

RESEARCH ARTICLE

Open Access



Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*

Nguyen H. Nguyen^{1*} , H. K. A. Premachandra¹, Andrzej Kilian² and Wayne Knibb¹

Abstract

Background: Genomic prediction using Diversity Arrays Technology (DArT) genotype by sequencing platform has not been reported in yellowtail kingfish (*Seriola lalandi*). The principal aim of this study was to address this knowledge gap and to assess predictive ability of genomic Best Linear Unbiased Prediction (gBLUP) for traits of commercial importance in a yellowtail kingfish population comprising 752 individuals that had DNA sequence and phenotypic records for growth traits (body weight, fork length and condition index). The gBLUP method was used due to its computational efficiency and it showed similar predictive performance to other approaches, especially for traits whose variation is of polygenic nature, such as body traits analysed in this study. The accuracy or predictive ability of the gBLUP model was estimated for three growth traits: body weight, folk length and condition index.

Results: The prediction accuracy was moderate to high (0.44 to 0.69) for growth-related traits. The predictive ability for body weight increased by 17.0% (from 0.69 to 0.83) when missing genotype was imputed. Within population prediction using five-fold across validation approach showed that the gBLUP model performed well for growth traits (weight, length and condition factor), with the coefficient of determination (R^2) from linear regression analysis ranging from 0.49 to 0.71.

Conclusions: Collectively our results demonstrated, for the first time in yellowtail kingfish, the potential application of genomic selection for growth-related traits in the future breeding program for this species, *S. lalandi*.

Keywords: Kingfish, Genetic improvement, Genomic prediction, Genomic selection and genotype by sequencing

Background

Yellowtail kingfish (YTK) *Seriola lalandi* has emerged as an important marine finfish for aquaculture, not only in Australia but also in many other parts of the world including Japan, Latin America (Chile, Mexico), the United States and New Zealand [1]. In Australia, the hatchery and production technologies have been well established for this species, and a genetic improvement program has started since 2010, using the founder stocks collected from the wild in South Australia [2]. To date, the selection program for YTK has been practised solely for growth rate. Body condition of the fish was also assessed when selection decision was made because there is a demand for well conformed fish from overseas markets [3].

Genetic evaluation of the first two generations of family selection showed that there was a significant improvement in fry performance by approximately 8% per generation [2]. Performance testing of growers produced in the second generation (G2) in sea cages however showed lower magnitude of selection response when the selected fish were compared with wild stocks or with offspring of the G1 parents (Premachandra et al., submitted). The positive response to selection achieved in the present YTK population was consistent with those reported for other aquaculture species, with an average genetic gain ranging from 5 to 15% per generation [4–9]. In the YTK population of this study, moderate heritabilities ($h^2 = 0.15$ to 0.41) were estimated for body and carcass traits [2, 3].

Along with the conventional selective breeding approach based on phenotype and pedigree, we are interested in exploring potential use of molecular and genomic information to accelerate genetic gain in this line of YTK, especially

* Correspondence: NNguyen@usc.edu.au

¹The University of the Sunshine Coast, Maroochydcyre DC, QLD 4558, Australia

Full list of author information is available at the end of the article



improving traits that are difficult or expensive to measure or lowly heritable. To date, high density SNP array-based chips are not commercially available for yellowtail kingfish *S. lalandi*. Further, the cost of whole genome sequencing is still high, especially for non-model species. Genotyping by sequencing (GBS) has been used as an alternative to whole genome sequencing approach due to its low cost of genotyping per animal, large amount of high quality genetic markers and the suitability of the markers for genomic prediction/selection. Among the Restricted-site Associated DNA sequencing (RAD-seq) methods, DArTseq™ sequencing technology represents a combination of a DArT complexity reduction methods and next generation sequencing platforms [10–13]. Similar to DArT methods based on array hybridisations, the technology is optimized for each organism and application by selecting the most appropriate complexity reduction method (both the size of the representation and the fraction of a genome selected for assays) and it can generate a large number of markers at a reasonable price per sample (approx. 35–50 AU\$/sample by the time of this report). To the best of our knowledge, the DArTseq™ sequencing technology has not been reported for any aquatic species. However, many studies have applied variations of DArT genotyping by sequencing methods in aquaculture, including those used for genomic prediction. A synthesised result from the literature shows that there is potential for the application of genomic selection in salmonids [14, 15] or gilthead seabream [16]. To date, there is no published information whether DNA markers obtained from next generation sequencing platforms can be used to predict breeding values for complex traits in yellowtail kingfish (*S. lalandi*).

Therefore, the principal aim of this study was to investigate the predictive ability of genomic Best Linear Unbiased Prediction (gBLUP) method for commercial traits of economic importance in yellowtail kingfish (*Seriola lalandi*), using the DArT sequencing technology. It is the most common method used to perform genomic prediction for complex traits whose variation is of polygenic nature. Specifically, we report the gBLUP prediction for growth traits (body weight, fork length and condition index), using five-fold across validation conducted in the same population. The effects of the number of markers on the predictive ability of gBLUP model were also estimated.

