Research
Breeding Engineering—Article

Development of a 50K SNP Array for Japanese Flounder and Its Application in Genomic Selection for Disease Resistance

Qian Zhou [a,b,#], Ya-dong Chen [a,#], Sheng Lu [a], Yang Liu [a], Wen-teng Xu [a,b], Yang-zhen Li [a], Lei Wang [a], Na Wang [a,b], Ying-ming Yang [a], Song-lin Chen [a,b,*]

[a] Key Laboratory for Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266373, China

ABSTRACT

Single nucleotide polymorphism (SNP) arrays are a powerful genotyping tool used in genetic research and genomic breeding programs. Japanese flounder (*Paralichthys olivaceus*) is an economically-important aquaculture flatfish in many countries. However, the lack of high-efficient genotyping tools has impeded the genomic breeding programs for Japanese flounder. We developed a 50K Japanese flounder SNP array, "Yuxin No. 1," and report its utility in genomic selection (GS) for disease resistance to bacterial pathogens. We screened more than 42.2 million SNPs from the whole-genome resequencing data of 1099 individuals and selected 48 697 SNPs that were evenly distributed across the genome to anchor the array with Affymetrix Axiom genotyping technology. Evaluation of the array performance with 168 fish showed that 74.7% of the loci were successfully genotyped with high call rates (> 98%) and that the polymorphic SNPs had good cluster separations. More than 85% of the SNPs were concordant with SNPs obtained from the whole-genome resequencing data. To validate "Yuxin No. 1" for GS, the arrayed genotyping data of 27 individuals from a candidate population and 931 individuals from a reference population were used to calculate the genomic estimated breeding values (GEBVs) for disease resistance to *Edwardsiella tarda*. There was a 21.2% relative increase in the accuracy of GEBV using the weighted genomic best linear unbiased prediction (wGBLUP), compared to traditional pedigree-based best linear unbiased prediction (ABLUP), suggesting good performance of the "Yuxin No. 1" SNP array for GS. In summary, we developed the "Yuxin No. 1" 50K SNP array, which provides a useful platform for high-quality genotyping that may be beneficial to the genomic selective breeding of Japanese flounder.

1. Introduction

Single nucleotide polymorphism (SNP) arrays are a high-quality and convenient genotyping platform. With a SNP array, tens of thousands of SNPs per sample can be simultaneously detected, allowing high-throughput and high-efficiency for genetic research and breeding programs. SNP arrays have been successfully used for germplasm characterization, complex trait dissection, marker-assisted selection (MAS), and genomic selection (GS) in several economic species. GS uses genetic markers spanning the entire genome to predict the genomic estimated breeding value (GEBV); animals with high GEBVs are selected for breeding [1]. The application of GS in breeding programs has been very successful. For example, in most countries, dairy breeding programs rely on GS with commercial cattle SNP arrays [2,3].

During the past six years, whole-genome sequencing projects for more than 20 fish species in China have been completed [4]. The availability of whole-genome sequences for numerous cultured fishes has promoted the development of GS programs and SNP arrays. More recently, several SNP arrays have been developed for aquaculture species, such as Atlantic salmon (*Salmo salar*) [5,6], common carp (*Cyprinus carpio*) [7], rainbow trout (*Oncorhynchus mykiss*) [8], and catfish (*Ictalurus punctatus* and *Ictalurus furcatus*) [9]. However, to our knowledge, there are no available SNP arrays for any flatfish species or disease-resistance selection in fish. Japanese flounder, *Paralichthys olivaceus*, is an important aquaculture species in many countries, including China, Republic of Korea, and Japan. Selective breeding for Japanese flounder began in the

---

* Corresponding author.
  E-mail address: chensl@ysfri.ac.cn (S. Chen).
# These authors contributed equally to this work.

early 1970s in Japan and in the 1990s in China. However, the sustainability of Japanese flounder aquaculture has faced multiple challenges, such as germplasm degradation, frequent infectious diseases, and a lack of superior breeding strains. Therefore, advanced genomic breeding programs are urgently needed to develop high-quality strains to optimize production and culture management. Several prior studies have attempted to facilitate the selective breeding and aquaculture of Japanese flounder. For example, a microsatellite marker for lymphocystis disease resistance has been reported (*Poli9-8TUF*) and used for MAS [10], and quantitative trait loci (QTLs) of *Vibrio anguillarum* disease resistance were identified with a SNP genetic linkage map [11]. These results have improved our understanding of the genetic architecture for disease resistance. However, the small number of markers limits the ability of selective breeding programs. GS, selection based on genome-wide SNPs, is urgently needed for disease resistance (a complex trait controlled by several loci) [12,13].

