4, SEQUOIA	Yellowtail kingfish were sampled from a communal rearing tank, and parentage was estimated using SNPs and microsatellite markers. By altering the subsets of markers used, the authors investigated the influence of number of SNPs and minor allele frequency on parentage assignment. They recommend using about 500 SNPs with moderate minor allele frequencies.
Thrasher et al. (2018)	Variegated Fairy-Wren (bird)	12	ddRAD-seq	411	CERVUS 3.0.7	After stringently filtering the ddRAD-seq dataset to 411 marker loci, these SNPs show greater power than 12 microsatellites. The SNPs slightly outperform the microsatellites in a study of extra-pair paternity.

Table 4. The leading software packages for parentage analysis.

Name of Software	Parentage Analysis Technique	Are Errors Taken into Account?	Graphical or Command Line?	Preferred Marker Type	Approx. Run Speed	Operating Systems	Reference
CERVUS	Assignment	Yes	Graphical	Micro-satellites or SNPs	Average	Windows, Mac	Marshall et al. (1998)
SNPPIT	Assignment	Yes	Command line	SNP	Fast	Windows, Mac	Anderson (2010; 2012)
COLONY	Sibship Reconstruction	Yes	Graphical	Micro-satellites or SNPs	Average	Windows (no GUI on Mac or Unix)	Jones and Wang (2010)
SEQUIOA	Sibship Reconstruction	Yes	Command line, in R	Micro-satellites or SNPs	Fast	Windows, Mac, Unix	Huisman (2017)
MASTERBAYES	Bayesian	Yes	Command line, in R	Micro-satellites or SNPs	Slow	Windows, Mac, Unix	Hadfield et al. (2006)

Table 5. Niche software packages.

Name of Software	Parentage Analysis Technique	Description	Reference
ACCURASSIGN	Assignment	Assigns parent-pairs using maximum-likelihood based approach similar to CERVUS.	Boichard (2014)
FAMOZ	Assignment	A likelihood-based parentage assignment program that can use codominant, dominant, and cytoplasmic markers. FAMOZ also does not require a priori information on the proportion of sampled candidate parents.	Gerber et al. (2003)
FAPS	Sibship	Given half-sibling progeny arrays with known mothers, sibships and paternities are determined through a hierarchical clustering approach.	Ellis et al. (2018)
FRANZ	Sibship	Can make use of prior information, such as sub-pedigrees, sex, and age.	Riester et al. (2009)
GERUD2.0	Parental Reconstruction	Reconstructs parental genotypes when no parents are known.	Jones (2005)
GRANDPARENT FINDER	Grandparent Identification	Match offspring with missing parents in the sample to grandparents using exclusion.	Christie et al. (2011)
ORCHARD	Exclusion/ Assignment	Combines exclusion and likelihood-based parentage assignment for autotetraploids.	Spielmann et al. (2015)
PARFEX	Assignment	Implements exclusion and likelihood-based parentage assignment in Microsoft Excel.	Sekino and Kakehi (2012)
POLYPATEX	Exclusion	Applies exclusion methods to polyploids with 4*n*, 6*n*, or 8*n* duplication.	Zwart et al. (2016)
SOLOMON	Bayesian	Designed for situations where only a small fraction of all candidate parents can be sampled. In such cases, SOLOMON uses Bayes' theorem to determine the probability of parent-offspring pairs being false given the frequencies of shared alleles. See also Anderson and Ng (2014).	Christie et al. (2013)
VITASSIGN	Exclusion	Allows for mismatches at one or more allele to recover assignment power.	Vandeputte et al. (2006)

Figure 1. A decision tree with price estimates for various SNP genotyping based on availability of existing resources. The prices assume a study with 192 offspring and a parental pool of 192 adults (i.e., 384 total individuals genotyped). Prices do not include DNA extraction, consumable plasticware, or reagents used for DNA quantification. Microsatellite prices were developed assuming the use of 6 microsatellites, following the protocol described in Jones and Avise (1997a). PCR-SNP costs were based on Broccanello et al. (2018) and SNP-chip costs were based on the Bovine SNP50 DNA Analysis Bead Chip (Illumina). RAD-seq costs were updated from Peterson et al. (2012) for Illumina HiSeq pricing. Targeted capture costs were based on Hoffberg et al. (2016). Note that prices may vary dramatically depending on availability of core facilities and contract pricing. Asterisks (*) denote costs assuming the appropriate machines are already available. The full breakdown of the cost estimates is available as Supplementary File 1.