Heredity (2015) 115, 206–215

www.nature.com/hdy

ORIGINAL ARTICLE

High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab *Eriocheir sinensis*

Z Cui[1,2,6], M Hui[1,6], Y Liu[1,6], C Song[1,3], X Li[1,3], Y Li[1], L Liu[1], G Shi[1,3], S Wang[4], F Li[1], X Zhang[1], C Liu[1], J Xiang[1] and KH Chu[5]

The sex determination system in crabs is believed to be XY-XX from karyotypy, but centromeres could not be identified in some chromosomes and their morphology is not completely clear. Using quantitative trait locus mapping of the gender phenotype, we revealed a ZW-ZZ sex determination system in *Eriocheir sinensis* and presented a high-density linkage map covering ~98.5% of the genome, with 73 linkage groups corresponding to the haploid chromosome number. All sex-linked markers in the family we used were located on a single linkage group, LG60, and sex linkage was confirmed by genome-wide association studies (GWAS). Forty-six markers detected by GWAS were heterozygous and segregated only in the female parent. The female LG60 was thus the putative W chromosome, with the homologous male LG60 as the Z chromosome. The putative Z and W sex chromosomes were identical in size and carried many homologous loci. Sex ratio (5:1) skewing towards females in induced triploids using unrelated animals also supported a ZW-ZZ system. Transcriptome data were used to search for candidate sex-determining loci, but only one LG60 gene was identified as an *ankyrin-2* gene. *Double sex- and mab3-related transcription factor 1* (*Dmrt1*), a Z-linked gene in birds, was located on a putative autosome. With complete genome sequencing and transcriptomic data, more genes on putative sex chromosomes will be characterised, thus leading towards a comprehensive understanding of the sex determination and differentiation mechanisms of *E. sinensis*, and decapod crustaceans in general.
Heredity (2015) **115,** 206–215; doi:10.1038/hdy.2015.26; published online 15 April 2015

INTRODUCTION

With the advent of next-generation sequencing technologies and high-throughput genotyping platforms, high-resolution linkage mapping (<1 cM) with sequence-based markers, especially single-nucleotide polymorphism (SNP), has developed rapidly. Accurate genetic linkage maps are crucial for the identification of the genomic loci related to phenotypic differences, mapping of quantitative trait loci (QTLs), and even the assembly of genome sequences. Among crustaceans that constitute a major group of Arthropoda, the largest phylum in the animal kingdom, only a high-density linkage map for giant tiger shrimp *Penaeus monodon* (Baranski *et al.*, 2014) has been reported and none has been constructed for crabs.

The Chinese mitten crab *Eriocheir sinensis* H. Milne Edwards, 1853 (Crustacea: Decapoda: Brachyura) is an anadromous species widely cultured in different regions of China, with more than 600 000 tons per year. Beyond its native range in East Asia, *E. sinensis* was accidentally introduced to Europe (Herborg *et al.*, 2005) as well as North America (Rudnick *et al.*, 2003), where it has led to various ecological and economic damages, and is listed by International Union for Conservation of Nature as one of the world's 100 worst invasive alien species (Lowe *et al.*, 2000). Hence, as an important ecological and economic species, *E. sinensis* could serve as a model species of decapod crustaceans. Genetic information, especially at the genomic level, would enhance the understanding of its biology with relevance to ecology and aquaculture.

In comparison with vertebrates, sex determination mechanisms in crustaceans are more diverse, and gender is often plastic, and affected by environmental factors (Ford, 2008). Decapoda are an advanced group of crustaceans, including diverse species of shrimps and crabs. In crabs, the sex determination system was believed to be XY-XX, based on karyotype studies of four species (*E. japonicus*, *Hemigrapsus sanguineus*, *H. penicillatus* and *Plagusia dentipes*) (Niiyama, 1937, 1938, 1959; Ginsburger and Charniaux, 1982; Lécher *et al.*, 1995). However, a recent study of *E. sinensis* and *E. japonicus* questioned the reliability

[1]Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China; [2]National and Local Joint Engineering Laboratory for Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China; [3]University of Chinese Academy of Sciences, Beijing, China; [4]Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao, China and [5]Simon FS Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Correspondence: Professor Dr Z Cui or Professor Dr J Xiang, Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, 7 Nanhai Road, Qingdao 266071, China.
E-mail: zhxcui@qdio.ac.cn or jhxiang@qdio.ac.cn
or Professor Dr KH Chu, Simon FS Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China.
E-mail: kahouchu@cuhk.edu.hk
[6]These authors contributed equally to this work.
Received 30 October 2014; revised 21 January 2015; accepted 23 February 2015; published online 15 April 2015

of karyotype analysis for inferring sex determination, as the centromeres could not be identified in some of the chromosomes (Lee *et al.*, 2004). Several genes related to sex determination and differentiation were obtained previously (Zhang and Qiu, 2010; Ma *et al.*, 2012; Wang *et al.*, 2012a, b, 2013; Shen *et al.*, 2014), but no conclusion can be reached on the genetic mechanism of sex determination in crabs.

Recent genetic studies based on high-density linkage maps have provided new insights into sex determination systems and genome organization of sex chromosomes (Jones *et al.*, 2013; Palaiokostas *et al.*, 2013; Robinson *et al.*, 2014). In addition, transcriptomic data assist in identifying the genes of sex chromosome and sex differentiation pathways. For example, the linkage map in combination with transcriptomes of the flatfish successfully facilitates the assembly and sequence assignment of chromosomes, providing the basis for understanding the evolution of sex chromosomes and the sex determination mechanism (Chen *et al.*, 2014). Determination of full-length transcripts of W-chromosome genes and their expression profiles in early avian embryos provide a complete annotation of the W chromosome (Ayers *et al.*, 2013). Genes in sex determination pathway of *Caenorhabditis elegans* are found to be poorly conserved in the cyst nematode *Globodera pallid*, verifying a different sex determination mechanism in *G. pallid* (Cotton *et al.*, 2013).

Although linkage mapping is relatively straightforward for many organisms, accurate high-density linkage map construction is usually difficult for many crustacean species, especially decapods, because of their large number of chromosomes. Herein, we construct a high-density (0.49 cM average marker interval) linkage map of *E. sinensis* with 10 358 SNP markers. The linkage map covers ~98.5% of the complete genome with 73 linkage groups corresponding to the chromosome number ($2n = 146$). A large number of genes related to sex determination and differentiation are identified from transcriptomic data and mapped by marker annotation. Putative sex chromosomes and the ZW-ZZ sex determination system in *E. sinensis* are proposed for the first time based on QTL mapping and genome-wide association studies (GWAS), and the system is also confirmed by triploid induction experiments.

MATERIALS AND METHODS

Mapping families

Thirty F1 full-sib families were established by single-pair crossing of *E. sinensis* from Panjin, China. One of these families with 120 progenies (59 females and 61 males) and high phenotypic variation was chosen for linkage mapping analysis. The gender of offspring was recorded.

Library preparation, sequencing and *de novo* genotyping

Given that there is no reference genome for *E. sinensis*, our genotyping used *r*estriction-site *a*ssociated DNA (RAD) sequencing. 2b-RAD libraries of two parents and 120 progenies were constructed by using type IIB restriction enzymes, following the protocol developed by Wang *et al.* (2012). In general, library preparation began with the digestion of 100–200 ng genomic DNA in a 15 μl reaction using 4 U *Bsa*XI at 37 °C for 3 h. Then, 12 μl of a ligation master mix containing 0.2 μM library-specific adaptors, 1 mM ATP and 800 U T4 DNA ligase was added to the digestion product and incubated at 4 °C for 16 h. For the parents, standard *Bsa*XI libraries were constructed with adaptors of fully degenerate cohesive ends, whereas for the progenies, reduced representation libraries were constructed using modified adaptors with 5'-NNG-3' and 5'-NNC-3' overhangs to target a subset of all *Bsa*XI fragments. In the second PCR amplification, specific adapter-barcode was incorporated into each library, and then all libraries were pooled for single-end sequencing (1×50 bp) using an Illumina HiSeq 2000 sequencer. The 2b-RAD sequence data set was submitted to the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/Traces/sra/) under the accession number SRX472293.

