Accepted refereed manuscript of: Li R, Bekaert M, Lu J, Lu S, Zhang Z, Zhang W, Shi O, Chen C, Mu C, Song W, Migaud H & Wang C (2020) Mapping and validation of sex-linked SNP markers in the swimming crab Portunus trituberculatus. *Aquaculture*, 524, Art. No.: 735228. https://doi.org/10.1016/j.aquaculture.2020.735228

Research Article

Mapping and validation of sex-linked SNP markers in the swimming crab *Portunus trituberculatus*

Ronghua Li[a,b,c,1], Michaël Bekaert[c,1], Junkai Lu[a,b], Shaokun Lu[a,b], Zhouyi Zhang[a], Weijia Zhang[a], Ouwen Shi[a], Chen Chen[a], Changkao Mu[a,b], Weiwei Song[a,b], Herve Migaud[a,b,c,*] and Chunlin Wang[a,b,*]

[a] Key Laboratory of Applied Marine Biotechnology, Ministry of Education, Ningbo University, Ningbo 315211, China

[b] Collaborative Innovation Center for Zhejiang Marine High-efficiency and Healthy Aquaculture, Ningbo University, Ningbo 315211, China

[c] Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, United Kingdom

[*] Corresponding author.

E-mail address: herve.migaud@stir.ac.uk (H. Migaud); wangchunlin@nbu.edu.cn (C. Wang).

[1] These authors contributed equally to this work.

Abstract

Portunus trituberculatus is one of the most commercially important marine crustacean species for both aquaculture and fisheries in Southeast and East Asia. Production of monosex female stocks is attractive in commercial production since females are more profitable than their male counterparts. Identification and mapping of the sex-linked locus is an efficient way to elucidate the mechanisms of sex determination in the species and support the development of protocols for monosex female production. In this study, a sex-averaged map and two sex-specific genetic maps were constructed based on 2b-restriction site-associated DNA sequencing. A total of 6,349 genetic markers were assigned to 53 linkage groups. Little difference was observed in the pattern of sex-specific recombination between females and males. Association analysis and linkage mapping identified 7 markers strongly associated with sex, four of which were successfully mapped on the extremity of linkage group 22. Females were homozygous and males were heterozygous for those 7 markers strongly suggesting an XX/XY sex determination system in this species. Three Markers were successfully validated in a wild population of *P. trituberculatus* and exhibited a specificity ranging from 93.3% to 100%. A high-resolution melting based assay was developed for sex genotyping. This study provides new knowledge and tools for sex identification which will help the development of protocols for monosex female production of *P. trituberculatus* and support future genomic studies.

Keywords: genetic linkage map, sex marker, sex determination, QTL mapping, high-resolution melting (HRM)

Highlights

- A genome survey and restriction site-associated DNA sequencing were combined.
- A high-density genetic map was created with 53 linkage groups.
- XX/XY sex determination system was validated.
- Male-specific alleles were identified and validated for 7 SNP markers.
- A PCR-based genetic sex identification method was successfully validated.

Introduction

The swimming crab, *Portunus trituberculatus*, naturally distributed along the coastal waters of temperate western Pacific Ocean, is a commercially important marine crustacean species for both aquaculture and fisheries in East Asian countries. In China, the farming of *P. trituberculatus* developed quickly and reached a production of 116,251 tons in 2018 (Fisheries Bureau of Ministry of Agriculture, 2019). Despite the growing commercial interest in this species, research effort has been limited so far with recent literature focused on the species biology including ovarian development (Che et al., 2018; Liu et al., 2018), larviculture (Shi et al., 2019), population genetics (Liu et al., 2012), genomics of growth traits (Lv et al., 2017; Feng et al., 2018), nutrition (DHA/EPA, Hu et al., 2017; phospholipids, Song et al., 2019) and immune function (Ren et al., 2017; Wu et al., 2019).

Sexual dimorphisms, especially in growth rates, between female and male individuals have been reported in many aquaculture species, making the production of monosex stocks particularly attractive for the aquaculture industry (Scott et al., 1989: Budd et al., 2015). Sexual dimorphisms has also been reported in *P. trituberculatus* with female crabs exhibiting higher growth rate and greater body weight (Wang et al., 2018). In addition, female crabs that reached sexual maturation are more profitable with a higher market value due to the accumulation of vitellogenin in the ovary. However, no protocol is available for the production of monosex females and the identification of sex in *P. trituberculatus* seedlings is difficult and can only be done reliably after 3-4 months of culture in ponds based on the shape of the abdomen. Therefore, in culture, male crabs are usually removed continuously during the growing period and female kept until harvest. This “sex grading” method is labour-intensive and very inefficient. The use of monosex female juveniles would greatly improve productivity and profitability of the sector. However, the prerequisites to any animal monosexing methodologies are the elucidation of the species sex determination system and the identification of potential sex-associated markers. These would help to fast track the establishment of all-female *P. trituberculatus*.

Linkage mapping and association mapping are two major strategies to identify genetic loci and markers for traits of commercial interest (Yu et al., 2017). The combination of high-throughput sequencing (HTS) and restriction digestion enzymes enable the

rapid discovery of genome-wide genetic markers, facilitating the construction of high-density genetic linkage maps and sex-specific quantitative trait loci (QTL) mapping, which narrow down the sex-determining region and can provide useful sex-linked markers (Palaiokostas et al., 2013a, 2013b; Cui et al., 2015; Palaiokostas et al., 2015; Yu et al., 2017; Shi et al., 2018; Waiho et al., 2019; Wang et al., 2019). Recently, a high-density linkage map based on specific length amplified fragment sequencing (SLAF-seq) has been constructed for *P. trituberculatus*, in which 10 growth-related QTLs and a significant QTL for sex were identified (Lv et al., 2017, 2018). This provided valuable genomic resources that can be used for marker-assisted selection and breeding, however, the candidate sex-specific markers identified were not completely linked (Lv et al., 2018). More detailed studies are required to identify informative markers associated with sex in *P. trituberculatus*.

In the present study, a genome survey was first conducted to provide basic characteristics of *P. trituberculatus* genome to be used as a reference for the following marker validation. Then, 2b-restriction site-associated DNA (2b-RAD) sequencing was employed to construct a high-density genetic linkage map and association analysis of sex-related QTL for *P. trituberculatus*. Finally, new sex-linked markers were validated based on the genome survey analysis aiming at increasing the reliability and accuracy of sex identification and elucidating the sex determination system in *P. trituberculatus*. Overall, the new knowledge gained will help the development of monosexing protocols in the species and support the expansion of the swimming crab aquaculture.

Materials and Methods

Genome sequencing survey and analysis

Genomic DNA of one female *P. trituberculatus* collected from wild population of Zhejiang Province, China was extracted from muscle for genome survey analysis (low coverage, whole genome sequencing). Two libraries with insert size of 350 nucleotides were constructed from randomly fragmented genomic DNA. The construction, sequencing and assembly of the genome survey libraries were performed by Novogene Co. (Beijing, China). Sequencing was performed on an Illumina HiSeq 2500 sequencing platform with paired-end 150 bp reads. Reads of low quality (*i.e.* with an average quality score less than 20) or having ambiguous bases or adaptors were clipped using Trimmomatics v0.38 (Bolger et al., 2014) as standard pre-processing methods. Jellyfish v2.2.10 (Marçais and Kingsford, 2011) was use for K-mer and the genome size estimation. SOAPdenovo r241 (Luo et al., 2012) was used for the *de novo* genome assembly. Due to the low coverage, no annotation was pursued, but completeness of the gene captured was assessed using BUSCO v3 (Waterhouse et al., 2018) using the Metazoa dataset. Reads were deposited at the EBI European Nucleotide Archive (ENA) projects PRJEB32999.

