Journal of *Marine Science and Engineering*

MDPI

Article

Effectiveness of SNPs for Sibship Assignment in Farmed Banana Shrimp (*Penaeus merguiensis*)

Chontida Phuthaworn [1,2,*], Nguyen Hong Nguyen [3,*] and Wayne Knibb [1,†]

[1] Centre for Bioinnovation, University of the Sunshine Coast, Locked Bag 4, Maroochydore DC, QLD 4558, Australia; wknibb@usc.edu.au
[2] Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngamwongwan Rd, Chatuchak, Bangkok 10900, Thailand
[3] School of Science, Engineering and Technology, University of the Sunshine Coast, Locked Bag 4, Maroochydore DC, QLD 4558, Australia
* Correspondence: chontida.phut@ku.th (C.P.); nnguyen@usc.edu.au (N.H.N.)
† Retired.

Abstract: Pedigrees are essential components in selective breeding programs to manage genetic diversity and obtain accurate genetic parameter estimates to ensure long-term response to selection in captive populations. High throughput and cost-effective sequencing technologies has offered opportunities of using single nucleotide polymorphisms (SNPs) to resolve penaeid shrimp pedigrees from mass spawning cohorts and communal rearing. Effects of SNPs for sibship assignment were investigated on 546 shrimp using two software programs, Colony and Sequoia. Assignment rates and accuracies using SNP subsets with six different minor allele frequencies (MAFs), four sets of SNPs, and five genotyping error rates were compared to the microsatellite-based pedigree established in a previous study. High MAFs and numbers of SNPs contributed to significant increases in assignment rates and accuracies, whereas genotyping error rates showed negligible impacts on assignment results. Sibship assignments achieved rates and accuracies of 98% and 83%, respectively, with a minimum number of 91 SNPs (average MAF $\geq$ 0.14), and the two different programs exhibited similar resulting patterns for different SNP subsets. High consistencies between SNP-based and microsatellite-based pedigrees showed that accurate pedigrees could be achieved by using SNPs and thus contribute to the long-term response to selection in farmed banana shrimp.

Keywords: parentage assignment; SNP; selective breeding; banana shrimp; colony; sequoia

Citation: Phuthaworn, C.; Nguyen, N.H.; Knibb, W. Effectiveness of SNPs for Sibship Assignment in Farmed Banana Shrimp (*Penaeus merguiensis*). *J. Mar. Sci. Eng.* **2023**, *11*, 1281. https://doi.org/10.3390/jmse11071281

Academic Editor: Ka Hou Chu

Received: 16 May 2023
Revised: 21 June 2023
Accepted: 21 June 2023
Published: 24 June 2023

1. Introduction

Shrimp farming is one of the fast-growing sectors in aquaculture. Selective breeding has been practiced to produce genetically improved stocks for shrimp species and selection requires accurate pedigree information to manage inbreeding, as well as to obtain accurate genetic parameters to maximize genetic responses [1].

In genetic improvement programs for aquaculture species, pedigrees, apart from controlled crosses and separate rearing, can be maintained using physical tags such as passive integrated transponder (PIT) and visible implant elastomer (VIE) tags, or genetic markers, such as DNA microsatellites. Physical tagging, however, only works successfully with a high survival rate and low tag loss once animals reach a certain size or weight as tiny larvae cannot be tagged [2,3]. Physical tagging often incurs high costs and requires intensive labor to maintain animals in separate rearing facilities until they reach a suitable size for tagging, hence inducing significant full-sib family or tank effects [4,5]. The number of families in a given pedigree for breeding programs, would thus be limited by the availability of rearing facilities [6] and in the long term this may result in increased inbreeding depression if insufficient families are bred each generation. The limitations of physical tagging may be resolved by using genetic markers (biological tags) to infer relatedness between individuals.

An advantage of the genetic-based pedigree is that animals can be reared communally soon after birth to offer a uniform culture environment for all individuals in a population, thereby increasing the accuracy of genetic parameter estimations while reducing the operational and labor costs compared to separate family rearing required with physical tagging [7].

Pedigree reconstructions in penaeid shrimp have used microsatellites due to their high polymorphism and accuracy to resolve genealogical relationships between individuals and has been commonly used for parentage assignment in framed shrimp [8,9]. However, the development of microsatellites is rather costly and time-consuming; microsatellites need to be developed de novo for species newly examined as their locations are usually within non-coding regions where the nucleotide substitution rates are high [10]. Microsatellites also suffer from reproducibility between studies due to their high polymorphisms [11]. Using the same primers to amplify the same microsatellites from the same species cannot guarantee the same polymorphism results; it gives a complication which limit transferable data between studies and requires substantial effort to standardize genotyping results. In recent years, however, the reduction in genome sequencing cost has generated an interest in applying SNP genotyping in human and plant genetics, and more recently in aquaculture [12,13]. Many studies have shown the effectiveness of SNPs for parentage assignment in wild fish population [14,15] and aquaculture species [16–18] as well as for brood-stock differentiation [19] and traceability [20]. Effective applications of SNPs have led to the recent development of SNP arrays in high-value farmed species, including Atlantic salmon [21], rainbow trout [22] and Pacific white shrimp [23,24]. In farmed shrimp, SNPs have been reported to show better performances than microsatellites for parentage assignment in black tiger shrimp (*Penaeus monodon*) [18] and Pacific white shrimp (*Penaeus vannamei*) [9]; however, comparisons between microsatellite and SNP markers for parentage assignment in banana shrimp remained unknown.