Raw reads were trimmed to produce high-quality reads for genotyping, which was performed using the RADtyping program v.1.0 (http://www2.ouc.edu.cn/mollusk/detailen.asp?Id = 727) under default parameters (Fu *et al.*, 2013). Only sequences with at least 25 × coverage in either parental genome or any progeny genome were used to ensure the genotyping accuracy. Sequencing the standard libraries for the parents resulted in an average of 47.4 million high-quality reads with sequencing depth of 58–60 × (Supplementary Table S1). After clustering the parental reads and filtering low-quality sequences, 32 433 parent-shared and 65 072 parent-specific sequences with variants remained, serving as representative reference tags for the following genotyping step. For the offspring, an average of 7.8 million high-quality reads were produced by sequencing the reduced representation libraries with the sequencing depth ranging from 25 to 42 ×.

Segregation and linkage analysis

Only markers with the expected Mendelian segregation ratios (assessed by χ^2 test yielding $P \geqslant 0.05$) were included in the linkage analysis. Sex-specific maps were first constructed for each parent using the two-way pseudo-testcross strategy (Grattapaglia and Sederoff, 1994). Maternal and paternal data sets were created using the function of 'Create Maternal and Paternal Population Nodes' in the JoinMap 4.0 program (Stam, 1993), which was also used to partition 1:2:1-type data into 1:1 female-type and 1:1 male-type data. The markers heterozygous in either the dam or the sire were tagged as 'f' or 'm', with markers heterozygous in both parents (bi-parental markers) tagged as 'h' and the dominant markers labeled as 'df' or 'dm'. A minimum logarithm of odds (LOD) score of 6.0 was used to assign the markers to linkage groups. The regression mapping algorithm was used for the map construction. Marker distances in centiMorgans (cM) were calculated using Kosambi's mapping function. Linkage groups were drawn visually with the MapChart software (Voorrips, 2002).

A consensus map was estimated by integrating the female and male maps using the shared markers with the MergeMap software (Wu *et al.*, 2011). Two total genetic map lengths were estimated based on the consensus map: G_{e1} (Fishman *et al.*, 2001) and G_{e2} (Chakravarti *et al.*, 1991), and their average was used as the predicted total genetic map length (G_e). The genome coverage was determined by G_{of}/G_e, where G_{of} was the observed genetic map length (the sum of the map lengths of all linkage groups).

QTL mapping and GWAS

The QTL analyses were performed in MapQTL 5.0 (Van Ooijen, 2004) by interval mapping using 1000 permutations based on the consensus linkage map. The phenotype was treated as a binary trait (0 for females and 1 for males). The LOD scores were first analyzed using the interval mapping method and then genome- and chromosome-wide LOD significance thresholds at the 95% level were determined by a 1000-permutation test for the trait, treating QTLs with LOD scores greater than the chromosome threshold at 95% as significant (Piepho, 2001; Li, 2011).

As a complementary approach to QTL mapping, the relationship between SNP markers covering most of the genome and the gender phenotypic variations was further tested by GWAS in the same family used for QTL mapping. It was implemented by using an R package called 'GWAF' (Chen and Yang, 2010), which conducted association tests between a batch of SNPs and a binary or continuous trait. A linear mixed-effects model taking account of the within full-sib pedigree correlation was used to test the genetic association between markers and phenotypes to avoid false positives due to unexpected familial correlation. The functions 'lme.batch' and 'gee.lgst.batch' were implemented to perform a global test (i.e., Wald χ^2 test) for genotype effects. A sequential Bonferroni correction was applied to counteract errors in multiple comparisons.

Triploidy induction

In addition to QTL mapping and GWAS for gender phenotyping in the single mapping family, triploidy induction in unrelated individuals instead of the mapping family was also performed to deduce the sex determination mechanism of *E. sinensis*. At 5 h after spawning, *E. sinensis* individuals carrying eggs (ovigerous) were immersed into 0.2% potassium chloride solution for 4 h

at 15 °C for triploidy induction by retention of the second polar body. Then, the treated ovigerous crabs were cultured in a similar manner as untreated ovigerous females. The pregnant crabs were removed after spawning and the zoeae were diluted to 1 50 000 individuals per m^3. Rotifers cultured with marine *Chlorella* were fed to zoeae I–II and *Artemia* nauplii were fed to zoeae III–V. Adults of *Artemia* were supplied to megalopal stage animals as food for growing up to the juvenile stage. The juvenile crabs were then transferred to earth ponds for grow-out. They were collected for analysis after 15 months when the crabs reached maturity. The ploidy level was determined on a PARTEC CCA-II flow cytometry machine (PARTEC) following Cui *et al.* (2004). The gender of each triploid crab was recorded.

Twenty zoeae reproduced from different ovigerous crabs were treated with 500 μg ml^{-1} colchicine for 1 h, and then transferred to 0.075 mol l^{-1} KCl for 30 min. After the hypotonic treatment, the zoeae were fixed in Carnoy's solution (methanol:acetic acid = 3:1). The solution was changed two to three times during a total of 1 h of fixation. Several individuals were put on clean slides and pressed gently. The preparations were air-dried and stained with Giemsa. They were observed under the microscope for photography.

Transcriptome assembly and functional annotation

To obtain candidate genes for sex determination and differentiation, we sequenced and searched the transcriptomes of *E. sinensis* at different developmental stages. We previously obtained transcriptomic unigenes of *E. sinensis* at the first zoea stage (Z1, 4.52 Gb total nucleotides assembled into 1 00 252 unigenes) (Cui *et al.*, 2013). In the present study, we obtained assembled and annotated new transcriptome sequences from the whole-body of *E. sinensis* at several developmental stages (the fifth zoea stage, Z5; megalopa stage, M; the first juvenile instar, J1; the third juvenile instar of female, J3F and male, J3M). This followed the method of Li *et al.* (2013). Raw data from the Illumina sequencing were deposited in an NCBI (National Center for Biotechnology Information) short read archive database (GenBank: SRX495513 for Z5, SRX495634 for M, SRX495635 for J1, SRX554564 for J3F and SR554562 for J3M).

Our Illumina high-throughput second-generation sequencing produced a total of 36.39×10^9 nucleotides. The assembly of full-length transcripts at different developmental stages yielded 98 021 from Z5, 127 983 from M, 107 899 from J1, 151 128 from J3F and 1 31 826 from J3M. Annotation of the transcripts was performed first by using the BlastX algorithm (*E*-value $< 1E-05$) against the NCBI non-redundant database and unigenes were obtained after clustering the top-hit results. Combined with the previously reported transcripts of *E. sinensis*, 1 41 524 unigenes were assembled altogether.

Unigenes associated with sex determination and differentiation were identified according to the non-redundant annotation and through further consulting published literatures and the UniProt Knowledgebase (http://www.uniprot.org/). The expression of sex-related genes was analyzed based on mean RPKM (reads per kb of exon model per million mapped reads) values across the two replicate samples for each developmental stage. Briefly, the RPKM value was calculated based on the the number of reads mapped to each gene and the length of the gene (Mortazavi *et al.*, 2008).

Annotation of the mapped markers

To detect whether the mapped markers were within genes and what kind of genes were found, we compared our marker sequences with all available transcriptome data from *E. sinensis*, including the transcriptomic sequences obtained above and those from adult animals published in GenBank. In all, 141 926 transcriptomic sequences were from males' testis and accessory sex gland (GenBank: GE339624–GE342613; JR707930–JR778295; KA660105–KA728674), and the other sequences were from individuals of unknown sex (73 445 from hepatopancreas (FG357325–FG360469; SRA068878), and 10 574 from hemocytes (FG981330–FG984370; FL568844–FL576376)). The sequences were first assembled into unigenes, and annotated as described above. RAD sequences that included mapped markers were then aligned to the unigenes using the SOAP 2.0 program (parameters –M 4, –v 2) (Li *et al.*, 2009). Markers that matched annotated genes were assigned gene names based on the annotation and classified into different biological processes. For the genes with sex-linked markers detected by QTL and GWAS, gene expression was analyzed at the different developmental stages studied (see above).