Mapping family preparation and sample collection for QTL analysis

An F1 full-sib family for linkage map construction was created by two parents from the wild population of Zhejiang Province, China. The full-sib family was reared at XinYi Aquatic Products Limited Company in 2015. A total of 118 progenies were randomly selected after being reared for 4 months. Muscle tissue (claw) from both parents and 118 offspring were collected and preserved in 95% ethanol until DNA extraction.

Muscle tissues of an additional 30 wild individuals (15 males and 15 females) from the Zhejiang Province, China were collected for the QTL mapping. Genomic DNA was extracted from the above samples using a standard phenol/chloroform protocol.

Library preparation and sequencing

The construction of 2b-RAD library using *Bsa*XI restriction enzymes was performed and sequenced by Oebiotech Co., Ltd. (Shanghai, China) following the methodology proposed by Wang *et al.* (2012, 2016) and using the multiplexing structure detailed in Supplementary Table S2. Reads were deposited at the EBI European Nucleotide Archive (ENA) projects PRJNA371532 (family cross) and PRJEB32947 (wild animals).

Genotyping 2b-RAD alleles

The 2b-RAD sequence data from the 150 individuals (120 full-sib family with parents, and 30 wild individuals; Table S1) were pre-processed to discard low quality reads (*i.e.* with a quality score less than 20), missing tag structure or ambiguous bases. The clean reads were demultiplexed (Table S2) as described in Wang *et al*. (2016). Six offspring from the full-sib family (9, 25, 26, 65, 73, 75) were found to have substantially lower quality reads and were excluded from all further analyses. Resulting reads were assembled *de novo*, sorted into loci and genotypes using Stacks v2.3 (Catchen et al., 2013). The key parameter values employed were: a minimum stack depth of 6, a maximum of 2 mismatches allowed in a locus in an individual and up to 1 mismatch between loci when building the catalogue. Informative markers were kept only when presenting at least two alleles with a minor allele frequency (MAF) above 0.01 and were present in at least 75% of the samples. Only one SNP was reported.

Construction of the linkage maps

Based on the SNP alleles obtained, a linkage map was constructed with LepMap3 (Rastas, 2017). SNPs deviating from the expected Mendelian segregation ($P < 0.001$) were excluded. Based on available karyotyping data (Zhu et al., 2005), the number of linkage groups was set to 53 (logarithm of odds, LOD = 9). The total length of the map in centimorgans (cM) was estimated using Kosambi mapping functions. Maps generated with OrderMarker2 module were checked for contiguous sequence (contig) continuity. Data processing was automatised using scripts available from

https://github.com/pseudogene/radmap. Genetic maps were drawn using Genetic-Mapper v0.11 (Bekaert, 2016). Recombination rates were calculated using Lepmap3 output as the number of recombination event divided by the total number of individuals for each chromosome for each sex class.

Identification and validation of sex associated markers

Using the phenotypic gender data from 30 wild individuals and parents of the full-sib family, an association analysis was performed within the package R/SNPassoc v1.9-2 (González et al., 2007), using the "codominant" model for the QTL analysis. The sequences of the predicted sex associated markers from QTL results were aligned against scaffold of genome survey database using Blast+ v2.8.1 (Altschul et al., 1990). Primers were designed on flanking regions of the SNPs using the Primer3 v2.40 (Untergasser et al., 2012).

Muscles from an extra 60 wild adult *P. trituberculatus* (30 males, 30 females) were collected from Xiangshan, Zhejiang Province, China, for the validation of these predicted sex associated markers. Genomic DNA was extracted from the muscle tissue by using a genomic DNA extraction kit（BioTeke, Beijing, China）following the manufacturer protocols. Polymerase chain reaction (PCR) was performed in 10 μL volumes containing 2× Power Taq PCR Master Mix (BioTeke, Beijing, China) 5 μL, 1 μM of each primer set, and about 100 ng template DNA. PCR was performed on a Master-cycler gradient thermal cycler (Eppendorf) with the following program: 3 min at 94 °C; 35 cycles of 1 min at 94 °C, annealing for 1 min, 72 °C for 1 min per cycle; followed by 5 min at 72 °C. PCR products were sequenced in both directions on the ABI3730 platform (Applied Biosystems). Alignment of the sequenced fragments was performed using Vector NTI 10.0 (Invitrogen, Carlsbad, CA) for the confirmation of predicted SNPs. High-resolution melting (HRM) analysis was also applied to the discrimination of predicted sex associated SNPs. PCR amplification and HRM analysis were performed on a LightCycler®480 real-time PCR instrument (Roche Diagnostics) as previously described (Chen et al., 2018).

Results

Genome survey summary

From the two 350 bp short paired-end DNA libraries constructed for the genome survey analysis of *P. trituberculatus*, a total of 165,950,266 raw paired-end reads were generated by sequencing. After removing low quality and adapters, 157,193,348 paired-end reads remained (94.7%). The estimated genome size using SOAPdenovo K-mer module (K = 17) was 1.083 Gb (Fig. 1A) with high heterozygosity (1.02%). A relatively high percentage of repetitive sequences (58.50%) was detected in the *P. trituberculatus* genome (Table 1). After assembly of this genome survey with SOAPdenovo (K = 41), a total of 1,910,434 scaffolds were generated with N50 of 1,212 nt, for a total length of 892,095,304 nt with a GC-value of 42.02% (Table 1), and 81.4% BUSCO metazoa genes were at least partially recovered (Fig. 1B).

2b-RAD data analysis and SNP calling

High throughput sequencing of the 150 animals produced 253,650,280 paired-end reads in total (two runs). After the removal of low-quality (QC < 20) and incomplete reads, 75.5% of the total reads were retained (191,576,072 paired-end reads). Demultiplexing generated 957,880,360 sequences with a length of 27 bp. The sequences were assembled *de novo* and genotypes for all samples were obtained using Stacks, yielding 496,426 unique SNP loci. Average coverage of each loci was 18.5×, 92.5× and 27.3× for offspring, parents and wild population, respectively. A total of 48,862 loci were polymorphic-allelic markers shared by at least 75% of the samples with a minor allele frequency above 0.01. All of these markers were subsequently used to construct genetic linkage maps and to perform an association analysis.