Thus, this study aimed to investigate effects of the (i) minor allele frequency of SNPs, (ii) numbers of SNPs, and (ii) genotyping error rates on assignment performance in farmed banana shrimp population.

2. Materials and Methods

2.1. Microsatellite-Based Pedigree

A pedigree of 562 shrimp comprising 48 full-sib families with a family size from 4 to 13 offspring was considered in the present study (Supplementary Table S1) and was part of a pedigree previously established using 10 DNA microsatellites for 1957 shrimp of sixty full-sib families [25,26]. The microsatellite-based pedigree was constructed using Colony version 2.0.5.0 [27] with the confidence probability of at least 95%. Shrimp were cultured communally in the same grow-out pond without physical tagging and were harvested after 140 days of growing out. In previous studies using other banana shrimp datasets, it was previously demonstrated that the family configuration of banana shrimp pedigrees using the same DNA microsatellite markers were highly consistent with maternal lines indicated by mitochondrial DNA (mtDNA) haplotypes [28]. The microsatellite-based pedigree was thus regarded as a standard configuration to which the full-sib pedigrees using SNPs were compared for assignment rates and accuracies.

2.2. Single Nucleotide Polymorphism (SNP)

The hepatopancreas samples of 562 banana shrimp were outsourced to Diversity Array Technology Pty Ltd. (DArT) in Canberra, Australia for genomic DNA extraction and SNP genotyping based on the genome complexity reduction method (DArTseq). The DArTseq protocol used for our study is similar to the one previously given in [29]. Individual SNPs were first filtered out if SNP call rate was less than 95% (i.e., SNP genotyped in less than 95% of individuals) and deviated from Hardy-Weinberg equilibrium (HWE) based on chi-square test at false discovery rate (FDR) less than 0.01. Remaining SNPs were then checked on individual samples for sample call rate at 95% (i.e., sample was genotyped with at least 95% of the remaining SNPs); average autosomal heterozygosity (i.e., an average

heterozygous genotypic frequency across SNP loci; significantly higher than a population mean, FDR < 0.01, may indicate sample contaminations) and identical-by-state (IBS) less than 95%, i.e., average kinship coefficients between two individuals across SNP loci; high IBS may indicate sample duplications. All quality checks were conducted using the R package 'GenABEL' [30]. The SNPs after quality checks were then screened for high linkage disequilibrium (LD) using the PriorityPruner version 0.1.4 and discarded if a squared correlation coefficient between SNPs (r^2; a statistical measure of the strength of association between two SNPs) was greater than 0.5, which indicated non-random segregation between SNP loci.

2.3. Sibship Assignment

Sibship assignment was conducted by two programs, Colony version 2.0.6.3, which improved computational efficiency for large datasets [27] and the R package 'Sequoia' [31].

2.3.1. Colony

Sibship assignments in Colony were conducted using the full-likelihood (FL) method with high precision assuming monogamy for both male and female. An assumption for monogamous mating was based on our previous study in the same banana shrimp population, from which only one sire genotype from spermatophore was observed for each dam and sibship groups and dam-offspring groups showed almost the same mtDNA haplotypes when pedigree assignments assumed monogamy using microsatellites [28]. Sibship assignments were completed with confidence probability of at least 95%.

2.3.2. Sequoia

The R package 'Sequoia' was designed particularly for parentage and sibship assignment using SNP data by using the likelihood ratio (LR) method to infer relationships between animals. Sibship assignments were conducted assuming monogamous mating with confidence probability estimated following the simulation method (repeated 10 times) in the Appendix S2 of [31].

Colony and Sequoia were both used for sibship assignment on SNP datasets with six different MAFs (0.02, 0.07, 0.14, 0.24, 0.35, and 0.45), four subsets of SNPs (50, 91, 150, and 200), and five genotyping error rates (0.0001, 0.001, 0.01, 0.05, and 0.1) (Supplementary Table S2). In this study, assignment rate was the proportion (%) of total offspring that were assigned to full-sib families; assignment accuracy was the proportion (%) of the assigned offspring to the full-sib families that were consistent with those of the microsatellite-based pedigree.

