JOURNAL OF
EMERGING INVESTIGATORS

Article

Examining the accuracy of DNA parentage tests using computer simulations and known pedigrees

Eileen Wang[1], Hong Yao[2]
[1]Upper Arlington High School, Upper Arlington, Ohio
[2]The Ohio State University South Centers, Piketon, Ohio

SUMMARY

DNA parentage tests are widely used for identifying family in humans, as well as controlling family representation and inbreeding in animals. But, how accurate are these tests? In this study, we hypothesized that current parentage tests are reliable if the analysis involves only one or a few families. We tested the hypothesis by simulating and performing DNA parentage testing on yellow perch fish *Perca flavescens* with known pedigrees. We found that 100% of the offspring matched their recorded/true parents when one true family was analyzed. However, only 47.25 % of offspring matched their true parents when a group of 20 families were evaluated. Our results suggest that DNA parentage tests are reliable as long as the right methods are used, since these tests involve only one family in most cases, and that the results from parentage analyses of large populations can only be used as a reference.

INTRODUCTION

DNA parentage tests are used to determine whether an individual is the biological parent of another using a DNA fingerprint or profile. In parentage analysis, the genotype of an offspring is compared with that of a candidate parent to determine if the latter could be a parent of the former (1). This analysis is used by many companies and laboratories for family identification (2).

DNA parentage tests are widely used in the justice system as legal evidence for child support, inheritance, social welfare benefits, and family-based immigration and adoption purposes. Paternity tests, which are more common than maternity tests, are used to determine the biological father of a child. In some circumstances, maternity tests are used to determine the biological mother of a child, such as in cases of an adopted child attempting to reunify with his or her biological mother (3).

Parentage analyses are also useful for studying breeding in animals with multiple mating in agricultural and natural systems. When conducting selective breeding programs, pedigree information is essential for identifying genetic relationships among individuals over multiple generations. Lack of pedigree information can potentially confound the effects for which breeding programs are designed (4-5). Practically, clear pedigree information permits maintenance of the maximum amount of genetic diversity and genetic response for economic traits and minimizes the potential detrimental effects of accumulated inbreeding (6-7). The alternative to parentage analysis, the physical tagging approach, is both labor- and space-intensive. The major advantage of parentage analysis is the ability to establish molecular pedigrees to control family representation and inbreeding in mass spawning species.

Molecular techniques for parentage analysis have been used for over two decades. These techniques include polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP), and microsatellite markers (8-11). Microsatellites were introduced as a tool in parentage analyses in the middle of the 1990s (12). The introduction of microsatellite markers, in combination with the proliferation and refinement of statistical techniques, is one of the most important technological breakthroughs for analysis of parentage data (13). Use of microsatellite markers allows close to 100% assignment success and offers new ways to develop parentage testing and agricultural breeding strategies.

Most companies that offer DNA parentage tests state that the probability of accurately determining parentage is typically 99.99%. Many factors affect the accuracy of parentage analysis, such as the type and number of genetic markers, computing software, and the size of the tested population or family. Currently, to our knowledge, there is no existing study testing the accuracy of DNA parentage analysis using large-scale known pedigrees (3, 14), since it is very difficult to set up experiments requiring a large population and different mating sets in humans and agricultural animals.

In this study, we evaluated the accuracy of parentage tests using yellow perch fish *Perca flavescens* as a model system. We chose to use the fish for this study, because *P. flavescens* can easily produce plenty of offspring and families with different mating sets, making them the ideal model for determining the accuracy of DNA parentage tests at a large scale. To our knowledge, this is the first study that evaluates the accuracy of parentage tests using known pedigrees at a large scale. We mated 50 sets of fish and recorded their pedigrees. We hypothesized that current parentage tests are

reliable when analysis involves a small number of families.

