

Letter to the editor

Open Access

## Whole-genome resequencing infers genomic basis of giant phenotype in Siamese fighting fish (*Betta splendens*)

Understanding the genetic basis of phenotypes is of importance in evolutionary biology and for genetic improvement of economically valuable animals. The giant phenotype of the fighting fish, *Betta splendens*, provides a unique opportunity to explore the genetic architecture of overgrowth in body size. As such, we re-sequenced and analyzed the genomes of 54 fighting fish. Genome-wide  $F_{ST}$  and selective sweep analyses using 3 582 429 DNA variants revealed three genomic regions at chr1, chr9, and chr11 that were associated with the giant phenotype. With a total length of ~3.5 Mb, these regions showed high divergence between the giant and non-giant bettas. In contrast, no signature of selection was detected in the wild-type fish. Transcriptome analysis of brain and muscle samples from giant and normal bettas identified 14 candidate genes that were likely responsible for the giant phenotype. Overall, our data provide novel insights into the genetic basis of body size variation. The genome sequences, transcriptome sequences, DNA sequence variants, and candidate genes for body size provide valuable resources for further biological and evolutionary studies, as well as for rapid improvement in growth-related traits.

Domestication provides an excellent resource for understanding the genetic basis of phenotypic variations. Genetic studies on domestication have primarily focused on agronomic crops, livestock, and pets. Domesticated fish species, particularly fighting fish and goldfish, have been bred into many stable forms, providing additional strains for elucidating the bridge between genetics and diverse phenotypes. The Siamese fighting fish (*B. splendens*), which originated from the Mekong River basin, has been selected for both ornamental and competitive purposes for several hundred years, leading to considerable diversity in

pigmentation pattern, fin shape, behavior, and body size. Certain traits, including double tails, fin spots, albinism, and elephant ears, are likely determined by major-effect loci (Wang et al., 2021). However, for the giant fighting fish, which is three times larger than the wild-type strain (Figure 1A), the molecular mechanism underlying its giant body size phenotype remains unknown.

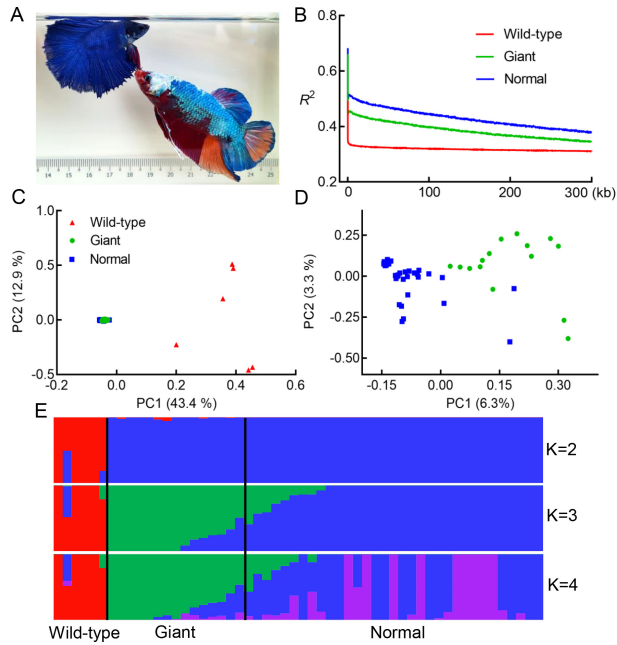
In the current study, we resequenced the genomes of 54 fighting fish and detected 3 582 429 DNA variants (Supplementary Methods). Based on these variants, the nucleotide diversity of the giant fighting fish was 0.00073 in a 100 kb window size, slightly higher than that of normal ornamental bettas ( $P_i=0.00067$ ,  $t$ -test,  $P<1.5\times 10^{-10}$ ), but four times lower than that of the wild-type strain ( $P_i=0.00271$ ,  $t$ -test,  $P<1.3\times 10^{-33}$ ). The significant loss of genetic diversity in the giant fighting fish compared to the wild-type fighting fish suggests founder events during domestication of this particular strain (Wang et al., 2021). However, genetic diversity of the giant fighting fish was slightly higher than that of the ornamental bettas, implying that the giant fighting fish may have a more recent wild origin or resulted from repeated selection of giant-related alleles from wild-type mutants. We observed that linkage disequilibrium (LD) decayed more rapidly in the wild-type fish than in the normal ornamental bettas (Figure 1B). Both the giant and ornamental bettas clearly diverged from their putative wild-type ancestors, as well as from each other (Figure 1C, D). This differs from our previous observations, where individuals with the same traits were not always assigned to the same genetic clusters (Wang et al., 2021). Here, admixture analysis revealed that, at  $K=2$ , the wild-type and ornamental strains were assigned to two distinct genetic clusters, whereas at  $K=3$ , the wild-type, giant, and ornamental strains were classified into three independent clusters, respectively (Figure 1E). However, considerable admixture was observed between seven giant fighting fish and nine normal ornamental bettas, and several giant fighting fish