We used next-generation sequencing (NGS) technology to complete the whole-genome sequencing and assembly of the Japanese flounder [14]. We also accomplished the GS to bacterial disease resistance based on large-scale genome re-sequencing data [15]. Here, we designed, characterized, and validated a 50K SNP array "Yuxin No. 1" for Japanese flounder. We used genome resequencing data from 1099 individuals to select high-quality and informative SNPs for the array and validated their genotyping performance. We also obtained good segregation of the population structure and high accuracy of the GEBV estimation when applying the "Yuxin No. 1" for GS of bacterial-resistance. These results demonstrate the potential of this array in genomic breeding programs for high disease resistance, as well as other economically-important traits. The SNP array is publicly available to any interested parties.

2. Material and methods

2.1. SNP identification

The SNPs used for the SNP array were from whole-genome resequencing data of 1099 Japanese flounders, including 931 individuals reported by Liu et al. [15] and 168 individuals sequenced in this study (NCBI SRA accession No. SRP253464). Briefly, genomic DNA was extracted from fin tissues and then subjected to Illumina pair-ended sequencing library construction according to the standard protocol (Illumina, USA). Raw reads were produced on the Illumina HiSeq 2000 sequencing platform and quality filtered with quality control (QC)-Chain [16] to remove the low-quality reads, adaptor sequences, and ambiguous nucleotides (Ns). The resultant clean reads were aligned to the Japanese flounder reference genome (NCBI accession No. GCA_001904815.2) with Burrows–Wheeler aligner [17]. SNP calling was performed with the Genome Analysis Toolkit (GATK), using default parameters [18], and the putative SNPs were identified by minimal mapping quality values of 20, SNP quality scores of 20, and read base qualities of 30.

2.2. SNP selection

The initial SNP set was filtered with multiple steps and parameters. First, we calculated the minor allele frequencies (MAFs) and missing rates using PLINK (v1.07) [19], and the SNPs with MAF $\leq$ 0.05 and missing rate $\geq$ 0.1 were excluded. We tested the Hardy–Weinberg Equilibrium (HWE) using VCFtools (v0.1.14) [20] with the -hwe option and removed SNPs that severely departed from HWE ($p < 0.01$). Next, the selected SNPs with upstream and downstream flanking sequences of 35 base pairs (bp) were submitted to the Affymetrix Axiom® myDesign

GW bioinformatics pipeline (Thermo Fisher Scientific Inc., USA) for probe design. In this pipeline, a *p*-convert value (between 0 and 1) was assigned to each SNP, representing the probability that the given SNP converts to a reliable SNP assay on the Affymetrix Axiom array system. This score considers the SNP sequence, binding energies, expected degree of nonspecific binding, and hybridization to multiple genomic regions. From the *p*-convert values and several other QC metrics implemented in the pipeline, the variants were classified as "recommended" (*p*-convert value > 0.6, no interfering polymorphisms (wobble) and poly count = 0), "not recommended" (*p*-convert value < 0.4, or no wobble $\geq$ 3, or poly count > 0, or duplicate count > 0), "not possible" (on a given strand if we cannot build a probe to interrogate the SNP in that direction, respectively) and "neutral" (others). Only probes designated as "recommended" or "neutral" were retained for further analysis. The flanking sequences of the candidate SNPs were also required to have no other variations or repetitive elements. The GC contents of the flanking sequences ranging from 30% to 70% were retained.

We also filtered the SNPs to ensure an even distribution across the genome. We excluded most A/T and C/G SNP transversions since these markers take up twice as much space on the Affymetrix Axiom array platform. Probes for the final selected SNP panel were anchored on the SNP array with 2000 dish quality control (DQC) probes (negative controls). Finally, we used SNPeff (v4.2) [21] to predict the functional effects of the SNPs on the predicted Japanese flounder gene models.

2.3. Evaluation of the SNP array performance

To assess the performance of the "Yuxin No. 1" SNP array, we genotyped 168 Japanese flounder individuals, including 96 that were randomly selected from the genome resequencing sample (for initial SNP discovery) and 72 from the reference population of a GS program [15].