Initially, we performed a simulation study with hypothetical parentage-pairs and pseudo-offspring and found that parentage analysis using the selected loci sets was effective. We then tested our hypothesis by mating the previously selected parentage-pairs and comparing the DNA test results of actual offspring to their recorded pedigrees. Using different computer programs, genetic marker numbers, and family size, we found that 100% of the offspring matched their recorded parents when one true family was analyzed. However, only 47.25 % of offspring matched their true parents when a group of 20 families were evaluated. Our work concluded that the current DNA parentage tests are reliable if the right methods are used, since these tests involve only one family in most cases.

RESULTS

Highly polymorphic microsatellite panel characterization

An eight microsatellite marker panel was used and evaluated for this study. Characteristics of the highly polymorphic panel of 8 microsatellite loci, based on the allele frequencies of 175 broodstock, are presented in **Table 1**. The 8 loci were informative and the number of alleles in the 8 selected loci varied from 5 to 23, with an average expected heterozygosity (i.e. gene diversity of a locus) of 0.81 and polymorphism information content (PIC) score of 0.78. PIC value is often used to measure the informativeness of a genetic marker. The expected heterozygosity ranged from 0.6367 to 0.9131, while the observed heterozygosity varied from 0.4262 to 0.9107. These loci displayed a very high level of PIC, ranging from 0.6212 to 0.9032 with an average of 0.7911, suggesting these markers are powerful for parentage analysis. The values of *Fis* were small, indicating that a very low level of inbreeding occurred in the broodstock.

Computer simulation of assumed parents and offspring

Simulated assignment with various markers and families

The simulation examined the feasibility of parentage analysis using selected loci sets, based on the allele frequencies and the number of candidate parents to be tested.

Table 1: Characteristics of the microsatellite panel used for parentage analysis of 175 broodstock and 465 progeny

Locus	No. alleles	HObs	HExp	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI	HW	F(Null)
YP109	26	0.943	0.900	0.892	0.332	0.199	0.061	0.018	0.305	**	-0.0259
YP60	10	0.864	0.863	0.848	0.434	0.275	0.112	0.034	0.327	***	-0.0008
YP41	12	0.865	0.839	0.818	0.493	0.323	0.151	0.046	0.343	***	-0.0156
YP96	24	0.835	0.824	0.806	0.501	0.331	0.147	0.048	0.350	***	-0.0108
YP49	11	0.879	0.810	0.784	0.549	0.373	0.192	0.062	0.361	***	-0.0418
YP78	19	0.895	0.790	0.763	0.570	0.394	0.203	0.071	0.373	***	-0.0709
YP73	24	0.785	0.789	0.769	0.558	0.379	0.182	0.064	0.372	**	-0.0022
YP30	5	0.769	0.643	0.574	0.788	0.638	0.481	0.196	0.478	***	-0.0947
Mean	16.38	0.854	0.807	0.782	4.90E-03	2.07E-04	5.10E-07	7.40E-11	2.86E-04		

Key: observed heterozygosity (HObs), expected heterozygosity (HExp), polymorphism information content (PIC), significance of deviation from Hardy-Weinberg equilibrium (HW), frequency of null alleles (F(Null)), and average non-exclusion probability (NE). ** = significant at the 1% level, *** = significant at the 0.1% level.

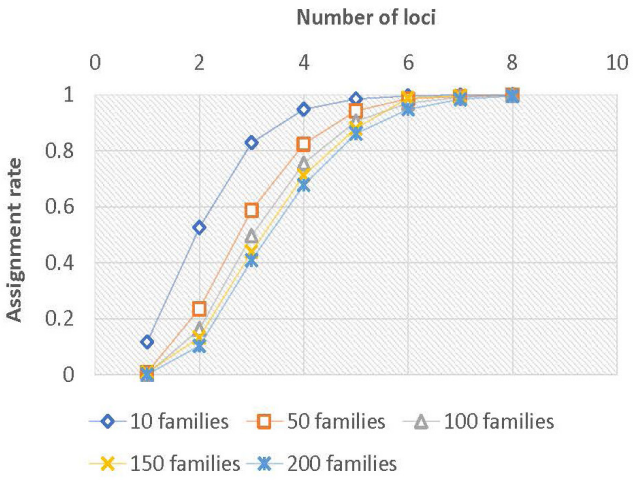

Figure 1. Simulated assignments with varying numbers of markers and families.