3. Results

3.1. Descriptive Statistics of SNP Data

The SNP data contained a total of 9472 genotyped SNPs from 562 individuals and were filtered according to the above criteria outlined in the methods section. After all quality checks, a total number of 2757 SNPs from 546 individuals remained before they were categorized into 6 SNP datasets based on MAF; the number of SNPs of each dataset ranged from 91 SNPs (average MAF = 0.45) to 1409 SNPs (average MAF = 0.02) (Table 1). SNP datasets had average heterozygosity ranging from 0.05 to 0.48 and average pairwise correlation between SNPs (r^2) close to zero, i.e., SNPs were not in high linkage disequilibrium, indicating that most of SNPs were independent from one another. Each of SNP datasets (in Table 1) was further subjected to random sampling for four subsets based on the number of SNPs (N = 50, 91, 150, 200); the four SNP subsets had the same average MAF as their original SNP datasets in Table 1.

Table 1. Descriptive parameters of SNP datasets.

Dataset †	MAF int ‡	N_{SNP}	Avg MAF	Avg Het §	Min r^2	Avg r^2	Max r^2
MAF0.02	<0.05	1409	0.02	0.05	0.0000	0.0015	0.4847
MAF0.07	0.05–≤0.1	608	0.07	0.13	0.0000	0.0026	0.2477
MAF0.14	>0.1–≤0.2	361	0.14	0.24	0.0000	0.0028	0.1176
MAF0.24	>0.2–≤0.3	172	0.24	0.36	0.0000	0.0030	0.0696
MAF0.35	>0.3–≤0.4	116	0.35	0.45	0.0000	0.0060	0.0698
MAF0.45	>0.4–≤0.5	91	0.45	0.48	0.0000	0.0110	0.1057

† SNP dataset: named based on an average MAF (Avg MAF); ‡ MAF interval for SNP dataset; § Observed heterozygosity: heterozygous genotypic frequency per SNP locus.

3.2. Sibship Assignments Using SNPs

3.2.1. Variations in Minor Allele Frequency (MAF)

MAF of SNPs (0.02, 0.07, 0.14, 0.24, 0.35, and 0.45) had significant impacts on assignment rate and accuracy in this banana shrimp population (Table 2). The assignment rates were higher than 98% for all six levels of MAF SNP datasets, and most with assignment confidences of at least 98% in Sequoia, except for the MAF0.02 dataset of which the assignment rate was achieved with confidence probability less than 95% (Table 2). An increase in assignment accuracy was observed when using SNPs with a higher MAF, and the number of full-sib families become closer to that of the 48 full-sib families of the microsatellite-based pedigree (Table 2).

Table 2. Assignment results using SNP datasets with different minor allele frequencies and the number of SNPs (N_{SNP}) in Colony and Sequoia.

N_{SNP}	Assignment	Colony						Sequoia					
		SNP Dataset						SNP Dataset					
		MAF0.02	MAF0.07	MAF0.14	MAF0.24	MAF0.35	MAF0.45	MAF0.02	MAF0.07	MAF0.14	MAF0.24	MAF0.35	MAF0.45
50													
	Rate (%)	100	100	100	100	100	100	† 97.62	†99.63	† 99.08	98.53	98.53	98.90
	Accuracy (%)	19.78	52.56	72.71	80.04	81.32	80.40	33.58	54.41	78.84	80.48	82.16	83.24
	# Full-sib families	436	169	95	70	65	70	52	52	49	49	45	50
91													
	Rate (%)	100	100	100	100	100	100	† 98.72	99.45	98.90	99.08	99.27	99.27
	Accuracy (%)	45.79	79.30	83.88	84.62	84.62	84.62	44.53	79.74	84.81	85.03	85.42	85.61
	# Full-sib families	205	65	51	49	49	49	40	45	46	45	46	45
150													
	Rate (%)	100	100	100	100	-	-	† 87.36	99.27	99.08	99.08	-	-
	Accuracy (%)	61.72	84.62	84.62	84.62	-	-	66.25	84.50	85.21	85.21	-	-
	# Full-sib families	127	49	49	49	-	-	39	45	45	45	-	-
200													
	Rate (%)	-	-	100	-	-	-	-	-	99.27	-	-	-
	Accuracy (%)	-	-	84.62	-	-	-	-	-	85.24	-	-	-
	# Full-sib families	-	-	49	-	-	-	-	-	45	-	-	-

Genotyping error rate at 0.05 per locus; † sibship assignment with confidence probability below 95%.

3.2.2. Variations in SNP Number

Four SNP subsets (n = 50, 91, 150, and 200 SNPs) were used for sibship assignment. Every level of SNP number was tested for all six levels of MAFs (Table 2). Assignment rates were 100% for all SNP datasets in Colony, while ranging between 87–99% in Sequoia (Table 2). At a given MAF, assignment accuracies increased with an increasing number of SNPs, and this trend was similarly observed in both Colony and Sequoia (Table 2). When a subset of fifty SNPs was used for sibship assignments, the number of full-sib families deviated from that of the microsatellite-based pedigree for all six levels of MAFs. The discrepancies in number of full-sib families, however, were improved by using the SNPs with the higher MAF (Table 2). We did not report assignment results for subsets of 200 SNPs (MAF0.02 and MAF0.07) because Colony and Sequoia considered offspring

as duplicated samples due to their identical genotypes in many SNP loci, and eventually stopped the analyses.