RESULTS

Sex-specific high-density linkage map construction

The markers were grouped into 73 linkage groups at an LOD threshold of 6.0, corresponding to the haploid chromosome number of *E. sinensis*. The female linkage groups ranged from 8.6 to 85.9 cM in length, with marker numbers from 4 to 163 across the linkage groups, whereas the male linkage groups ranged from 14.5 to 88.4 cM in length and marker numbers from 9 to 148. In total, the female map included 5655 markers spanning 4338.3 cM with an average marker interval of 0.77 cM, whereas the male map consisted of 5606 markers spanning 4392.7 cM and a 0.79 cM average marker interval (Supplementary Table S2).

Our consensus map included 10 358 markers spanning 5125.53 cM with an average marker interval of 0.49 cM (Supplementary Table S3 and Supplementary Figure S1). The average female/male recombination rate in the two maps was 0.99, but different linkage groups differed considerably (from 0.56 to 1.77). The estimated genetic map length based on the G_{e1} method was 5198.53 cM, whereas the length was 5207.44 cM based on the G_{e2} method. The average of the two estimates (5202.99 cM) was taken as the total genetic map length (Supplementary Table S1). The estimated coverage of the consensus map was 98.5%. Given the estimated genome size of ~1.68 Gb for *E. sinensis* (Zhu *et al.*, 2008), the average recombination rate across all the linkage groups was ~3.09 cM Mb^{-1}.

QTL mapping and GWAS of sex-linked markers detection on LG60

A highly significant QTL for gender (LOD > 19.1 at the whole genomic level; Figure 1a) was mapped on integrated LG60 with a confidence interval of 15.15 cM (0–15.15 cM, m6616-h1802; Figure 2). In total, 69 markers were detected in this QTL region, including 33 markers segregating in the female parent, 24 segregating in the male and 12 markers heterozygous in both parents. The LG60 markers were distributed at intervals of 0–14.50 cM in the map from the female parent (67 markers, spanning 52.1 cM), and 0–12.86 cM in that from the male parent (56 markers, spanning 49.7 cM). We found no significant sex difference in the recombination rate for LG60 (female:male = 1.05).

GWAS revealed 46 sex-associated markers with high statistical significance ($P < 1E-6$) (Figure 1b), all in female LG60 (Figure 2). These markers (tagged with 'f' or 'h') were all genotyped as heterozygous (Aa) and segregated in the female parent, conforming to the expected ZW-ZZ marker patterns 1–3. It indicated that the sex determination in *E. sinensis* might be ZW-ZZ system and females were the heterogametic sex. These markers all fell into the region from 0 to 17.2 cM of the female LG60, which is thus the putative W chromosome. Biparental markers also identified the homologous male linkage group (male LG60) as the putative Z chromosome (Figure 2). However, the segregation patterns of the detected markers, especially pattern 2, showed partial sex linkage instead of complete sex linkage, indicating genes/markers linked to the sex locus, but not fully, which could not be determined by single family (Charlesworth, 2013).

Tendency of feminization in triploidy

The chromosome number for larvae (219 for triploid and 146 for diploid crabs; Figure 3a) and the histogram of DNA relative content for individual (diploid: triploid = 2:3, Figure 3b) confirmed the success of triploid induction. However, most crabs died when they grew to the third to fifth juvenile instars for unknown reasons. The survival rate

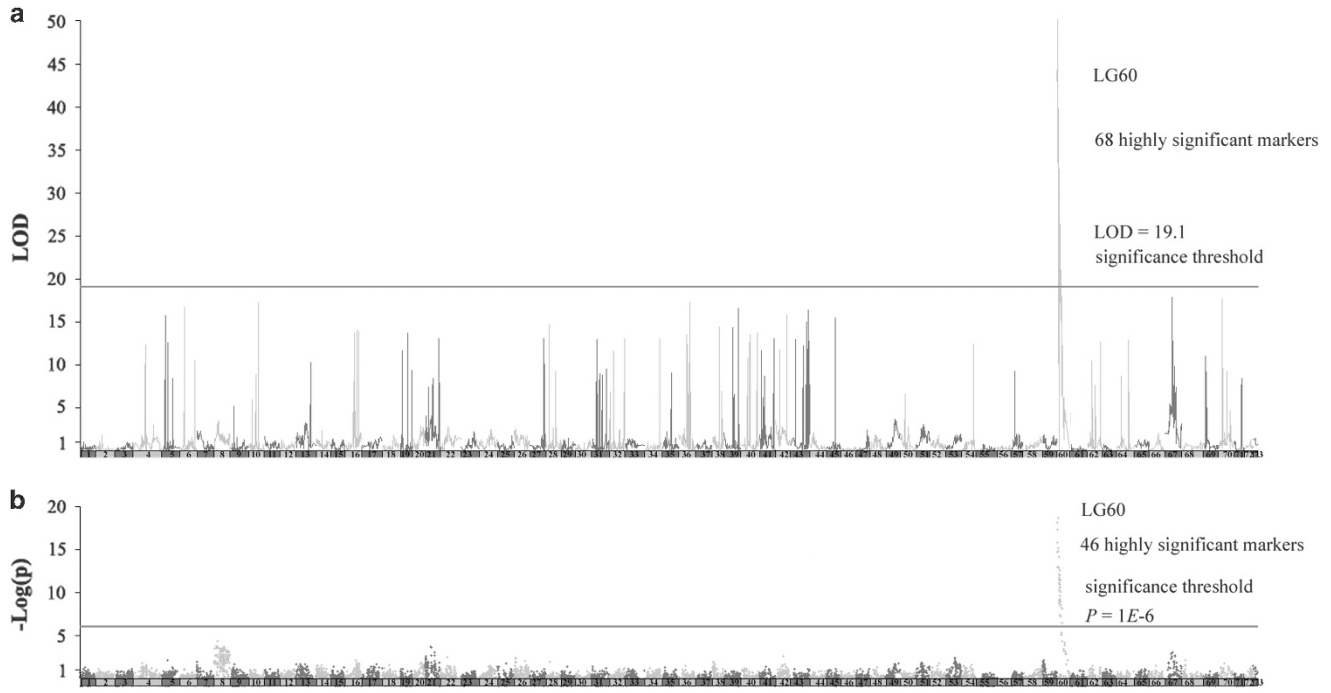


Figure 1 LOD values of QTL mapping (**a**) and significance test of GWAS (**b**) for gender phenotype in *E. sinensis*. The red lines in (**a**) and (**b**) show genome-wide significance thresholds of QTL mapping and association analysis, respectively. The two horizontal axes represent linkage groups 1–73. −Log (*p*), minus logarithm of *P*-value. A full color version of this figure is available at the *Heredity* journal online.

was <5%, and of the survivors, 1800 crabs at maturity were checked. Triploid adults of *E. sinensis* were firstly obtained with sex ratio of 5:1 (female triploids: male triploids). The mechanism of sex skew in the induced triploids as explained by the ZW-ZZ system was demonstrated in Figure 3c. The occurrence of few surviving males might be because of the possibility of recombination between the sex determination locus and centromere, which could skew the sex ratios between 50 and 100% of females (Devlin and Nagahama, 2002).