Methods

Fish samples and phenotypic data

In this study, a total of 752 DNA samples of yellowtail kingfish (YTK) were sequenced using DArT-seq technology (see section “[Library preparation and DArT sequencing](#)”). The animals originated from a selective breeding program for YTK at Clean Sea Tuna Ltd. in South Australia. The fish samples analysed in this study were

offspring of 35 families produced from the first generation (G1) and wild brood stock parents. The average number of offspring per family was 17. Breeding was conducted in tank, comprising three males and three females. In total, 16 sires and 31 dams successfully produced offspring in this study. After a nursing/rearing period of about 120 days in tanks, fingerlings were transferred to culture in sea cages. When the fish reached an average body weight of 3 kg, they were harvested and then anaesthetised using clove oil (40 mg/L) and cold water before morphometric measurements and fin tissue collection. The morphometric and phenotypic measurements included body weight (W), fork length (L, measured from the tip of the snout to the end of the caudal fin rays). The condition index (factor) was calculated as $k = W/L^3$. Deformities [17] included a range of measures, namely deformed snout, water belly – a condition where the belly is distended, deformed tail, deformed operculum and lower jaw). Skin fluke is due to the monogenean fluke parasite *Benedenia seriola*; this fluke inhabits the skin and fins of *Seriola spp.* and feeds on mucus and epithelia cells. Both deformity and fluke were recorded as binary traits depending on their presence or absence on the body of the fish at harvest (~3 kg) and coded as 1 and zero, respectively. The incidence of skin fluke and deformity recorded under field condition in this population was low (4.3 and 17.6%, respectively) and hence, results from genomic prediction for these traits were not tabulated in this paper.

The fin caudal tissue sample was collected from caudal fin of each fish and stored in 80% ethanol for DNA extraction at a laboratory of the University of the Sunshine Coast (USC). Fish biometrics, fluke abundance and deformities were recorded before they were released back to the cage.

A detailed description of the population is given in our earlier studies [2, 3, 17].

DNA extraction, genotyping and parentage analysis

Fin tissue samples of a total of 1568 fish were used to extract total genomic DNA (gDNA) using DNeasy Blood and Tissue kits (Qiagen, Germany) and NucleoMag® 96 Tissue kit (MACHEREY-NAGEL GmbH & Co. KG, Germany). A panel of eight microsatellite markers that consisted of five newly developed candidates from *S. lalandi* transcriptome sequences (YTK002, YTK008, YTK011, YTK017 and YTK019) [3] and three loci (Sdu21, Sdu32 and Sdu46) validated from the literature [18, 19] were used to genotype the experiment fish for parentage analysis in this study. Genotyping the broodstock fish was completed in our previous study [2]. All of these markers were proved and validated to use with yellowtail kingfish previously [2, 3].

DNA amplification was achieved using Qiagen Multiplex PCR PLUS Kits (Qiagen, Germany) in 13.5 μ L reactions, each containing 1.25 μ L of 10 \times primer mix, 6.25 μ L of Multiplex PCR Master Mix, 2.75 μ L of RNase free water, 1.25 μ L of Q-Solution and 2.0 μ L of approximately 20 ng template gDNA. Amplification was performed using an Eppendorf Mastercycler nexus (Hamburg, Germany) with cycling conditions as follows: initial denaturation at 95 $^{\circ}$ C for 15 min, followed by 35 cycles of 95 $^{\circ}$ C for 30 s, 57 $^{\circ}$ C for 90 s, and 72 $^{\circ}$ C for 30 s; with a final extension at 68 $^{\circ}$ C for 30 min.

PCR products were separated by capillary electrophoresis on an AB 3500 Genetic Analyser (Applied Biosystems) at the University of the Sunshine Coast. Fragment sizes were determined relative to an internal lane standard (GS-600 LIZ; Applied Biosystems) using GENE-MARKER v1.95 software (SoftGenetics; State College, USA) and double-checked manually. Individuals with low or missing peaks were amplified/genotyped a second time and checked for evidence of large allele dropout, extreme stuttering and null alleles, based on 1000 bootstraps and a 95% confidence interval. Tests for HWE at each locus and genotypic linkage equilibrium among pairs of loci, numbers of alleles and the observed and expected heterozygosities of each locus were determined using GENALEX v6.5, while polymorphic information content (*PIC*) was computed in CERVUS v3.0. Parentage assignment was completed using COLONY software [20] with confidence scores of above 95%. The pedigree included 65 full-sib groups (16 dams and 31 sires), with the family size of 3 to 108 offspring. A total of 1568 offspring out of 1998 were assigned to full sib families and the family size. Based on this pedigree, large size families (35 full-sib families and averaging 17 fish per family) were chosen to send to DArT Ptd Ltd. in Canberra, Australia for sequencing.

Library preparation and DArT sequencing

Four methods of DArTseq™ complexity reduction were tested in *Seriola lalandi* (data not presented) and the PstI-SphI method was selected. DNA samples were processed in digestion/ligation reactions principally as per Kilian et al. [10] but replacing a single PstI-compatible adaptor with two different adaptors corresponding to two different Restriction Enzyme (RE) overhangs. The PstI-compatible adapter was designed to include Illumina flowcell attachment sequence, sequencing primer sequence and “staggered”, varying length barcode region, similar to the sequence reported by Elshire et al. [21]. Reverse adapter contained flowcell attachment region and SphI-compatible overhang sequence.

Only “mixed fragments” (PstI-SphI) were effectively amplified in 30 rounds of PCR using the following reaction

conditions: 94 $^{\circ}$ C for 1 min, then 30 cycles of 94 $^{\circ}$ C for 20 s, 58 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 45 s and 72 $^{\circ}$ C for 7 min.