3.2.3. Variations in Genotyping Error Rate

Assignment rates were 97–100% in Colony and Sequoia, irrespective of assumed genotyping error rates (Table 3). Assignment accuracies increased slightly when genotyping errors increased from 0.0001 to 0.01 per locus and plateaued thereafter (Table 3). The number of assigned full-sib families tended to reduce with increasing error rates, less deviated from the 48 full-sib families of the pedigree derived from microsatellites markers (Table 3). However, the assignment accuracies and the number of assigned full-sib families showed only slight differences between levels of genotyping error rates when a large number of SNPs (200 SNPs) was used (Table 3).

Table 3. Assignment results using SNPs with different genotyping error rates in Colony and Sequoia (average MAF = 0.14).

N_{SNP}	Assignment	Colony					Sequoia				
		Genotyping Error Rate					Genotyping Error Rate				
		0.0001	0.001	0.01	0.05	0.1	0.0001	0.001	0.01	0.05	0.1
91											
	Rate (%)	100	100	100	100	100	97.25	97.25	97.99	98.90	99.27
	Accuracy (%)	78.94	81.50	83.33	83.88	83.70	78.91	79.66	83.93	84.81	83.95
	# Full-sib families	77	69	57	51	51	60	58	48	46	46
200											
	Rate (%)	100	100	100	100	100	97.99	98.35	99.08	99.27	99.82
	Accuracy (%)	83.15	83.88	84.62	84.62	84.62	81.87	84.36	85.21	85.24	84.59
	# Full-sib families	56	53	49	49	47	52	47	45	45	46

3.2.4. Variations in Parentage Assignment Programs

Colony and Sequoia produced assignment results in similar trends for most of SNP properties. Nevertheless, using either SNPs with low MAF (<0.14) or low number of SNPs (<91) in sibship assignment clearly reduced assignment accuracies regardless of the software used. It was particularly affecting the performance of Colony in assigning full-sib families, which were greatly deviated from the family configuration of the microsatellite-based pedigree; the confidence probabilities of sibship assignments in Sequoia were all below 95% (Table 2).

When comparing assignment results between Colony and Sequoia instead of comparing to the microsatellite-based pedigree, SNPs with low MAF (<0.14), low number (< 91), or low genotyping error rates (0.0001 and 0.001) remarkably reduced family configuration consistencies between the two software programs to less than 95% (Table 4). These consistencies, however, were improved by using SNPs with higher MAF (≥0.14), larger number (≥91), or higher genotyping error rates (0.01, 0.05 and 0.1) (Table 4).

Table 4. Assignment consistencies (%) between Colony and Sequoia.

N_{SNP}	Colony vs. Sequoia						Colony vs. Sequoia				
	SNP Dataset						Genotyping Error Rate				
	MAF0.02	MAF0.07	MAF0.14	MAF0.24	MAF0.35	MAF0.45	0.0001	0.001	0.01	0.05	0.1
50	5.13	48.1	83.97	89.74	91.03	91.58	-	-	-	-	-
91	33.15	89.93	97.80	98.90	98.72	99.27	87.91	90.29	96.15	97.62	97.80
150	62.82	93.77	98.90	99.08	-	-	-	-	-	-	-
200	-	-	-	99.27	-	-	94.14	97.80	99.08	99.27	99.27

4. Discussion

4.1. MAF and SNP Number for Sibship Assignment

Our study showed that efficiency of sibship assignment was improved when datasets with high number of SNPs exhibiting high MAF were used. Our results agreed with those from a simulation study in Chinook salmon (*Oncorhynchus tshawytscha*) where SNPs with high MAF were expected to have a higher power than those with low MAF for parentage assignment [32]. However, using high-MAF SNPs does not always guarantee the best assignment performance unless an optimal number of SNPs is used, i.e., a number of SNPs to obtain the highest assignment rate and accuracy. As shown in our results, the SNP dataset with the highest MAF average of 0.45 (MAF0.45), performed less efficient at 50 SNPs in comparison to that of 91 SNPs at the same MAF. This may be because lower numbers of SNPs reduce the total power of SNP dataset to assess relationships between individuals, particularly for closely related individuals, i.e., full sibs. In addition, more SNPs are required to achieve an accurate pedigree inference if the number of potential parents and offspring (sibling groups) are high [32]; this conclusion is consistent with the simulation study in Pacific oyster (*Crassostrea gigas*) that showed more than 55 polymorphic SNPs (average observed heterozygosity = 0.284, sires = 6, dams = 6) would be required to obtain assignment accuracy over 85%, if more than 25 families presented in parentage analysis [16]. Indeed, sibship assignment in this study required 91 SNPs (MAF $\geq$ 0.14) to achieve 83% assignment accuracy for the 48 full-sib families determined by DNA microsatellites. The assignment accuracy of 83% here represented the concordance between assignment results based on SNP and microsatellite markers which could also imply that there was the discrepancy of 17% between the assignment results of the two markers.