Genomic DNA samples were extracted from each fish and labeled according to the Affymetrix Axiom® 2.0 Assay protocol, with a final concentration of 50 ng·μL^{-1} and a volume of 10 μL. DNA hybridization and array scanning were completed on the Affymetrix GeneTitan Multi-Channel Instrument (Thermo Fisher, USA), which generated raw data in CEL files. These files were imported into the Axiom Analysis Suite software for quality control and genotyping; the sample QC parameters were DQC value $\geq$ 0.82, call rate (CR) $\geq$ 0.97, percent of passing samples $\geq$ 95%, and average CR for passing samples $\geq$ 98.5% (following the "best practices workflow"). The default SNP QC thresholds were also used to filter the genotypes.

The probe conversion quality of the variants was evaluated by the signal intensity and cluster separation, and the number of heterozygous/homozygous genotype calls were calculated. Based on these metrics, the variants were classified into six categories: "PolyHighResolution" (SNPs had good cluster resolution and at least two examples of the minor allele), "MonoHighResolution" (SNPs had good SNP clustering but less than two samples had the minor allele), "NoMinorHom" (SNPs had good cluster resolution but no samples had the minor allele), "OffTargetVariation (OTV)" (an off-target variant cluster was called), "CallRateBelowThreshold" (SNPs had call rates CR below the threshold, but the other properties were above the threshold), and "Other" (more than one cluster property was below the threshold) [22].

To further test the genotyping quality and accuracy of the SNP array, we randomly selected 96 individuals from the genome resequencing sample and compared the genotypes from the two methods.

2.4. Population structure analysis

Using the genotyping data of 168 individuals from the SNP array, we conducted a principal component analysis (PCA) in GCTA [23] and the first and second components were plotted.

2.5. Application of "Yuxin No. 1" in GS for disease resistance

We previously completed a GS program for *Edwardsiella tarda* (*E. tarda*) disease resistance using whole-genome resequencing data [15]. In this study, we genotyped 72 individuals using "Yuxin No. 1," 27 of which (14 males and 13 females) were the parents of 16 families. GEBV was estimated using the weighted genomic best linear unbiased prediction (wGBLUP), and the mid-parental GEBV was used as the GEBV of the corresponding family. Estimated breeding value (EBV) were also estimated by a pedigree-based best linear unbiased prediction (ABLUP) involving five generations. The model used for (G)EBV estimation was defined as

$$\boldsymbol{y} = X\boldsymbol{b} + Z\boldsymbol{g} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the phenotypic vector, which contains the observations (0 for dead/1 for alive), $\boldsymbol{b}$ is the fixed effect vector (intercept, different batches of the challenge experiment, and different ages when challenged), $\boldsymbol{g}$ is the random effect vector, and $\boldsymbol{e}$ is the random residual. For wGBLUP [24], the random effect vector is assumed to follow $N\left(0, G^*\sigma_g^2\right)$, where G^* is the weighted genomic relationship matrix (weighted G matrix) that is estimated from the iterative algorithm [25]. For ABLUP, g is assumed to follow $N\left(0, A\sigma_g^2\right)$, where A is the numerator relationship matrix with a four-generation pedigree, σ_g is the additive genetic variance. X and Z are incidence matrices that relate the phenotypes to the effects and individuals, respectively. The weighted G matrix was constructed using an in-house R script, and the (G)EBVs were estimated in R-ASReml [26].

Sixteen families were challenged with *E. tarda*, so the predictive accuracy of wGBLUP and ABLUP could be assessed by the family-level GEBV and survival rate. The area under the receiver operator characteristic curves (AUC) [27] was used as an index to illustrate the accuracies of wGBLUP and ABLUP. To estimate the AUC, the survival rates of the 16 families were converted into the binary statistics; we assigned 1 or 0 to the families with survival rates higher or lower than the average value (44.33%), respectively (i.e., families valued 1 may be used for breeding). The AUC was analyzed in R-pROC [28].