Two computer programs, CERVUS (15) and PAPA (16), were used for simulation. Simulation of assignment success was based on genotypes of 10,000 pseudo-offspring at varying amounts of microsatellite loci with an error rate of 0.01. With four markers, the percentage of parentage-pair assignment was ~95% in 100 families and ~82% in 50 families, while, in contrast, the percentages were close to 100% when the number of markers increased to 8 (**Figure 1**).

Simulated assignment with different computer programs

Assignment rates taken from simulations performed by CERVUS and PAPA using 50 parent-pair sets at an error rate of 0.01 were compared. For CERVUS, the parent-pair assignment rate was only 3.00% when using four markers, and 97.00% when using eight markers (**Figure 2**), while the rate of parentage assignment reached 83.38% when using four markers, and 99.82% when using eight markers for PAPA (**Figure 3**).

Actual assignment with actual parents and offspring

Microsatellite profiles from 8 loci selected were used to identify the most likely parents among the 175 potential parents for 465 offspring in 50 mating sets. Parentage assignment of the actual offspring based on the likelihoods of each potential parent pair showed the correct assignment of 465 progeny to their parents, including 107 fathers, 68 mothers, and maximum of 7276 parent pairs based on genotyping errors at 0.01 and 0.05 (**Figure 4**). When using five or less loci, PAPA's assignment rates were over three times higher compared to CERVUS; assignment rates for both computer programs were similar when 7-8 loci were used (**Figure 4**).

Comparison of actual and simulated assignments

The actual assignment results were compared with the simulated assignment to examine how close the actual

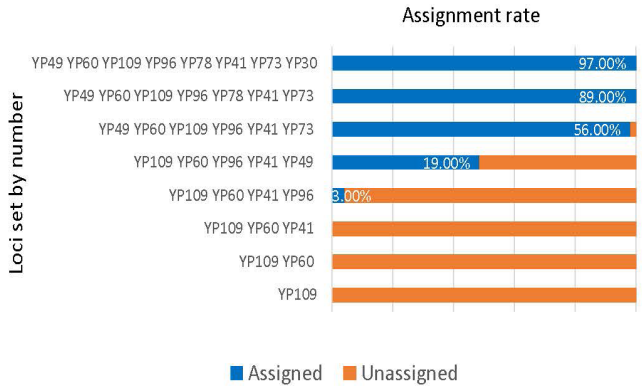

Figure 2. Parentage assignment with the best loci set selected by CERVUS simulation of 10000 pseudo-offspring and at 0.01 of error rate.

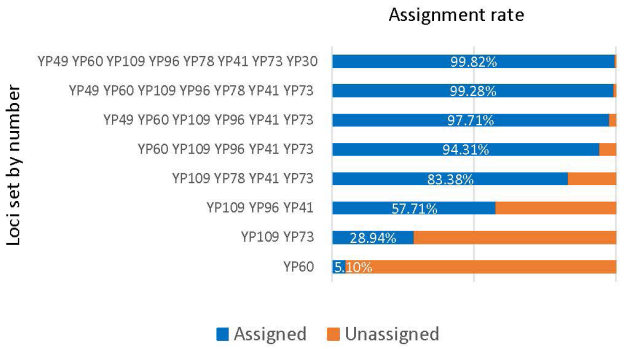

Figure 3. Parentage assignment with the best loci set selected by PAPA simulation of 10000 pseudo-offspring and at 0.01 of error rate.

assignment rates were to simulated assignment rates. Simulations were run based on 10,000 offspring, and actual assignments were on pooled data of offspring. Results of assignment rates for actual and simulation assignments with different numbers of microsatellites using CERVUS at error rates=0.01 and 0.05 are shown in **Table 2**. The results showed that when 7-8 selected markers were used, the actual assignment rates were remarkably close to simulated assignment rates at a 0.01 error rate.