Sex determination and differentiation-related genes in transcriptomes and the expression of crucial genes in the sex differentiation pathway

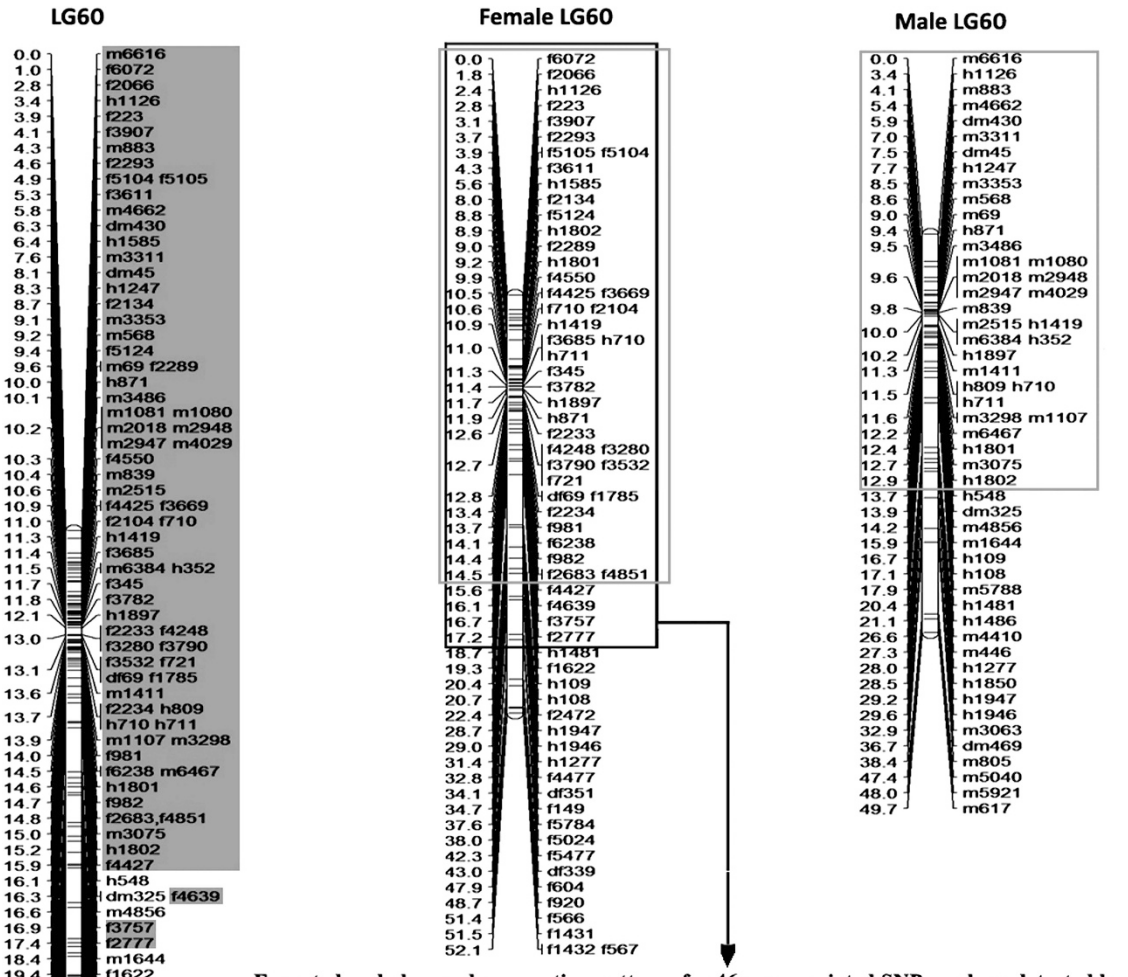
Several crucial genes with similarity to those involved in the sex differentiation pathway of *Drosophila melanogaster*, *C. elegans* and chicken were identified from the transcriptomic data of *E. sinensis*, including putative orthologs of *sex-lethal* (*Sxl*), *transformer 2* (*tra-2*), *double sex* (*dsx*), *feminization 1* (*fem-1*), *aromatase*, *wingless-Type MMTV integration site family, member 4* (*Wnt4*), *β-catenin* and *double sex-and mab3-related transcription factor 1* (*Dmrt1*) genes (Figure 4a). To infer whether these genes were related to sex determination and differentiation in *E. sinensis*, their expression levels (the RPKM values) across different larval stages as well as female and male were calculated and compared (Figure 4b). In general, the expression of *dsx*, *tra-2*, *Wnt4* and *Dmrt1* was lower as compared with *Sxl*, *fem-1*, *aromatase* and *β-catenin*. Unlike *Sxl* in *D. melanogaster* showing sex-specific expression (Penalva and Sánchez, 2003), the *Sxl* in this study was found to be expressed in both males and females. The absence of sex-specific transcripts was also reported by Shen *et al.*, 2014. The *dsx* gene was detected only in M and J1, whereas *Dmrt1* was restricted to J1. Notably, the expression of *fem-1* gene was significantly higher in Z5 than in other stages.

Identification of genic markers and their annotation and gene expression

Mapping all the genetically mapped markers to these *E. sinensis* unigenes revealed a total of 564 markers in, or closely linked to, transcriptomic sequences. These genic markers were found on all linkage groups, except for LG38, LG42 and LG73, with 1–17 markers per linkage group (Supplementary Table S4). 393 sequences (69.7%) corresponded to genes with known or putative annotations, of which 188 can be categorized into 13 functional groups, including metabolism (19.7%), growth and development (17.6%), transcription regulation (14.9%), transport protein (13.3%), immunity (9.04%), signal-transduction (6.91%), ubiquitin system (6.91%), RNA processing (5.32%), cell junction (4.79%), neuroendocrine system (4.79%), reproduction (3.72%), membrane protein (3.19%) and sex differentiation (1.06%).

Nine markers on seven linkage groups were in genes with potential functions in sex determination and differentiation (Figure 5a). Several genes, *spermatogenesis associated 11-like* and *centrin-1*, were expressed in all developmental stages tested, while *maternal protein pumilio* and *jumonji/arid domain-containing protein* were not detected in Z5 (Figure 5b). No expression for *gonadal-histone H2B* was found at stage J1, whereas *Dmrt1* was expressed only at this stage. The *histone-lysine N-methyltransferase PRDM9* was detected at Z1, J3F and J3M. However, none of the genes were found to be specifically expressed in early developmental stages or gender-specific.

Interestingly, one gene with a possible sex function, an *ankyrin-2* gene (marker f4427), was located on the putative sex chromosome (LG60: 15.86 cM position). It was expressed in all the developmental stages tested, but at significantly high levels in Z1, J3F and J3M ($P<0.01$; Figure 5). The RPKM value of *ankyrin-2* was higher in female than in male (female:male = 1.38), but the difference was not significant ($P>0.05$). Considering its location on putative sex



Expected and observed segregation patterns for 46 sex associated SNP markers detected by GWAS under the assumption of either a ZW–ZZ or a XX–XY sex-determination system

Pattern	Dam genotype	Sire genotype	Female offspring genotype	Male offspring genotype	No. of markers
ZW-ZZ					
1	Aa	AA	Aa	AA	20
2	Aa	aa	aa	Aa	17
3	Aa	Aa	Aa or aa	AA or aA	9
4	AA	Aa	AA or aA	AA or aA	0
XX-XY					
5	AA	Aa	AA	Aa	0
6	aa	Aa	aA	aa	0
7	Aa	Aa	AA or aA	Aa or aa	9
8	Aa	AA	AA or aA	AA or aA	0

Figure 2 Sex linkage groups of *E. sinensis*. Female LG60 corresponds to the putative W chromosome and homologous male LG60 corresponds to the putative Z chromosome. GWAS markers linked to sex are indicated in black frames on female LG60. QTL markers linked to sex are indicated in red frames on both female and male LG60. All possible sex-linked markers are shaded in gray in the integrated LG60. The markers heterozygous in either the sire or the dam are tagged as 'f' or 'm', with markers heterozygous in both parents (bi-parental markers) tagged as 'h' and the dominant markers labeled as 'df' or 'dm'. The expected and observed segregation patterns for markers in ZW-ZZ or XX-XY system are shown in the table (according to Staelens *et al*., 2008). Segregation patterns of all 46 markers detected by GWAS conform to patterns 1–3 within the frame. 'A' and 'a' are symbols and do not refer to 'dominant' or 'recessive'. A full color version of this figure is available at the *Heredity* journal online.

