

넙치(*Paralichthys olivaceus*)의 성장형질 연관 유전자 변이 탐색을 위한 전장유전체연관분석(GWAS) 알고리즘 비교 분석 연구

윤상원[†] · 이희건^{1†} · 박종원 · 정민환 · 이다인 · 정효선 · 김주란 · 양혜림 · 이승환² · 이정호^{*}

국립수산과학원 육종연구센터, ¹충남대학교 바이오시융합학과, ²충남대학교 동물자원과학부

Comparison of Genome-wide Association Study (GWAS) Algorithms for Detecting Genetic Variants Associated with Growth Traits in Olive Flounder *Paralichthys olivaceus*

Sangwon Yoon[†], Heegun Lee^{1†}, Jong-Won Park, Minhwan Jeong, Dain Lee, Hyo Sun Jung, Julan Kim, Hye-Rim Yang, Seung Hwan Lee² and Jeong-Ho Lee^{*}

Genetics and Breeding Research Center, National Institute of Fisheries Science, Geoje 53334, Republic of Korea

¹Department of Bio-AI Convergence, Chungnam National University, Daejeon 34134, Republic of Korea

²Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Republic of Korea

Genome wide association studies (GWAS) identify genetic loci associated with quantitative traits in genomic selection. Although several studies have compared performance of various algorithms, no study compares them in olive flounder *Paralichthys olivaceus*. This study compared the GWAS results of four mixed linear model (MLM) algorithms and one Fixed and random model Circulating Probability Unification (FarmCPU) algorithm in olive flounder. Considering gender and genetic association matrices as fixed and random effects, the MLM had stable performance without inflation for λ_{GC} (genomic inflation factor) of $-\log_{10}P$. The FarmCPU algorithm had some appropriate λ_{GC} of $-\log_{10}P$, and an upward tail was identified in quantile-quantile plots. Therefore, the models were suitable for detecting genetic variants associated with olive flounder growth traits. Moreover, significant genotypes appeared several times at chromosome 22, around which quantitative trait loci are expected to exist. Finally, in both models, some of the most genetic variants were found in genes related to growth traits, confirming their reliability. These results will be helpful when applied to the genomic selection of olive flounder growth traits in the future.

Keywords: *Paralichthys olivaceus*, Growth traits, Genome wide association study, Mixed linear model

서론

생물 유전체 정보 분석능력 기술이 급속도로 발전하면서 2000년대 중반 도입된 NGS (next generation sequencing) 기술을 기반으로 넙치, 전복, 참돔 등 국내 주요 양식생물의 유전체 해독에 성공함에 따라 수산생물 유전체 분석 및 활용을 위한 연구가 가속화되고 있다(Behjati and Tarpey, 2013; Nam et al., 2017; Shao et al., 2017; Shin et al., 2018). 세계적으로 수산물의 수요가 지속적으로 증가하며 수산종자 개발의 중요성이 증대되는 가운데, 빠르게 구축되고 있는 양식생물의 유전체 빅데이터에

따른 대용량의 유전자 디지털 정보는 경제형질 육종을 위한 유전자 탐색 및 개발에 적극 활용되고 있다(Zenger et al., 2019).

넙치(*Paralichthys olivaceus*)는 2022년 국내 양식생산량의 50.6% (45,801 metric ton)를 차지하는 대표적인 양식어종으로, 국민हित감으로 소비량이 높다(KOSIS, 2022). 우리나라에서 1982년 넙치 양식을 위한 인공종자 생산 개발이 시작된 이후로 넙치 양식이 본격적인 산업으로 발전하여 현재 제주도과 완도 지역을 중심으로 전국 곳곳에서 양식이 이루어지고 있다(Park et al., 2022). 또한, 어류의 유전적 개량량 증가를 위한 효과적인 방법으로 알려진 선발 육종이 넙치의 성장, 체형, 체색, 내병

*Corresponding author: Tel: +82. 55. 639. 5800 Fax: +82. 55. 639. 5809

E-mail address: jhlee@korea.kr [†]Contributed equally.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

<https://doi.org/10.5657/KFAS.2023.0411>

Korean J Fish Aquat Sci 56(4), 411-418, August 2023

Received 31 May 2023; Revised 24 July 2023; Accepted 16 August 2023

저자 직위: 윤상원(연구원), 이희건(대학원생), 박종원(연구원), 정민환(연구원), 이다인(연구원), 정효선(연구원), 김주란(연구원), 양혜림(연구원), 이승환(교수), 이정호(연구관)

성 형질 개량에 적극 활용됨에 따라 양식산업의 지속적인 발전과 경쟁력 향상을 위한 우량품종 개발이 진행되고 있다(Ogata et al., 2002; Fuji et al., 2006, 2007; Min et al., 2010; Kim et al., 2011).

유전체 선발기술은 유전체 분석 기술의 급속한 발전에 힘입어 등장한 최신 육종기술로, 품종 개량에 오랜 시간이 소요되는 유전자 표지 기반의 선발육종기술의 한계를 극복할 수 있는 혁신기술로 주목받고 있으며(Boudry et al., 2021), 개체의 유전체 정보를 이용하여 수정란 단계에서도 유전능력을 평가할 수 있어 조기선발이 가능하며 양적형질 개량에 적합하다(Garcia et al., 2018; D'Agaro et al., 2021). 또한 유전체 선발기술은 single nucleotide polymorphism (SNP) 칩과 전장유전체연관분석(genome wide association study, GWAS) 등의 기반기술을 바탕으로 수산분야 육종에 활발하게 적용되고 있다(Zenger et al., 2019; Fraslin et al., 2022).