Comparison of actual and recorded pedigrees

Microsatellite DNA marker-based pedigrees were compared with the physical pedigrees to evaluate the accuracy of the DNA parentage analyses. Based on the assignment analysis of one true family, 100% of the offspring matched their true or recorded pedigree/parents, while 95.65% matched to both true parents and 100% matched to half true parents (i.e. true mother or father) in the analysis of 5 true families. In the analysis of 10 true families, 75% matched both true parents and 91.67% matched to half parents. However, when the number of true families reached 20, only 47.25% of the offspring matched their true parents and 71.43 % matched their half true parents (**Figure 5**).

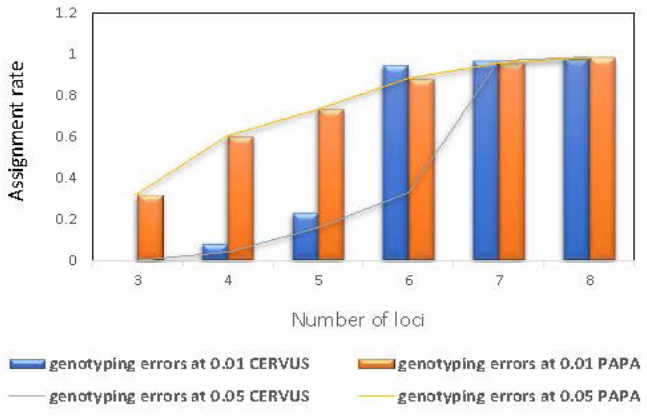

Figure 4. Parentage assignment of the actual offspring based on the likelihoods of each potential parent pair using CERVUS and PAPA.

DISCUSSION

Based on the comparison of actual assignments versus recorded pedigrees, findings from the present study showed that 100% of the offspring matched their recorded or true parents when one true family was analyzed and 95.65% of offspring matched both true parents, with a 100% match to the half parents when a set of 5 families was analyzed together. These results suggest that the DNA parentage tests that are widely adapted by court or other organizations as legal evidence for child support, inheritance, social welfare benefits, and adoption purposes should be very reliable as long as the right methods are used, since these tests only involve one family in most cases and since the principles and methods for human and animal parentage tests are similar.

Developing and selecting the best panel of genetic markers is critical when studying breeding strategies in agricultural animals. Results showed that 75% of assignments matched both true parents and 91.67% matched half true parents when a set of 10 families was analyzed using a panel of 8 microsatellite markers. This result is acceptable for breeding programs to avoid inbreeding in agricultural animals. However, only 47.25 % of offspring matched their true parents and 71.43 % matched their half true parents when a group of 20 families were evaluated, suggesting that the results from parentage analyses of large populations of nature or ecological systems can only be used as a reference

Table 2. Assignment rate for actual (simulation) assignment with different numbers of microsatellites using CERVUS at error rates 0.01 and 0.05

No. loci	Genotyping error=0.01			Genotyping error=0.05		
	Father	Mother	Both Parents	Father	Mother	Both parents
3	3% (5%)	3% (8%)	0% (1%)	4% (4%)	6% (6%)	1% (1%)
4	19% (26%)	16% (37%)	8% (33%)	14% (12%)	11% (19%)	4% (12%)
5	35% (51%)	36% (71%)	23% (80%)	31% (36%)	26% (45%)	16% (47%)
6	54% (88%)	98% (100%)	95% (99%)	40% (65%)	48% (80%)	35% (85%)
7	98% (100%)	99% (100%)	97% (100%)	63% (90%)	92% (100%)	97% (100%)
8	99% (100%)	99% (100%)	100%(100%)	73% (95%)	99% (100%)	99% (100%)