The mismatches between assignment results from SNP and microsatellite markers may be due to missing parental genotypes. Lacking parental information may shifted pedigree inferences to solely rely on statistical estimations for relatedness between offspring and confidence probability in sibship assignment. Full-sib pedigrees from SNP and microsatellite markers were both established with a confidence probability at 95% which implied that the pedigrees could be incorrectly inferred at 5% by chance, i.e., at maximum 5% of the offspring could be incorrectly assigned to the full-sib groups. Thus, the nature of stochastic errors from the statistical estimations may have contributed to the observed discrepancies between the pedigrees from the two markers in the present study. It was also suggested that parental information could dramatically improve pedigree inferences and always providing them in parentage analyses if available (Wang, 2004).

Using 60–100 SNPs was reported to be sufficient for parentage assignment in Chinook salmon [32]. It was estimated that 59–122 SNPs were sufficient for correct parentage assignment over 99% in black tiger shrimp [18] similar to the high assignment accuracy (>98%) when using 50 SNPs (MAF $\geq$ 0.3, 162 broodstock) for parentage assignment in Pacific white shrimp [9]. Perez-Enriquez and Max-Aguilar (2016) also demonstrated that using 76 SNPs for parentage assignment resulted in higher assignment rates than microsatellites (SSR), both in simulated data (SSR: 23–83%, SNP: 85–98%) and empirical data (SSR: 15–76%, SNP: 94–96%). It was reported that the minimum number of 36 SNPs (average MAF = 0.42, 85 sires, 162 dams) was able to correctly assign more than 92% of offspring to their true parents, and the assignment accuracy reached 100% when using 68 SNPs (average MAF = 0.37) in farmed rainbow trout (*Oncorhynchus mykiss*) [22]; a similar study showed that 72–188 polymorphic SNPs were able to obtain assignment accuracy of above 95% in farmed steelhead [33]. The effectiveness of SNPs for parentage assignment using small SNP subsets was also proved in Pacific white shrimp when 88 SNPs with an average MAF of 0.38 (44 sires, four dams) were able to completely assign offspring to correct sires and dams with a confidence higher than 95% [24]. High assignment accuracy, however, should not be the only criterion to assess effectiveness of SNPs for sibship assignment unless assignment rate has also been considered. Sibship assignment would not benefit much if the high accuracy (i.e., corrected assignment rate) was derived from a low assignment rate,

which would eventually reduce an overall assignment performance (assignment power) of the SNPs.

Our results showed that the minimum criteria of the SNP subset used for sibship assignment with assignment accuracy greater than 95% was 91 SNPs with an average MAF of 0.14; however, it was noteworthy that the assignment results using SNPs with an average of 0.07 (MAF0.07) were also on par with those of other SNP subsets when the number of SNPs reached 150 loci (Table 2). It could imply that the low MAF could still be used for sibship assignment, provided that there were sufficient numbers of SNPs. This was similarly suggested by the previous studies in Portuguese oysters (18 sires, 15 dams), wherein 1352 SNPs (minimum MAF = 0.1) were as informative as 400 SNPs (minimum MAF $\geq$ 0.3) for parentage assignment, with an accuracy higher than 98% [34].

4.2. Genotyping Error Rate for Sibship Assignment

At a given MAF and number of SNPs, overall, we found that genotyping error rates did not have major impacts on assignment rate, assignment accuracy, and the number of full-sib groups. Genotyping errors assumed for parentage assignment are generally 0.005–0.01 per locus for SNPs [17,22] and 0.01–0.05 per locus for microsatellites in mollusks and fish [8,9]. The actual genotyping error rates, which can be caused by scoring errors, allelic dropouts, mutations, or null alleles, were typically unknown unless SNPs or microsatellites were re-genotyped; typing errors estimated after re-genotyping have been reported between 0.0034–0.0044 for SNPs, and 0.02 for microsatellites in fish [14,33]. In this study, genotyping error rates were assumed based on those reported in aquaculture species as we did not re-genotype our SNPs. It is more preferable if actual genotyping errors were used for pedigree inferences; however, it may not be applicable when using genetic markers developed de novo for the population, i.e., no prior information on any error types that contribute to genotyping error rates. Thus, it is maybe better to assume a range of genotyping error rates and investigate the performance of SNPs for pedigree inferences in this situation.