GWAS는 양적형질과 기민한 연관성을 가지는 유전자 좌위를 확인하기 위한 분석 기술이며, 표현형과 유전형 간의 연관성을 비교하는 방법으로 다양한 모델이 소개되었다. 그 중 주성분과 kinship을 고려한 혼합선형모델 혹은 SNP 효과만을 고려한 일반선형모델이 주로 활용되고 있다(Liu et al., 2016; Bonnafous et al., 2018; Kaler et al., 2019). Liu et al. (2016)는 기존의 주로 사용되는 두 방식이 위양성(false positive)을 효율적으로 줄이는 반면 실제 유의미한 SNP도 마찬가지로 유의미하지 않다고 판단하는 위음성(false negative) 오류를 확인하였고, 이에 multilocus를 기반으로 한 FarmCPU (fixed and random model circulating probability unification) 모델을 기존의 혼합선형모델에서 집단 구조와 kinship을 고려한 모델보다 한층 효율적이고 위양성 문제를 더 잘 해결하는 모델로 제시하였다(Liu et al., 2016). 하지만 FarmCPU 모델을 활용한 GWAS는 이전 모델들에 비교하여 그 신뢰성 및 효율성이 아직 다방면으로 입증되지 않았기 때문에 많이 활용되고 있지 않다.

본 연구에서는 우리나라 주요 양식품종인 넙치를 대상으로 성장형질(표현형)과 유전형의 연관성 분석을 위하여 GWAS의 여러 모델을 비교 분석하였다. 이를 위해 4가지 혼합선형 모델과 FarmCPU 모델을 구성하여 총 5가지 다른 방식의 모델을 활용한 GWAS 결과를 비교하였다. 이를 통하여 넙치의 GWAS를 위한 최적 모델을 탐색하고 성장형질과 연관성이 높은 유전자 변이를 도출하여 넙치 유전체 선발 및 넙치 생산성 향상을 위한 기초 자료를 제공하고자 하였다.

재료 및 방법

시험어 생산, 성장 측정 및 유전자형 분석 개체 선정

본 연구에서는 국립수산물과학원 울릉연구센터의 넙치 핵집단 6세대와 7세대를 이용하였다. 핵집단 6세대는 유전적 다양성이 확보된 친어 집단에서 각 개체의 유전적 유연관계를 고려하여

최대의 선발효과와 근친도의 적정 균형을 유지하는 최적 교배 지침에 따라 2017년 암수 1:1 인공수정을 통해 245가계로 구성되어 생산되었다. 핵집단 7세대는 핵집단 6세대를 친어로 활용하여 각 개체의 유전적 유연관계를 고려하여 최대의 선발효과와 근친도의 적정 균형을 유지하는 최적 교배지침에 따라 2019년 암수 1:1 인공수정을 통해 351가계로 구성되어 생산되었다. 사육 수조에 따른 환경 효과를 최소화하기 위하여 전 가계를 혼합사육 하였으며 부화 후 11개월령, 18개월령, 22개월령에 체중, 전장을 측정하였고 비만도는 [체중(g)/전장(cm)] × 100의 값으로 산출하였다. 본 연구에 활용할 개체 선발을 위하여 넙치 핵집단 6세대, 7세대의 친자확인을 수행하였으며, 가계별 성장형질 순위와 가계 내 개체수를 고려하여 넙치 핵집단 6세대 52마리와 7세대 399마리를 유전자형 분석 대상으로 선정하였다.

DNA 추출 및 유전자형 분석

넙치 가슴지느러미 조직을 채취하여 후 MagExtractor DNA Multi-Prep Kit (TNT research)을 이용하여 genomic DNA를 추출하였다. 조직 시료의 lysis는 55°C에서 overnight으로 수행하였으며, 이후 RNase A (100 mg/mL) 3 µL를 넣고 37°C에서 15분 반응시켰다. 이후 KingFisher Flex Purification System의 제공되는 프로토콜에 의거하여 genomic DNA를 정제하였다. 추출된 DNA는 전기영동을 통해 1차 QC (Quality control) 과정을 진행 후 Picogreen assay로 2차 QC를 진행하였다. 최종적으로 QC가 통과된 시료의 DNA 150 ng을 이용하여 Axiom Mini 96 myDesign Custom Array 플랫폼 기반으로 60K SNP chip을 활용하여 개체별 유전자형을 분석하였다. 유전자형 정보는 단일염기다형성(SNP)으로 구성되며, 유전자형 정보의 QC를 위해 Plink 1.9 (Purcell et al., 2007)를 이용하여 유전자형 결측이 5% 이상, minor allele frequency가 5% 미만, 하디-와인버그 평형 검정(Hardy-Weinberg equilibrium test)의 P-value가 0.0001 미만인 SNP을 제거함으로써 최종적으로 총 42,034개의 SNPs을 GWAS에 이용하였다.

전장유전체연관분석(GWAS)

성장형질 연관 유전자 변이를 탐색하기 위하여 1) 성별을 고정효과로 고려한 혼합선형모델; 2) 1)의 모델에 추가적으로 주성분을 고정효과로 고려한 혼합선형모델(Purcell et al., 2007); 3) 1)의 모델에 유전연관행렬을 임의효과로 고려한 혼합선형모델; 4) 2)의 모델에 유전연관행렬을 임의효과로 고려한 혼합선형모델(Yang et al., 2011a); 5) FarmCPU (Yin et al., 2021) 총 다섯 가지 모델의 GWAS를 수행하였다.