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 18 September 2021; Accepted: 03 December 2021; Online: 05 December 2021

Foundation items: This research was supported by the Internal Funds of the Temasek Life Sciences Laboratory (5020)



**Figure 1 Population structure of fighting fish (*Betta splendens*)**

A: Giant and normal bettas. B: LD decay among wild-type, giant, and normal bettas. C, D: Population structure within betta fish and between giant and normal bettas, respectively, as revealed by PCA. E: Ancestral genetic clusters among wild-type, giant, and normal bettas, as revealed by admixture analysis.

showed a considerable proportion (~2%) of ancestral genetic clusters from the wild-type strain (Figure 1E). These data suggest that giant fighting fish may contain specific genetic factors that determine the giant phenotype.

Whole-genome scanning based on a 30 kb window-size  $F_{ST}$  between giant and non-giant bettas identified three genomic regions at chr1: 22.55–23.85 Mb, chr9: 1.78–2.85 Mb, and chr11: 4.37–5.47 Mb, which were significantly associated with the giant phenotype, with a  $F_{ST}$  cutoff value of 0.36 in the upper 1% percentile of null distribution (Supplementary Figure S1A). Locus-specific  $F_{ST}$  identified five regions with  $F_{ST}>0.8$ : i.e., chr1: 22.55–23.85 Mb, chr2: 10.22–10.32 Mb, chr9: 1.78–2.85 Mb, chr11: 4.37–5.47 Mb, and chr15: 6 798 647 bp (Supplementary Figure S1B). Based on both  $F_{ST}$  methods, the three genomic regions at chr1, chr9, and chr11 were consistently associated with the giant phenotype, and showed more highly differentiated single nucleotide polymorphisms (SNPs) ( $F_{ST}>0.8$ ) than the background (Supplementary Figure S1B) as well as significant differences in allele frequency between the giant and non-giant bettas (Supplementary Figure S1C). Tajima's  $D$  identified selection signatures in the giant and normal ornamental bettas in three main genomic regions (i.e., chr1, chr9, and chr11), but not in the wild-type strain (Supplementary Figure S2A). Consistently, H-scan analysis detected selective sweeps at the same three genomic regions. In particular, the genomic region at chr9 was under selection in both the giant and normal bettas (Supplementary Figure S2B). In accordance with Tajima's  $D$  results, we did not find any evidence of selection in the wild-type strain at these genomic regions (Supplementary Figure S2B).