High-density linkage map construction

Using genotype information from the 48,862 loci of the 114 individuals (112 offspring and 2 parents), sex-averaged, female, and male genetic linkage maps were constructed with LepMap3 at the LOD threshold of 9.0 (Table 2 and Table S3 & S4). The maps were constructed using 6,349 informative SNPs to 53 linkage groups (each

comprising at least 10 SNPs), spanning a total distance of 2,960.5 cM, 2,728.3 cM, 3,237.7 cM for sex-averaged, female, and male map respectively (Fig. 2 and Fig. S1 & S2). The average marker interval of the sex-averaged map is 0.47 cM. The male-specific linkage map is slightly longer than the female-specific map with an average male to female ratio of 1.19. The sex-specific recombination was studied, and little difference was observed in the recombination rates between sex, with the average female/male ratio of 0.97 over all the linkage groups (linkage group 46-52 were excluded from the recombination calculations because of low marker numbers).

QTL mapping and validation of sex associated markers

Using both the 48,862 informative markers and 6,349 mapped markers, R/SNPassoc was used to conduct a quantitative trait locus (QTL) mapping analysis for the sex determination association. A total of 7 markers strongly associated with sex (100% specificity, Table 3) were identified, four of which were successfully mapped on the extremity of LG 22, where a highly significant QTL for gender (Fig. 3A, peak LOD = 8.48, at the whole genomic level) was located in the region ranging from 4.874 to 6.201 cM (Fig. 3B) suggested to be a potential sex determining region. Females were homozygous and males were heterozygous for those 7 markers. The 2b-RAD sequences and the scaffold sequences of all these 7 markers identified from the survey database were provided in Table S5. Orthologous sequences of these scaffolds were searched in the GenBank through NCBI BLAST (Johnson et al., 2008), however, no hits were detected.

Primers were designed for the seven potential sex-specific markers according to the corresponding scaffolds of the genome survey, and validated in another wild population (30 males, 30 females). Three of them were successfully amplified (Table 4), two of which were located in the genetic linkage map (Ptr67655, Ptr138136). Sequencing results demonstrated that the specificity of these 3 markers in the discrimination of the wild population were 93.3%, 100% and 100%, respectively for markers PS7, PS8 and PS11 (Table 5). Interestingly, more than one SNP was found in PS11, and the predicted male-specific SNP (C/T) appeared to be C in all the 60 individuals, but six new SNPs in this fragment were detected, females were homozygous and males were heterozygous for these six new SNPs.

All three markers were tested using HRM analysis to develop a sex genotyping assay. Only primer set PS11 was found to be clearly distinct between sex in the melting profiles from the HRM analysis (Fig. 4) and T_m calling analysis showed that the melting points were 80.5 °C and 81.9 °C, respectively for male and female.

Discussion

High-density genetic linkage maps are important genomic tools for fine mapping of quantitative trait loci, genome assembly, functional gene localization, comparative genome and analysis of sex chromosome evolution. The 2b-RAD technology has been widely applied for the construction of high-density linkage maps in many aquaculture species due to its relative simplicity (no requirement for prior genomic data), cost-effectiveness, even distribution on the genomes, uniform fragments and adjustable genome coverage (Shi et al., 2014; Cui et al., 2015; Tian et al., 2015; Feng et al., 2018; Liu et al., 2018; Wang et al., 2018). In the present study, three genetic linkage maps were constructed for *P. trituberculatus* using 2b-RAD, covering a total distance of 2,960.5 cM, 2,728.3 cM and 3,237.7 cM for sex-averaged, female, and male maps, respectively. The average marker interval of the sex-averaged map was 0.47 cM, which was comparable to previous reports in many crustacean species (0.4 -1.51 cM), including *Litopenaeus vannamei* (0.7 cM, Yu et al., 2015); *Penaeus monodon* (1.51 cM, Guo et al., 2019); *Scylla paramamosain* (0.81 - 0.92 cM, Waiho et al., 2019; Zhao et al., 2019); *Eriocheir sinensis* (0.49 - 0.81 cM, Cui et al., 2015; Qiu et al., 2017) and *P. trituberculatus* (0.51 cM, Lv et al., 2017). The high resolution of the new SNP-based linkage map will facilitate further studies of detailed QTL mapping.

Monosex populations in aquatic animals are generally produced by combining sex reversal and progeny testing. The ability to determine rapidly the genetic sex of an individual through sex-linked markers can fast track the development of monosex lines in economically important cultured species (Chen et al., 2007). In the present study, association analysis and linkage mapping identified 7 markers strongly associated with sex, four of which were successfully mapped onto the extremity of LG 22, where a significant QTL for gender (peak LOD = 8.48, at the whole genomic level) was located in the region ranging from 4.874 to 6.201 cM. Females were homozygous and males were heterozygous for those 7 markers confirming the suggested XY sex determination system of *P. trituberculatus* (Lv et al., 2018). All the

scaffolds containing the 7 sex-linked markers were further searched against GenBank to identify candidate gene(s) for sex determination, however, no informative blast hits were detected. Therefore, it is possible that these sex-linked loci found in *P. trituberculatus* represent novel genes or regulatory elements. Further studies are needed to characterise the function of these new sex-specific loci.

Chromosomal karyotype analysis showed that the diploid chromosome number of *P. trituberculatus* is 106 (2n = 106), however, no heteromorphism has been observed for sex chromosomes (Zhu et al., 2005). Sex chromosomes are derived from ordinary autosomes, which is thought to involve recombination suppression in sex determining regions, followed by the accumulation of deleterious mutations and the degeneration of the sex-specific (*e.g.* Y) chromosome (Graves, 2006). It is generally thought that the morphological differentiation between sex chromosomes is a by-product of the degeneration of the chromosome that is present only in the heterogametic sex (*i.e.* Y or W) and is thus completely sheltered from genetic recombination (Bachtrog, 2006). In our study, there was little difference in the recombination rates between sex, with an average female/male ratio of 0.97, recombination suppression has not been detected in any specific linkage groups (including LG 22 where the sex-determining region is located). The lack of non-recombining regions together with the apparent existence of a sex-determining region may suggest "proto-sex chromosomes" in *P. trituberculatus*. In such a sex determination system, a chromosome carries a newly arisen sex-determining gene or a newly evolved sex-determining region, but recombination suppression has not yet evolved and therefore there is no sex chromosome heteromorphism (Charlesworth and Mank, 2010).

Validation of the 3 successfully amplified sex associated markers in another wild population of *P. trituberculatus* exhibited a specificity from 93.3% to 100%. Only markers Ptr138136 (primer set PS8) and Ptr62530 (primer set PS11) were strongly associated with sex. This is especially true for the latter locus (Ptr62530) in which the predicted SNP (C/T) appeared to be C in all individuals screened from the validation population, however, six new SNPs in this fragment were also found to be associated with sex which could also be easily separated by HRM analysis. The inconsistency of identified sex-linked markers in different populations has also been reported in previous species including *L. vannamei* and *P. trituberculatus* (Yu et al., 2017; Lv et al., 2018). This may be partly explained by the lack of recombination suppression in regions that include and flank the sex determining mutation in these species, so that

the detected sex associated markers may frequently cross over during meiosis. Nevertheless, the marker Ptr138136 (primer set PS8) was found to be strongly associated with sex in both populations of *P. trituberculatus*, and it was mapped to the most significant sex associated region in the LG 22 with a peak LOD of 8.48. Although marker Ptr62530 was not located on the genetic map, it was also found to be linked with sex. The combination of these sex associated markers should significantly improve the reliability and accuracy of sex identification in *P. trituberculatus*.