3. Results and discussion

This study aimed to develop a high-quality and standardized genome-wide SNP array for Japanese flounder and to validate the genotyping performance of the array for GS breeding programs. Several factors can affect the design and quality of SNP arrays, such as the quality of the initial SNP set, the SNP filtering and selection parameters, and the SNP array production technology. The two most commonly used SNP array platforms are provided by Affymetrix and Illumina. Both platforms use target hybridization to loci-specific probes, and the probe intensities reflect the abundance of the respective alleles [29]. In the Affymetrix array, probes for specified positions are tiled on the array surface to obtain the SNP information (i.e., chips), whereas the Illumina array uses microscopic beads to anchor the probes. These SNP genotyping platforms have been used in a wide range of genetic studies. High-throughput NGS is a powerful technique to identify genome-wide SNPs, which can be used to select informative SNPs for SNP arrays.

3.1. Sequencing and SNP calling

We conducted whole-genome resequencing of 168 fish and generated 974.9 Gb of sequencing data after quality filtering. These data were combined with the sequencing data of 931 individuals from 90 breeding families that had pedigree information and were phenotypically diverse for disease resistance [15]. Finally, 3.54 Tb of sequencing data from 1099 individuals were aligned to the reference genome (Table S1 in Appendix A), and more than 42.2 million SNPs were identified. Large-scale genome resequencing of different families allowed us to identify a high-quality and diverse candidate SNP set, which was advantageous for subsequent SNP selection.

3.2. SNP selection and array design

The initial identified SNPs were subjected to several selection steps. First, we filtered the SNPs for MAF $\geq$ 0.05, missing rate < 0.1, and significant departure from HWE ($p < 0.01$). The MAF filtering excluded SNPs with little variation, and the SNPs with high missing rates suggested limited numbers of that genotype in the population. HWE filtering excluded SNPs arising from sequencing errors and natural selection. Thus, these filters removed low-quality SNPs that may have influenced the results. Filtering resulted in a SNP panel of 3 410 891 candidate SNPs, which were submitted to the Affymetrix *in silico* probe design pipeline; 959 651 SNPs passed the *p*-convert evaluation and were retained. Finally, we selected SNPs that were evenly distributed across the genome, and 48 697 SNPs were anchored on the SNP array, with an average *p*-convert value of 0.684.

3.3. Characterization of the "Yuxin No. 1" SNP array

For the 48 697 SNPs tiled on the array, 48 768 probes were synthesized—48 626 SNPs had one probe, and 71 had two probes. The SNPs on the array had an average MAF of 0.177 (Fig. 1(a)), and the median MAF ranged from 0.115 to 0.189 across 24 chromosomes (Fig. 1(b)). Moreover, 20% of the markers exhibited high variability (MAF > 0.3), and 64.3% had MAF values greater than 0.1 (Fig. 1(a)). This MAF distribution indicates a high level of polymorphism, which is desirable for genetic analysis, such as genome-wide association studies (GWASs) and linkage mapping. More than 35% of the MAF values were less than 0.1. The general MAF patterns were consistent with the source individuals for genome resequencing (i.e., from the breeding population and families). This will aid future attempts to improve breeding efficiency and accuracy using genotyping data.

To assess the SNP distribution across the genome, we aligned the loci on the array to the Japanese flounder reference genome and calculated the inter-locus distances. The SNP distribution exhibited a wide but even inter-spacing spectrum. The average spacing between adjacent loci was 9.6 kb (Fig. 2(a)), and the inter-SNP spacings were: 5125 SNPs spaced < 6 kb, 5175 (10.8%) spaced 6–7 kb, 6315 (13.1%) spaced 7–8 kb, 5471 (11.4%) spaced 8–9 kb, 5546 (11.6%) spaced 9–10 kb, 6017 (12.5%) spaced 10–11 kb, 6557 (13.7%) spaced 11–12 kb, and 5964 (12.4%) spaced 12–13 kb. Cumulatively, ~96% of the gaps between adjacent SNPs were larger than 13 kb. Moreover, the variants were evenly distributed throughout the genome, with an average median distance of 9.8 kb across 24 chromosomes (Fig. 2(b)). For most of the regions with large distances between SNPs, only a few SNPs met the selection criteria.