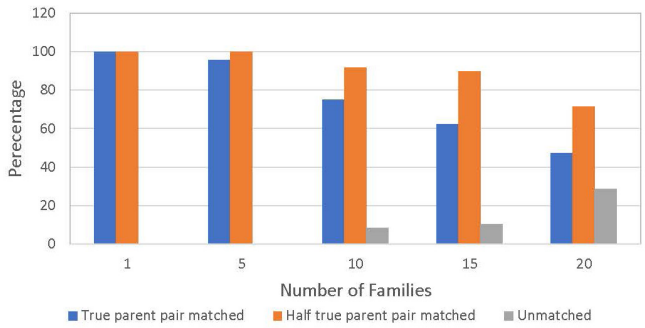

Figure 5. True parent match rates of actual assignment (DNA marker-based pedigrees) comparing recorded pedigree.

for population management.

Many factors affect the accuracy of parentage analysis, such as the type and number of genetic markers, computing software, and population or family size. Microsatellite DNA marker-based methods allow for accurate parentage analysis in human and animal species (6, 17). In order to develop a powerful microsatellite parentage tool, some properties of microsatellites with relevant influence in parentage assignment, such as polymorphism, which is usually measured by number of alleles, expected heterozygosity, which is a fundamental measure of genetic variation in a population, and polymorphism information content, which is used to measure the informativeness of a genetic marker, and distribution of alleles and null allele (i.e. a nonfunctional allele caused by a genetic mutation) frequency, should be checked (6).

The feasibility of parentage analysis of animal species using 5-6 microsatellites has been demonstrated in several studies (6, 8, 11). In this study, 3-8 loci were selected as the microsatellite DNA marker panel for the evaluation of parentage analysis, and high percentage rates were observed when 6-8 microsatellite loci were used. These loci showed a high exclusion potential and an appropriate technical resolution in both the computer simulation and the actual assignment with real offspring. The results were supported by the comparative analysis of the actual assignment with a known pedigree. This was shown by the high indices in panel of YP49, YP109, YP96, YP60, YP30, YP73, YP41, and YP78, indicating that this panel is powerful for parentage analysis in this model species.

Null alleles usually constitute a source of incompatibilities in microsatellite parentage analysis (11, 18-19). Several studies suggested that a widespread presence of null alleles seemed to be one characteristic in the analysis (19). For example, in three mapping families of the Pacific oyster, 51% of microsatellite loci contain at least one null allele (20). Identification of null alleles is critical since frequencies above 5% are considered to compromise pedigree inference (11, 15). In the present study, the null allele frequencies of all 8 loci estimated by CERVUS in broodstocks based on heterozygote deficiency were below the 5% threshold (**Table 1**). This may explain the increased accuracy (almost 100% assignment) when many families were simulated.

The accuracy of parentage analysis is also affected by different computer programs. Findings showed that there were slight differences in the actual assignment rates between CERVUS and PAPA because of differences in their statistical approaches. CERVUS is a parental allocation program (15), which is the choice of candidate parents of the same sex, and the other sex are represented by a single known individual. PAPA is a parental pair allocation program (17) and allocation on this basis involves choosing among putative parents of both sexes.

In conclusion, the general concept of parentage analysis is to match candidate parents with offspring depending on their genetic similarity. Microsatellite DNA marker-based methods are an ideal tool for accurate parentage analysis in human and animal species. Current parentage tests have been found to be reliable, since 100% of assigned offspring match their true or recorded parents when analysis only involves one family. The molecular pedigree from parentage analysis of 10 to 15 families is acceptable for breeding programs to avoid inbreeding in agricultural animals, since approximately 75% of assigned offspring could match both true parents and 92% could match the half true parents when the families were analyzed using 8 microsatellite markers. However, the results from parentage analyses of large populations should only be used as a reference for population reference, as in an evaluation of 20 families, only 47.25% offspring matched their true parents and 71.43% matched their half true parents.