The confirmation of the sex-determining region and system will greatly help further studies to search for sex-determining genes and apply molecular sexing methods of *P. trituberculatus* while also benefitting developmental and evolutionary biology fields. Diverse sex determination systems have evolved in the animal and plant kingdoms, including the genetic sex determination (GSD), the environmental sex determination (ESD) especially with regards to temperature (TSD), and the interaction between the two (Korpelainen, 1990; Charlesworth and Mank, 2010; Palaiokostas et al., 2013a). In fish, the complex interaction between genetic and environmental factors has been observed in several species such as Nile tilapia, *Oreochromis niloticus* (Cáceres et al., 2019), sea bass, *Dicentrarchus labrax* (Piferrer et al., 2005) and flatfish species (Luckenbach et al., 2009), with male/female heterozygous system (XX/XY or ZZ/ZW). Sex determination in crustacea can also show plasticity, being influenced by environmental variables such as light and temperature (Ford, 2008). Despite the male heterozygous system shown in *P. trituberculatus*, the influence of environmental variables on sex determination can not be ruled out and should be investigated.

The genome survey performed in the present study provides useful background for the species given the lack of whole-genome resources. In our study, the 2b-RAD library generated tags with 27 bp in length, which was short to design primers for PCR validation or function analysis, However, after blasting against the genome survey data, the 7 potential sex-specific markers identified matched a significant scaffold sequence extending the flanking region for primer design and PCR validation. Besides, a BUSCO completeness assessment of *P. trituberculatus* genome survey recovered 81.4% of the metazoa genes, which confirmed the overall robustness of the sequencing and assembly of the genome survey, providing valuable resources for future genetic studies in this species. Genome survey could also estimate some basic genomic characteristics which can then help to determine the best sequencing

strategies and most suitable assembly algorithms for whole genome studies. The estimated genome size of *P. trituberculatus* in our study was 1.083 Gb with high heterozygosity (1.02%), and a relatively high percentage of repetitive sequences (58.50%) detected. The complexity of crustacean genomes has been acknowledged previously and it makes the assembly of the whole genome sequence challenging when only based on data generated by Illumina sequencing (Yu et al., 2015; Lv et al., 2017). Future sequencing efforts for crustacean species with such a complex genomes should integrate other advanced technology such as the PacBio long reads sequencing platform and algorithms to reduce assembly errors.

Ethical statement

This study was approved by the Ethics Committee of Ningbo University, and conducted according to relevant national and international guidelines.

Author Contributions

Conceptualization, R.L. and C.L.; Methodology, R.L. and M.B.; Formal analysis, M.B., R.L. and H.M.; Investigation, J.L., S.L., Z.Z., W.Z., O.S., C.C., C.M. and W.S.; Writing - original draft preparation, R.L. and M.B.; Writing - review and editing, funding acquisition, R.L., C.W. and H.M.

Acknowledgments

This research was supported by National Key R&D Program of China (2018YFD0900303), National Key R&D Program of China (2018YFD0901304), Natural Science Foundation of Zhejiang Province (No. LY17C190005, No. LY18H110003), Ministry of Agriculture of China & China Agriculture Research System (no: CARS-48), Ningbo Science and Technology Project (2017C110007), CSC Scholarship (201708330421) and K.C. Wong Magana Fund in Ningbo University.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bachtrog, D., 2006. A dynamic view of sex chromosome evolution. Current Opinion in Genetics and Development 16, 578–585. https://doi.org/10.1016/j.gde.2006.10.007

Bekaert, M., 2016. Genetic-Mapper: vectorial genetic map drawer. F1000Research 5. https://doi.org/10.7490/f1000research.1112266.1

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Budd, A., Banh, Q., Domingos, J., Jerry, D., 2015. Sex control in fish: approaches, challenges and opportunities for aquaculture. Journal of Marine Science and Engineering 3, 329–355. https://doi.org/10.3390/jmse3020329

Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., Soto, J., Salas, D., Yáñez, J.M., 2019. Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile Tilapia (Oreochromis niloticus L.). G3 (Bethesda, Md.) 9, 3213–3223. https://doi.org/10.1534/g3.119.400297

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: An analysis tool set for population genomics. Molecular Ecology 22, 3124–3140. https://doi.org/10.1111/mec.12354

Charlesworth, D., Mank, J.E., 2010. The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. Genetics 186, 9–31. https://doi.org/10.1534/genetics.110.117697

Che, J., Liu, M., Dong, Z., Hou, W., Pan, G., Wu, X., 2018. The growth and ovarian development pattern of pond-reared swimming crab *Portunus trituberculatus*. Journal of Shellfish Research 37(3), 521–528.

Chen, S.L., Li, J., Deng, S.P., Tian, Y.S., Wang, Q.Y., Zhuang, Z.M., Sha, Z.X., Xu, J.Y., 2007. Isolation of female-specific AFLP markers and molecular

identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). Marine Biotechnology 9, 273–280. https://doi.org/10.1007/s10126-006-6081-x

Chen, X., Li, R., Wang, C., Mu, C., Song, W., Liu, L., Shi, C., Zhan, P., 2018. An effective method for identification of three mussel species and their hybrids based on SNPs. Conservation Genetics Resources. https://doi.org/10.1007/s12686-018-1051-y

Cui, Z., Hui, M., Liu, Y., Song, C., Li, X., Li, Y., Liu, L., Shi, G., Wang, S., Li, F., Zhang, X., Liu, C., Xiang, J., Chu, K.H., 2015. High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab *Eriocheir sinensis*. Heredity 115, 206–215. https://doi.org/10.1038/hdy.2015.26

Feng, X., Yu, X., Fu, B., Wang, X., Liu, H., Pang, M., Tong, J., 2018. A high-resolution genetic linkage map and QTL fine mapping for growth-related traits and sex in the Yangtze River common carp (*Cyprinus carpio haematopterus*). BMC Genomics 19, 1–13. https://doi.org/10.1186/s12864-018-4613-1

Fisheries Bureau of Ministry of Agriculture, 2019. China fishery statistical yearbook. China Agriculture Press, Beijing, China.

Ford, A.T., 2008. Can you feminise a crustacean? Aquatic Toxicology 88, 316–321. https://doi.org/10.1016/j.aquatox.2008.04.013

González, J.R., Armengol, L., Solé, X., Guinó, E., Mercader, J.M., Estivill, X., Moreno, V., 2007. SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23, 644–645. https://doi.org/10.1093/bioinformatics/btm025

Graves, J.A.M., 2006. Sex chromosome specialization and degeneration in mammals. Cell 124, 901–914. https://doi.org/10.1016/j.cell.2006.02.024

Guo, L., Xu, Y.-H., Zhang, N., Zhou, F.-L., Huang, J.-H., Liu, B.-S., Jiang, S.-G., Zhang, D.-C., 2019. A high-density genetic linkage map and QTL mapping for sex in black tiger shrimp (*Penaeus monodon*). Frontiers in Genetics 10, 1–9. https://doi.org/10.3389/fgene.2019.00326

Hu, S., Wang, J., Han, T., Li, X., Jiang, Y., Wang, C., 2017. Effects of dietary DHA/EPA ratios on growth performance, survival and fatty acid composition of juvenile swimming crab (*Portunus trituberculatus*). Aquaculture Research 48, 1291–1301. https://doi.org/10.1111/are.12971