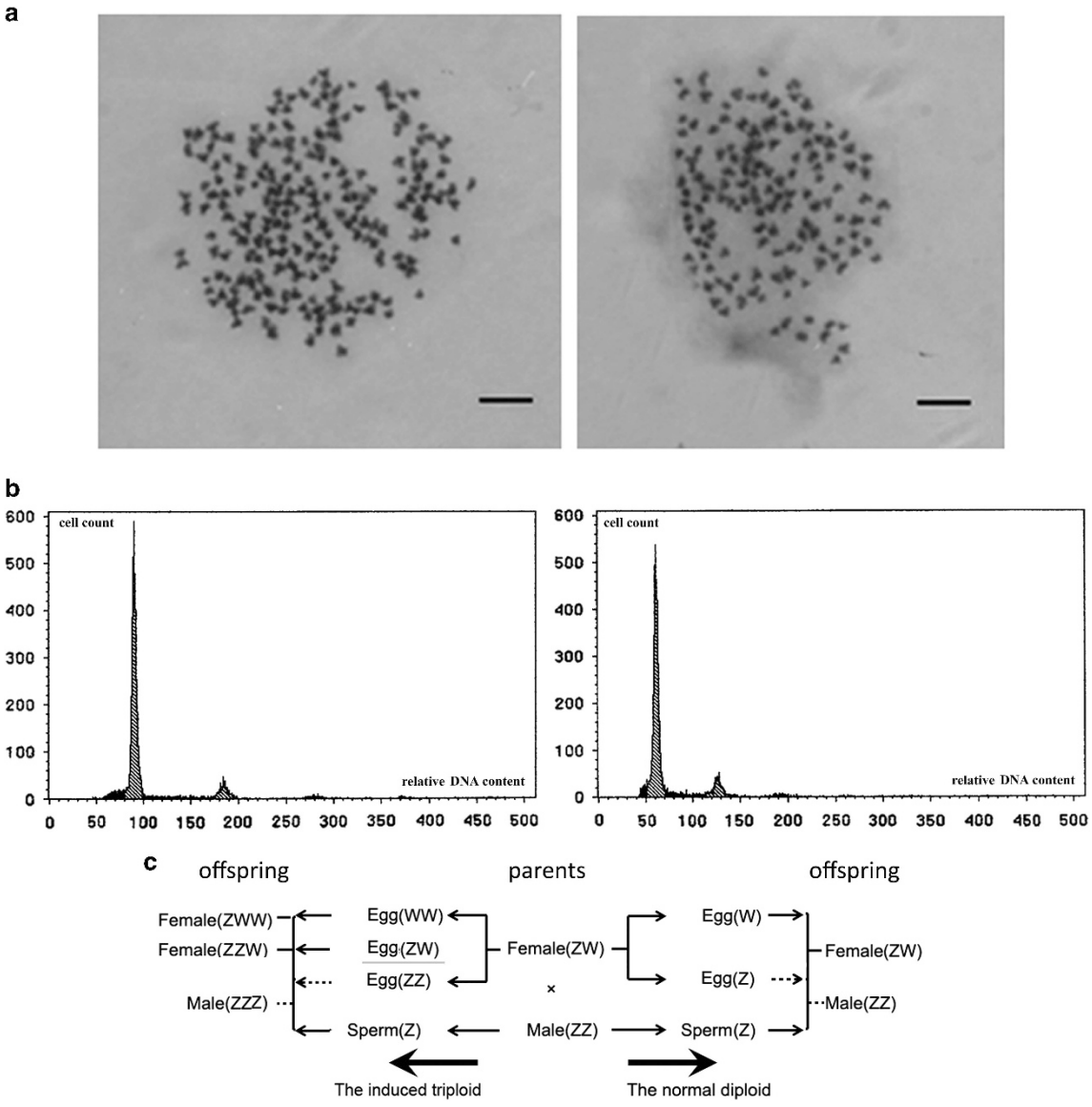


Figure 3 Triploidy induction in *E. sinensis*. (**a**) Photomicrographs of chromosomes of triploid (left, $3n=219$) and diploid (right, $2n=146$) in *E. sinensis*, bar = 10 μm. (**b**) Histogram of relative DNA content in triploid (left) and diploid (right) *E. sinensis* individuals. The higher peaks in each case show the relative DNA contents of triploid (90) and diploid (60), respectively. (**c**) Schematic flow chart of the production of normal diploids and induced triploids. The egg (ZW) underlined is resulted from the possibility of recombination.

chromosome, this gene might function in sex determination or differentiation in *E. sinensis*.

DISCUSSION

Given the high chromosome number of *E. sinensis* ($2n=146$), construction of an accurate high-resolution linkage map has been regarded as a challenge. Our linkage analysis successfully identified all 73 linkage groups. The linkage map provides high resolution, and the average recombination rate across all linkage groups in *E. sinensis* is ~3.09 cM Mb^{-1}, which is much higher than those reported in many other animals, such as mouse (0.5 cM Mb^{-1}), catfish (1.65 cM Mb^{-1}), zebrafish (1.35 cM Mb^{-1}), scallop (1.3 cM Mb^{-1}) and shrimp (1.9 cM Mb^{-1}) (Shimoda *et al.*, 1999; Jensen-Seaman *et al.*, 2004; Kucuktas *et al.*, 2009; Huang *et al.*, 2011; Baranski *et al.*, 2014; Jiao *et al.*, 2014). This suggests that QTLs, if identified, can be narrowed down to quite small genomic region in *E. sinensis*. However, many markers could not be annotated as genes possibly because some of them could not be mapped to exon sequences, and in other cases, where they might be mapped to gene sequences, no annotation is possible since the genes have not been sequenced and characterised in this species or its closely related species.

Sex determination is an integral part of reproduction and a crucial process in genome evolution. In crab aquaculture, sex determination is also of importance. As female crabs are of higher economic value, monosex culture is more profitable. The QTL and GWAS results suggest that LG60 corresponds to the sex chromosome. The female and male LG60 are similar in size, and the two sex chromosomes have a large number of homologous genes. And no difference is found in the recombination rates (female:male = 1.05) (Supplementary Table S3). These suggest that presumably either the fully sex-linked region forms only a small part of this linkage group or there are no fully sex-linked genes at all. The identification of complete sex linkage markers