MATERALS AND METHODS

Source of biological materials

The biological materials used in this study were acquired from The Ohio State University's aquatic genetics lab. The fish species yellow perch was used as the model system for this study. A total of 50 sets of mating were designed and the female in each set was mated to 1 to 4 males, which made for a maximum of 175 potential families. A fin clip of each parent and their 5-day-old progeny were sampled and stored in 95% ethanol at -80°C prior to DNA analysis. The genetic relationship between all parents and progeny were recorded as a physical pedigree.

Microsatellite markers and genotyping

An eight microsatellite marker panel (YP49, YP109, YP96, YP60, YP30, YP73, YP41, and YP78) developed and optimized by Li *et al* (21) was used and evaluated for this study. For amplification, the PCR conditions of these microsatellite DNA markers were re-optimized by temperature gradient PCRs, and thermal cycling was performed in the Temperature-Gradient Thermal Cycler System (PTC 2000, Bio-Rad). The markers with superior amplification and clear background were adopted for further analysis. Additionally, the parents were used to calculate the genetic parameter of polymorphic information content for marker selection. The

markers with PIC values of higher than 0.6, were selected for further analysis to choose the best marker panel.

Total genomic DNA was extracted from fin tissue (parents) or whole body (progeny) according to the methods described by Li *et al* (21) based on 96-well plate model. Amplification of microsatellite loci was performed with three primers in PTC-200 thermal cyclers (MJ Research) using an initial denaturation at 94°C for 2 min, followed by 35 cycles of 30 sec denaturation at 94°C, 30 sec annealing at a locus-specific temperature, 30 sec extension at 72°C, and a final 5 min extension at 72°C (13). Amplification products were genotyped using an ABI 3130 DNA genetic analyzer and the results were analyzed using Genemap(R) 4.0 software.

Computer simulation and parentage assignment

Two computer programs, CERVUS 3.07 (15) and PAPA 2.0 (16) were used for simulation, parentage assignment, and related evaluation. Simulation of assignment success was based on genotypes of 10,000 pseudo-offspring at varying amounts of microsatellite loci with an error rate of 0.01. 10 to 200 families were simulated and calculated based on likelihood of parental pair. Microsatellite profiles from 8 loci selected were used to identify the most likely parents among the 175 potential parents for 465 offspring in 50 mating sets. 5 larvae were randomly selected from each sampled family for microsatellite genotyping. For the statistical analysis, critical values with confidence intervals at 80% and 95% were generated by 100,000 cycles of bootstrapping from the allele frequencies of the parents, and a default error rate of 1% was used, assuming that all possible parents were sampled (9). All parental individuals were included as putative candidate parents. Additionally, genetic diversity (PIC and number of alleles) were also calculated from allele frequencies of the parents using CERVUS 3.03.

The offspring assigned by computer software were visually checked against the corresponding recorded physical pedigree, and microsatellite DNA marker-based pedigrees were compared with the physical pedigrees to evaluate the accuracy of the DNA parentage analyses.

ACKNOWLEDGMENTS

This study was supported by the National Institute of Food and Agriculture (NIFA), U.S. Department of Agriculture, under Agreement No. 2010-38879-20946. The authors wish to thank Paul O'Bryant and Dean Rapp for their assistance in fish breeding and pedigree recording.