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic acids research 36, 5–9. https://doi.org/10.1093/nar/gkn201

Korpelainen, H., 1990. Sex ratios and conditions required for environmental sex determination in animals. Biological Reviews of the Cambridge Philosophical Society 65, 147–184. https://doi.org/10.1111/j.1469-185x.1990.tb01187.x

Liu, H., Pang, M., Yu, X., Zhou, Y., Tong, J., Fu, B., 2018. Sex-specific markers developed by next-generation sequencing confirmed an XX/XY sex determination system in bighead carp (*Hypophthalmichthys nobilis*) and silver carp (Hypophthalmichthys molitrix). DNA Research 25, 257–264. https://doi.org/10.1093/dnares/dsx054

Liu, L., Li, J., Liu, P., Zhao, F., Gao, B., Du, Y., 2012. A genetic linkage map of swimming crab (*Portunus trituberculatus*) based on SSR and AFLP markers. Aquaculture 344–349, 66–81. https://doi.org/10.1016/j.aquaculture.2012.01.034

Liu, M., Pan, J., Liu, Z., Cheng, Y., Gong, J., Wu, X., 2018. Effect of estradiol on vitellogenesis and oocyte development of female swimming crab, *Portunus trituberculatus*. Aquaculture 486, 240–245. https://doi.org/10.1016/j.aquaculture.2017.12.034

Luckenbach, J.A., Borski, R.J., Daniels, H. V., Godwin, J., 2009. Sex determination in flatfishes: Mechanisms and environmental influences. Seminars in Cell and Developmental Biology 20, 256–263. https://doi.org/10.1016/j.semcdb.2008.12.002

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Yong, Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, Jian, Lam, T.-W., Wang, Jun, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1, 18. https://doi.org/10.1186/2047-217X-1-18

Lv, J., Gao, B., Liu, P., Li, J., Meng, X., 2017. Linkage mapping aided by de novo genome and transcriptome assembly in *Portunus trituberculatus*: applications in growth-related QTL and gene identification. Scientific Reports 7, 1–13. https://doi.org/10.1038/s41598-017-08256-8

Lv, J., Sun, D., Huan, P., Song, L., Liu, P., Li, J., 2018. QTL mapping and marker identification for sex-determining: Indicating XY sex determination system in

497 the swimming crab (*Portunus trituberculatus*). Frontiers in Genetics 9, 1–9.
498 https://doi.org/10.3389/fgene.2018.00337
499 Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel
500 counting of occurrences of k-mers. Bioinformatics 27, 764–770.
501 https://doi.org/10.1093/bioinformatics/btr011
502 Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M.E., Oral, M., Taggart, J.B.,
503 Gharbi, K., McAndrew, B.J., Penman, D.J., Migaud, H., 2013a. Mapping the sex
504 determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using
505 RAD sequencing. BMC Genomics 14, 1–12. https://doi.org/10.1186/1471-2164-
506 14-566
507 Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew,
508 B.J., Penman, D.J., 2015. A novel sex-determining QTL in Nile tilapia
509 (*Oreochromis niloticus*). BMC Genomics 16, 1–10.
510 https://doi.org/10.1186/s12864-015-1383-x
511 Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew,
512 B.J., Penman, D.J., 2013b. Mapping and Validation of the Major Sex-
513 Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) Using RAD
514 Sequencing. PLoS ONE 8, 1–9. https://doi.org/10.1371/journal.pone.0068389
515 Piferrer, F., Blázquez, M., Navarro, L., González, A., 2005. Genetic, endocrine, and
516 environmental components of sex determination and differentiation in the
517 European sea bass (*Dicentrarchus labrax* L.). General and Comparative
518 Endocrinology 142, 102–110.
519 Qiu, G.F., Xiong, L.W., Han, Z.K., Liu, Z.Q., Feng, J. Bin, Wu, X.G., Yan, Y.L.,
520 Shen, H., Huang, L., Chen, L., 2017. A second generation SNP and SSR
521 integrated linkage map and QTL mapping for the Chinese mitten crab *Eriocheir*
522 *sinensis*. Scientific Reports 7, 1–11. https://doi.org/10.1038/srep39826
523 Rastas, P., 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole
524 genome sequencing data. Bioinformatics 33, 3726–3732.
525 https://doi.org/10.1093/bioinformatics/btx494
526 Ren, X., Gao, B., Liu, X., Li, J., Liu, P., 2017. Comparison of immune responses and
527 antioxidant status of different generations of growth-selected *Portunus*
528 *trituberculatus* families. Aquaculture Research 48, 1315–1326.
529 https://doi.org/10.1111/are.12973

Scott, A.G., Penman, D.J., Beardmore, J.A., Skibinski, D.O.F., 1989. The “YY” supermale in *Oreochromis niloticus* (L.) and its potential in aquaculture. Aquaculture 78, 3–4.

Shi, X., Waiho, K., Li, X., Ikhwanuddin, M., Miao, G., Lin, F., Zhang, Y., Li, S., Zheng, H., Liu, W., Aweya, J.J., Azmie, G., Baylon, J.C., Quinitio, E.T., Ma, H., 2018. Female-specific SNP markers provide insights into a WZ/ZZ sex determination system for mud crabs *Scylla paramamosain*, *S. tranquebarica* and *S. serrata* with a rapid method for genetic sex identification. BMC Genomics 19, 1–12. https://doi.org/10.1186/s12864-018-5380-8

Shi, Y., Wang, S., Gu, Z., Lv, J., Zhan, X., Yu, C., Bao, Z., Wang, A., 2014. High-density single nucleotide polymorphisms linkage and quantitative trait locus mapping of the pearl oyster, *Pinctada fucata martensii* Dunker. Aquaculture 434, 376–384. https://doi.org/10.1016/j.aquaculture.2014.08.044

Song, D., Shi, B., Ding, L., Jin, M., Sun, P., Jiao, L., Zhou, Q., 2019. Regulation of dietary phospholipids on growth performance, antioxidant activities, phospholipid metabolism and vitellogenesis in prereproductive phase of female swimming crabs, *Portunus trituberculatus*. Aquaculture 511, 734230. https://doi.org/10.1016/j.aquaculture.2019.734230

Tian, M., Li, Y., Jing, J., Mu, C., Du, H., Dou, J., Mao, J., Li, X., Jiao, W., Wang, Y., Hu, X., Wang, S., Wang, R., Bao, Z., 2015. Construction of a high-density genetic map and quantitative trait locus mapping in the sea cucumber *Apostichopus japonicus*. Scientific Reports 5, 1–11. https://doi.org/10.1038/srep14852

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research 40, e115–e115. https://doi.org/10.1093/nar/gks596

Waiho, K., Shi, X., Fazhan, H., Li, S., Zhang, Y., Zheng, H., Liu, W., Fang, S., Ikhwanuddin, M., Ma, H., 2019. High-density genetic linkage maps provide novel insights Into ZW/ZZ sex determination system and growth performance in mud crab (*Scylla paramamosain*). Frontiers in Genetics 10, 1–16. https://doi.org/10.3389/fgene.2019.00298