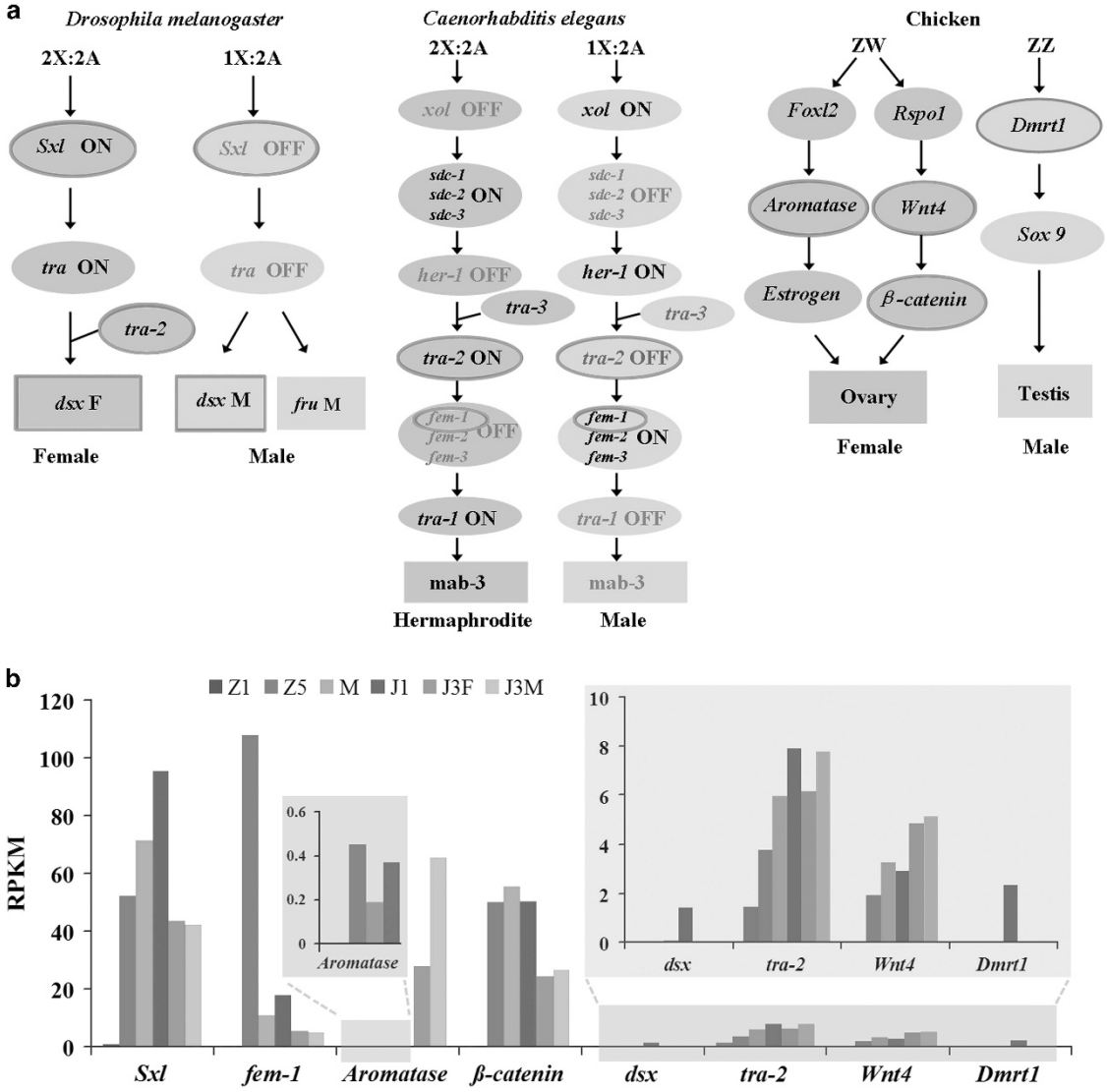


Figure 4 Primary sex determination pathway-related genes detected in *E. sinensis* transcriptomes. (**a**) Sex determination pathways in *Drosophila melanogaster*, *Caenorhabditis elegans* and chicken. Active gene products and gene interactions are in bold and marked with 'ON', whereas the inactive ones are in gray and marked with 'OFF'. The orthologs identified from *E. sinensis* transcriptomes are enclosed in red frames. The pathways are modified based on Hansen and Pilgrim (1999),Chue and Smith (2011) and Kopp (2012). *Sxl, sex-lethal*; *tra, transformer*; *dsx, double sex*; *fru, fruitless*; *xol, XO-lethal*; *sdc, sex and dosage compensation*; *her, hermaphroditization*; *fem, feminization*; *Foxl2, forkhead box protein L2*; *Rspo1, R-spondin-1*; *Wnt4, wingless-type MMTV integration site family, member 4*. (**b**) Expression levels of the identified orthologs of sex determination genes at different developmental stages of *E. sinensis*: Z1, Z5, M, J1, J3F and J3M. A full color version of this figure is available at the *Heredity* journal online.

or regions requires further test in multiple families. However, the ZW-ZZ system is clearly supported by the segregation pattern of the 46 sex-associated markers. The sex ratio of triploids induced by the inhibition of the second polar body skewing towards female also contributes to supporting the putative ZW-ZZ sex determination mechanism (Sellars *et al.*, 2010).

Among the many genes identified with potential functions in sexual development in other animals, all but one are located on autosomal linkage groups of *E. sinensis*, and only an *ankyrin-2* gene with sex-associated marker f4427 is mapped on the putative sex chromosomes (Figure 5a). Ankyrins are well-known adapter proteins that mediate protein–protein interaction through the ANK repeat motif. Some ankyrin-repeat-containing genes are essential for the process of sex determination, such as *ankyrin 6* in *Arabidopsis thaliana* (Yu *et al.*, 2010), *fem-1* in *C. elegans* (Spence *et al.*, 1990) and *ankyrin pk2* in *Wolbachia pipientis* (Pichon *et al.*, 2012). *Ankyrin 6* in *A. thaliana* has a central role in male–female gamete recognition, possibly by regulating mitochondrial gene expression. *Fem-1* in *C. elegans* encodes an ankyrin-repeat-containing protein that is required in hermaphrodites for sperm production. The specific expression of one *ankyrin pk2* allele only in the feminizing *Wolbachia* strains suggests that *Wolbachia* feminization might be related to ankyrin-repeat encoding genes. Considering its location on the linkage map, *ankyrin* gene might also has a role in sex determination of *E. sinensis*.

In contrast to a previous report on a *Dmrt*-like gene in *E. sinensis* (Zhang and Qiu, 2010), four *Dmrt* genes are newly found (*Dmrt1*, *Dmrt2*, *DmrtA2* and *Dmrt11E*). *Dmrt1*, corresponding to marker h1284 on LG3 (Figure 5b), is an important downstream sex determination gene conserved in animals with diverse sex-determining mechanisms, including nematode, fruit fly, fish, amphibians, reptiles,