GWAS를 통해 도출된 각 SNP의 P-value의 유의수준은 0.05로 설정하였으며, 다중검정에 대한 보정으로 본페로니 교정(Holm, 1979) (Bonferroni correction)을 적용하여 최종적으로 $-\log_{10} P > 5.92$ 인 SNPs을 유의미한 연관성을 가지는 유전자 변이라 판단하였다. GWAS 결과의 적합도는 Quantile-quantile plot (Q-Q plot)으로 확인하였다. Q-Q plot의 분포를 수직으로 나타내

어 전체 관측된 P-value 분포의 inflation/deflation 여부를 보여주는 genomic inflation factor (λ_{GC})의 산출은 아래의 공식을 이용하였다.

$$\lambda_{GC} = \frac{\log_{10}(p_{obs}[0.5])}{\log_{10}(0.5)}$$

(Yamamoto and Matsunaga, 2021)

여기서 $p_{obs}[0.5]$ 는 관찰된 P-value의 0.5 분위를 나타낸다. 각 모델별로 산출된 λ_{GC} 을 기준으로 알고리즘 적합도를 판단하여 넙치를 대상으로 하는 최적 GWAS 모델을 선정하였다.

유전자 기능 조사

넙치를 대상으로 하는 최적 GWAS 모델을 이용해 도출된 성장형질 연관 유전자 변이의 기능을 조사하기 위하여, 선정된 최적 GWAS 모델을 이용하였을 때 각 형질별로 가장 연관성이 높게 나타난 유전자 변이 염기와 양 옆 500 bp을 포함한 총 1001 bp를 하나의 flanking 서열로 취급하여 BLAST (v2.12.0) (Altschul et al., 1990)의 query로 활용하고, nt 데이터베이스에서 e-value<10⁻²⁰의 cut-off를 적용하여 해당 서열과 일치하는 유전자를 nucleotide 수준에서 확인한 후 성장형질과의 관련성을 Uniprot (www.uniprot.org)의 통합 유전자 검색을 통해 조사하였다.

결과 및 고찰

표현형 정보

GWAS에 활용한 넙치 451마리의 성장형질 기초 통계량은 Table 1과 같다. 11개월령, 18개월령 및 22개월령에 측정된 체중(body weight), 전장(total length), 비만도(condition factor)의 평균 및 표준편차는 각각 601.90 ± 96.96 g, 1413.86 ± 324.75 g, 2190.49 ± 591.06 g, 36.55 ± 1.88 cm, 48.83 ± 3.68 cm, 53.87 ± 4.30 cm, 12.24 ± 0.73, 11.93 ± 0.81, 13.66 ± 1.28이었다. 체중 및 전장과는 다르게 비만도는 18개월령에서 11개월령보다 더 낮았으며 22개월령에서 가장 높은 것으로 나타났다.

GWAS Quantile-quantile plot 비교

본 연구에서는 singlelocus 모델을 대표하는 혼합선형모델과, multilocus 모델이며 콩(*Glycine max* L.)과 옥수수(*Zea mays* L.)에서 여러 모델과 성능 비교가 검증된(Kaler et al., 2019) FarmCPU를 넙치의 9가지 형질에 대해 총 5가지 모델을 설정하여 GWAS를 수행하고 그 결과를 비교하였다. 이를 위해 성별을 고정효과로 고려한 혼합선형모델(모델 1), 성별 및 상위 10개의 주성분을 고정효과로 고려한 혼합선형모델(모델 2), 성별 및 유전연관행렬을 각각 고정효과 및 임의효과로 고려한 혼합선형모델(모델 3), 성별 및 상위 10개의 주성분을 고정효과로, 유전연관행렬을 임의효과로 고려한 혼합선형모델(모델 4), FarmCPU 모델(모델 5)을 구성하였다. 모델 4는 multilocus 모델, 나머지 세 모델은 모두 singlelocus 모델이다. Multilocus 모델은 모든 유전자 변이를 동시에 측정하는 방법을 통해 singlelocus 모델을 활용한 GWAS 결과가 polygenic background의 영향을 받는 quantitative trait locus (QTL)이 본페로니 threshold에 의해 걸러지는 결점을 보완한 방법이다. Multilocus 모델은 1차적으로 잠재적인 형질 연관 유전자를 선별하고, 2차적으로 그 안에서 최대우도추정법을 통해 실제로 유의미한 QTL만 골라내는 알고리즘을 활용한다(Zhong et al., 2021). 또한, 모델의 GWAS 적합도를 확인하기 위해 보편적으로 쓰이는 방법으로 quantile-quantile plot (Q-Q plot)이 있다(Stich and Melchinger, 2009; Riedelsheimer et al., 2012; Kristensen et al., 2018). Q-Q plot의 x축은 표준 정규분포를 따르는 기대 P-value, y축은 실제 관측된 P-value에 각각 -log₁₀를 취한 값으로 그려진다. 관측된 P-value 분포의 inflation/deflation 여부를 보여주는 λ_{GC} 값이 1.1보다 작다면 집단층화(population stratification)가 거의 없이 이상적인 분포로서 판단하고, 1.1을 넘어가면 이는 집단층화에 의한 것이거나, 자료처리 오류, 배치효과 또는 비임의적 유전자형 분석 실패에 의해 야기된 것으로 여겨진다(Kim and Hong, 2011; Yang et al., 2011b). 이를 바탕으로 5가지 모델의 적합성을 비교하기 위해 GWAS 결과의 Q-Q plot과 λ_{GC} 를 확인하였다(Fig 1). 모델 1은 모든 형질에서 λ_{GC} 이 높았으며(1.2-1.69), 과도한 inflation이 발생하였다. 비슷하게, 모델 2는 모델 1보다는 λ_{GC} 이 낮은 수준이었지만(1.1-1.35) 22개월령 체중을 제외

Table 1. Summary statistics for growth traits in olive flounder *Paralichthys olivaceus*

Month	Body weight (g)			Total length (cm)			Condition factor		
	11	18	22	11	18	22	11	18	22
Count	451	451	451	451	451	451	451	451	451
Mean	601.90	1413.86	2190.49	36.55	48.83	53.87	12.24	11.93	13.66
SD	96.96	324.75	591.06	1.88	3.68	4.30	0.73	0.81	1.28
Minimum	358	660	990	30.0	39.2	42.5	10.2	10.0	11.0
Maximum	897	2231	3742	42.5	57.2	63.5	14.9	14.6	19.6
Range	539	1571	2752	12.5	22.0	21.0	4.7	4.6	8.6

SD, Standard deviation.