After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on Illumina HiSeq2500. The sequencing (single read) was run for 77 cycles.

Sequences generated from each lane were processed using proprietary DArT analytical pipelines. In the primary pipeline the fastq files were first processed to filter away poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence (minimum Phred pass score = 30 and minimum pass percentage = 75). In that way the assignments of the sequences to specific samples carried in the “barcode split” step were very reliable. Approximately 2,500,000 sequences per barcode/sample were identified and used in marker calling. Finally, identical sequences were collapsed into “fastqcoll files”. The fastqcoll files were “groomed” using DArT PL’s proprietary algorithm which corrects low quality base from singleton tag into a correct base using collapsed tags with multiple members as a template. The “groomed” fastqcoll files were used in the secondary pipeline for DArT PL’s proprietary SNP and SilicoDArT (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14).

For SNP calling all tags from all libraries included in the DArTsoft14 analysis are clustered using DArT PL’s C++ algorithm at the threshold distance of 3, followed by parsing of the clusters into separate SNP loci using a range of technical parameters, especially the balance of read counts for the allelic pairs. Additional selection criteria were added to the algorithm based on analysis of approximately 1000 controlled cross populations. Testing for Mendelian distribution of alleles in these populations facilitated selection of technical parameters discriminating well true allelic variants from paralogous sequences. In addition, multiple samples were processed from DNA to allelic calls as technical replicates and scoring consistency was used as the main selection criteria for high quality/low error rate markers. Calling quality was assured by high average read depth per locus (Average across all markers was over 30 reads/locus). Regarding whole read quality, the minimum Phred pass score was set at 10 and the minimum pass percentage was 50. The average SNP calling rate was 92% and the fish sample calling rate was 95%. The raw sequence data (accession number SRP130211) is available at <https://www.ncbi.nlm.nih.gov/sra/SRP130211> (Release date: 2019-02-28) Release date: 2019-02-28.

GBS data analysis

Sample and markers statistics is given in Additional file 1: Tables S1 and S2, respectively. In total, there were 14,448

SNP markers. The PIC value for the SNPs was 0.16 under additive genetic model, whereas it was substantially higher, 0.46 under codominant model. The average proportion of missing genotype (SNPs) was only 14.8%. The frequency of minor allele was 0.29. The missing genotype was also imputed using theDARTsoft14 analysis pipeline.

Pedigree based analysis of heritability

Restricted maximum likelihood (REML) method applied to a linear (animal) mixed model was used to estimate heritability for traits studied. The model included the fixed effect of stock origin (wild and F1 fish) and the additive genetic random effect of individual fish in the pedigree. Our preliminary analysis using logarithmic likelihood ratio test (LRT) indicated that the common full-sib effect was not significant ($P > 0.05$). This has been a result of early communal rearing of all families soon after birth.

Heritability (h^2) for the traits studied was estimated as $h^2 = \sigma_A^2 / \sigma_P^2$ where σ_A^2 is the additive genetic variance, σ_P^2 the phenotypic variance ($\sigma_P^2 = \sigma_A^2 + \sigma_E^2$) and σ_E^2 is the residual variance. Pedigree based analysis of heritability were carried out using the ASREML software package [22].

Genomic prediction

Genomic best linear unbiased prediction (gBLUP) method was used to assess the predictive ability of the model for traits studied, using SVS Suite (Golden Helix, 2016). This method (gBLUP) generally shows similar predictive capacity to other non-linear approaches, especially for growth traits as used in this study. Theoretical statistics of the gBLUP method was discussed in earlier reports [23]. In brief, the statistical model and assumptions regarding SNP distribution and its variance are given below.

The (gBLUP) method uses genomic relationship matrix (G) derived from DNA marker information to calculate genomic breeding values for each individual in the pedigree Clark et al., [24]. The model is written in a matrix notation as follows:

$$y = m + Xb + Za + e$$

where y is the vector of phenotypic observations (body weight, length and condition index), m = overall mean, X is the incidence matrix containing fixed effects (i.e. stock origin in this study) in b . The matrix Z relates records to genomic values, a is the SNP effect and e is the residual error. In this model, gBLUP assumes equal variance σ_z^2 for all loci/SNPs:

$$Var(g) = ZZ' \sigma_z^2 = G\phi \sigma_z^2 = G \sigma_g^2$$

where σ_g^2 is the genetic variance and ϕ is the normalisation factor.

The pseudo heritability was calculated from the gBLUP model as $h_p^2 = \hat{\sigma}_g^2 / (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)$ where $\hat{\sigma}_g^2$ is the genomic variance and $\hat{\sigma}_e^2$ is the component of error variance.

The accuracy or predictive ability of the gBLUP model was defined as the correlations between the predicted breeding values and actual phenotypes ($r_{y,\hat{y}}$), divided by square root of the trait heritability.

Within population prediction

We conducted within population prediction using five-fold cross validation approach to assess how well our model can predict the phenotype. With this approach, the training data consisting of 752 samples with both phenotype and genotype were randomly partitioned into five equal sub-samples (paternal half-sibs were present in all sub-sets), and in each round (e.g. k-1), one sub-set was selected as a test (validation) and the model fitted with four folds to predict the validation set.