Wang, L., Xie, N., Shen, Y., Ye, B., Yue, G.H., Feng, X., 2019. Constructing high-density genetic maps and developing sexing markers in northern snakehead

(*Channa argus*). Marine Biotechnology 348–358. https://doi.org/10.1007/s10126-019-09884-z

Wang, S., Liu, P., Lv, J., Li, Y., Cheng, T., Zhang, L., Xia, Y., Sun, H., Hu, X., Bao, Z., 2016. Serial sequencing of isolength RAD tags for cost-efficient genome-wide profiling of genetic and epigenetic variations. Nature Protocols 11, 2189–2200. https://doi.org/10.1038/nprot.2016.133

Wang, S., Meyer, E., Mckay, J.K., Matz, M. V., 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. Nature Methods 9, 808–810. https://doi.org/10.1038/nmeth.2023

Wang, Xinhua, Fu, B., Yu, X., Qu, C., Zhang, Q., Tong, J., 2018. Fine mapping of growth-related quantitative trait loci in Yellow River carp (*Cyprinus carpio haematoperus*). Aquaculture 484, 277–285. https://doi.org/10.1016/j.aquaculture.2017.11.016

Wang, X., Wang, Y., Ye, T., Lu, W., Zhou, C., 2018. A preliminary analysis of the growth characteristics of *Portunus trituberculatus*. Transactions of Oceanology and Limnology 2, 131–136.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution 35, 543–548. https://doi.org/10.1093/molbev/msx319

Wu, X., Huang, Y., Yu, Z., Mu, C., Song, W., Li, R., Liu, L., Ye, Y., Shi, C., Wang, C., 2019. An MBT domain containing anti-lipopolysaccharide factor (PtALF8) from *Portunus trituberculatus* is involved in immune response to bacterial challenge. Fish and Shellfish Immunology 84, 252–258. https://doi.org/10.1016/j.fsi.2018.10.016

Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., Huang, L., Zheng, H., Xiang, J., 2015. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. Scientific Reports 5, 1–14. https://doi.org/10.1038/srep15612

Yu, Y., Zhang, X., Yuan, J., Wang, Q., Li, S., Huang, H., Li, F., Xiang, J., 2017. Identification of sex-determining loci in Pacific White Shrimp *Litopeneaus vannamei* using linkage and association analysis. Marine Biotechnology 19, 277–286. https://doi.org/10.1007/s10126-017-9749-5

596 Zhao, M., Wang, W., Chen, W., Ma, C., Zhang, F., Jiang, K., Liu, J., Diao, L., Qian,
597 H., Zhao, J., Wang, T., Ma, L., 2019. Genome survey, high-resolution genetic
598 linkage map construction, growth-related quantitative trait locus (QTL)
599 identification and gene location in *Scylla paramamosain*. Scientific Reports 9, 1–
600 10. https://doi.org/10.1038/s41598-019-39070-z
601 Zhu, D., Wang, C., Li, Z., 2005. Karyotype analysis on *Portunus trituberculatus*.
602 Journal of Fisheries of China 29(5), 649–653.
603

Figures (for draft only)

Figure 1 (2 columns – 146 mm x 75 mm)

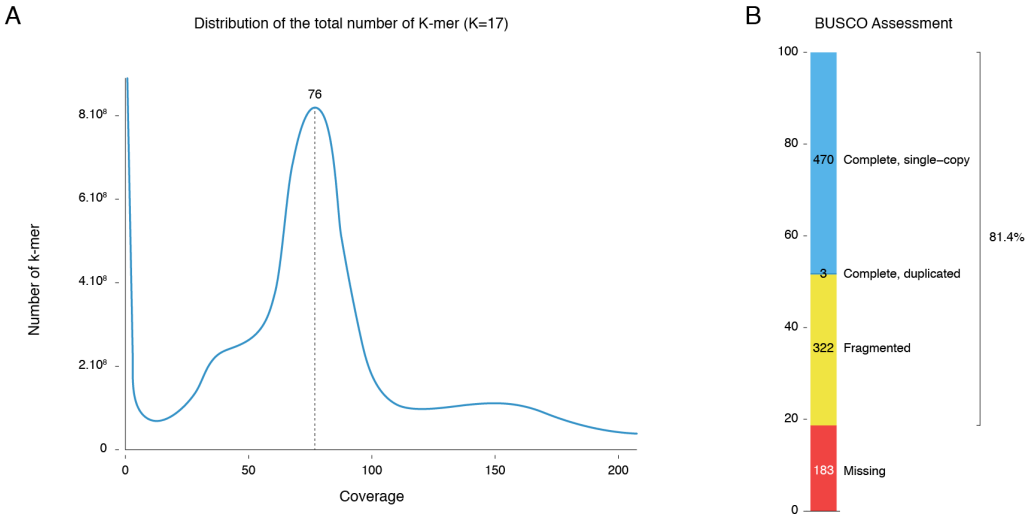

Figure 2 (2 columns/full page – 170 mm x 257 mm)