Received: December 17, 2019
Accepted: June 7, 2020
Published: July 13, 2020

REFERENCES

1. Sefc, Kritsina M., and Stephan Koblmüller. "Assessing Parent Numbers from Offspring Genotypes: The Importance of Marker Polymorphism." *Journal of Heredity*, vol. 100, no. 2, 2008, pp. 197–205.
2. Phillips, Andelka. "Only a click away—DTC genetics for ancestry, health, love...and more: A view of the business and regulatory landscape." *Applied & Translational Genomic*, vol. 8, 2016, pp. 16–22.
3. Yu, Neng *et al*. "Disputed Maternity Leading to Identification of Tetragametic Chimerism." *New England Journal of Medicine*, vol. 346, no. 20, 2002, pp. 1545–1552.
4. Herbinger, Christophe M. *et al*. "Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks." *Aquaculture*, vol. 173, no. 1-4, 1999, pp. 105–116.
5. Gilk, Sara E. *et al*. "Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhyncus gorbuscha*, populations: marine survival, homing ability, and variability in family size." *Enviromental Biology of Fishes*, vol. 69, no. 1, 2004, pp. 287–297.
6. Castro, Jaime *et al*. "A microsatellite marker tool for parentage assessment in gilthead seabream (*Sparus aurata*)." *Aquaculture*, vol. 272, Supp. 1, 2007, pp. 210–216.
7. Gray, Andrew K. *et al*. "Unanticipated departures from breeding designs can be detected using microsatellite DNA parentage analyses." *Aquaculture*, vol. 280, no. 1-4, 2008, pp. 71–75.
8. Jackson, Timothy R. *et al*. "Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock." *Aquaculture,* vol. 220, no. 1-4, 2003, pp. 245–259.
9. Hayes, Ben *et al*. "Evaluation of three strategies using DNA markers for traceability in aquaculture species." *Aquaculture*, vol. 250, no. 1-2, 2005, pp. 70–81.
10. Porta, Javier *et al*. Genetic structure and genetic relatedness of a hatchery stock of Senegal sole (*Solea senegalensis*) inferred by microsatellites. *Aquaculture*, vol. 251, no. 1, 2006, pp. 46–55.
11. Castro, Jaime *et al*. "A microsatellite marker tool for parentage assessment in Senegal sole (*Solea senegalensis*): genotyping errors, null alleles and conformance to theoretical assumptions." *Aquaculture*, vol. 261, no. 4, 2006, pp. 1194–1203.
12. Norris, Ashie *et al*. "Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers." *Aquaculture*, vol. 182, no. 1-2, 2000, pp. 73–83.
13. Jones, Adam *et al*. "A practical guide to methods of parentage analysis: technical review" *Molecular Ecology Resources*, vol. 10, no. 1, 2010, pp. 6–30.
14. Wenk, Robert *et al*. "Empowering sibship analyses with reference pedigrees." *Transfusion*, vol. 52, no. 12, 2012, pp. 2614 – 2619.
15. Marshall, Tristan *et al*. "Statistical confidence for likelihood-based paternity inference in natural populations." *Molecular Ecology*, Vol. 7, no. 5, 1998, pp. 639–655.
16. Duchesne, Pierre *et al*. "PAPA (package for the analysis

of parental allocation): a computer program for simulated and real parental allocation." *Molecular Ecology Notes,* vol. 2, no. 2, 2002, pp. 191 – 193.
17. Liu, Zhanjiang *et al*. "DNA marker technologies and their applications in aquaculture genetics." *Aquaculture*, vol. 238, no. 1-4, 2004, pp. 1–37.
18. Callen, David *et al*. "Incidence of null alleles in the (AC) n microsatellite markers." *American Journal of Human Genetics*, vol. 52, no. 5, 1993, pp. 922–927.
19. Hedgecock, Dennis *et al.* "Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster *Crassostrea gigas*." *Journal of Shellfish Research,* vol. 23, no. 2, 2004, pp. 379-385.
20. Hubert, Sophie *et al*. "Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*." *Genetics*, vol. 168, no. 1, 2004, pp. 351–362.
21. Li, Li. *et al*. "Isolation and characterization of microsatellites in yellow perch (*Perca flavescens*)." *Molecular Ecology Notes*, vol. 7, no. 4, 2007, pp. 600-603.