The gBLUP five-fold cross validation model used was the same as described above. The allele substitution effects (ASE) and fixed effect coefficients obtained from iterations and k-folds that gave the largest R^2 -value were used to predict phenotypes with the following model:

$$\hat{y} = X\hat{\beta} + M\hat{\alpha}$$

where \hat{y} are the predicted phenotypes, X is the fixed effects matrix, $\hat{\beta}$ are the fixed effect coefficients, M is the genotype matrix, and $\hat{\alpha}$ are the ASE values.

The five-fold cross validation analyses for all the traits were conducted in SVS Suite (Golden Helix, 2016). We compared the actual and predicted phenotypes by association test and linear regression analysis. The coefficient of determination (R^2) from the regression analysis (e.g. VanRaden et al. [25]) was also used to evaluate the predictive ability of the gBLUP model.

Results

Animals, phenotype and heritability

A total of 752 fish sampled from a population of 1568 individuals having phenotype data was sequenced using DART marker technology. The number of observations, mean and standard deviation) for three traits studied are given in Table 1. Restricted maximum likelihood (REML) analysis applied to a single trait mixed model that included the fixed effect of stock origin and the additive genetic effect of individual animal as a random factor showed that the growth-related traits (i.e. body weight, fork length and condition index) were moderately heritable ($h^2 = 0.11$ to 0.42).

The pseudo-heritability (h_p^2) of traits estimated from the marker data using gBLUP genomic relationship matrix is

Table 1 Phenotypic information of 752 fish sequenced and the heritability estimated from a pedigreed population ($h_a^2 \pm$ standard errors) and pseudo genomic heritability (h_p^2) for traits estimated from the markers, using gBLUP kinship matrix

Traits	Unit	N	Mean	SD	$h_a^2 \pm$ s.e.	$h_p^2 \pm$ s.e.
Weight	Kg	752	3.0	0.35	0.42 \pm 0.10	0.47 \pm 0.180
Length	Cm	752	58.2	2.10	0.42 \pm 0.10	0.43 \pm 0.230
Condition index	Unit	752	15.0	1.02	0.11 \pm 0.05	0.21 \pm 0.223

Pedigree based analysis of heritability was conducted on genotyped animals
SD Standard deviation

consistent with those estimated from the pedigree (Table 1). Growth related traits (weight, length and condition factor) had moderate pseudo-heritability ($h_p^2 = 0.21$ to 0.47). Magnitudes of the pseudo-heritability estimates obtained from genomic data (SNPs) are generally similar among statistical methods used.

Predictive ability of gBLUP model

The accuracy or predictive ability of genomic prediction for growth related traits (i.e. body weight, fork length and condition index) was moderate to high (0.44 to 0.69) (Table 2).

In addition, within population prediction using five-fold cross validation approach showed that the correlation between the predicted and actual phenotypes for all traits especially for weight and length were very high, $r_{y,\hat{y}} = 0.76-0.89$. The coefficient of determination (R^2) from the linear regression analysis of the actual on predicted phenotypes ranged from 0.49 to 0.71 across the growth traits. The 5-fold cross validations using gBLUP for body weight and length are given in Fig. 1.

Predictive ability using imputation data

The proportion of missing genotype was only about 14% in this population; however, imputations were made to have a complete genotype (without missing value). The imputation slightly increased predictive ability of genomic prediction for growth related traits (Fig. 2 vs. Table 2). For body weight, the predictive ability increased from 0.69 to 0.83 when analysis of the imputed data was compared with the original genotype provided by DArT.