Identifying candidate genes and causal mutations for the giant phenotype is of importance for understanding the mechanisms underlying overgrowth and genetic modification to improve growth performance in food animals. However, the five genomic regions had a total length of ~3.5 Mb and harbored 251 predicted protein-coding genes, making it difficult to accurately distinguish candidate genes for the giant phenotype. Studies on gene expression patterns can help to identify candidate genes, based on the assumption that changes in the abundance of gene products are linked to phenotypic variations or changes in body size, although this approach will fail to identify candidate genes with altered functions. We identified a total of 14 differentially expressed genes (DEGs) in the brain and/or muscle transcriptomes of the giant and wild-type bettas. Analysis of the brain transcriptomes identified 11 DEGs located in the five genomic regions, with two, six, and three located at chr1, chr9, and chr11, respectively (Supplementary Figure S3). Five genes, including *rubcn* at chr1, *clec4f* and *plec* at chr9, and *rcn1* and *fbx120* at chr11, were up-regulated in the brain of giant bettas, whereas the remaining six genes, i.e., *hes1*, *stard3*, *psmb6*, *hepatic lectin*, *nhe*, and *ccne2*, were down-regulated in the brain of giant bettas (Supplementary Figure S3). Transcriptome sequencing of muscle samples identified five DEGs located in the five genomic regions, including *ppp1r2* and unannotated gene (*BSCG00000019285*) in chr1 and chr11, respectively, which were up-regulated in the giant bettas, and *psmb6*, *hepatic lectin*, and *nfe2l1* in chr9, which were down-regulated in the giant bettas (Supplementary Figure S4). Two genes, *psmb6* and *hepatic lectin*, were down-regulated in both the brain and muscle samples of giant bettas.

We then analyzed structural variation within the genomic regions likely associated with the giant phenotype. We compared genomic sequences between normal (Wang et al., 2021) and giant bettas (Fan et al., 2018) and found evidence of duplications of three genes, i.e., *znf706*, *bmp8a*, and *gdf6b*, within the chr9 region in the giant fighting fish (Supplementary Figure S5A). We then analyzed the genome sequence of a wild-type betta (Kwon et al., 2021), but found no evidence of duplication of the three genes. The overall expression of *bmp8a* in the brain transcriptome was significantly higher in the giant fighting fish than in the normal wild-type strain (Supplementary Figure S5B). *Bmp8a* is associated with bone differentiation and morphogenesis (Wozney, 2002). In domestic chickens and ponies, mutations in bone morphogenetic protein genes can cause body size variation (Nanaei et al., 2020; Wang et al., 2016). Thus, our data suggest that *bmp8a* may be a candidate gene for the giant phenotype.

Finally, we annotated peak SNPs within the genomic regions and identified two missense SNPs (chr9: 121 817, G/A, Arg/Gln and chr9: 2 642 023, G/A, Ala/Thr) in *col11a1* and *kdm1b*, respectively. The ~100 kb long genomic fragment (chr9: 76 791–162 802 bp) was likely misassembled. In both giant and wild-type betta genome assemblies (Fan et al., 2018; Kwon et al., 2021), this fragment is joined with the genomic region chr9: 1.78–2.85 Mb. Thus, with a missense mutation in the giant fighting fish, *col11a1* may be another

potential candidate gene. Recent studies in humans, mice, and chickens have revealed that variations in the *col11a1* expression levels and sequences are associated with skeletal overgrowth and/or dwarfism (Chu et al., 2021; Shen et al., 2016; Wang et al., 2017). However, transcriptome sequencing is not sufficient to capture candidate genes that determine growth in the early developmental stages and are expressed in other tissues, except for the herein studied brain and muscle. To overcome this challenge, we searched for candidate genes within the genomic regions based on literature mining, and found that *umur1*, *mfsd8*, *kdm4b*, *slc39a7*, *pou2f2*, *znf652*, *ca10*, *gjc1*, *igfbp1*, and *slc16a3/mct4* are likely associated with either growth in animals or human height. Interestingly, a cluster of homeobox genes, including *hoxb1a*, *hoxb2a*, *hoxb3a*, *hoxb4*, *hoxb5a*, *hoxb6a*, and *hoxb8*, which specify regions of the body plan along the head-tail axis in animals (Pineault et al., 2015), was identified in the chr11 genomic region, indicating that *hox* genes may also be responsible for the giant phenotype.