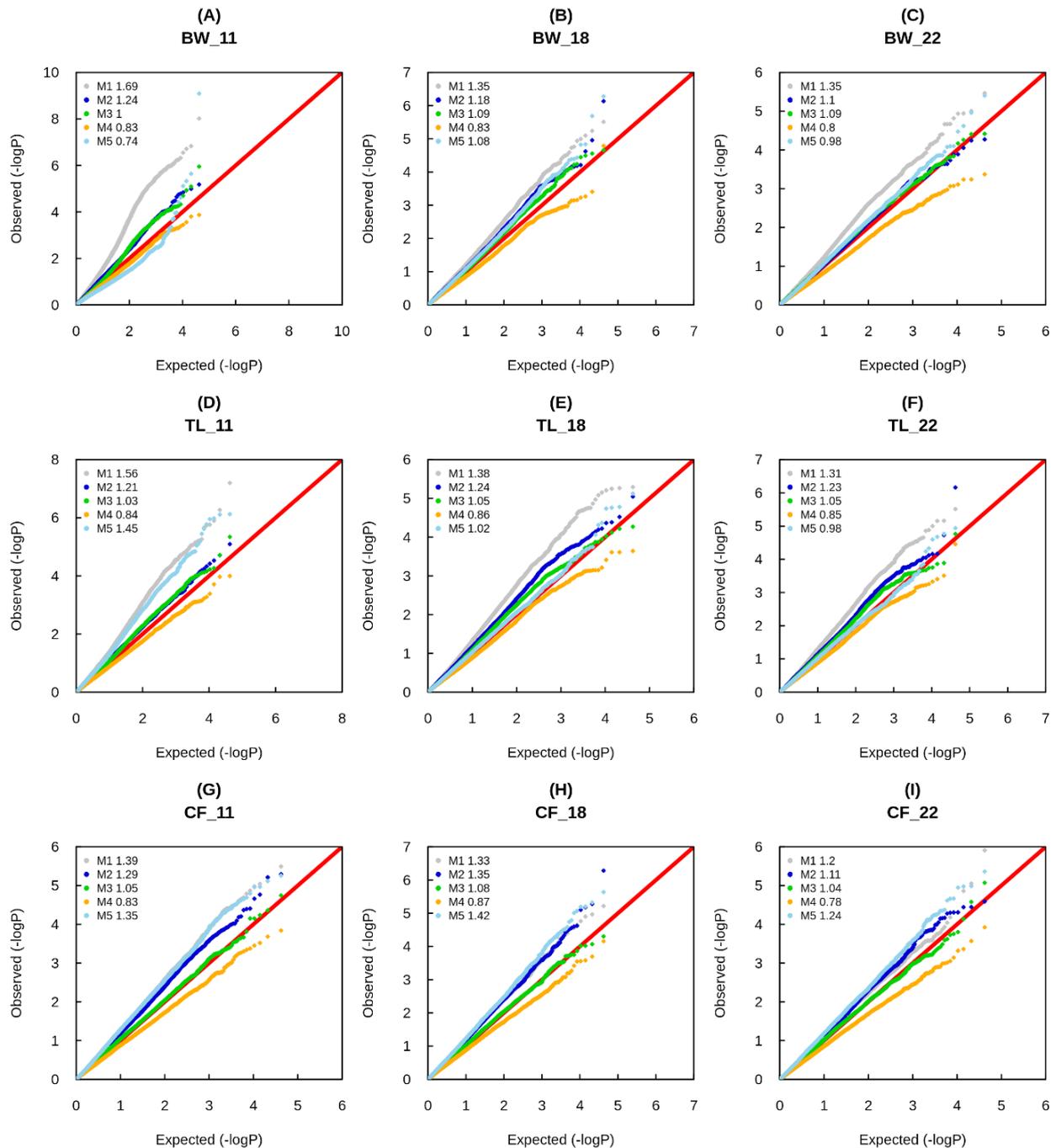


Fig. 1. P-values distribution of quantile-quantile (Q-Q) plots of the five models for nine quantitative traits in olive flounder *Paralichthys olivaceus*. BW_11, 11 months body weight (A); BW_18, 18 months body weight (B); BW_22, 22 months body weight (C); TL_11, 11 months total length (D); TL_18, 18 months total length (E); TL_22, 22 months total length (F); CF_11, 11 months condition factor (G); CF_18, 18 months condition factor (H); CF_22, 22 months condition factor (I); M1, mixed linear model (MLM) corrected for gender as a fixed effect; M2, MLM corrected for gender and top 10 principal components (PC) as fixed effects, M3, MLM corrected for gender and genetic relationship matrix (GRM) as fixed and random effects, respectively; M4, MLM corrected for sex and top 10 PCs as fixed effects and GRM as random effects; M5, Fixed and random model Circulating Probability Unification (FarmCPU). X-axis indicates expected distribution of $-\log_{10}P$ based on chi-square distribution while y-axis indicates observed. Red line means 1:1 line of expected vs. observed $-\log_{10}P$. Numbers next to model in legend indicate λ_{GC} for each model.