All of the SNPs tiled on the array were annotated and classified into different genomic groups based on their predicted effects (Table 1). Among the 48 697 SNPs, 26 274 SNPs (53.9%) were in genic regions, which included exons, introns, splicing sites, and
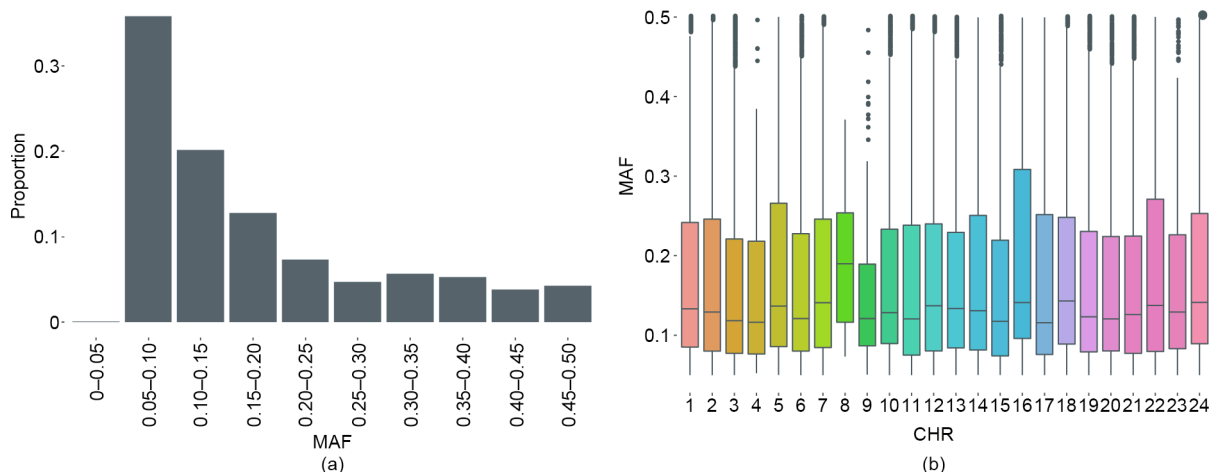
Fig. 1. MAFs of the SNPs in the Japanese flounder "Yuxin No. 1" SNP array. (a) Proportion of MAFs; (b) MAF distribution across 24 chromosomes.
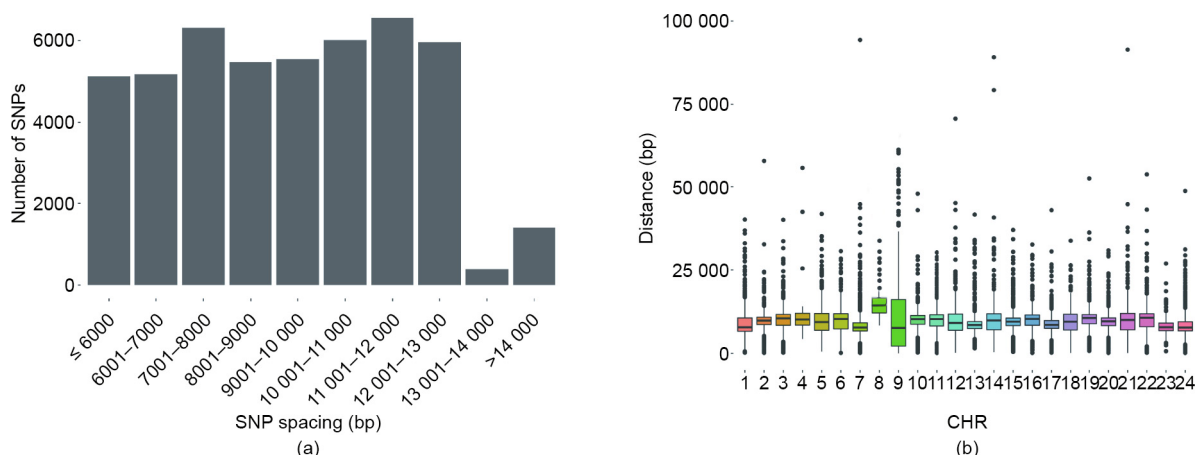

Fig. 2. Distribution of loci spacing in the Japanese flounder "Yuxin No. 1" SNP array. (a) Inter-loci spacing across 24 chromosomes; (b) distribution of SNPs with different inter-loci spacing.

Table 1
The effects of SNP tiling on the Japanese flounder "Yuxin No. 1" SNP array.