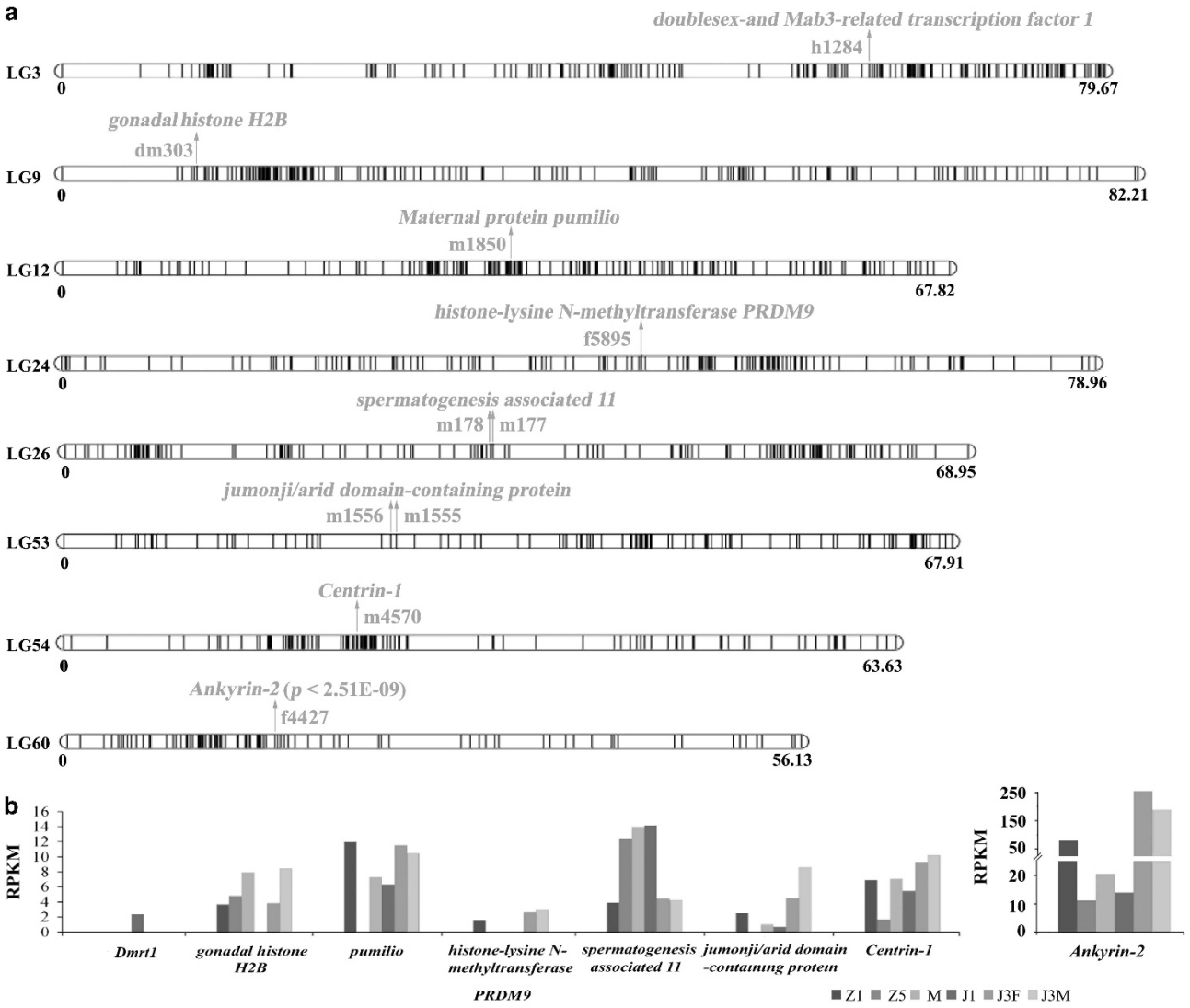


Figure 5 Distribution of putative sex-associated markers on linkage groups of *E. sinensis*. (**a**) Related markers with corresponding genes. Significant *P*-value for marker f4427 in GWAS is shown within parentheses. (**b**) Expression levels of the related genes located on the map at different developmental stages, Z1, Z5, M, J1, J3F and J3M. *Dmrt1*, double sex- and mab3-related transcription factor 1; *Pumilio*, maternal protein pumilio.

birds and mammals, and is expressed in the developing male gonad before morphological differentiation (Raymond *et al.*, 1999b; Marchand *et al.*, 2000; Moniot *et al.*, 2000; Smith *et al.*, 2003). In *E. sinensis*, however, *Dmrt1* is on a putative autosome, as in humans (Raymond *et al.*, 1999a) and amphibian *Rana rugosa* (Aoyama *et al.*, 2003), whereas in chicken and birds, it is Z-linked necessary for testis development (Smith *et al.*, 1999, 2003; Raymond *et al.*, 1999b). Moreover, in contrast to the organisms just mentioned, the expression of this gene in *E. sinensis* begins at J1, whereas gender is already determined at the earlier M (Lee *et al.*, 1994). This suggests that *Dmrt1* might not be involved in sex determination in *E. sinensis*, but the expression of this gene in earlier embryonic stages should be further investigated.

In the sex determination pathway of the three model species *D. melanogaster*, *C. elegans* and chicken, *fem-1* gene is the only one specifically expressed in the early developmental stages of *E. sinensis*. This gene has a signal-transduction role in the sex determination of *C. elegant* (Figure 4a; Hansen and Pilgrim, 1999), Recently, a strong association between SNPs of *fem-1* and sex is reported in the giant tiger shrimp, *P. monodon* (Robinson *et al.*, 2014). All these indicate that *fem-1* might be involved in the sex determination process of decapod crustaceans.

In summary, all our results support a ZW-ZZ sex determination system in *E. sinensis*, thus advancing our knowledge of crustacean sex determination. Sex-associated markers identified in this study could serve as useful candidate sex-specific markers using more families in determination. More than 500 markers on the linkage map are matched with functional genes, covering almost all linkage groups. Candidate genes related to sex determination and differentiation are identified from the transcriptomes, including particularly the crucial genes in sex determination pathway of model species. An *ankyrin* gene located on the putative sex chromosome and *fem-1* with high expression level in the earlier larval developmental stage might have roles in the sex determination of *E. sinensis*, whereas *Dmrt1* might be not. However, the functions of these genes need to be further elucidated, for example, by gene knockdown. Moreover, the genetic map provides an essential tool for the future assembly of whole genome sequences of *E. sinensis*. With more genomic information,

more genes on putative sex chromosomes will be characterised, thus leading towards a comprehensive understanding of the sex determination issue of *E. sinensis*, and decapod crustaceans in general.

DATA ARCHIVING

The 2b-RAD sequence data set was submitted to the SRA (http://www.ncbi.nlm.nih.gov/Traces/sra/) under the accession number SRX472293. Raw data from the Illumina sequencing were deposited in an NCBI short read archive database (GenBank: SRX495513 for Z5, SRX495634 for M, SRX495635 for J1, SRX554564 for J3F and SR554562 for J3M).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Drs Jerome Hui and David Wilmshurst (The Chinese University of Hong Kong) for their comments on the manuscript. This work was supported by grants from the Chinese National '863' Project (No. 2012AA10A409) and National Natural Science Foundation of China (31101924 and 41276165) to Dr Zhaoxia Cui, and a Collaborative Research Fund (C4042-14G) from the Research Grants Council, Hong Kong Special Administrative Region, China (to Dr KH Chu).

Aoyama S, Shibata K, Tokunaga S, Takase M, Matsui K, Nakamura M (2003). Expression of Dmrt1 protein in developing and in sex-reversed gonads of amphibians. *Cytogenet Genome Res* **101**: 295–301.

Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH *et al.* (2013). RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol* **14**: R26.

Baranski M, Gopikrishna G, Robinson NA, Katneni VK, Shekhar MS, Shanmugakarthik J *et al.* (2014). The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PLoS One* **9**: e85413.

Charlesworth D (2013). Plant sex chromosome evolution. *J Exp Bot* **64**: 405–420.

Chakravarti A, Lasher LK, Reefer JE (1991). A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* **128**: 175–182.

Chen MH, Yang Q (2010). GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**: 580–581.

Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P *et al.* (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**: 253–260.

Chue J, Smith CA (2011). Sex determination and sexual differentiation in the avian model. *FEBS J* **278**: 1027–1034.

Cotton JA, Lilley CJ, Jones LM, Kikuchi T, Reid AJ, Thorpe P *et al.* (2013). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol* **15**: R43.

Cui Z, Li X, Liu Y, Song C, Hui M, Shi G *et al.* (2013). Transcriptome profiling analysis on whole bodies of microbial challenged *Eriocheir sinensis* larvae for immune gene identification and SNP development. *PLoS One* **8**: e82156.

Cui Z, Xiang J, Zhou L, Cai N, Song L (2004). Improvement of polyploidy induction in *Eriocheir sinensis*. *Acta Oceanol Sin* **23**: 725–732.

Devlin RH, Nagahama Y (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208**: 191–364.

Fishman L, Kelly AJ, Morgan E, Willis JH (2001). A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701–1716.

Ford AT (2008). Can you feminise a crustacean? *Aquat Toxicol* **88**: 316–321.

Fu X, Dou J, Mao J, Su H, Jiao W, Zhang L *et al.* (2013). RADtyping: an integrated package for accurate *de novo* codominant and dominant RAD genotyping in mapping populations. *PLoS One* **8**: e79960.

Ginsburger VT, Charniaux CH (1982). Sex determination. In: Abele LG (eds). *The Biology of Crustacea*. Academic Press: Orlando, FL, USA, pp 257–281.

Grattapaglia D, Sederoff R (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**: 1121–1137.

Hansen D, Pilgrim D (1999). Sex and the single worm: sex determination in the nematode *C. elegans*. *Mech Dev* **83**: 3–15.