In summary, genomic sequences, transcriptomic sequences, genetic variants, and candidate genes for growth are valuable resources for understanding the genetic basis of body size and for in-depth exploration of the ecological and evolutionary processes that determine body size. These genomic resources can serve as selection targets for rapid improvement in growth-related traits in fish species by genome editing using the CRISPR/Cas9 system.

#### DATA AVAILABILITY

Raw sequencing reads were archived in the DDBJ Sequencing Read Archive (SRA) database with BioProject ID PRJDB7253.

#### SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

#### COMPETING INTERESTS

The authors declare that they have no competing interests.

#### AUTHORS' CONTRIBUTIONS

G.H.Y. initiated the study. L.W. and G.H.Y. designed the study. F.S., M.L., and L.W. prepared the samples and sequenced the genome and transcriptome. L.W. analyzed the

data. L.W. and G.H.Y. drafted the manuscript. All authors read and approved the final version of the manuscript.

Le Wang<sup>1</sup>, Fei Sun<sup>1</sup>, May Lee<sup>1</sup>, Gen-Hua Yue<sup>1,2,\*</sup>

<sup>1</sup> *Molecular Population Genetics & Breeding Group, Temasek Life Sciences Laboratory, Singapore 117604, Singapore*

<sup>2</sup> *Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore*

\*Corresponding author, E-mail: [genhua@tll.org.sg](mailto:genhua@tll.org.sg)

#### REFERENCES

- Chu FC, Hii LY, Hung TH, Lo LM, Hsieh TT, Shaw SW. 2021. A novel de novo mutation in *COL2A1* gene associated with fetal skeletal dysplasia. *Taiwanese Journal of Obstetrics and Gynecology*, **60**(2): 359–362.
- Fan GY, Chan J, Ma KL, Yang BR, Zhang H, Yang XW, et al. 2018. Chromosome-level reference genome of the Siamese fighting fish *Betta splendens*, a model species for the study of aggression. *Gigascience*, **7**(11): giy087.
- Kwon YM, Vranken N, Hoge C, Lichak MR, Francis KX, Camacho-Garcia J, et al. 2021. Genomic consequences of domestication of the Siamese fighting fish. *bioRxiv*, doi: 10.1101/2021.04.29.442030.
- Nanaei HA, Esmailizadeh A, Mehrgard AA, Han JL, Wu DD, Li Y, et al. 2020. Comparative population genomic analysis uncovers novel genomic footprints and genes associated with small body size in Chinese pony. *BMC Genomics*, **21**(1): 496.
- Pineault KM, Swinehart IT, Garthus KN, Ho E, Yao Q, Schipani E, et al. 2015. *Hox11* genes regulate postnatal longitudinal bone growth and growth plate proliferation. *Biology Open*, **4**(11): 1538–1548.
- Shen CB, Zheng XD, Gao J, Zhu CH, Ko R, Tang XF, et al. 2016. A rare variant in *COL11A1* is strongly associated with adult height in Chinese Han population. *Journal of Genetics and Genomics*, **43**(9): 549–554.
- Wang L, Sun F, Wan ZY, Ye BQ, Wen YF, Liu HM, et al. 2021. Genomic basis of striking fin shapes and colors in the fighting fish. *Molecular Biology and Evolution*, **38**(8): 3383–3396.
- Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, et al. 2016. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. *Journal of Molecular Cell Biology*, **8**(6): 542–552.
- Wang MS, Otecko NO, Wang S, Wu DD, Yang MM, Xu YL, et al. 2017. An evolutionary genomic perspective on the breeding of dwarf chickens. *Molecular Biology and Evolution*, **34**(12): 3081–3088.
- Wozney JM. 2002. Overview of bone morphogenetic proteins. *Spine*, **27**(16S): S2–S8.