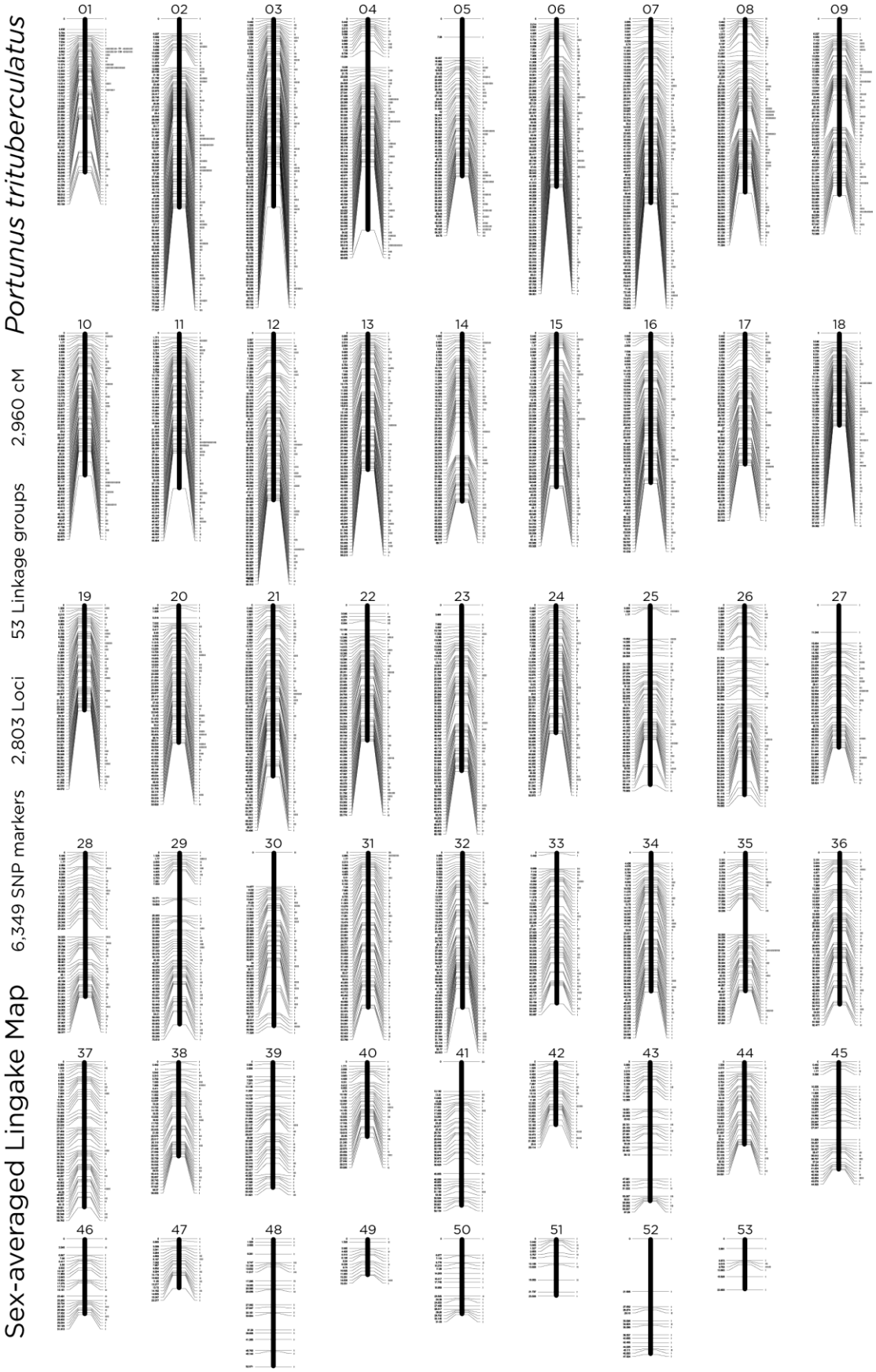

Figure 3 (1 column – 85 mm x 170 mm)