Category		Number of SNPs	Proportion
Exonic	Stop gain	8	0.02%
	Stop loss	2	0.00%
	Synonymous	1 912	3.93%
	Non-synonymous	756	1.55%
Intronic		23 475	48.21%
Splicing		4	0.01%
Upstream/downstream (1 kb)		116	0.24%
Upstream (1–5 kb)		1 684	3.46%
Downstream (1–5 kb)		1 754	3.60%
Intergenic		18 985	38.99%
Total		48 697	

1 kb regions of the upstream and downstream sequences of the genes. The two most abundant groups in the genic regions were intronic and synonymous coding SNPs, which contained 23 475 and 1912 SNPs, respectively. The non-genic SNPs included 1684 (3.46%) upstream (1–5 kb from the start codons), 1754 (3.60%) downstream (1–5 kb from the stop codons), and 18 985 (38.99%) intergenic SNPs.

3.4. Genotyping performance of "Yuxin No. 1"

The performance of the SNP array was evaluated by genotyping 168 DNA samples from the breeding families, 166 of which (98.2%) passed the sample QC and CR threshold of 97%. We investigated the conversion performance of the variants on the array for the genotype call rates, cluster separation, polymorphism in the panel, and concordance of the SNP genotypes between the array and resequencing.

Among the 48 697 SNPs on "Yuxin No. 1," 36 383 SNPs (74.71%) passed all of the SNP quality criteria. Of these genotyped loci, 41.07% were categorized as polymorphic (PolyHighResolution and NoMinorHom), and 33.64% were categorized as monomorphic (MonoHighResolution). The other loci had poor genotyping qualities, exhibited poor performance in the cluster properties, and were classified into as "OTV," "CallRateBelowThreshold," or "Others." A relatively large proportion of monomorphic SNPs were detected (33.64%); some may be false positives from the SNP discovery process or SNPs that cannot be effectively scored due to a lack of suitable assays. Also, the genotyping population is the same as the SNP discovery population, and the genotypes were very similar. Some of these SNPs may be polymorphic if more populations are genotyped.

We also compared the genotypes from "Yuxin No. 1" to those from the resequencing. Among a test cohort of 96 fish, 95 were

successfully genotyped by the array. For the successfully genotyped loci, 14 899 (41.0%) were 100% concordant with the variants called by resequencing; 4002 (11.0%), 3421 (9.4%), and 3162 (8.7%) SNPs had concordance rates within 0.95–0.99, 0.90–0.95, and 0.85–0.90, respectively. In summary, 70% of the SNPs had concordance rates ≥ 85%, indicating cross-validation of the genotyping results from the SNP array and genome resequencing.

3.5. PCA of the population structure

Investigations of the population structure are essential for many population genetic studies. To evaluate if the "Yuxin No. 1" SNP array can detect population segregation, we performed a PCA of the SNPs from 168 genotyped individuals. The samples clustered into two distinct groups, according to the first and second principal components (PC) (Fig. 3). This was consistent with their origins/sampling sites in Hebei and Shandong Provinces, China, respectively. This clustering analysis demonstrates the ability of "Yuxin No. 1" to characterize population structures.

3.6. Application of "Yuxin No. 1" to GS

Through selective breeding programs, significant genetic improvement can be achieved for economically-important traits in fish. We completed a GS program for disease resistance in Japanese flounder, based on different breeding families and artificial infection of the pathogen *E. tarda* [15]. To test the viability and performance of the "Yuxin No. 1" SNP array in GS, we genotyped the parents of 16 randomly selected families (27 individuals) via the SNP array and estimated the (G)EBVs using the reference population [15]. The average survival rate of seven families (namely disease-resistant family) was 61.13%, while the average survival rate of nine families (namely susceptible family) was 31.27%. The average GEBV (2.10) of the disease-resistant families was higher than the average GEBV (1.56) of the susceptible families (Table 2). The wGBLUP resulted in an 80% predictive accurate rate, which outperformed the ABLUP method (66%) (Fig. 4). Thus, our SNP array-based GS led to a 21.21% relative increase in predictive accuracy. The GEBVs predicted by wGBLUP were moderately correlated (Pearson's correlation = 0.70) with the EBVs estimated by ABLUP, indicating different accuracies in the (G)EBV estimations of wGBLUP and ABLUP. Our results are consistent with previous GS studies for disease resistance in fish—all GS methods performed better for GEBV estimation and resulted in 13% to 52% increases in predictive accuracy compared to ABLUP [30–32,34]. Together,
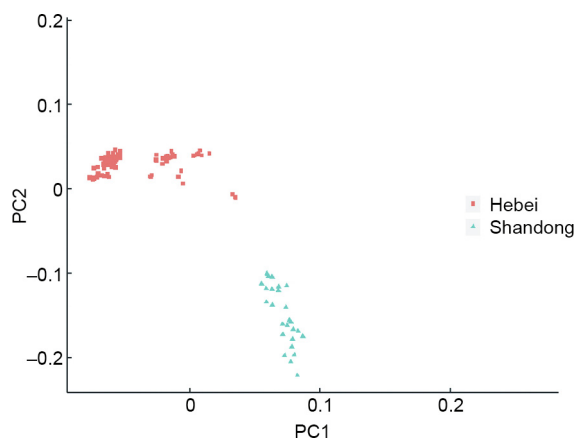