Table 2 Accuracy or predictive ability of gBLUP model for growth related traits

Traits	gBLUP
Weight	0.69
Length	0.67
Condition index	0.44

gBLUP Genomic best linear unbiased prediction

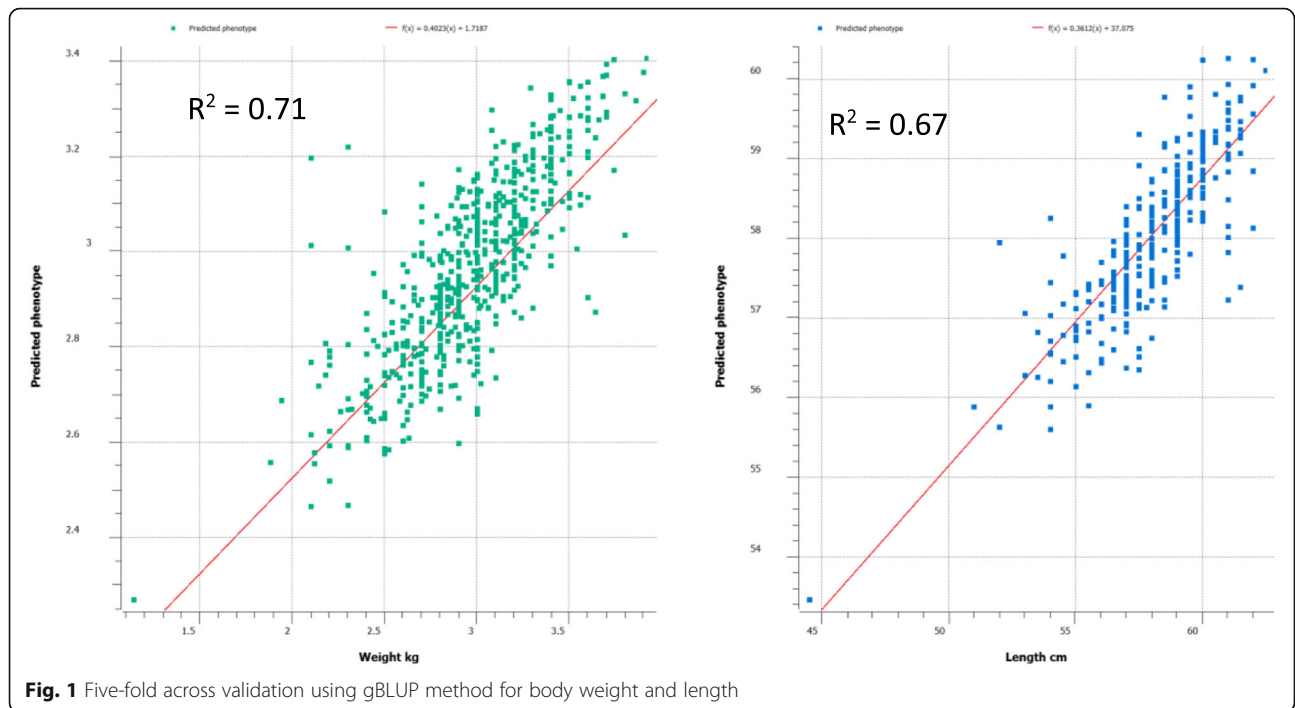
Effect of markers sub-sets

Figure 3 shows the predictive ability of gBLUP model for body weight when different marker sub-sets randomly sampled from the dataset (20%, 40, 60% and 80% of 14,448 SNPs) were analysed. The predictive ability of the gBLUP model decreased when random marker subsets were used. The reduction in the predictive ability ranged from 13 to 18% across important traits studied. A graphical presentation of the predictive ability using different marker subsets is given for body weight in Fig. 3.

Discussion

The central theme of this study was to determine the predictive ability of genomic prediction for traits of economic importance to enable the application of genomic selection in yellowtail kingfish (YTK). Our gBLUP analysis shows that the predictive ability obtained for growth-related traits (body weight, fork length and condition index) was moderate to high (0.44 to 0.69). A similar predictive ability was observed for body traits when numerator relationship matrix obtained from the pedigree was included in analyses. A linear regression analysis of the actual on predicted phenotypes also indicated a high predictive capacity of our models for the observed phenotype. The findings are particularly significant because one of the promising features of genomic selection is to reduce efforts of phenotyping in selection populations, and the performance of unphenotyped animals can be predicted from the training and validation populations from which they are derived. Our prediction within the same population using five-fold cross validation approach also confirmed that there was a high correlation between the training and validation datasets for body weight, $r_{y,\hat{y}} = 0.89$.

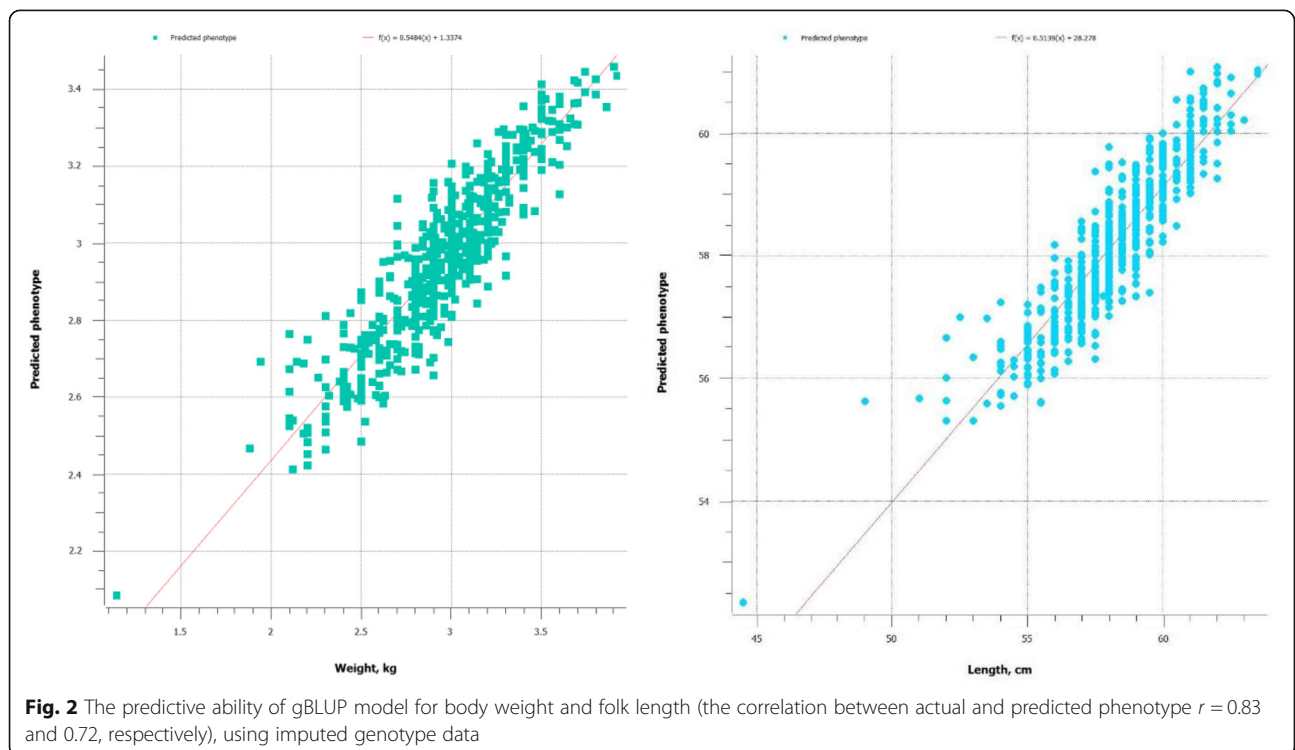
Overall our results are in line with other studies reported in the literature that based on computer simulations or using empirical data from practical genetic enhancement breeding programs in farmed animals [26] or aquaculture species [15, 27–29]. However, comparing the results from different studies and populations is not rigorous due to differences in genetic background and histories of the experimental animals. The predictive ability of statistical methods used for genomic selection generally depends on marker effects, frequency of rare allele in populations [30]. In the present study, we found that gBLUP had similar predictive ability to Bayesian methods (i.e. Bayes C or Bayes Cπ, results were omitted). This is as expected for growth related traits that are well known under control by many genes, each with small effects. With regard to commercial application, gBLUP is preferred to other methods because it was less computationally demanding. Furthermore, gBLUP can account for various variance structure (e.g. heterogeneous variance due to genotype by environment