4.3. Colony and Sequoia

Assignment results can vary due to differences in analytical methods (e.g., exclusion, pairwise, and likelihood), assumptions, and parameter settings. Both Colony and Sequoia apply the likelihood method for sibship analysis; however, Colony used the maximum likelihood algorithm, which reshuffled the family configuration based on a predefined repeat (iteration) for the final sibship configuration [35], whereas Sequoia instead used the log–likelihood ratio to infer the final sibship groups [31]. Different computation settings for confidence probability between Colony and Sequoia could affect assignment rates in this study. A colony, by default, would compute family configurations of at least 95% confidence, meaning that a family with only one member would be possible as long as there was a 95% confidence. Sequoia, on the other hand, set the confidence probability as floating; it simulated SNP genotypes based on MAF, SNP number, and genotyping error rates similar to that of the empirical SNP dataset, then compared pedigrees to that of the empirical SNP dataset before computing the confidence probability. These factors explain different assignment rates regardless of SNP properties in Colony (100%) but different values in Sequoia.

In this study, we compared assignment results produced from Colony and Sequoia to the microsatellite-based pedigree (Table 2) and compared the family configuration results between the two software programs (Table 4). High consistent family configurations (>95%) between Colony and Sequoia, when using $\geq$91 SNPs with high MAF ($\geq$0.14) and genotyping error rates at 0.01, 0.05 and 0.1, indicated the transferability of the SNP-based pedigrees. This means that the observed results and trends would be consistent regardless of conditions (i.e., software programs) when SNPs are used for sibship assignment.

Even though factors such as MAF, number, and genotyping error rates had impacts on the accuracy of SNP-based pedigrees and assignment consistencies between Colony and Sequoia, for this banana shrimp population, inbreeding may have also contributed to the

pedigree errors. We observed inbreeding coefficients above 1% in our shrimp population (around 7% of the 546 offspring). It was suggested that inbreeding coefficients exceeding 1% may affect the accuracy of molecular-based pedigrees [36]. Breeding closely related animals would eventually increase number of SNP loci with low MAF (i.e., uninformative SNPs), provided that the same SNP loci were genotyped. Low MAF indicates most individuals possessing homozygous genotypes for a given SNP locus, i.e., low the observed heterozygosity is. Homozygous genotypes did not add useful information to differentiate relatedness between individuals but rather introduce noise into assignment analyses, which would contribute to the pedigree errors.

For selective breeding perspectives, an inbreeding level greater than 1% reflects the suboptimal management on the breeding population, which could increase pedigree errors (i.e., reduced data accuracies for genetic parameter estimations) and contribute to reduced genetic gain in the long term. It was estimated that 5% of pedigree errors could cause 1–4% loss of genetic gain in cattle [37]. To avoid, or at least mitigate, such situations for this banana shrimp population, only outbred animals may be selected and mated with their unrelated counterparts guided by the SNP-based pedigree in order to maintain genetic diversity and minimizing inbreeding rate (<1% per generation) for the long-term breeding program in this population.

5. Conclusions

Sibship assignment rates and accuracies could be improved by using a high number of SNPs with high MAF. Genotyping error rates did not have major impacts on assignment results, especially when using a large number of SNPs. Our results showed that, using a minimum number of 91 SNPs with an average MAF of at least 0.14, we were able to achieve assignment rates and accuracies above 98% and 83%. This confirmed that sibship assignment using SNPs could effectively generate informative pedigrees and contribute to selective breeding programs in farmed banana shrimp.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jmse11071281/s1, Table S1: Composition of the 4 families; Table S2: Sibship assignments using SNPs with different minor allele frequencies, numbers and genotyping error rates.

Author Contributions: C.P. prepared tissue samples for partial genome sequencing, participated in the design, performed statistical analyses, and drafted the manuscript. N.H.N. participated in the design and data collection and contributed to the manuscript. W.K. conceived the study, participated in the design and data collection, and contributed to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available on request from the corresponding authors. The data are not publicly available due to privacy restrictions.

Acknowledgments: We acknowledge the support of the Genecology Research Centre via a Genecology HDR research grant to Chontida Phuthaworn (former at USC). We acknowledge the contributions from a Cooperative Research Centre grant (Australian Seafood CRC: Project No. 2009/724) to Wayne Knibb (former at USC), and Seafarm at Cardwell to the availability of banana shrimp samples. We thank Prof Abigail Elizur, Josephine Nocillado and Ngo Phu Thoa for their contributions to sample and data collection, Ido Bar for his advice during statistical analyses, and Diversity Array Technology Pty Ltd. (DArT) for their supportive services and generous offers on partial genome sequencing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nguyen, N.H.; Phuthaworn, C.; Knibb, W. Genomic prediction for disease resistance to Hepatopancreatic parvovirus and growth, carcass and quality traits in Banana shrimp Fenneropenaeus merguiensis. *Genomics* **2020**, *112*, 2021–2027. [CrossRef]
2. Godin, D.M.; Carr, W.H.; Hagino, G.; Segura, F.; Sweeney, J.N.; Blankenship, L. Evaluation of a fluorescent elastomer internal tag in juvenile and adult shrimp Penaeus vannamei. *Aquaculture* **1996**, *139*, 243–248. [CrossRef]