한 모든 형질에서 inflation이 발생하였다. 모델 3에서는 λ_{GC} 이 inflation되지 않고 안정적인 모습을 보였다(1.00–1.09). 모델 4는 1보다 작은 값의 λ_{GC} 를 가졌으며(0.78–0.87) 전체적으로 deflation된 결과가 나타났다. 마지막으로, FarmCPU를 적용한 모델 5에서는 전체 형질에 있어 부분적으로 안정된 λ_{GC} 값을 보였지만(Fig. 1B, 1C, 1E, 1F) 몇몇 형질에서 높거나(Fig. 1D, 1G, 1H, 1I) 낮은(Fig. 1A) 값을 보이며 큰 격차(0.74–1.45)를 나타내었다.

유의한 유전자 변이 수 비교

넙치의 9가지 성장형질에 대해 5가지 모델을 적용하고, 본페로니 교정을 통해 다중검정에 대한 보정을 진행하여 GWAS를 수행하였을 때 유의한 유전자 변이의 수는 Table 2와 같다. 모델 1, 모델 2 및 모델 5의 경우 다른 모델에 비해 도출된 유의한 유전자 변이의 수가 많았다. 모델 3은 11개월령 체중과 연관된 단 하나의 유의한 유전자 변이가 검출되었으며, 모델 4의 경우 성장형질과 연관된 유의한 유전자 변이가 검출되지 않았다. 종합적으로, 모든 모델에서 22개월령 체중, 18개월령 전장, 11개월령 비만도 및 22개월령 비만도와 연관된 유의한 유전자 변이가 검출되지 않았다. 이는 본페로니 교정이 다중검정에 대한 보정 방법 중에서 몹시 보수적이기 때문에 위음성이 나타날 가능성이 있어(Narum, 2006), 모델 간 유의한 유전자 변이의 차이를 명백하게 구분하기 위해 suggestive ($-\log_{10}P>3.5$)를 기준으로 설정하여 유의한 유전자 변이를 재해석하였다(Table 2). 모든 형질의 GWAS 결과가 높은 수준의 λ_{GC} 를 보였던 모델 1은 다른 모델에 비해 과도하게 많은 수의 유의한 유전자 변이가 검출되었다. 모든 형질의 GWAS 결과가 두 번째로 높은 수준의 λ_{GC} 를 보였던 모델 2는 안정된 λ_{GC} 값을 가지는 모델 3보다 많은 수의 유의한 유전자 변이가 검출되었다. 모든 형질의 GWAS 결과가 가장 낮은 수준의 λ_{GC} 를 보였던 모델 4는 다른 모델에 비해 유의한 유전자 변이의 수가 적게 나타났다. 모델 5의 경우

모든 형질의 GWAS 결과가 높은 수준의 λ_{GC} 를 보였던 11개월령 전장과 11, 18, 22개월령 비만도와 연관된 유의한 유전자 변이의 수가 많았으며, 낮은 수준의 λ_{GC} 를 보였던 11개월령 체중 및 22개월령 전장과 연관된 유의한 유전자 변이의 수가 적었다.

넙치 GWAS 적합모델 선정

Kaler et al. (2019)에 따르면, Q-Q plot에서 예상값(expected)와 관측값(observed)에 대해서 1:1 line에 가까우며 후반부에 우상향 꼬리가 있는 경우, 위양성과 위음성이 잘 통제되었으며 형질과 관련된 유전자 변이가 잘 탐지되었음을 의미한다. 이러한 관점에서 각 모델들이 넙치 성장형질 연관 유전자 변이 탐색에 적합한지 확인하기 위해 각 모델 및 형질별 Q-Q plot 결과를 비교해 보았다. 모델 1과 모델 2의 $-\log_{10}P$ 값은 모든 형질에서 다른 모델들에 비해 1:1 line과 멀리 분포하였으며 후반부에 상향하는 꼬리가 거의 발견되지 않았다(Fig 1). 또한 $-\log_{10}P$ 가 inflation된 λ_{GC} 값을 가지기 때문에 넙치 성장형질 연관 유전자 변이 탐색에 적합하지 못한 모델이라 판단하였다. 모델 4의 경우 모든 형질에서 deflation이 일어나 $-\log_{10}P$ 값이 1:1 line 아래로 형성되었고 후반부에 상향하는 꼬리가 거의 나타나지 않았을 뿐만 아니라, suggestive를 기준으로 유의한 유전자 변이를 구분하였을 때 그 수가 명백히 적었기 때문에 위음성이 발생하였을 가능성이 높을 것이라 판단하여 넙치 성장형질 연관 유전자 변이 탐색에 부적합한 모델이라 판단하였다. 한편, 모델 3의 $-\log_{10}P$ 값은 후반부에 상향하는 꼬리가 거의 나타나지 않았지만, 모든 형질에서 1.0–1.1 범위의 inflation되지 않은 λ_{GC} 값을 가지고 있고 1:1 line과 가깝게 위치하였기 때문에 모델 3을 넙치 성장형질 연관 유전자 변이 탐색에 적합한 모델로 선정하였다. 또한, 모델 5의 $-\log_{10}P$ 값은 λ_{GC} 의 편차가 크지만, 18개월령 체중 및 전장에서 1.0–1.1 범위를 가지는 λ_{GC} 값과 후반부에 우상향 꼬리가 발견되었기 때문에 모델 5를 넙치 성장형질 연관 유전자 변이 탐색에 적합한 모델로 선정하였다.