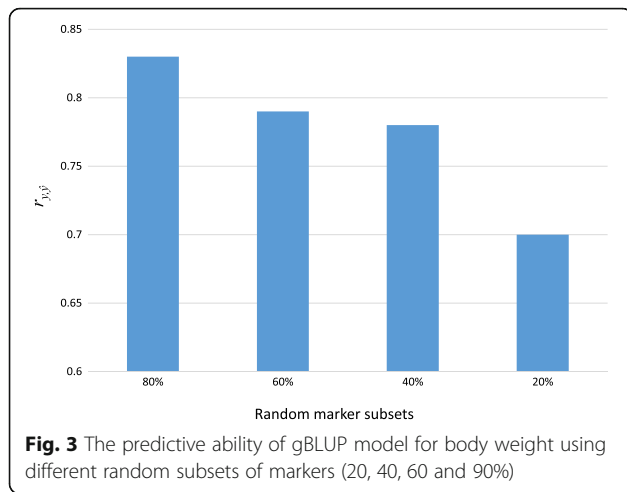


interaction effect) and capture genetic relationship among individuals in the pedigree that are more important than capturing LD.

In the present study, we also found that the predictive ability of the gBLUP model is affected by the number of

markers used in our analysis. A random sample of 20, 40, 60 and 80% of the total number of markers (14,448 SNPs) was analysed and the predictive ability of gBLUP model decreased by 13 to 18% across traits, in comparison with those obtained when the complete set of markers was used.





In addition, we performed gBLUP analysis on a subset of 7988 markers after quality control. The marker filtering was based on two main criteria: i) < 10% missing values and 2) a minor allele frequency > 0.05. Linkage disequilibrium (LD) pruning was also applied to remove highly correlated markers. However, the predictive ability of the five-fold cross validation model for body traits was somewhat higher than those obtained from the full marker set (14,448 markers), e.g. $r_{y,\hat{y}} = 0.90$ for body weight.

Our five-fold across validation within the same population indicated very high predictive ability of the gBLUP model used especially for body weight. In this study, the full data that comprised 752 fish originating from the generation in 2014 were split into training and validation subsets. It is however necessary, in a future study, to sequence offspring of this population produced in the latest generation and re-train the model to ensure that capacity of gBLUP prediction is maintained before genomic selection can be considered as an option for future genetic improvement in this YTK population. The capacity of genomic prediction is likely to be reduced in subsequent generations, mainly as a result of recombination that causes a breakdown of the linkage disequilibrium (LD) between markers and quantitative trait loci (QTL). The prediction ability also depends on the number of distant ancestral generations included in the training set; the further generations apart the lower accuracy may be observed [31].

Collectively, the results achieved from this study showed the potential for genomic selection in this population of YTK. However, to enable the application of genomic selection in different populations (e.g. those from other States in Australia), genotyping a larger sample of animals would be needed for further prediction and validation. Genomic prediction equations developed for one population may not be applicable to another due

to LD decays across generations as reported in farmed animals [32]. A larger training population together with a higher density SNP panel would also help to increase the accuracy of genomic prediction for traits with low heritability, namely skin fluke and deformity (results not tabulated). In dairy cattle, across-breed genomic selection used 300,000 SNPs and the accuracies of prediction ranged from 0.6 to 0.8 for economically important traits such as milk yield and fertility [33]. Whole genome sequencing also increased predictive power of genomic prediction, although the improvement was not large and varied with traits and studies [34]. Further, continuation of collecting more genomic data in future generations of this YTK population should be conducted to improve the reliability of genomic prediction for traits used here.