3. Soula, M.; Navarro, A.; Hildebrandt, S.; Zamorano, M.J.; Roo, J.; Hernández-Cruz, C.M.; Afonso, J.M. Evaluation of VIE (Visible Implant Elastomer) and PIT (Passive Integrated Transponder) physical tagging systems for the identification of red porgy fingerlings (*Pagrus pagrus*). *Aquac. Int.* **2012**, *20*, 571–583. [CrossRef]
4. Nguyen, N.H. A systematic review and meta-analysis of genetic parameters for complex quantitative traits in aquatic animal species. *bioRxiv* **2021**. [CrossRef]
5. Vu, N.T.; Phuc, T.H.; Nguyen, N.H.; Van Sang, N. Effects of common full-sib families on accuracy of genomic prediction for tagging weight in striped catfish Pangasianodon hypophthalmus. *Front. Genet.* **2023**, 13. [CrossRef]
6. Vandeputte, M.; Haffray, P. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* **2014**, *5*, 432. [CrossRef]
7. Ninh, N.H.; Ponzoni, R.W.; Nguyen, N.H.; Woolliams, J.A.; Taggart, J.B.; McAndrew, B.J.; Penman, D.J. A comparison of communal and separate rearing of families in selective breeding of common carp (*Cyprinus carpio*): Responses to selection. *Aquaculture* **2013**, *408–409*, 152–159. [CrossRef]
8. Dong, S.R.; Kong, J.; Zhang, T.S.; Meng, X.H.; Wang, R.C. Parentage determination of Chinese shrimp (Fenneropenaeus chinensis) based on microsatellite DNA markers. *Aquaculture* **2006**, *258*, 283–288. [CrossRef]
9. Perez-Enriquez, R.; Max-Aguilar, A. Pedigree traceability in whiteleg shrimp (Litopenaeus vannamei) using genetic markers: A comparison between microsatellites and SNPs. *Cienc. Mar.* **2016**, *42*, 227–235. [CrossRef]
10. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16. [CrossRef] [PubMed]
11. Seeb, J.E.; Carvalho, G.; Hauser, L.; Naish, K.; Roberts, S.; Seeb, L.W. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol. Ecol. Resour.* **2011**, *11*, 1–8. [CrossRef]
12. Boonyarit, H.; Mahasirimongkol, S.; Chavalvechakul, N.; Aoki, M.; Amitani, H.; Hosono, N.; Kamatani, N.; Kubo, M.; Lertrit, P. Development of a SNP set for human identification: A set with high powers of discrimination which yields high genetic information from naturally degraded DNA samples in the Thai population. *Forensic Sci. Int. Genet.* **2014**, *11*, 166–173. [CrossRef] [PubMed]
13. Zhao, Y.S.; Gowda, M.; Liu, W.X.; Würschum, T.; Maurer, H.P.; Longin, F.H.; Ranc, N.; Reif, J.C. Accuracy of genomic selection in European maize elite breeding populations. *Theor. Appl. Genet.* **2011**, *124*, 769–776. [CrossRef] [PubMed]
14. Hauser, L.; Baird, M.; Hilborn, R.A.Y.; Seeb, L.W.; Seeb, J.E. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Mol. Ecol. Resour.* **2011**, *11*, 150–161. [CrossRef] [PubMed]
15. Hess, J.E.; Campbell, N.R.; Docker, M.F.; Baker, C.; Jackson, A.; Lampman, R.; McIlraith, B.; Moser, M.L.; Statler, D.P.; Young, W.P.; et al. Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Mol. Ecol. Resour.* **2014**, *15*, 187–202. [CrossRef]
16. Jin, Y.-L.; Kong, L.-F.; Yu, H.; Li, Q. Development, inheritance and evaluation of 55 novel single nucleotide polymorphism markers for parentage assignment in the Pacific oyster (*Crassostrea gigas*). *Genes Genom.* **2014**, *36*, 129–141. [CrossRef]
17. Nguyen, T.T.; Hayes, B.J.; Ingram, B.A. Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture* **2014**, *420–421*, 295–301. [CrossRef]
18. Sellars, M.J.; Dierens, L.; McWilliam, S.; Little, B.; Murphy, B.; Coman, G.J.; Barendse, W.; Henshall, J. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquac. Res.* **2014**, *45*, 417–426. [CrossRef]
19. Liu, S.; Palti, Y.; Martin, K.E.; Parsons, J.E.; Rexroad, C.E. Assessment of genetic differentiation and genetic assignment of commercial rainbow trout strains using a SNP panel. *Aquaculture* **2017**, *468*, 120–125. [CrossRef]
20. Hess, J.E.; Matala, A.P.; Narum, S.R. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Mol. Ecol. Resour.* **2011**, *11*, 137–149. [CrossRef]
21. Holman, L.E.; de la Serrana, D.G.; Onoufriou, A.; Hillestad, B.; Johnston, I.A. A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture* **2017**, *476*, 59–64. [CrossRef]
22. Liu, S.; Palti, Y.; Gao, G.; Rexroad, C.E. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture* **2015**, *452*, 178–182. [CrossRef]
23. Garcia, B.F.; Bonaguro, Á.; Araya, C.; Carvalheiro, R.; Yáñez, J.M. Application of a novel 50K SNP genotyping array to assess the genetic diversity and linkage disequilibrium in a farmed Pacific white shrimp (*Litopenaeus vannamei*) population. *Aquac. Rep.* **2021**, *20*, 100691. [CrossRef]
24. Silva, N.M.L.; Ianella, P.; Yamagishi, M.E.B.; Rocha, J.L.; Teixeira, A.K.; Farias, F.G.; Guerrelhas, A.C.; Caetano, A.R. Development and validation of a low-density SNP panel for paternity and kinship analysis and evaluation of genetic variability and structure of commercial Pacific white shrimp (Litopenaeus vannamei) populations from Brazil. *Aquaculture* **2022**, *560*, 738540. [CrossRef]
25. Knibb, W.; Whatmore, P.; Lamont, R.; Quinn, J.; Powell, D.; Elizur, A.; Anderson, T.; Remilton, C.; Nguyen, N.H. Can genetic diversity be maintained in long term mass selected populations without pedigree information? —A case study using banana shrimp Fenneropenaeus merguiensis. *Aquaculture* **2014**, *428–429*, 71–78. [CrossRef]
26. Nguyen, N.H.; Quinn, J.; Powell, D.; Elizur, A.; Thoa, N.P.; Nocillado, J.; Lamont, R.; Remilton, C.; Knibb, W. Heritability for body colour and its genetic association with morphometric traits in Banana shrimp (*Fenneropenaeus merguiensis*). *BMC Genet.* **2014**, *15*, 132. [CrossRef]