Table 2. The number of significant SNPs identified by suggested thresholds in olive flounder *Paralichthys olivaceus*

Month	Model	Body weight			Total length			Condition factor			Total SNPs
		11	18	22	11	18	22	11	18	22	
Suggestive ($-\log_{10}P>3.5$)	1	504	80	66	241	107	105	80	48	25	1256
	2	63	49	17	22	50	41	49	53	37	381
	3	65	24	15	30	17	22	12	15	8	208
	4	3	1	0	3	3	2	3	5	2	22
	5	10	46	28	149	18	12	75	70	49	457
Bonferroni ($-\log_{10}P>5.92$)	1	16	0	0	2	0	0	0	0	0	18
	2	0	1	0	0	0	1	0	1	0	3
	3	1	0	0	0	0	0	0	0	0	1
	4	0	0	0	0	0	0	0	0	0	0
	5	1	1	0	4	0	0	0	0	0	6

SNP, Single nucleotide polymorphism.

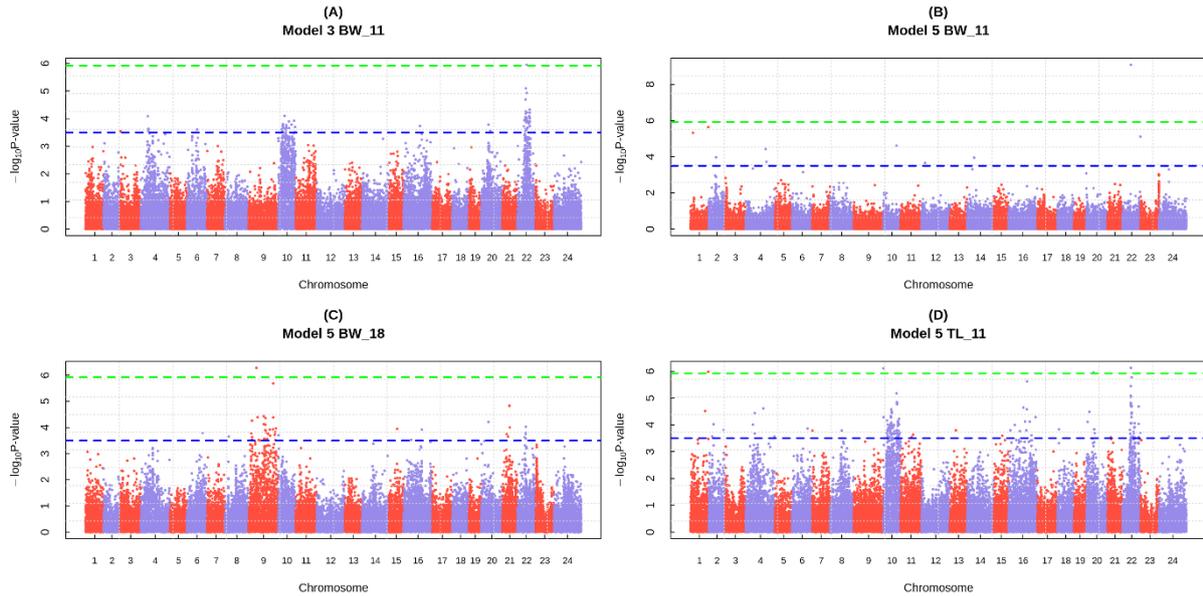


Fig. 2. Manhattan plot showing the genome-wide significant loci associated with growth traits of olive flounder *Paralichthys olivaceus*. Model 3 BW_11, 11 months body weight for mixed linear model (MLM) corrected for gender and genetic relationship matrix (GRM) as fixed and random effects, respectively (A); Model 5 BW_11, 11 months body weight for Fixed and random model Circulating Probability Unification (FarmCPU) (B); Model 5 BW_18, 18 months body weight for FarmCPU (C); Model 5 TL_11, 11 months total length for FarmCPU (D); X-axis indicates the 24 chromosomes while y-axis indicates $-\log_{10}P$. The green dash indicates Bonferroni ($-\log_{10}P > 5.92$) threshold. The blue dash indicates suggestive ($-\log_{10}P > 3.5$) threshold.

적합모델의 유의미한 유전자 변이 비교

5가지 모델 중 선별된 두 모델(모델 3, 모델 5)의 GWAS 결과로 도출된 유의한 유전자 변이를 비교하기 위해 전체 형질 중 본페로니 교정($-\log_{10}P > 5.92$)을 적용하였을 때 유의미하게 나온 유전자 변이를 살펴보았다. 그 결과, 모델 3의 11개월령 체중, 모델 5의 11개월령 체중, 18개월령 체중, 11개월령 전장과 관련된 유의미한 유전자 변이가 발견되었다(Fig. 2, Table 3). 이 중에서, 22번 염색체에서는 특히 모델 3의 11개월령 체중과 관련된 유의한 유전자 변이가 발견되었고, 모델 5 역시 11개월

령 체중과 전장과 연관된 유의한 유전자 변이가 같은 좌위에서 발견되었다(AX-605207160). 또한 해당 유전자 변이들의 좌위는 12,779,405-13,593,447 bp으로, 상당히 가까이 위치하였다.

모델 신뢰도 검증 및 결론

넙치를 대상으로 GWAS를 하기 적합한 모델로 선정된 두 모델에서, 모델 3은 $-\log_{10}P$ 값의 후반부에 상향하는 꼬리가 거의 나타나지 않았으며 모델 5는 $-\log_{10}P$ 값의 λ_{GC} 의 편차가 크다는 문제점이 존재한다. 해당 문제점에 대해 유의미하게 나온 유

Table 3. Significant SNPs associated with growth traits using Bonferroni correction in olive flounder *Paralichthys olivaceus*

	Trait	Chromosome	Position (bp)	Name	P-value
Model 3	BW_11	22	13593447	AX-604936826	1.12e-06
	BW_11	22	12779405	AX-605207160	8.13e-10
	BW_18	9	7830374	AX-605017214	5.53e-07
Model 5		22	12779405	AX-605207160	7.50e-07
	TL_11	10	80893	AX-604993421	7.80e-07
		1	30162663	AX-604983128	1.05e-06
		20	8517150	AX-604977869	1.12e-06

SNP, Single nucleotide polymorphism; BW_11, 11 months body weight; BW_18, 18 months body weight; TL_11, 11 months total length.