This study is the first reporting genomic prediction in marine yellowtail kingfish. Earlier studies involved computer simulation and showed potential benefits of genomic selection for aquatic animal species [35]. Nielsen et al. [36] demonstrated that the accuracy of genomic selection in a sib based breeding program can be 33% higher than selection based on conventional BLUP approach and genetic gain under optimal contribution selection based on genomic EBV can be improved by 81%. Implementation of genomic selection has been shown to reduce generation time, lower inbreeding and improve genetic gain for traits of commercial importance in agricultural species, e.g. in dairy cattle [37]. Recent studies in salmonids reported that the accuracy of genomic prediction ranged from 0.34 to 0.61 for growth traits [27] and from 0.45 to 0.71 for resistance to a range of diseases such as sea louse *Caligus rogercresseyi* [28], *Piscirickettsia* [29] or *Flavobacterium psychrophilum* [15]. A synthesised result from the literature together with our estimates obtained in this study suggest that genomic selection has the potential to accelerate genetic gain in the future breeding program for yellowtail kingfish. To improve the cost efficiency of genomic selection for this population, reduced sequence coverage in combination with imputation can be leveraged to genotype a larger number of training individuals to improve prediction accuracy and a larger number of selection candidates (validation population) to increase selection intensity and as a consequence, this can lead to increased genetic gain in YTK. A recent simulation study in plants [38] demonstrated that imputation can allow a reduced sequence coverage to as low as 1x with 10,000 markers and this can provide comparable genomic prediction accuracy as SNP arrays. Improved accuracy with imputation from a low- (256 SNPs) to high-density panel (250 K SNPs) is also shown in Atlantic salmon [39]. Return on investment from the low-cost genotyping by sequencing and imputation to enable a cost-effective genomic selection program merits future study in this species.

Conclusions

Main findings from our first genotyping by sequencing study in yellowtail kingfish include: i) a large number of markers were derived from DArTseq™ sequencing technology and they can be used for genomic prediction and selection, ii) the predictive ability of gBLUP model for growth-related traits (weight, length and condition index) was high to enable the application of genome-based selection in this population, and iii) the markers subsets (e.g. 80% of the full set used in this study) can provide a comparable predictive capacity for growth characters. However, genomic prediction for traits with low heritability (e.g. skin fluke or deformity, results omitted) using genotyping by sequencing technology and imputation deserves further study in this population of yellowtail kingfish. A large scale routine data recording from multiple generations in in-depth pedigree populations is also needed to improve accuracy of genomic breeding value estimation for other traits of commercial importance in this species *S. lalandi*.

Additional files

Additional file 1: Basic statistics of the sequence data used. Descriptive population genetic estimates and statistics of Single Nucleotide Polymorphisms (SNPs). (DOCX 27 kb)

Abbreviations

DArT: Diversity Arrays Technology; EBV: Estimated Breeding Values; gBLUP: Genomic best linear unbiased prediction; GBS: Genotyping by Sequencing; SNP: Single Nucleotide Polymorphism

Acknowledgements

We would like to thank Adam Miller, Trent D'Antignana and Clean Seas Ltd. in South Australia for providing support during the field trial, harvesting and data collection.

Funding

This study was funded by The Legacy of The Australian Seafood CRC Project No. 2013/700 to Prof Abigail Elizur and A/Prof Wayne Knibb from University of the Sunshine Coast. The funding body had no role in the design and collection, analysis and interpretation of the data.

Availability of data and materials

Supplementary files are included. The sequence data (SRP130211) is available at <https://www.ncbi.nlm.nih.gov/sra/SRP130211> (Release date: 2019-02-28) and the phenotypic data may be available upon request with permission of Clean Seas Ltd. in South Australia.

Authors' contributions

NHN and WK conceived and designed the experiments. NHN analysed the data and prepared the manuscript; AP conducted the genotyping, laboratory work and pedigree, AK ran DArT sequencing (DArTseq™), WK overseen the study, managed the pedigree and the breeding program. NHN, AP and WK participated in data and tissue sample collection. All authors read and approved the manuscripts.

Ethics approval and consent to participate

All the methods and experimental protocols of this study were performed in accordance with guidelines and regulations approved by the animal ethics committee of the University of the Sunshine Coast, Australia (approval number AN/A/10/51).

Consent for publication

Not Applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The University of the Sunshine Coast, Maroochydore DC, QLD 4558, Australia. ²Diversity Arrays Technology Pty Ltd, Kirinari St, Bruce, ACT 2617, Australia.