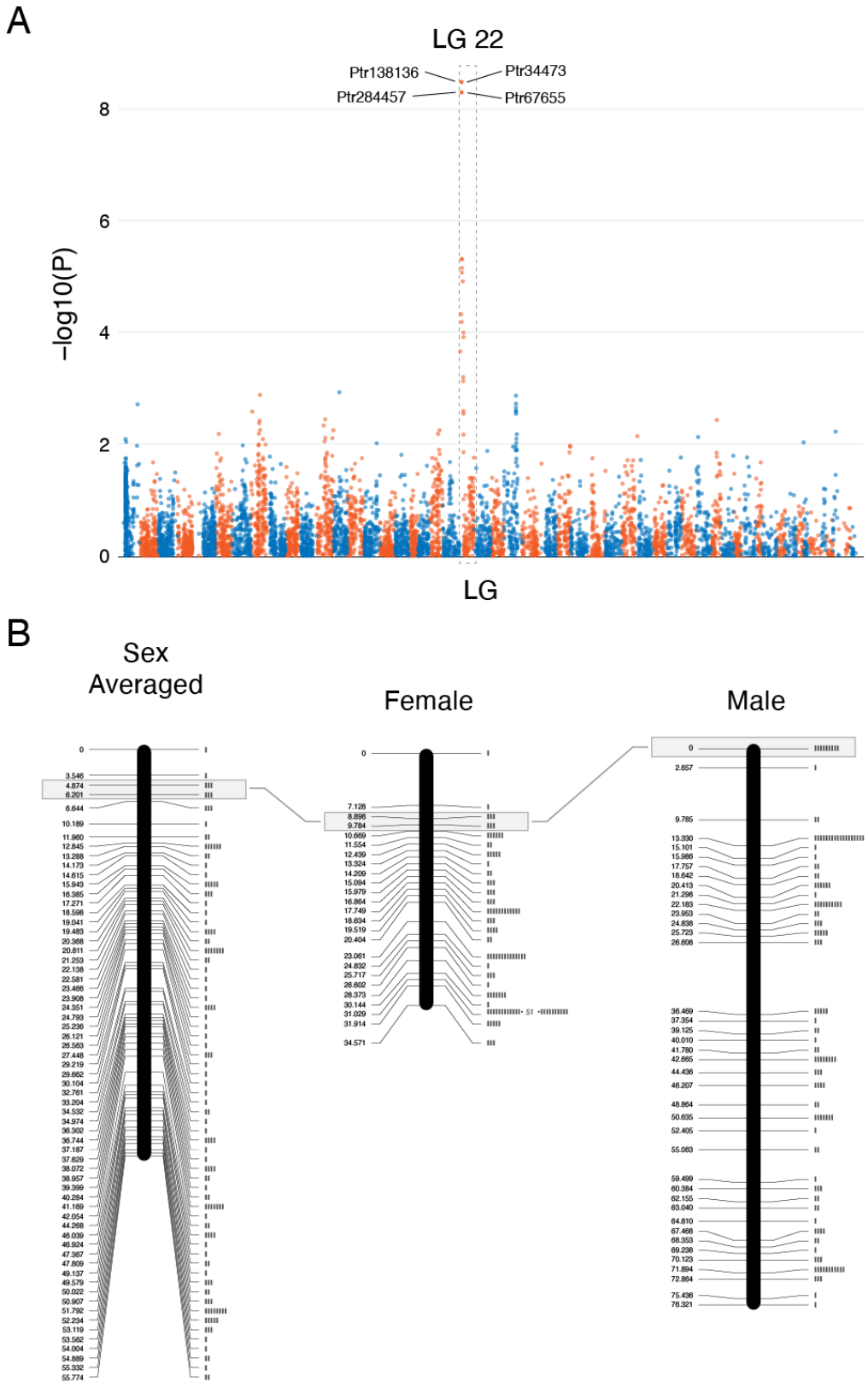

Figure 4

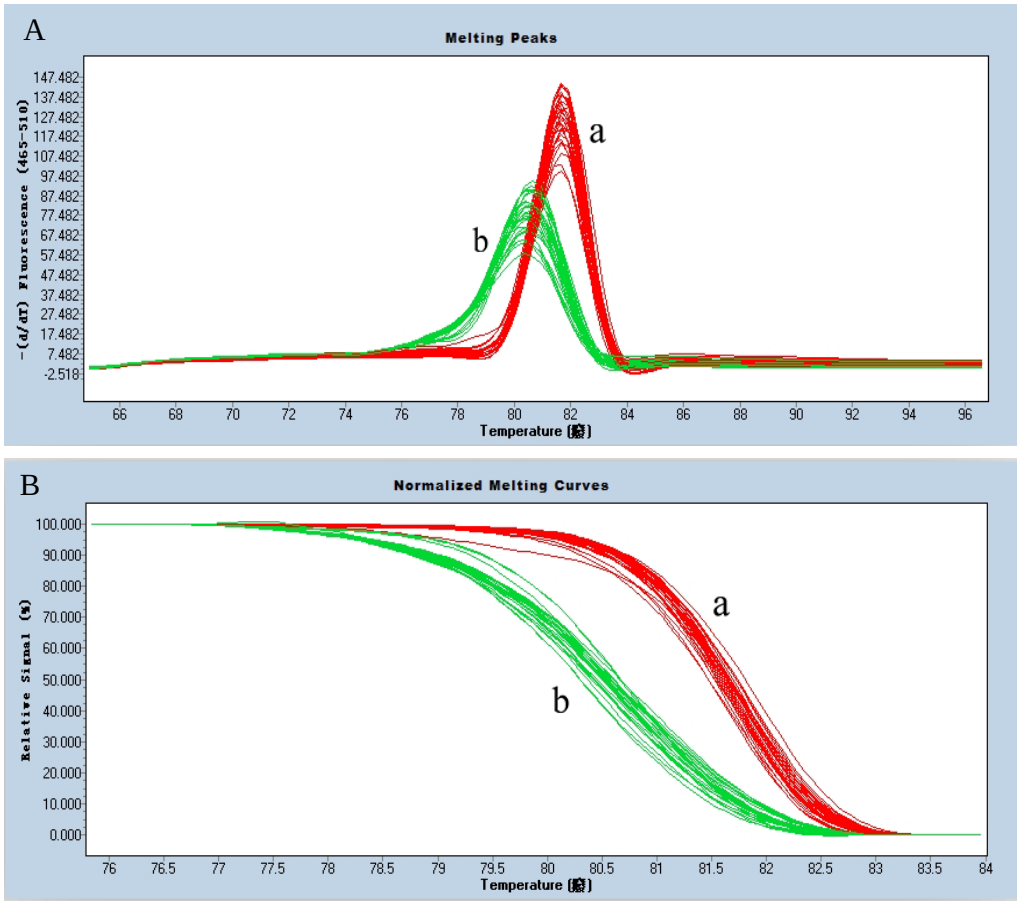
A
Melting Peaks
-(d/dT) Fluorescence (465-510)
147.482
137.482
127.482
117.482
107.482
97.482
87.482
77.482
67.482
57.482
47.482
37.482
27.482
17.482
7.482
-2.518
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
96
Temperature (℃)
a
b
B
Normalized Melting Curves
Relative Signal (%)
100.000
90.000
80.000
70.000
60.000
50.000
40.000
30.000
20.000
10.000
0.000
76
76.5
77
77.5
78
78.5
79
79.5
80
80.5
81
81.5
82
82.5
83
83.5
84
Temperature (℃)
a
b

Figure Captions

Figure 1. Genome coverage. (A) Distribution of the total number of k-mer analysed. The x-axis is the frequency or the number of times a given k-mer is observed in the sequencing data. The y-axis is the total number of k-mers with a given frequency. The peak is at K = 76. (B) BUSCO assessment (Metazoa database; number of BUSCO, 978), 81.4% of the genes were recovered.

Figure 2. Sex-Averaged linkage map, with linkage groups ordered by number of SNP markers. In each linkage group, numbers shown on the left provide position (in cM) of the respective locus on the chromosome while bars on the right indicates the relative number of SNP markers. Detailed data are provided in Supplementary Data Table S4. Female and Male linkage map are provided in Figure S1 & S2 respectively.

Figure 3. Markers associated with phenotypic sex. (A) Manhattan plot of the association P values for phenotypic sex. The $-\log_{10}(P)$ values for association of directly genotyped SNPs are plotted as a function of position of the genetic map. Each linkage group (LG) has been represented with a different colour. (B) Details of the LG22 and location of the markers associated with phenotypic sex. Female and Male linkage maps are reversed.

Figure 4. Discriminations of SNPs between different sex of *P. trituberculatus* using high-resolution melting analysis for PS11. (A) Melting peaks. (B) Normalised melting curves. (a) female and (b) male.

Tables

Table 1 – Summary statistics of sequencing and assembly of *P. trituberculatus* genome survey.

Category	**Number/length**
Total number of raw PE reads	165,950,266
Maximum read length (nt)	150
Cleaned PE reads	157,193,348
K-mer = 17	83,557,275,304
K-mer = 17 (peak)	76
Estimated genome size	1,083.38 Mb
Estimated repeat	58.50%
Estimated heterozygocity	1.02%
Number of scaffolds	1,910,434
Total length of scaffolds (nt)	892,095,304
Max length of scaffold (nt)	60,585
Scaffold N50 (nt)	1,212
GC value	42.02%
Genome coverage	0.82×

Table 2 – Summary of the linkage map. Full details for each linkage group in Table S3.

Category	Number/length
Number of linkage group	53
Total markers	6,349
Loci (Sex-averaged)	2,803
Loci (Female)	1,312
Loci (Male)	1,773
Length (Sex-averaged)	2,960.0 cM
Length (Female)	2,728.1 cM
Length (Male)	3,238.1 cM

Table 3 – SNP markers significantly associated with phenotypic sex. For each marker the expected association and LOD value are reported. Markers and contig sequences are provided in Table S5.

Marker ID	LOD	Genetic map location	Genome position	Female	Male
Ptr34473	8.48	22:4.874	contig1099313:1843-1869	A/A	A/G
Ptr62530	8.16	-	contig136081:198-224	T/T	C/T
Ptr84718	7.12	-	contig681633:749-775	C/C	C/T
Ptr138136	8.48	22:4.874	contig 325741:898-872	A/A	A/G
Ptr67655	8.29	22:6.201	contig1070277:1474-1448	T/T	C/T
Ptr284457	8.29	22:4.874	contig569169:778-804	T/T	A/T
Ptr326206	8.16	-	contig1048881:434-460	C/C	C/T

657 **Table 4** – Primers used for amplification of the three sex-specific SNP markers.
658 Annealing temperature (T_a): 60°C.
659

Marker ID	Primer ID		Primer sequence
Ptr67655	PS7	Forward	5' -TTAAGTTTGAGTATTGAGTATCCAC- 3'
		Reverse	5' -AATGAGAAGTATTGTAAATGATGTT- 3'
Ptr138136	PS8	Forward	5' -ATACCAGACAAGAGGGCTTC- 3'
		Reverse	5' -TCCCATATAGATATTAGTGTCATTC- 3'
Ptr62530	PS11	Forward	5' -CCGACAACACAGATCCACTAAC- 3'
		Reverse	5' -CGAGTGTGGAGAGAATGATTTTT- 3'

660
661

662 **Table 5** – Validation data of three sex markers in wild populations.

663

Primer ID	Marker ID	Genotype	Male	Female	Specificity (%)
PS7	Ptr67655	A/G	27	1	93.3
		A/A	3	29	
PS8	Ptr138136	T/C	30	0	100.0
		T/T	0	30	
PS11	Ptr62530	T/C A/G C/T A/C C/T T/A	30	0	100.0
		C/C G/G T/T T/T T/T A/A	0	30	

664

665

Supporting information captions

Table S1. Details of the samples analysed. (2b-RAD reads numbers, polymorphic marker, and gender are included).

Table S2. Demultiplexing key for 2b-RAD library (30 wild individuals). See Wang *et al.* (Wang et al., 2016) for details on usage.

Table S3. Summary statistic of the linkage groups.

Table S4. Details of marker position in the genetic maps. (A total of 6,349 markers including association with phenotypic sex and marker sequence).

Table S5. Details of the 7 sex-linked SNP markers. Sequence of 2b-RAD and corresponding contig (from genome survey).

Figure S1. Female linkage map, with linkage groups ordered by number of SNP markers.

Figure S2. Male linkage map, with linkage groups ordered by number of SNP markers.