전자 변이를 유전자의 기능 관점에서 확인하기 위해 BLAST를 진행하였다. 그 결과, 성장형질과 관련된 유전자가 모델 3, 모델 5에서 각각 3개, 1개가 발견되었다. 모델 3의 경우, 18개월령 전장과 관련된 유전자 변이가 *Cbl proto-oncogene B (cblb)* 영역에 존재하였으며 해당 유전자는 인간(*Homo sapiens*)에서 epidermal growth factor으로부터 유도된 세포의 성장을 억제하고(Ettenberg et al., 1999), 대서양 연어(*Salmo salar*)에서 단백질에 의한 분해를 촉진시키는 것으로 밝혀진 바 있다. 또한 22개월령 전장과 관련된 유전자 변이가 *nestin-like* 영역에 존재하였으며 해당 유전자는 연어에서 지질 결합 및 세포 내 단백질 운반에 관여하는 것으로 밝혀진 바 있다. 마지막으로, 11개월령 비만도와 연관된 유전자 변이는 *rho guanine nucleotide exchange factor 28-like* 영역 내에 존재하였으며 해당 유전자는 연어에서 세포 성장에 관여하는 것으로 밝혀진 바 있다. 모델 5의 경우 18개월령 전장과 연관된 유전자 변이가 *SOX14* 영역에 존재하였으며 해당 유전자는 제브라피쉬(*Danio rerio*)에서 번역과정의 negative regulator로 작용하고 세포 분화에 관여하는 것으로 밝혀진 바 있다. 모델 3 및 모델 5에서 모두 성장과 관련된 유전자가 확인되었으며 이는 모델 3과 모델 5가 넙치를 대상으로 하는 GWAS로써 신뢰할 만한 결과를 나타내었음을 뒷받침한다. 또한 22번 염색체의 12,779,405–13,593,447 bp 좌위에서 두 모델 모두 성장형질과 연관된 유의한 유전자 변이가 발견되어 성장형질에 대한 QTL이 존재할 것으로 예상되며, 추가 분석을 통해 해당 좌위와 연관된 유전자의 기능이 밝혀진다면 두 모델에 대한 신뢰도가 더욱 증가할 것으로 사료된다. 이러한 결과를 바탕으로, 넙치를 대상으로 하는 GWAS 알고리즘으로 고정효과와 임의효과를 각각 성별 및 유전연관행렬을 보정한 MLM 방법(모델 3)과 FarmCPU 방법(모델 5)을 활용한다면 넙치 성장 연관 유전자 변이를 효율적으로 검출하여 유전체 선발육종에 활용할 수 있을 것으로 기대한다.

사 사

본 연구는 2023년도 해양수산부 국립수산물과학원 수산과학 연구사업(디지털 정보기반 넙치 스마트육종 실용화 기술 개발, R2023031)의 지원에 의해 수행되었음.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Behjati S and Tarpey PS. 2013. What is next generation sequencing?. *Arch Dis Child Educ Pract* 98, 236-238. <https://doi.org/10.1136/archdischild-2013-304340>.

Bonnaïfous F, Fievet G, Blanchet N, Boniface MC, Carrere S, Gouzy J, Legrand L, Marage G, Bret-Mestries E, Munos S, Pouilly N, Vincourt P, Langlade N and Mangin B. 2018.

Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. *Theor Appl Genet* 131, 319-332. <https://doi.org/10.1007/s00122-017-3003-4>.

Boudry P, Allal F, Aslam ML, Bargelloni L, Bean TP, Brard-Fudulea S, Briec MS, Calboli FC, Gilbey J and Haffray P. 2021. Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. *Aquac Rep* 20, 100700. <https://doi.org/10.1016/j.aqrep.2021.100700>.

D'Agaro E, Favaro A, Matiussi S, Gibertoni PP and Esposito S. 2021. Genomic selection in salmonids: New discoveries and future perspectives. *Aquac Int* 29, 2259-2289. <https://doi.org/10.1007/s10499-021-00747-w>.

Ettenberg SA, Keane MM, Nau MM, Frankel M, Wang LM, Pierce JH and Lipkowitz S. 1999. Cbl-b inhibits epidermal growth factor receptor signaling. *Oncogene* 18, 1855-1866. <https://doi.org/10.1038/sj.onc.1202499>.

Fraslin C, Koskinen H, Nousianen A, Houston RD and Kause A. 2022. Genome-wide association and genomic prediction of resistance to *Flavobacterium columnare* in a farmed rainbow trout population. *Aquaculture* 557, 738332. <https://doi.org/10.1016/j.aquaculture.2022.738332>.

Fuji K, Hasegawa O, Honda K, Kumasaka K, Sakamoto T and Okamoto N. 2007. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* 272, 291-295. <https://doi.org/10.1016/j.aquaculture.2007.07.210>.

Fuji K, Kobayashi K, Hasegawa O, Coimbra MRM, Sakamoto T and Okamoto N. 2006. Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* 254, 203-210. <https://doi.org/10.1016/j.aquaculture.2005.11.024>.

Garcia AL, Bosworth B, Waldbieser G, Misztal I, Tsuruta S and Lourenco DA. 2018. Development of genomic predictions for harvest and carcass weight in channel catfish. *Genet Sel Evol* 50, 66. <https://doi.org/10.1186/s12711-018-0435-5>.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Statist* 6, 65-70.