Received: 16 February 2017 Accepted: 23 January 2018

Published online: 30 January 2018

References

- Symonds J, Walker S, Pether S, Gublin Y, McQueen D, King A, Irvine G, Setiawan A, Forsythe J, Bruce M. Developing yellowtail kingfish (*Seriola lalandi*) and hāpuku (*Polyprion oxygeneios*) for New Zealand aquaculture. *N Z J Mar Freshw Res.* 2014;48(3):371–84.
- Knibb W, Miller A, Quinn J, D'Antignana T, Nguyen NH. Comparison of lines shows selection response in kingfish (*Seriola lalandi*). *Aquaculture.* 2016;452:318–25.
- Whatmore P, Nguyen NH, Miller A, Lamont R, Powell D, D'Antignana T, Tubner E, Elizur A, Knibb W. Genetic parameters for economically important traits in yellowtail kingfish *Seriola lalandi*. *Aquaculture.* 2013;400–401(0):77–84.
- Hung D, Vu NT, Nguyen NH, Ponzoni RW, Hurwood DA, Mather PB. Genetic response to combined family selection for improved mean harvest weight in giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture.* 2013;412–413(0):70–3.
- Dong Z, Nguyen NH, Zhu W. Genetic evaluation of a selective breeding program for common carp *Cyprinus carpio* conducted from 2004 to 2014. *BMC Genet.* 2015;16(1):1–9.
- Oliveira CAL, Ribeiro RP, Yoshida GM, Kunita NM, Rizzato GS, Oliveira SN, Santos AI, Nguyen NH. Correlated changes in body shape after five generations of selection to improve growth rate in a breeding program for Nile tilapia *Oreochromis niloticus* in Brazil. *J Appl Genet.* 2016;57(4):487–93.
- Hamzah A, Ponzoni RW, Nguyen NH, Khaw HL, Yee HY, Nor SAM. Genetic evaluation of the Genetically Improved Farmed Tilapia (GIFT) strain over ten generations of selection in Malaysia. *J Trop Agric Sci.* 2014;37(4):411–29.
- Nguyen HN. Genetic improvement for important farmed aquaculture species with a reference to carp, tilapia and prawns in Asia: achievements, lessons and challenges. *Fish Fish.* 2016;17:483–506.
- Thoa NP, Ninh NH, Knibb W, Nguyen NH. Does selection in a challenging environment produce Nile tilapia genotypes that can thrive in a range of production systems? *Sci Rep.* 2016;6:21486. <https://doi.org/10.1038/srep21486>.
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol.* 2012;888:67–89.
- Courtois B, Audebert A, Dardou A, Roques S, Gheim-Herrera T, Droc G, Frouin J, Rouan L, Gozé E, Kilian A. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS One.* 2013;8(11):e78037.
- Von Mark VC, Kilian A, Dierig DA. Development of DArT marker platforms and genetic diversity assessment of the US collection of the new oilseed crop *lesquerella* and related species. *PLoS One.* 2013;8(5):e64062.
- Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N, Diffey S, Kadkol G, Edwards D, McCully M, et al. Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE.* 2014;9(7):e101673.
- Tsai H-Y, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, Gharbi K, Stear M, Matika O, Pong-Wong R, et al. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol.* 2016;48(1):1–11.
- Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, Fragomeni BO, Wiens GD, Palti Y. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet Sel Evol.* 2017;49(1):17.

16. Palaikostas C, Ferraresso S, Franch R, Houston RD, Bargelloni L. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3*. 2016;6(11):3693–700.
17. Nguyen N, Whatmore P, Miller A, Knibb W. Quantitative genetic properties of four measures of deformity in yellowtail kingfish *Seriola lalandi* Valenciennes, 1833. *J Fish Dis*. 2016;39:217–28.
18. Renshaw MA, Patton JC, Rexroad CE, Gold JR. Isolation and characterization of dinucleotide microsatellites in greater amberjack, *Seriola dumerili*. *Conserv Genet*. 2006;8(4):1009–11.
19. Renshaw MA, Patton JC, Rexroad CE, Gold JR. PCR primers for trinucleotide and tetranucleotide microsatellites in greater amberjack, *Seriola dumerili*. *Mol Ecol Notes*. 2006;6(4):1162–4.
20. Jones OR, Wang J. Molecular marker-based pedigrees for animal conservation biologists. *Anim Conserv*. 2010;13(1):26–34.
21. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS one*. 2011;6:e19379.
22. Gilmour AR, Gogel B, Cullis B, Thompson R, Butler D. ASReml user guide release 3.0. Hemel Hempstead: VSN International Ltd; 2009.
23. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193(2):327–45.
24. Clark SA, Hickey JM, Van der Werf JH. Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol*. 2011;43(1):18.
25. VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F. Invited review: reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*. 2009;92(1):16–24.
26. Daetwyler HD, Calus MP, Pong-Wong R, de los Campos G, Hickey JM. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*. 2013;193(2):347–65.
27. Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015;16:969.
28. Correa K, Bangera R, Figueroa R, Lhorente JP, Yáñez JM. The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genet Sel Evol*. 2017;49(1):15.
29. Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM. Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics*. 2017;18(1):121.
30. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136(2):245–57.
31. Weng Z, Wolc A, Shen X, Fernando RL, Dekkers JCM, Arango J, Settar P, Fulton JE, O'Sullivan NP, Garrick DJ. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genet Sel Evol*. 2016;48(1):1–10.
32. Hayes B, Bowman P, Chamberlain A, Verbyla K, Goddard M. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol*. 2009;41(1):51.
33. van Binsbergen R, Calus MP, Bink MC, van Eeuwijk FA, Schrooten C, Veerkamp RF. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genet Sel Evol*. 2015;47(1):1–13.
34. Iheshiulor OO, Woolliams JA, Yu X, Wellmann R, Meuwissen TH. Within-and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet Sel Evol*. 2016;48(1):1.
35. Sonesson A, Meuwissen T. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*. 2009;41(1):37.
36. Nielsen H, Sonesson A, Yazdi H, Meuwissen T. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture*. 2009;289:259–64.
37. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci*. 2016;113:E3995–4004.
38. Gorjanc G, Dumasy J-F, Gonen S, Gaynor RC, Antolin R, Hickey JM. Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. *Crop Sci*. 2017;57(3):1404–20.
39. Tsai H-Y, Matika O, Edwards SM, Antolín-Sánchez R, Hamilton A, Guy DR, Tinch AE, Gharbi K, Stear MJ, Taggart JB. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3*. 2017;7(4):1377–83.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