Herborg LM, Rushton S, Clare A, Bentley M (2005). The invasion of the Chinese mitten crab (*Eriocheir sinensis*) in the United Kingdom and its comparison to continental Europe. *Biol Invas* **7**: 959–968.

Huang SW, Lin YY, You EM, Liu TT, Shu HY, Wu KM *et al.* (2011). Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*. *BMC Genom* **12**: 242.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF *et al.* (2004). Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* **14**: 528–538.

Jiao W, Fu X, Dou J, Li H, Su H, Mao J *et al.* (2014). High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Res* **21**: 85–101.

Jones DB, Jerry DR, Khatkar MS, Raadsma HW, Zenger KR (2013). A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for gene localisation and marker-assisted selection. *BMC Genom* **14**: 810.

Kopp A (2012). *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet* **28**: 175–184.

Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z *et al.* (2009). Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics* **181**: 1649–1660.

Lécher P, Defaye D, Noel P (1995). Chromosomes and nuclear DNA of Crustacea. *Invertebr Reprod Dev* **27**: 85–114.

Lee TH, Naitoh N, Yamazaki F (2004). Chromosome studies on the mitten crabs *Eriocheir japonica* and *E. sinensis*. *Fisheries Sci* **70**: 211–214.

Lee TH, Yamauchi M, Yamazaki F (1994). Sex differentiation in the crab *Eriocheir japonicus* (Decapoda, Grapsidae). *Invertebr Reprod Dev* **25**: 123–137.

Li H (2011). A quick method to calculate QTL confidence interval. *J Genet* **90**: 355.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K *et al.* (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966–1967.

Li X, Cui Z, Liu Y, Song C, Shi G (2013). Transcriptome analysis and discovery of genes involved in immune pathways from hepatopancreas of microbial challenged mitten crab *Eriocheir sinensis*. *PLoS One* **8**: e68233.

Lowe S, Browne M, Boudjelas S, De Poorter M (2000). *100 Of the World's Worst Invasive Alien Species: A Selection From The Global Invasive Species Database*. Auckland, New Zealand: Invasive Species Specialist Group, p 12.

Ma A, Wang Y, Zou Z, Fu M, Lin P, Zhang Z (2012). *Erk2* in ovarian development of green mud crab *Scylla paramamosain*. *DNA Cell Biol* **31**: 1233–1244.

Marchand O, Govoroun M, D'Cotta H, McMeel O, Lareyre JJ, Bernot A (2000). *DMRT1* expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. *BBA-Gene Struct Expr* **1493**: 180–187.

Moniot B, Berta P, Scherer G, Südbeck P, Poulat F (2000). Male specific expression suggests role of *DMRT1* in human sex determination. *Mech Dev* **91**: 323–325.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621–628.

Niiyama H (1937). The problem of male heterogamety in the decapod Crustacea, with special reference to the sex-chromosomes in *Plagusia dentipes* de Haan and *Eriocheir japonicus* de Haan. *J Sci Hokkaido Univ Zool* **5**: 283–295.

Niiyama H (1938). The XY chromosomes of the shore-crab, *Hemigrapsus sanguineus* (de Haan). *Jpn J Genet* **14**: 34–38.

Niiyama H (1959). An XX-Y sex-mechanism in the male of a decapod crustacea *Cervimunida princeps* Benedict. *Bull Fac Fisheries Hokkaido Univ* **10**: 106–112.

Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB (2013). Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genom* **14**: 566.

Penalva LO, Sánchez L (2003). RNA binding protein sex-lethal (Sxl) and control of *Drosophila* sex determination and dosage compensation. *Microbiol Mol Biol Rev* **67**: 343–359.

Pichon S, Bouchon D, Liu C, Chen L, Garrett RA, Grève P (2012). The expression of one ankyrin pk2 allele of the WO prophage is correlated with the Wolbachia feminizing effect in isopods. *BMC Microbiol* **12**: 55.

Piepho HP (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* **157**: 425–432.

Raymond CS, Kettlewell JR, Hirsch B, Bardwell VJ, Zarkower D (1999b). Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev Biol* **215**: 208–220.

Raymond CS, Parker ED, Kettlewell JR, Brown LG, Page DC, Kusz K (1999a). A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. *Hum Mol Genet* **8**: 989–996.

Robinson NA, Gopikrishna G, Baranski M, Katneni VK, Shekhar MS, Shanmugakarthik J *et al.* (2014). QTL for white spot syndrome virus resistance and the sex-determining locus in the Indian black tiger shrimp (*Penaeus monodon*). *BMC Genom* **15**: 731.

Rudnick DA, Hieb K, Grimmer KF, Resh VH (2003). Patterns and processes of biological invasion: the Chinese mitten crab in San Francisco Bay. *Basic Appl Ecol* **4**: 249–262.

Sellars M, Li F, Preston N, Xiang J (2010). *Penaeid shrimp* polyploidy: global status and future direction. *Aquaculture* **310**: 1–7.

Shen SH, Hu YC, Zhou X (2014). Sex-lethal gene of the Chinese mitten crab *Eriocheir sinensis*: cDNA cloning, induction by eyestalk ablation, and expression of two splice variants in males and females. *Dev Genes Evol* **224**: 97–105.

Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**: 219–232.

Smith CA, Katz M, Sinclair AH (2003). *DMRT1* is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. *Biol Reprod* **68**: 560–570.

Smith CA, McClive PJ, Western PS, Reed KJ, Sinclair AH (1999). Evolution: conservation of a sex-determining gene. *Nature* **402**: 601–602.

Spence AM, Coulson A, Hodgkin J (1990). The product of *fem-1*, a nematode sex-determining gene, contains a motif found in cell cycle control proteins and receptors for cell-cell interactions. *Cell* **60**: 981–990.

Staelens J, Rombaut D, Vercauteren I, Argue B, Benzie J, Vuylsteke M (2008). High-density linkage maps and sex-linked markers for the black tiger shrimp (*Penaeus monodon*). *Genetics* **179**: 917–925.

Stam P (1993). Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J* **3**: 739–744.

Van Ooijen JW (2004). *MapQTL 5. Software for the Mapping of Quantitative Trait Loci in Experimental Populations*. Kyazma BV: Wageningen, p 63.

Voorrips R (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**: 77–78.

Wang J, Fang DA, Wang Y, Wang YL, Cheng L, He L *et al*. (2013). Cathepsin A protein from the accessory sex gland of the Chinese mitten crab (*Eriocheir sinensis*) plays a key role in spermatophore digestion. *J Insect Physiol* **59**: 953–960.

Wang Q, Chen L, Wang Y, Li W, He L, Jiang H (2012b). Expression characteristics of two ubiquitin/ribosomal fusion protein genes in the developing testis, accessory gonad and ovary of Chinese mitten crab *Eriocheir sinensis*. *Mol Biol Rep* **39**: 6683–6692.

Wang Q, Fang DA, Sun JL, Wang Y, Wang J, Liu LH (2012a). Characterization of the *vasa* gene in the Chinese mitten crab *Eriocheir sinensis*: A germ line molecular marker. *J Insect Physiol* **58**: 960–965.

Wang S, Meyer E, McKay JK, Matz MV (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods* **9**: 808–810.

Wu Y, Close TJ, Lonardi S (2011). Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Trans Comp Biol Bioinf* **8**: 381–394.

Yu F, Shi J, Zhou J, Gu J, Chen Q, Li J *et al*. (2010). ANK6, a mitochondrial ankyrin repeat protein, is required for male–female gamete recognition in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **107**: 22332–22337.

Zhang EF, Qiu GF (2010). A novel *Dmrt* gene is specifically expressed in the testis of Chinese mitten crab, *Eriocheir sinensis*. *Dev Genes Evol* **220**: 151–159.

Zhu Z, Jie Y, Yuhui S, Guowei L (2008). SYBR Green I Real-time PCR-based method for the estimation of genome sizes of *Eriocheir sinensis*. *Agric Sci Jiangsu* **5**: 164–166.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)