Kaler AS, Gillman JD, Beissinger T and Purcell LC. 2019. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front Plant Sci* 10, 1794. <https://doi.org/10.3389/fpls.2019.01794>.

Kim JH, Lee JH, Kim HC, Noh JK, Kang JH and Kim KK. 2011. Body shape and growth in reciprocal crosses of wild and farmed olive flounder, *Paralichthys olivaceus*. *J World Aquac Soc* 42, 268-274. <https://doi.org/10.1111/j.1749-7345.2011.00463.x>.

Kim SJ and Hong KS. 2011. Genome-wide association study

- in psychiatric disorders. *J Korean Neuropsychiatric Assoc* 50, 20-38.
- KOSIS (Korean Statistical Information Service). 2022. Cultivation Status by Fish Species by Farming Method by City/Si/Gun/Gu. Retrieved from https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1EZ0008&conn_path=I2 on May 22, 2023.
- Kristensen PS, Jahoor A, Andersen JR, Cericola F, Orabi J, Janss LL and Jensen J. 2018. Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Front Plant Sci* 9, 69. <https://doi.org/10.3389/fpls.2018.00069>.
- Liu X, Huang M, Fan B, Buckler ES and Zhang Z. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12, e1005767. <https://doi.org/10.1371/journal.pgen.1005767>.
- Min B, Kim HC, Lee JH, No J, An H, Park C, Choi S and Myeong J. 2010. Comparison of growth parameters in selected and unselected strains of olive flounder *Paralichthys olivaceus*. *Korean J Fish Aquat Sci* 43, 457-461. <https://doi.org/10.5657/kfas.2010.43.5.457>.
- Nam BH, Kwak W, Kim YO, Kim DG, Kong HJ, Kim WJ, Kang JH, Park JY, An CM, Moon JY, Park CJ, Yu JW, Yoon J, Seo M, Kim K, Kim DK, Lee S, Sung S, Lee C, Shin Y, Jung M, Kang BC, Shin GH, Ka S, Caetano-Anolles K, Cho S and Kim H. 2017. Genome sequence of pacific abalone (*Haliotis discus hannai*): The first draft genome in family Haliotidae. *GigaScience* 6, gix014. <https://doi.org/10.1093/gigascience/gix014>.
- Narum SR. 2006. Beyond bonferroni: Less conservative analyses for conservation genetics. *Conserv Genet* 7, 783-787. <https://doi.org/10.1007/s10592-005-9056-y>.
- Ogata HY, Oku H and Murai T. 2002. Growth, feed efficiency and feed intake of offspring from selected and wild Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* 211, 183-193. [https://doi.org/10.1016/S0044-8486\(01\)00798-0](https://doi.org/10.1016/S0044-8486(01)00798-0).
- Park J, Lee D, Jung HS, Kim J, Yang H, Kim H and Lee J. 2022. Estimation of genetic parameter for growth traits of olive flounder *Paralichthys olivaceus* on the 8th generation of selective breeding using multiple traits animal model. *Korean J Fish Aquat Sci* 55, 549-556. <https://doi.org/10.5657/KFAS.2022.0549>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575. <https://doi.org/10.1086/519795>.
- Riedelsheimer C, Lisek J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L and Melchinger AE. 2012. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc Natl Acad Sci* 109, 8872-8877. <https://doi.org/10.1073/pnas.1120813109>.
- Shao C, Bao B, Xie Z, Chen X, Li B, Jia X, Yao Q, Ortí G, Li W, Li X, Hamre K, Xu J, Wang L, Chen F, Tian Y, Schreiber AM, Wang N, Wei F, Zhang J, Dong Z, Gao L, Gai J, Sakamoto T, Mo S, Chen W, Shi Q, Li H, Xiu Y, Li Y, Xu W, Shi Z, Zhang G, Power DM, Wang Q, Schartl M and Chen S. 2017. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet* 49, 119-124. <https://doi.org/10.1038/ng.3732>.
- Shin GH, Shin Y, Jung M, Hong JM, Lee S, Subramaniyam S, Noh ES, Shin EH, Park EH, Park JY, Kim YO, Choi KM, Nam BH and Park CI. 2018. First draft genome for red sea bream of family Sparidae. *Front Genet* 9, 643. <https://doi.org/10.3389/fgene.2018.00643>.
- Stich B and Melchinger AE. 2009. Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. *BMC Genomics* 10, 94. <https://doi.org/10.1186/1471-2164-10-94>.
- Yamamoto E and Matsunaga H. 2021. Exploring efficient linear mixed models to detect quantitative trait locus-by-environment interactions. *G3 Genes Genom Genet* 11, jkab119. <https://doi.org/10.1093/g3journal/jkab119>.
- Yang J, Lee SH, Goddard ME and Visscher PM. 2011a. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76-82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'Connell JR, Mangino M, Magi R, Madden PA, Heath AC, Nyholt DR, Martin NG, Montgomery GW, Frayling TM, Hirschhorn JN, McCarthy MI, Goddard ME, Visscher PM and Consortium G. 2011b. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 19, 807-812. <https://doi.org/10.1038/ejhg.2011.39>.
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X and Liu X. 2021. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinform* 19, 619-628. <https://doi.org/10.1016/j.gpb.2020.10.007>.
- Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR and Raadsma HW. 2019. Genomic selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front Genet* 9, 693. <https://doi.org/10.3389/fgene.2018.00693>.
- Zhong H, Liu S, Sun T, Kong W, Deng X, Peng Z and Li Y. 2021. Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol* 21, 364. <https://doi.org/10.1186/s12870-021-03146-8>.