Fig. 3. PCA to investigate the population structure of the Japanese flounders genotyped using the "Yuxin No. 1" SNP array. "Hebei" and "Shandong" indicate the individuals collected in Hebei and Shandong Provinces, China, respectively.

Table 2
Average survival rate after *E. tarda* infection and estimated breeding values of 16 Japanese flounder families.

Item	Disease-resistant family	Susceptible family
No. of family	7	9
Average survival rate (%)[a]	61.13	31.27
Average GEBV	2.10	1.56
Average EBV	0.18	0.16

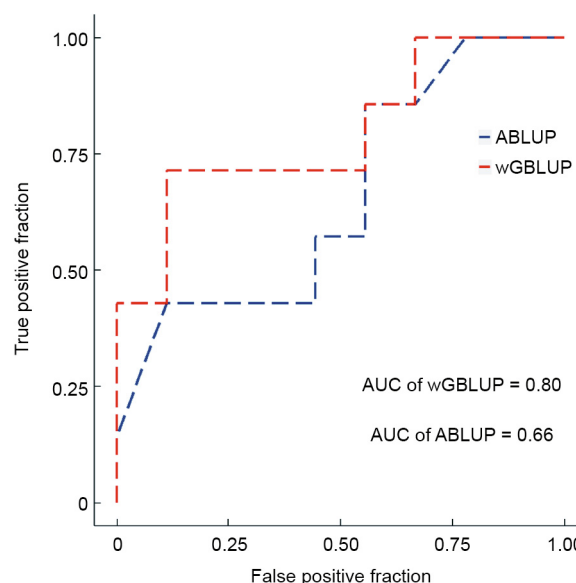[a] Survival rate: survival rate after *E. tarda* infection.

Fig. 4. The predictive accuracies of wGBLUP and ABLUP for GS, evaluated using receiver operator characteristic curves.

these results suggest that "Yuxin No. 1" can be used for the GS of disease-resistant germplasm. Nevertheless, the number of individuals used for GEBV estimation in this study was limited and does not fully simulate a GS program for *E. trada* resistance in Japanese flounder. Additional animals are required to fully evaluate the array for GS, and we are working to increase the sample sizes of the reference and candidate populations for genotyping with the SNP array.

4. Conclusions

Here, we report the design and development of a Japanese flounder 50K SNP array "Yuxin No. 1." Leveraging wfwhole-genome sequencing data from 1099 individuals, we identified a starting set of more than 42.2 million SNPs and selected 48 697 SNPs to anchor the array based on MAF, genomic location, and probe design recommendations from Thermo Fisher Axiom® technology. Using "Yuxin No. 1," we generated high-quality genotyping data for a test panel of 168 fish, and the performance of the genotyping data in GS for disease resistance was consistent with prior studies. This indicates that "Yuxin No. 1" is applicable for GS of economically-important traits and provides a robust platform for genotyping and selective breeding programs.

Authors' contribution

Song-lin Chen obtained the funding, and conceived and instructed the study. Qian Zhou performed the SNP selection and probe design for the SNP array. Ya-dong Chen and Yang Liu prepared the DNA sample. Qian Zhou, Sheng Lu, and Ya-dong Chen performed the SNP array scanning and analyzed the genotyping data. Sheng Lu performed GEBV calculation. Yang-zhen Li, Lei Wang, and Ying-ming Yang performed the family construction and bacterial challenging experiment. Wen-teng Xu and Na Wang participated the project managements. Qian Zhou, Sheng Lu, and Song-lin Chen wrote the manuscript. All authors reviewed the manuscript.