27. Jones, O.R.; Wang, J. COLONY: A program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* **2010**, *10*, 551–555. [CrossRef]
28. Knibb, W.; Quinn, J.; Lamont, R.; Whatmore, P.; Nguyen, N.H.; Remilton, C. Reproductive behaviour of captive Fenneropenaeus merguiensis: Evidence for monogamy and high between family variances for offspring number. *Aquaculture* **2014**, *426–427*, 60–65. [CrossRef]
29. Nguyen, N.H.; Vu, N.T.; Patil, S.S.; Sandhu, K.S. Multivariate genomic prediction for commercial traits of economic importance in Banana shrimp Fenneropenaeus merguiensis. *Aquaculture* **2022**, *555*, 738229. [CrossRef]
30. Aulchenko, Y.S.; Ripke, S.; Isaacs, A.; Van Duijn, C.M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **2007**, *23*, 1294–1296. [CrossRef]
31. Huisman, J. Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.* **2017**, *17*, 1009–1024. [CrossRef] [PubMed]
32. Anderson, E.C.; Garza, J.C. The Power of Single-Nucleotide Polymorphisms for Large-Scale Parentage Inference. *Genetics* **2006**, *172*, 2567–2582. [CrossRef]
33. Steele, C.A.; Anderson, E.C.; Ackerman, M.W.; Hess, M.A.; Campbell, N.R.; Narum, S.R.; Campbell, M.R. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. *Can. J. Fish. Aquat. Sci.* **2013**, *70*, 1046–1054. [CrossRef]
34. Vu, S.V.; Premachandra, H.K.A.; O'connor, W.; Nguyen, N.T.H.; Dove, M.; Van Vu, I.; Le, T.S.; Vendrami, D.L.J.; Knibb, W. Development of SNP parentage assignment in the Portuguese oyster Crassostrea angulata. *Aquac. Rep.* **2021**, *19*, 100615. [CrossRef]
35. Wang, J.L. Sibship Reconstruction from Genetic Data with Typing Errors. *Genetics* **2004**, *166*, 1963–1979. [CrossRef]
36. Blonk, R.J.W.; Komen, J.; Kamstra, A.; Crooijmans, R.P.M.A.; van Arendonk, J.A.M. Levels of inbreeding in group mating captive broodstock populations of Common sole, (Solea solea), inferred from parental relatedness and contribution. *Aquaculture* **2009**, *289*, 26–31. [CrossRef]
37. García-Ruiz, A.; Wiggans, G.R.; Ruiz-López, F.J. Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. *J. Dairy Sci.* **2019**, *102*, 1806–1810. [CrossRef]