Compliance with ethics guidelines

Qian Zhou, Ya-dong Chen, Sheng Lu, Yang Liu, Wen-teng Xu, Yang-zhen Li, Lei Wang, Na Wang, Ying-ming Yang, and Song-lin Chen declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2020.06.017.

References

[1] Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001;157(4):1819–29.

[2] Pryce JE, Goddard ME, Raadsma HW, Hayes BJ. Deterministic models of breeding scheme designs that incorporate genomic selection. J Dairy Sci 2010;93(11):5455–66.

[3] Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 2009;4(4):e5350.

[4] Chen SL, Xu WT, Liu Y. Fish genomic research: decade review and prospect. J Fish China 2019;43:1–14 Chinese.

[5] Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics 2014;15(1):90.

[6] Poćwierz-Kotus A, Bernaś R, Kent MP, Lien S, Leliūna E, Dębowski P, et al. Restitution and genetic differentiation of salmon populations in the southern Baltic genotyped with the Atlantic salmon 7K SNP array. Genet Sel Evol 2015;47(1):39.

[7] Xu J, Zhao ZX, Zhang XF, Zheng XH, Li JT, Jiang YL, et al. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). BMC Genomics 2014;15(1):307.

[8] Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour 2015;15(3):662–72.

[9] Zeng QF, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S, et al. Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. Sci Rep 2017;7(1):40347.

[10] Fuji K, Hasegawa O, Honda K, Kumasaka K, Sakamoto T, Okamoto N. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). Aquaculture 2007;272(1–4):291–5.

[11] Shao CW, Niu YC, Rastas P, Liu Y, Xie ZY, Li HD, et al. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. DNA Res 2015;22(2):161–70.

[12] Robledo D, Gutiérrez AP, Barría A, Lhorente JP, Houston RD, Yáñez JM. Discovery and functional annotation of quantitative trait loci affecting resistance to sea lice in Atlantic salmon. Front Genet 2019;10:56.

[13] Zhou Q, Su ZC, Li YZ, Liu Y, Wang L, Lu S, et al. Genome-wide association mapping and gene expression analyses reveal genetic mechanisms of disease resistance variations in *Cynoglossus semilaevis*. Front Genet 2019;10:1167.

[14] Shao CW, Bao BL, Xie ZY, Chen XY, Li B, Jia XD, et al. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. Nat Genet 2017;49(1):119–24.

[15] Liu Y, Lu S, Liu F, Shao CW, Zhou Q, Wang N, et al. Genomic selection using BayesCπ and GBLUP for resistance against *Edwardsiella tarda* in Japanese flounder (*Paralichthys olivaceus*). Mar Biotechnol 2018;20(5):559–65.

[16] Zhou Q, Su XQ, Wang AH, Xu J, Ning K. QC-Chain: fast and holistic quality control method for next-generation sequencing data. PLoS ONE 2013;8(4):e60234.

[17] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754–60.

[18] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297–303.

[19] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–75.

[20] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics 2011;27(15):2156–8.

[21] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila–melanogaster* strain w[1118]; *iso-2; iso-3*. Fly 2012;6(2):80–92.

[22] Gao H, Pirani A, Webster T, Shen MM, inventors; Affymetrix, Inc., assignee. Systems and methods for SNP characterization and identifying off target variants. United States patent US 20140274749. 2014 Sep 18.

[23] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.

[24] Zhang XY, Lourenco D, Aguilar I, Legarra A, Misztal I. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. Front Genet 2016;7:151.

[25] Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res 2012;94(2):73–83.

[26] Gilmour ARGB, Gogel B, Cullis BR, Thompson R. ASReml user guide release 3.0. Hemel Hempstead: VSN International Ltd; 2009.

[27] Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet 2010;6(2):e1000864.

[28] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf 2011;12(1):77.

[29] Lindgren D, Höglund M, Vallon-Christersson J. Genotyping techniques to address diversity in tumors. Adv Cancer Res 2011;112:151–82.

[30] 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.

[31] Odegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwissen THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front Genet 2014;5:402.

[32] Palaiokostas C, Cariou S, Bestin A, Bruant JS, Haffray P, Morin T, et al. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. Genet Sel Evol 2018;50(1):30.

[34] Zhang Y, Liu Z, Li H. Genomic prediction of columnaris disease resistance in catfish. Mar Biotechnol 2020;22(1):145–51.